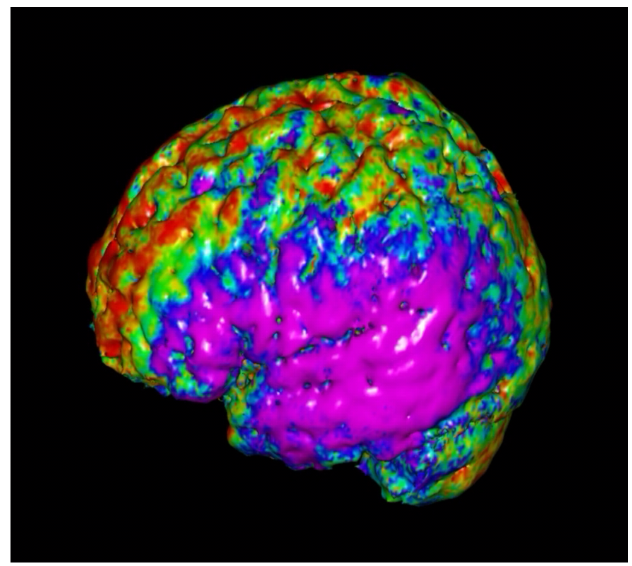
Intrinsic Semiparametric and Nonparametric Models for Symmetric Positive Definite Matrices

Hongtu Zhu, Ph.D.

Department of Biostatistics and Biomedical Research Imaging Center

University of North Carolina at Chapel Hill



Outline

Medical Imaging

Regression Models for Symmetric Positive Definite Matrices

Nonparametric Models for Symmetric Positive Definite Matrices

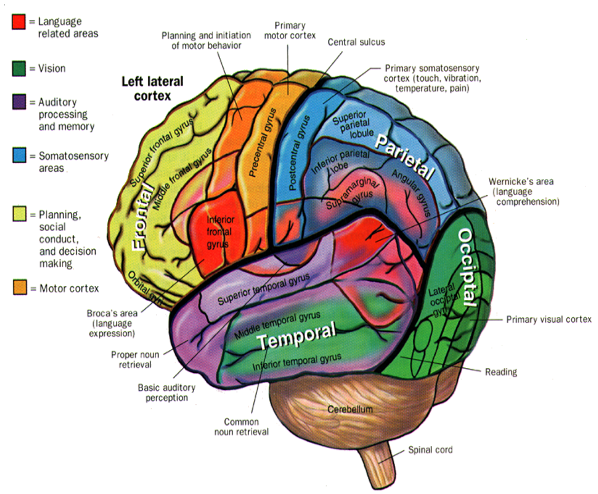
Simulation Studies and Real Data Analysis

Future Work

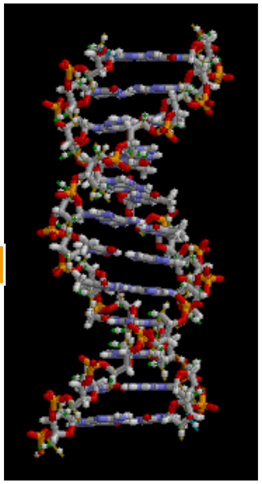
Medical Imaging

Study function and development of brain functional and structural connectivity.

BRAIN

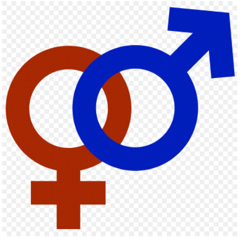


DNA



BEHAVIOR

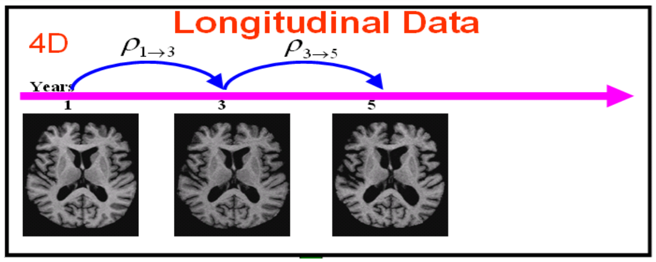
GENDER

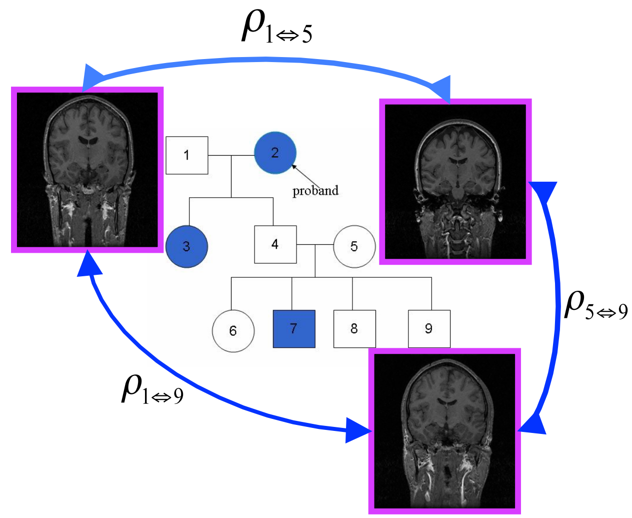


Age

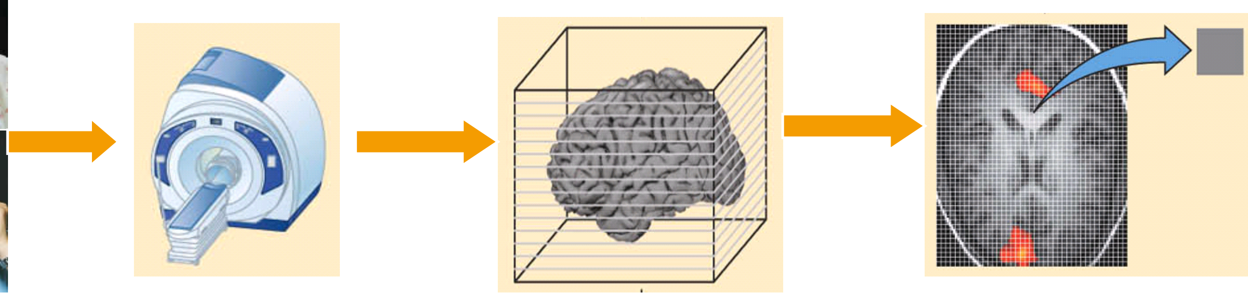
Medical Imaging

Study Design





Neuroimaging Data

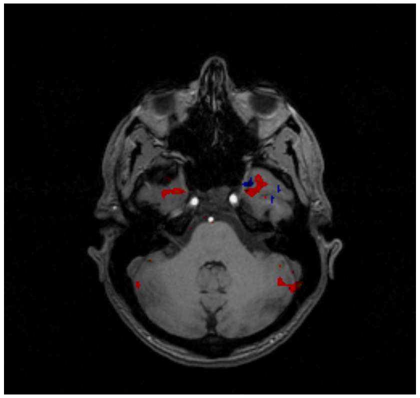


www.guysandstthomas.nhs.uk/.../T/Twins400.jpg

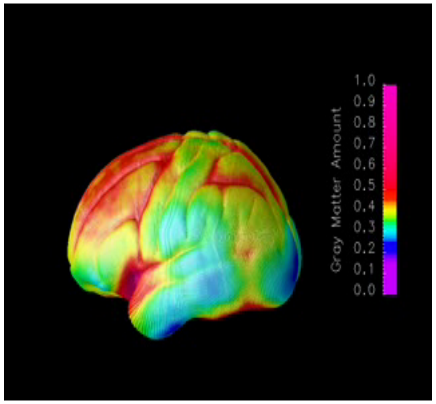
Medical Imaging

Euclidean-valued Data

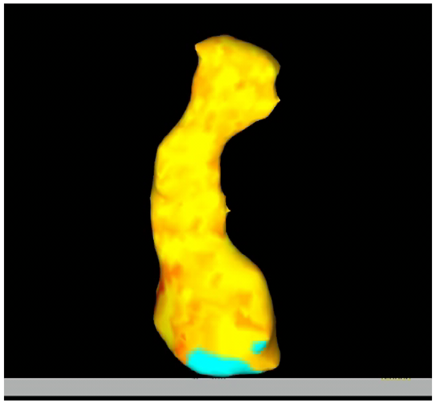
intensity, fMRI, volume, grey matter density, SPHARM, invariant measure, signed-Euclidean distance, …



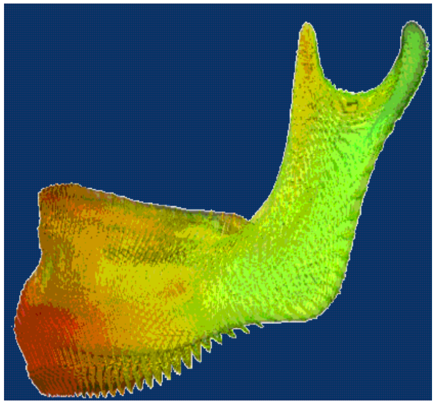
fMRI



Grey matter density



Signed-Euclidean distance

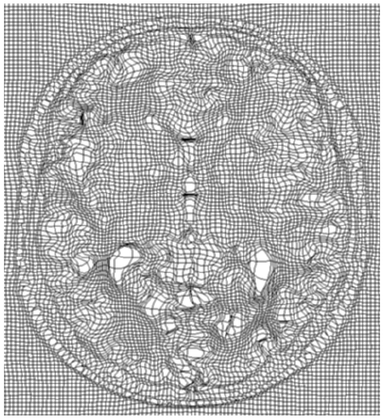


SPHARM

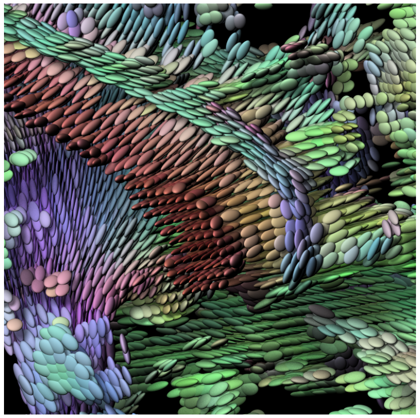
Medical Imaging

Manifold-valued Data

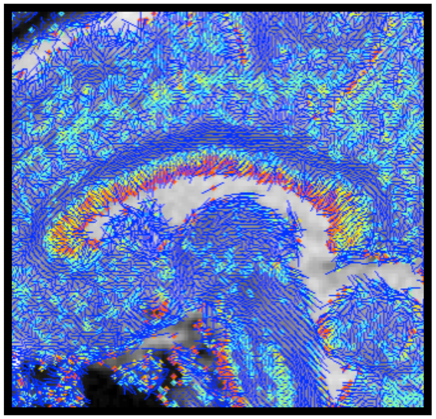
Directional data, deformation tensors, diffusion tensors, principal directions, medial representation, projections, orientation, rigid motion,



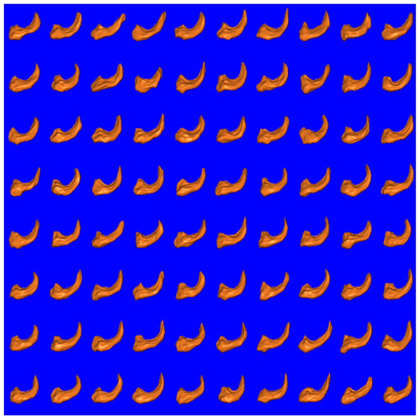
Deformation Tensor



Diffusion Tensor



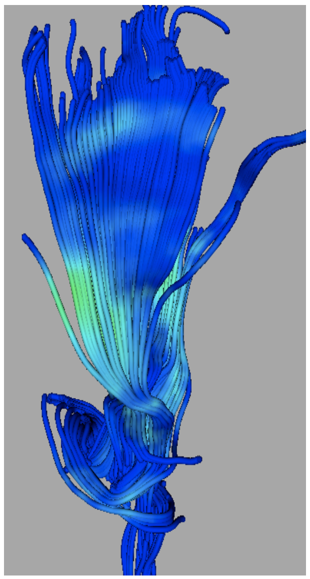
Principal Direction



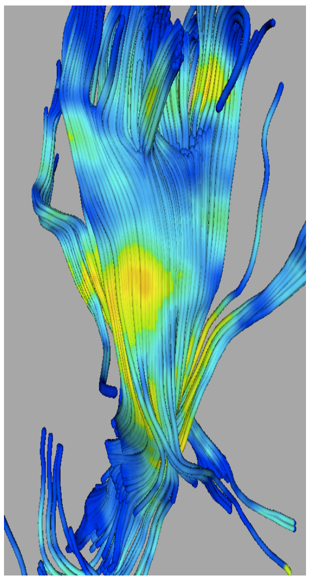
Medial Representation

Medical Imaging

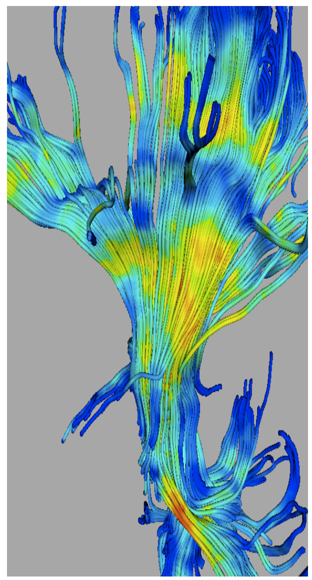
White Matter Maturation



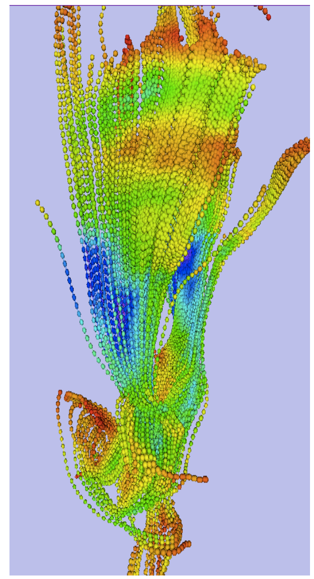
Week 2



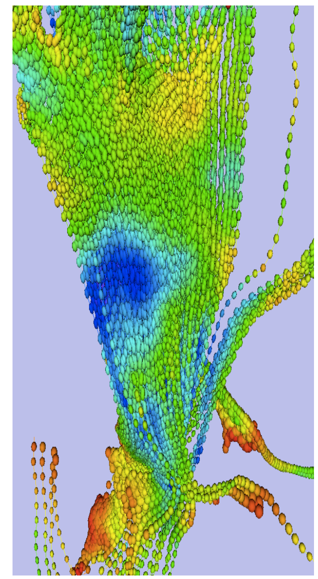
Year 1



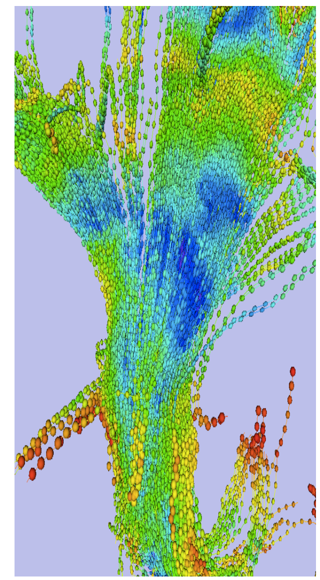
Year 2



Week 2



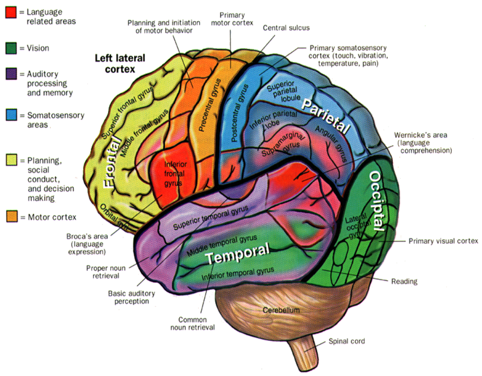
Year 1



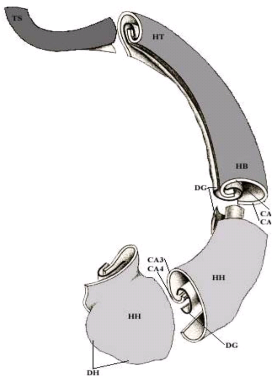
Year 2

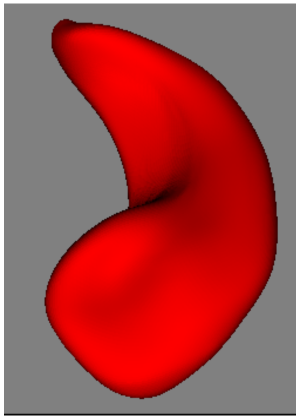
Medical Imaging

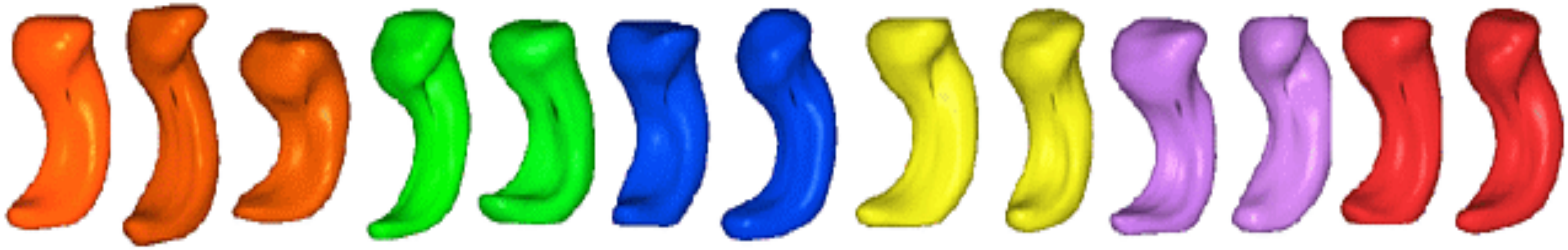
Volumetric Measures



Shape Representation

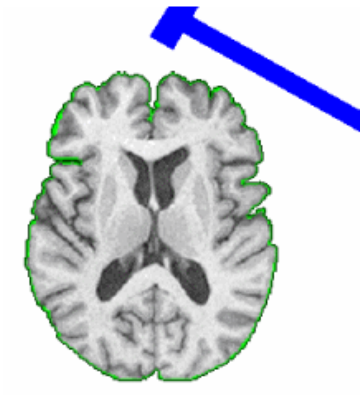


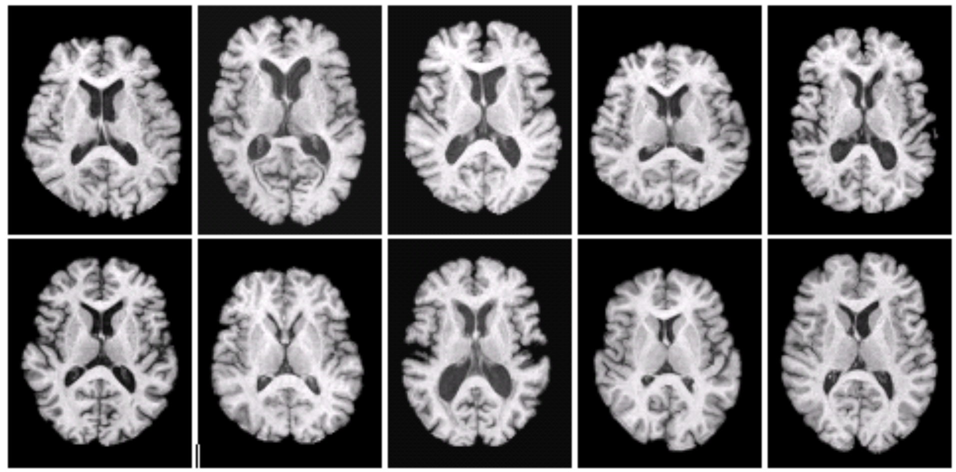


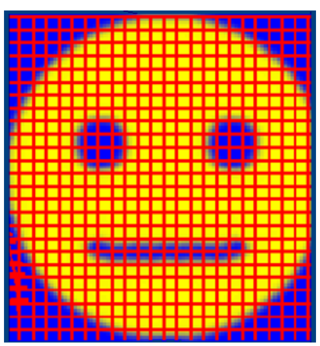


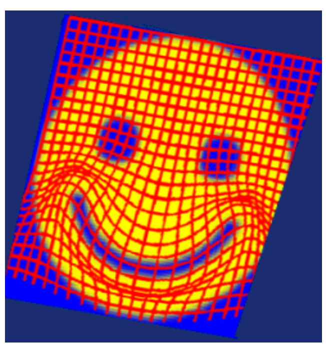
Medical Imaging

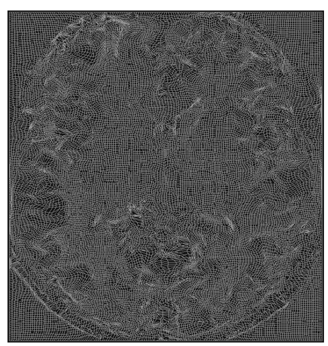
Deformation Field

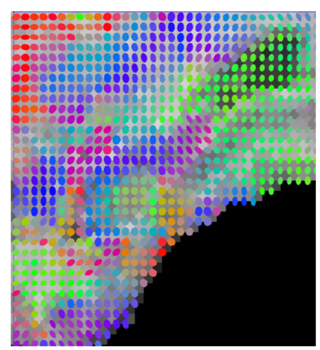


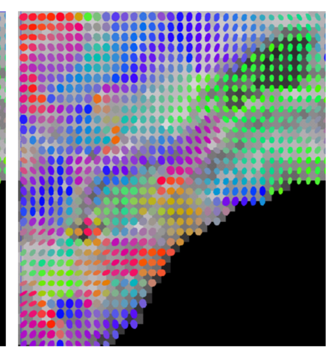






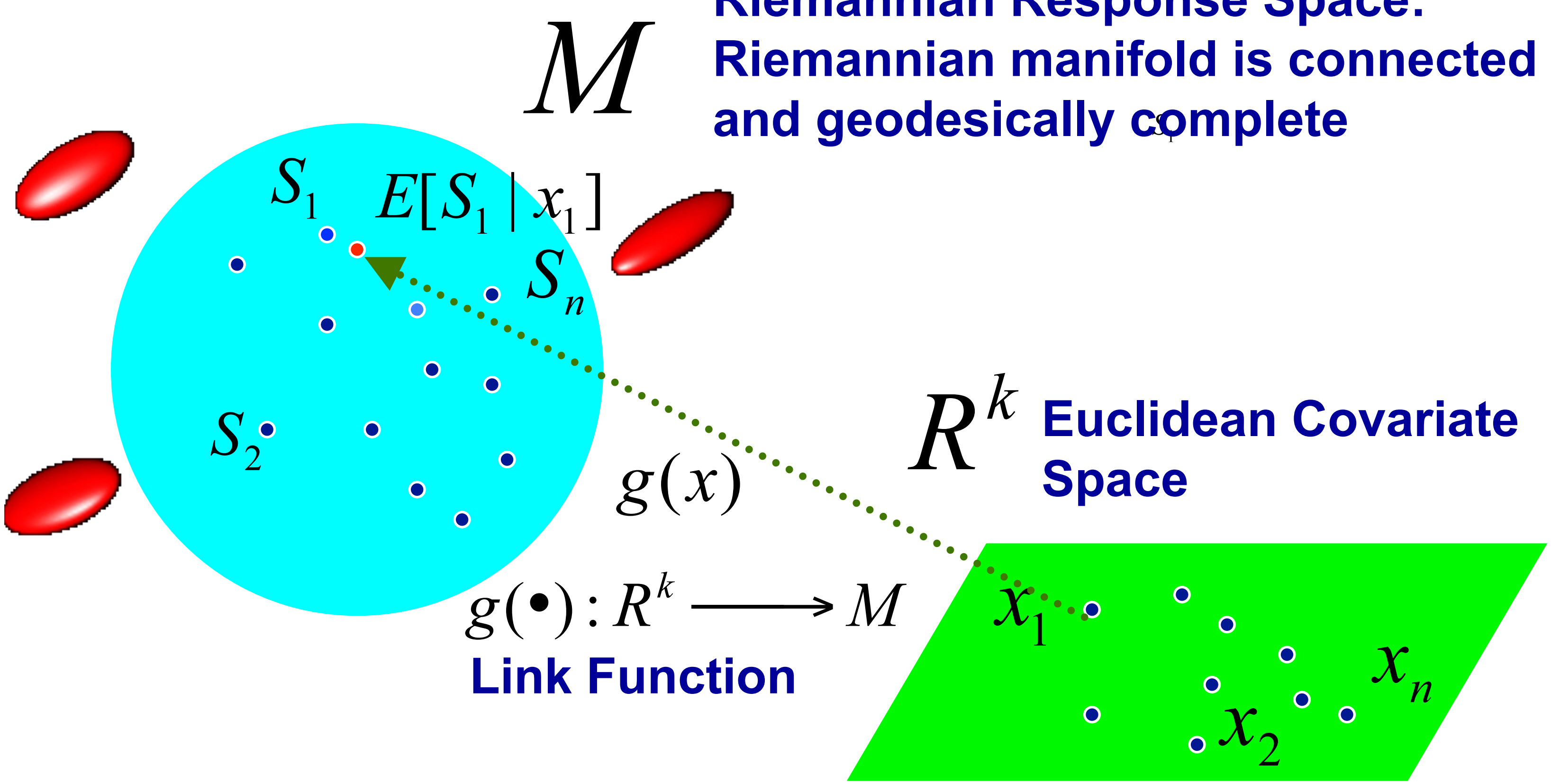






Medical Imaging

Manifold-valued Data



Medical Imaging

Existing Literature in Statistics

Parametric/Semiparametric Statistical Inference in Euclidean Space:

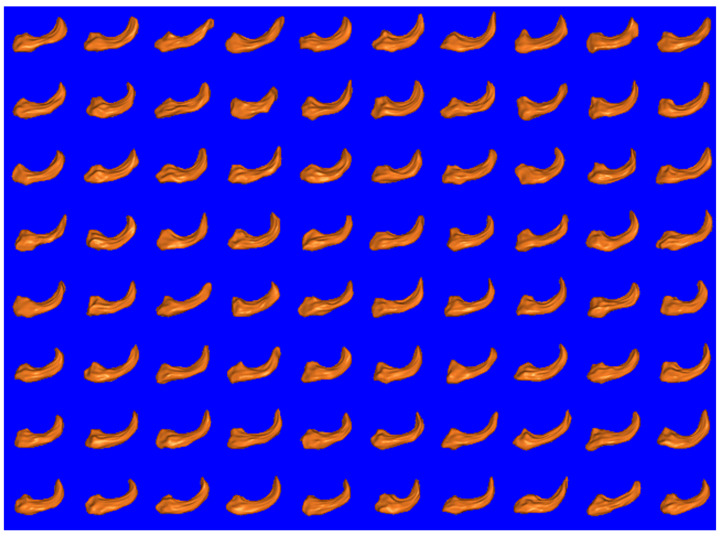
Rao (1945), Efron (1975), Amari (1985), Cook (1986), McCullagh (1987), Barndorff-Nielsen and Cox (1994), Wei (1988), …

Statistics for Manifold-valued Data:

- Directional Statistics: Fisher (1953), Fisher (1993), Kent (1977), Watson (1983), Mardia and Jupp (1999), …
- Axial and Shape Spaces: Kendall (1977, 1984), Dryden and Mardia (1998), Kendall, Barden, Carne, and Le (1999), …
- Diffusion Tensors: Armin Schwartzman (2006, 2008), Fletcher and Joshi (2007), Dryden et al. (2009), Zhu et al. (2009), …
- Riemmanian Manifold: Bhattacharya and Patrangenaru (2003a, b), …
- Data Mining: Huckemann, S., Hotz, T., Munk, A. (2010), Huckemann et al. (2006), ..
- Bayesian methods: Jermyn (2005), Angers and Kim (2005), Bhattacharya and Dunson (2010), …

Medical Imaging

Semiparametric and Nonparametric Regression for Manifold-valued Response from Cross-sectional, Longitudinal and Family-based Neuroimaging Studies



M

$$= g(x, \theta, f) \oplus \varepsilon$$

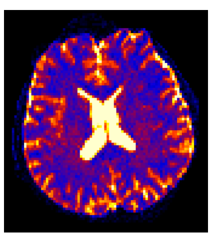
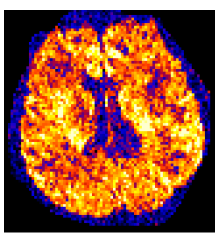
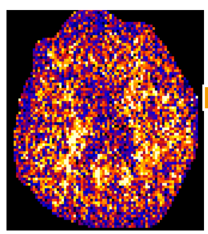
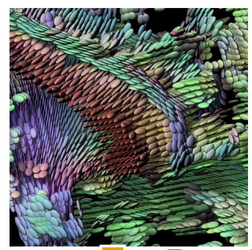
$$x \in R^k, \theta \in \Theta \subset R^p, f \in F$$

$$g : R^k \times R^p \times F \to M$$

Regression Models for SPDs

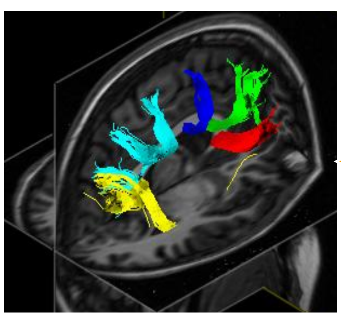
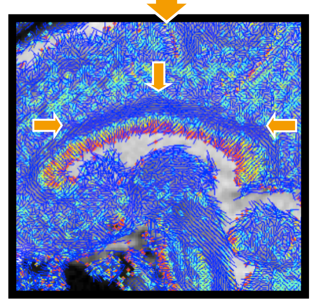
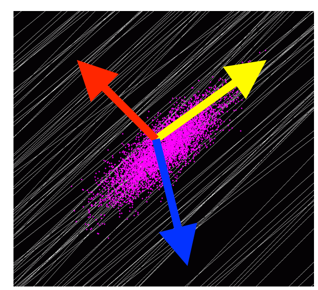
Symmetric Positive Definite Matrix (SPD)

- Diffusion Tensors in DTI are 3x3 SPDs. DTI is an imaging modality that allows measurement of fiber-tract trajectories *in vivo* in soft tissues.
- Covariance Matrices: Multivariate analysis, Longitudinal data, Spatial data, ...
- Network Data:

$$D = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{pmatrix} > 0$$

$$D = \lambda_1 \vec{\eta}_1 \vec{\eta}_1^T + \lambda_2 \vec{\eta}_2 \vec{\eta}_2^T + \lambda_3 \vec{\eta}_3 \vec{\eta}_3^T$$

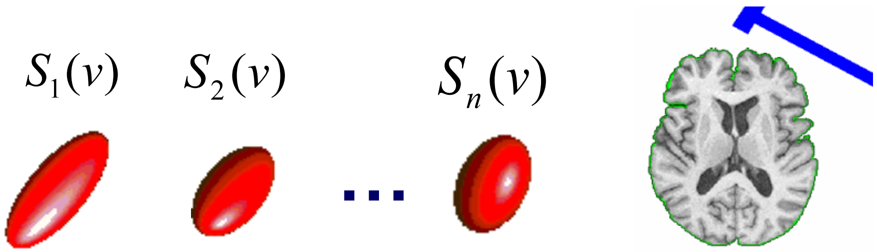




$\lambda_1 \vec{\eta}_1$

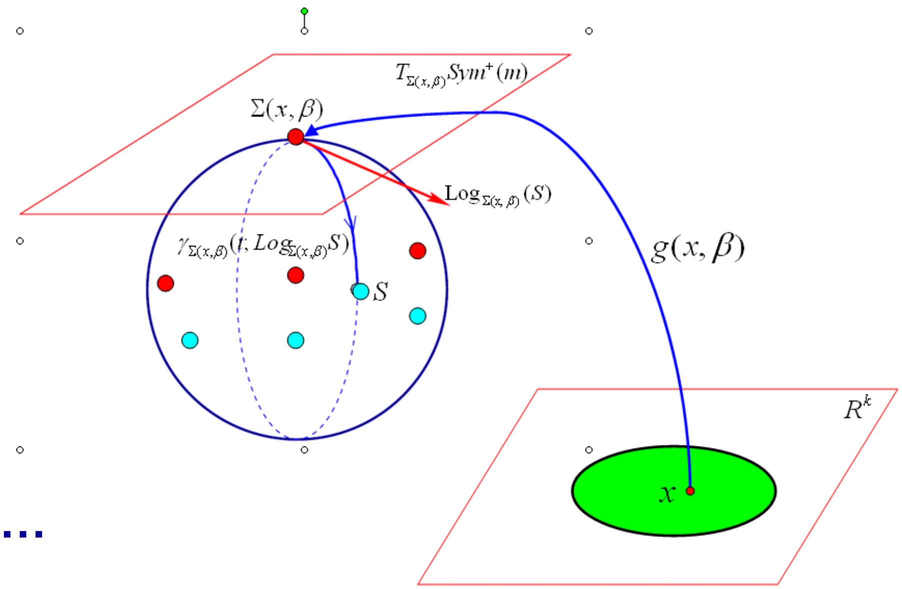
Regression Models for SPDs

Diffusion Tensors

$S_1(v)$ $S_2(v)$ … $S_n(v)$



x_1 x_2 x_n



Age, Gender, Race, Diagnostic Status, …

An appropriate statistical analysis of SPD matrices is important for understanding normal brain development, the neural bases of neuropsychiatric disorders, and the joint effects of environmental and genetic factors on brain structure and function.

Euclidean Space

Regression Models for SPD

A formal statistical framework for using a set of covariates in a Euclidean space to predict SPD matrices as responses:

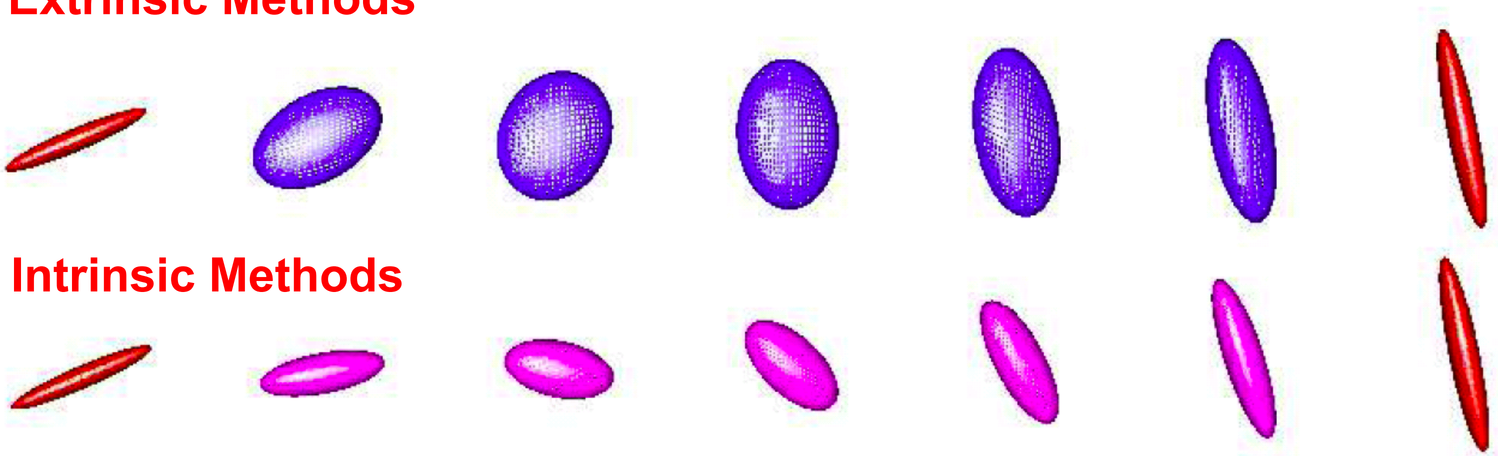
- **Extrinsic Methods:** **Ignore the fact that SPD matrices are in a nonlinear space and then directly apply classical multivariate regression.** Schwartzman and Taylor (2008), …

- **Intrinsic Methods:** **Several parametric models for SPD matrices.**

Regression Models for SPD

Extrinsic Methods

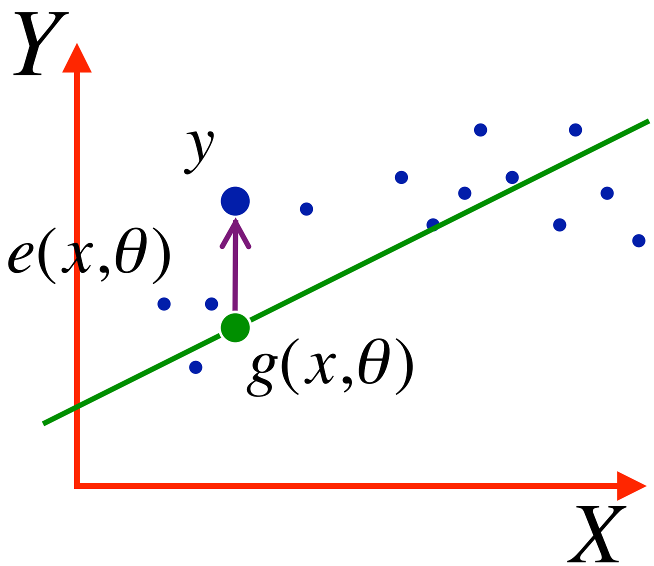
Intrinsic Methods



Dryden, I.L., Koloydenko, A. and Zhou, D. (2009).

Regression Models for SPD

$$\{(x_i, y_i) : i = 1, \cdots, n\}$$



- **residual**

$$e(x,\theta) = y - g(x,\theta)$$

- **link function**

$$g(x,\theta) : R^k \times \Theta \to \underline{R}$$

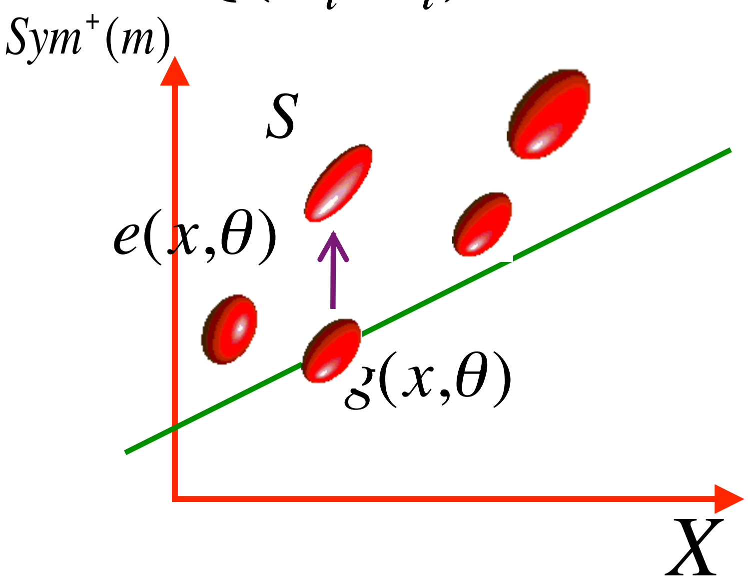
Space of Y

Conditional mean

$$E[e(x,\theta) \mid x] = E[y - g(x,\theta) \mid x] = 0$$

Regression Models for SPD

$$\{(x_i, S_i) : i = 1, \cdots, n\}$$



Questions:

- **How to define residual ?**

$$e(x,\theta) = S - g(x,\theta)$$

- **How to define link function ?**

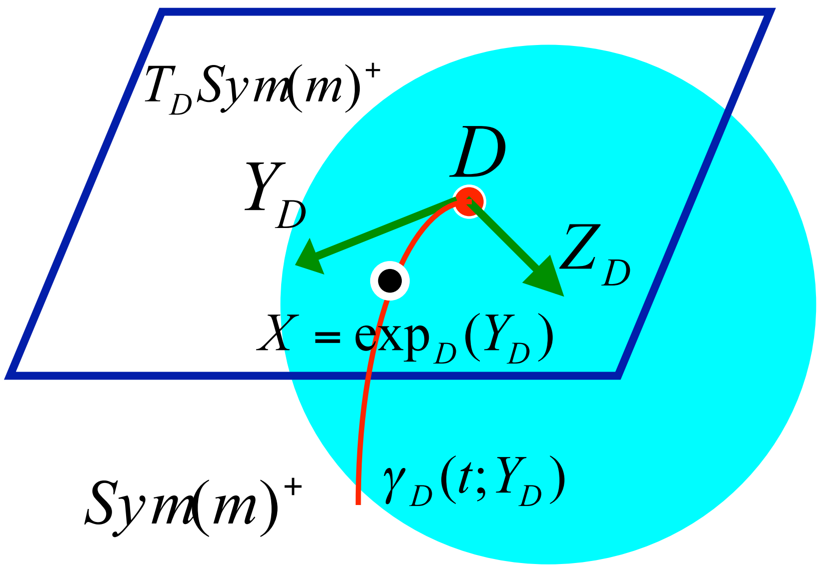
$$g(x,\theta) : R^k \times \Theta \rightarrow Sym^+(m)$$

Space of S

- **How to define conditional mean ?**

$$E[e(x,\theta) \mid x] = E[S - g(x,\theta) \mid x] = 0$$

Regression Models for SPD



Scale Frobenius inner product

$$< Y_D, Z_D > = tr(Y_D D^{-1} Z_D D^{-1})$$

Geodesic

$$\gamma_D(t; Y_D) = G \exp(t G^{-1} Y_D G^{-T}) G^T$$

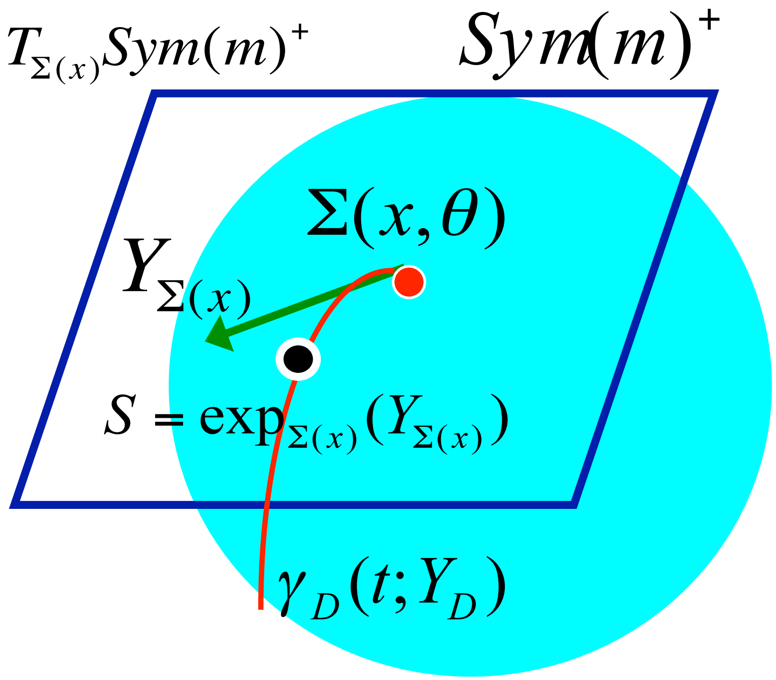
$$D = G G^T$$

Riemannian exponential/logarithm maps

$$X = Exp_D(Y_D) = \gamma_D(1; Y_D) = G \exp(G^{-1} Y_D G^{-T}) G^T$$

$$Y_D = Log_D(X) = G \log(G^{-1} X G^{-T}) G^T$$

Regression Models for SPD



Riemannian logarithm map

$$Y_{\Sigma(x)} = Log_{\Sigma(x)}(S) = G(x)\log(G(x)^{-1}SG(x)^{-T})G(x)^T$$

$$\Sigma(x) = \Sigma(x,\theta) = G(x,\theta)G(x,\theta)^T = G(x)G(x)^T$$

- **Use Riemannian logarithm map to construct residuals**
- **Rotate residuals to the same tangent plane (parallel transport)**

Q1: Residual $e(x,\theta) = Log_{\Sigma(x,\theta)}(S) = \log(G(x,\theta)^{-1}SG(x,\theta)^{-T})$

Regression Models for SPD

Q2: Link functions

$$\Sigma(x,\theta) : R^k \times \Theta \to Sym(m)^+$$

- **Cholesky decomposition**

$$\Sigma(x,\theta) = G(x,\theta)G(x,\theta)^T$$

$$G(x,\theta) = \begin{pmatrix} g_{11}(x,\theta) & 0 & 0 \\ g_{21}(x,\theta) & g_{22}(x,\theta) & 0 \\ g_{31}(x,\theta) & g_{32}(x,\theta) & g_{33}(x,\theta) \end{pmatrix}$$

$$g_{ii}(x,\theta) \geq 0$$

Regression Models for SPD

Link functions

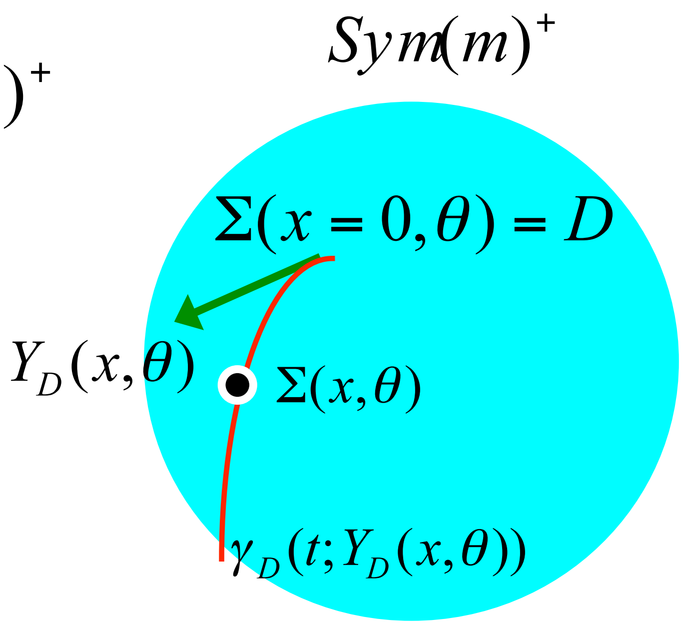
$$\Sigma(x,\theta): R^k \times \Theta \to Sym(m)^+$$

- **matrix logarithm link**

$$\log(\Sigma(x,\theta)) = g(x,\theta)$$

- **geodesic link**

$$\Sigma(x=0,\theta) = D,$$
$$\Sigma(x,\theta) = \gamma_D(t(x), Y_D(x,\theta)),$$
$$t(x)Y_D(x,\theta) = g(x,\theta)$$



Regression Models for SPD

Q3: Conditional Moment Model

$$E[e(x,\theta)\,|\,x] = E[Log_{\Sigma(x,\theta)}(S)\,|\,x] = 0$$

Intrinsic least square estimator (ILSE)

$$\hat{\theta} = \arg\min G_n(\theta) = \arg\min \sum_{i=1}^{n} tr(Log_{\Sigma(x_i,\theta)}(S_i) Log_{\Sigma(x_i,\theta)}(S_i))$$

ILSE includes the intrinsic mean as a special case.

$$\sum_{i=1}^{n} tr(Log_{\Sigma(x_i,\theta)}(S_i) Log_{\Sigma(x_i,\theta)}(S_i)) = \sum_{i=1}^{n} d(S_i, \Sigma(x_i,\theta))^2$$

Regression Models for SPD

Annealing Optimization Algorithm

- **Gradient algorithm for computing ILSE is relatively sensitive to the starting point.**

$$\theta^{(r+1)} = \theta^{(r)} + \rho\{-\nabla^2 G_n(\theta^{(r)})\}^{-1}\nabla G_n(\theta^{(r)}) \qquad \theta \in R^p$$

$$Hess\{G_n(\theta^{(r)})\}\delta_r = -\nabla G_n(\theta^{(r)}) \qquad \theta \in M$$

$$\theta^{(r+1)} = R_{\theta^{(r)}}(\delta_r)$$

- **Gibbs sampler** $\exp(-G_n(\theta)/\tau)$

Annealing Evolutionary Stochastic Approximation Monte Carlo Algorithm

Liang (2010)

Regression Models for SPD

Estimation Theory

- **Consistency** $\hat{\theta} \xrightarrow{p} \theta_*$

- **Asymptotic Normality**

$$(E[\sum_{i=1}^{n}\{\partial_\theta tr(e(x_i,\hat{\theta})^2)\}^{\otimes 2}])^{-1/2} E\{-\nabla^2 G_n(\hat{\theta})\}(\hat{\theta}-\theta_*) \xrightarrow{L} N(0, I_{p*})$$

Regression Models for SPD

Testing Linear Hypothesis

$$H_0 : A\theta = A_0 \qquad v.s. \qquad H_0 : A\theta \neq A_0$$

- **Wald/Score test statistics**

- **Resampling method/false discovery rate to correct for multiple comparisons**

Regression Models for SPD

Simulation Studies

Cholesky decomposition

$$G(x,\theta) = \begin{pmatrix} x_i^T\beta_1 & 0 & 0 \\ x_i^T\beta_2 & x_i^T\beta_3 & 0 \\ x_i^T\beta_4 & x_i^T\beta_5 & x_i^T\beta_6 \end{pmatrix} \qquad x_i = (1, z_i)^T$$

Data model

$$S_i = G(x_i,\theta)\exp(E_i)G(x_i,\theta)^T$$

$$E_i \sim MN(0,\Omega)$$

Correlation

$$\Omega_1 = \begin{pmatrix} 0.6 & 0 & 0 \\ 0 & 0.6 & 0 \\ 0 & 0 & 0.6 \end{pmatrix} \qquad \Omega_2 = \begin{pmatrix} 0.6 & 0.3 & 0.3 \\ 0.3 & 0.6 & 0.3 \\ 0.3 & 0.3 & 0.6 \end{pmatrix}$$

Regression Models for SPD

Simulation Study I

Table 1. Bias ($\times 10^{-2}$), RMS ($\times 10^{-2}$), SE ($\times 10^{-2}$), SD–SE ($\times 10^{-2}$), and RS of all 12 parameters under Ω_1 and Ω_2. BIAS denotes the bias of the mean of the ILSE estimates; RMS denotes the root-mean-square error; SE denotes the mean of the standard deviation estimates; SD–SE denotes the standard deviation of the standard deviation estimates; RS denotes the ratio of RMS over SD. Two different sample sizes {20, 80} and 500 simulated datasets were used for each case

	$n = 20$					$n = 80$				
	BIAS	RMS	SE	SD–SE	RS	BIAS	RMS	SE	SD–SE	RS
					Ω_1					
β_1	2.60	6.10	6.73	2.20	1.10	0.58	3.59	3.37	0.68	0.94
β_2	1.78	6.10	6.61	1.72	1.08	0.06	3.90	3.61	0.55	0.92
β_3	1.88	7.06	6.96	1.32	0.98	0.69	3.91	3.56	0.35	0.91
β_4	1.15	6.86	6.89	1.30	1.01	0.35	3.83	3.51	0.33	0.92
β_5	3.83	15.34	17.24	4.22	1.12	1.08	8.35	8.58	1.20	1.02
β_6	2.83	15.07	16.97	3.76	1.12	0.54	8.45	8.46	0.95	1.01
β_7	1.43	8.75	8.07	1.42	0.92	−0.37	4.19	4.10	0.39	0.98
β_8	0.48	8.32	7.98	1.40	0.96	−0.44	4.12	4.06	0.38	0.98
β_9	5.14	29.38	32.06	7.57	1.09	1.6	14.84	16.07	2.23	1.08
β_{10}	3.97	28.88	31.59	7.14	1.09	1.05	14.84	15.85	1.77	1.07
β_{11}	3.62	20.32	19.91	3.87	0.98	1.00	10.63	10.15	0.85	0.96
β_{12}	2.69	20.11	19.68	3.83	0.98	0.53	10.48	10.03	0.84	0.96
					Ω_2					
β_1	2.76	6.87	6.73	2.21	0.97	0.59	4.00	3.7	0.57	0.93
β_2	1.81	6.72	6.61	1.63	0.98	0.46	3.97	3.7	0.53	0.93
β_3	1.96	7.74	7.23	1.27	0.93	0.23	3.72	3.54	0.36	0.95
β_4	1.24	7.43	7.15	1.26	0.96	0.03	3.67	3.5	0.36	0.95
β_5	3.08	11.63	12.61	2.85	1.08	1.02	6.87	6.33	0.73	0.92
β_6	1.87	11.78	12.41	2.31	1.05	0.86	6.9	6.3	0.7	0.91
β_7	1.56	8.44	8.26	1.46	0.98	0.11	4.5	4.2	0.34	0.93
β_8	0.62	8.09	8.17	1.44	1.01	0.04	4.5	4.1	0.34	0.91
β_9	2.75	18.90	19.75	4.14	1.04	1.06	10.74	9.93	1.09	0.93
β_{10}	1.33	18.84	19.46	3.68	1.03	0.86	10.79	9.82	0.93	0.91
β_{11}	3.87	16.68	15.49	2.54	0.93	−0.58	7.45	7.44	0.66	0.99
β_{12}	2.88	16.46	15.31	2.51	0.94	−0.91	7.22	7.37	0.65	1.02

Regression Models for SPD

Simulation Study II

Table 2. Comparisons of the rejection rates for score test statistics under Ω_1 and Ω_2. Three differing sample sizes {20, 40, 80} and 1,000 simulated datasets were used for each case and two significance levels, 5% and 1%, were considered. The null and alternative hypotheses were, respectively, given by $H_0: \boldsymbol{\beta}_{\cdot,2} = \mathbf{0}$ and $H_1: \boldsymbol{\beta}_{\cdot,2} \neq \mathbf{0}$. Two methods including the resampling method (RE) and χ^2 distribution [$\chi^2(6)$] were used to calculate the rejection rates

	$n=20$				$n=40$				$n=80$			
	5%		1%		5%		1%		5%		1%	
$\boldsymbol{\beta}_{\cdot,2}$	RE	$\chi^2(6)$	RE	$\chi^2(6)$	RE	$\chi^2(6)$	RE	$\chi^2(6)$	RE	$\chi^2(6)$	RE	$\chi^2(6)$
						Ω_1						
$0 \times \mathbf{1}_6$	0.143	0.031	0.037	0	0.067	0.043	0.026	0.007	0.067	0.037	0.017	0.003
$0.2 \times \mathbf{1}_6$	0.513	0.177	0.253	0.011	0.957	0.883	0.796	0.461	1	1	0.991	0.971
$0.4 \times \mathbf{1}_6$	0.597	0.213	0.293	0.022	0.993	0.951	0.832	0.481	1	1	1	1
$0.6 \times \mathbf{1}_6$	0.773	0.442	0.520	0.042	1	1	1	0.983	1	1	1	1
						Ω_2						
$0 \times \mathbf{1}_6$	0.126	0.023	0.037	0	0.063	0.037	0.017	0.003	0.061	0.033	0.013	0.003
$0.2 \times \mathbf{1}_6$	0.581	0.221	0.302	0.010	0.977	0.953	0.851	0.491	1	1	0.991	0.991
$0.4 \times \mathbf{1}_6$	0.602	0.227	0.321	0.032	0.991	0.981	0.871	0.562	1	1	1	1
$0.6 \times \mathbf{1}_6$	0.903	0.51	0.611	0.051	1	1	1	0.981	1	1	1	1

Regression Models for SPD

Simulation Study III

Table 3. Comparisons of the family-wise error rates and average powers for the test procedure under two different correlations $\rho = 0.0$ and 0.5. We considered two different sample sizes $\{40, 80\}$ and 100 simulated datasets for each case at the 5% significance level. In all voxels, the null and alternative hypotheses were, respectively, given by $H_0 : \boldsymbol{\beta}_{\cdot,2}(d) = \mathbf{0}_6$ and $H_1 : \boldsymbol{\beta}_{\cdot,2}(d) \neq \mathbf{0}_6$. We considered four different $\boldsymbol{\beta}_{\cdot,2}(d)$ $\{0.0 \times \mathbf{1}_6, 0.3 \times \mathbf{1}_6, 0.6 \times \mathbf{1}_6, 0.9 \times \mathbf{1}_6\}$ for all voxels within the region of interest, whereas we set $\boldsymbol{\beta}_{\cdot,2}(d) = \mathbf{0}_6$ for all voxels outside the region of interest. FWR denotes the family wise error rate and Apower denotes the average rejection rate for voxels inside the region of interest

	$n = 40$				$n = 80$			
	$\rho = 0.0$		$\rho = 0.5$		$\rho = 0.0$		$\rho = 0.5$	
$\boldsymbol{\beta}_{\cdot,2}(d)$	FWR	Apower	FWR	Apower	FWR	Apower	FWR	Apower
$0.0 \times \mathbf{1}_6$	0.12	0.00	0.06	0.00	0.08	0.00	0.07	0.00
$0.3 \times \mathbf{1}_6$	0.18	0.10	0.12	0.10	0.06	0.56	0.06	0.57
$0.6 \times \mathbf{1}_6$	0.14	0.67	0.06	0.68	0.02	1.00	0.03	1.00
$0.9 \times \mathbf{1}_6$	0.12	0.83	0.10	0.85	0.08	1.00	0.06	1.00

Regression Models for SPD

HIV Neuroimaging Data (PI: Colin Hall)

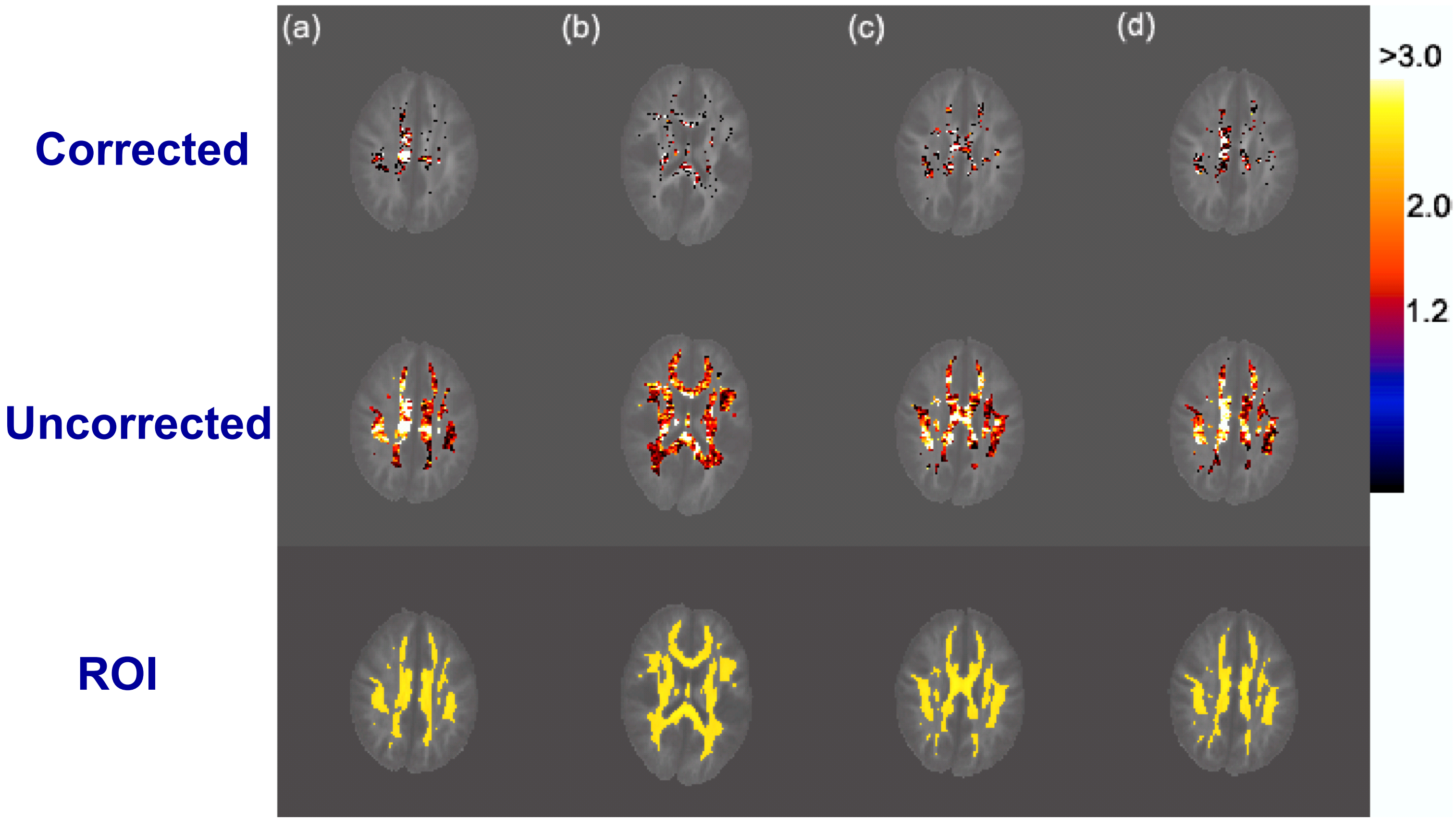
Objective: Assess diagnosis and age on the integrity of white matter in a cross-sectional study of human immunodeciency virus (HIV).

Participants: All 47 subjects with 29 HIV+ subjects (21 males and 8 females) and 18 healthy (9 males and 9 females) controls were studied. We limited the statistical analysis within the major white matter regions (FA>0.4).

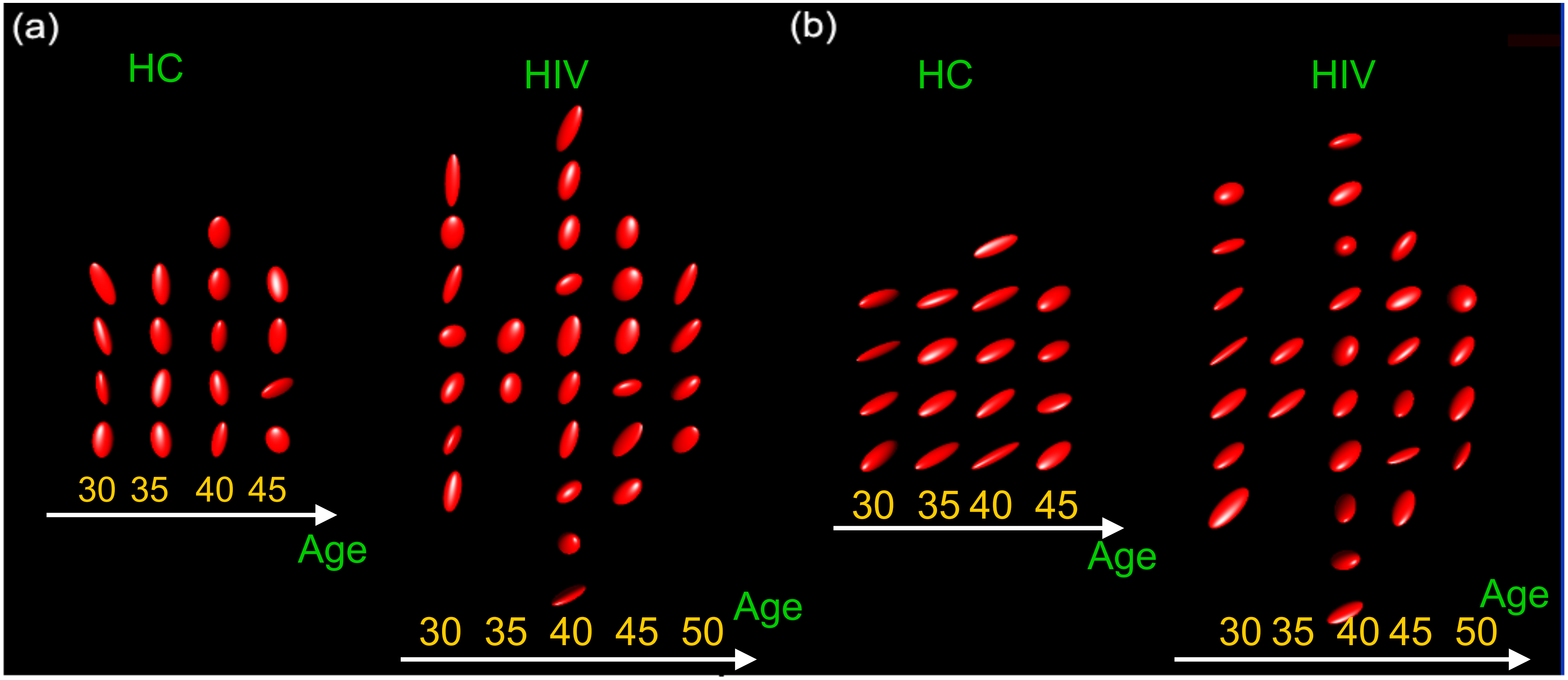
Results: We observe statistically significant diagnosis effects in superior internal capsule area and age effects in inferior longitudinal fasciculus.

Regression Models for SPD

Group Effect



Regression Models for SPD



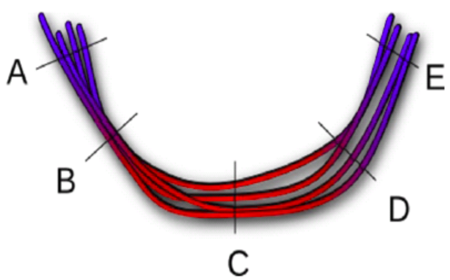
Diagnostic Effect

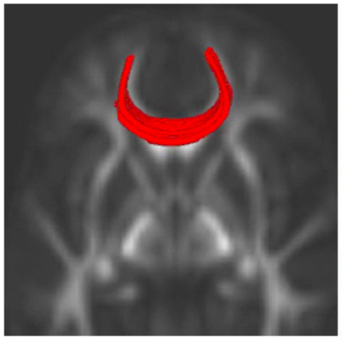
Age Effect

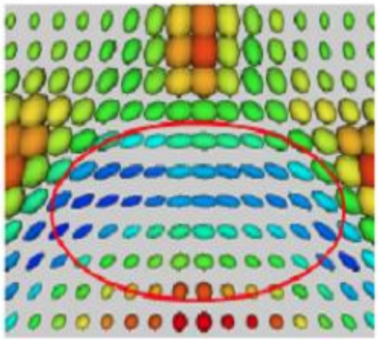
Local Polynomial Kernel Regression for SPD

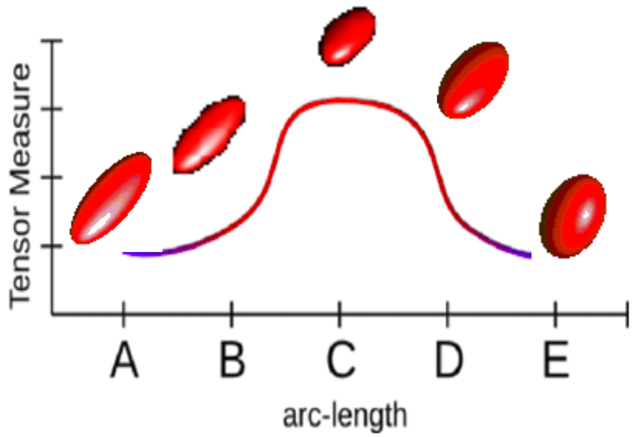
Data $(x_1, S_1), \cdots, (x_n, S_n)$

$S(x)$

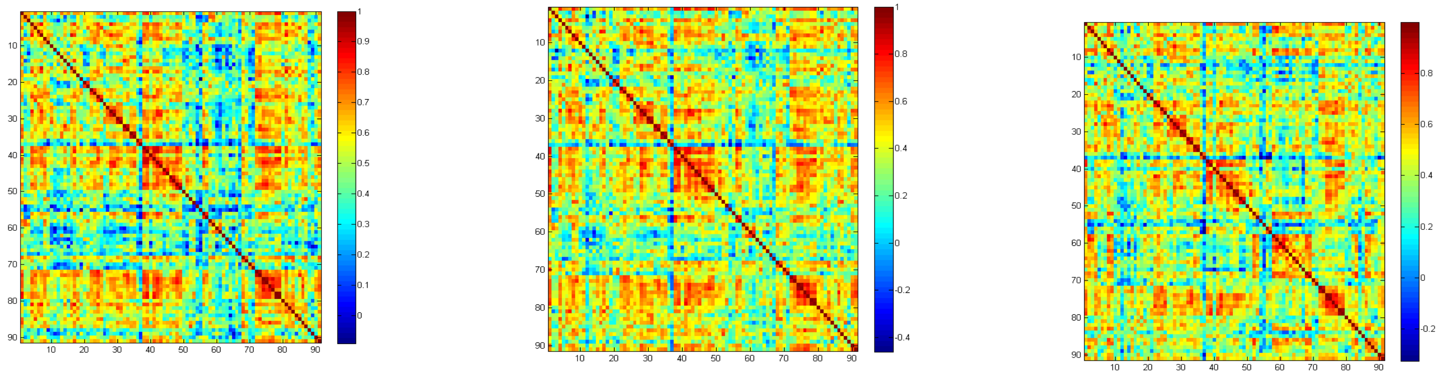






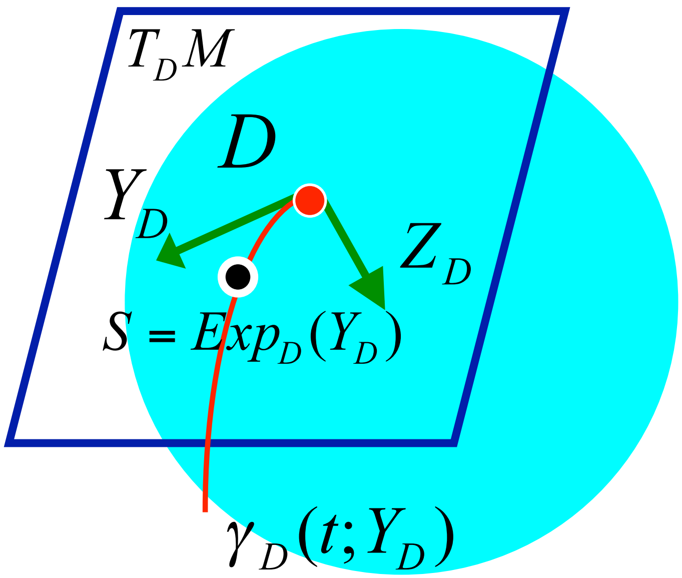


Styner, M. (2008).



Local Polynomial Kernel Regression for SPD

$$M = Sym(m)^+$$



Inner product $<< Y_D, Z_D >>$

Geodesic

Riemannian exponential/logarithm maps

- **Affine invariant metric**

$$<< Y_D, Z_D >>_{D,R} = \mathrm{tr}(Y_D D^{-1} Z_D D^{-1})$$

- **Log-Euclidean metric**

$$<< Y_D, Z_D >>_{D,L} = \mathrm{tr}(R_D(Y_D) R_D(Z_D))$$

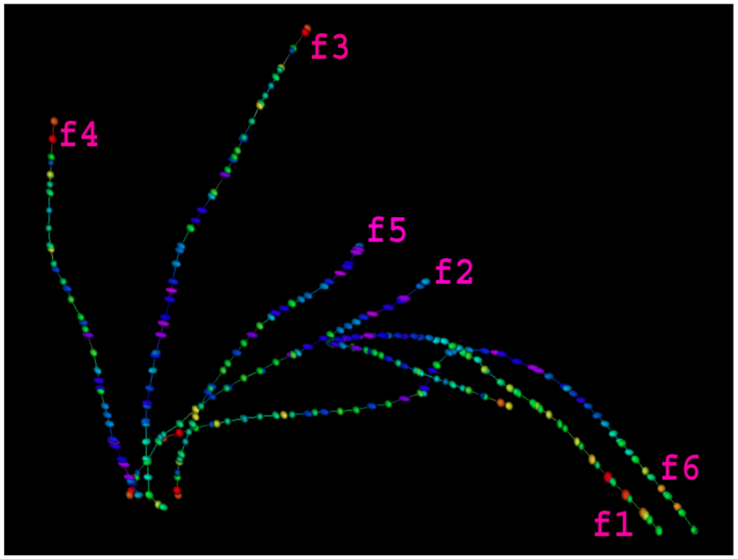
$$R_D : T_D M \to T_{I_m} M$$

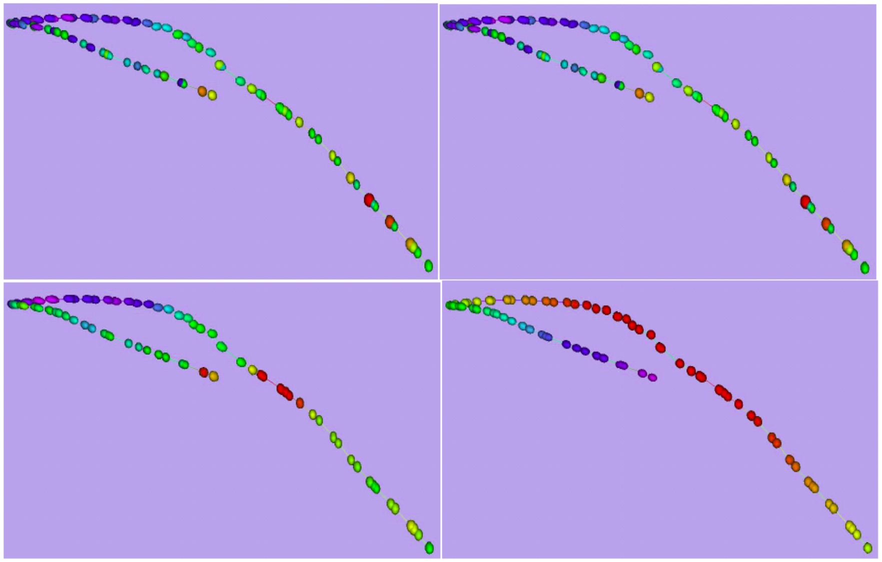
Dryden et al. (2009)
Ying, Zhu, Lin and Marron (2010)

Local Polynomial Kernel Regression for SPD

Questions:

- How to define local polynomial kernel regression to nonparametrically estimate an intrinsic mean of S given x?
- Whether local linear regression performs better than local constant regression?
- How much statistical inferences depend on a specific inner product defined on the tangent space?





Local Polynomial Kernel Regression for SPD

Conditional Expectation

$$D(x) = E[S \mid X = x]$$

$$E[e(x) \mid X = x] = E[S - D(x) \mid X = x] = 0$$

Intrinsic Conditional Expectation

$$e_{D(x)} = \mathrm{Log}_{D(x)}(S) \in T_{D(x)} Sym^{+}(m)$$

$$\mathrm{E}[e_{D(x)} \mid X = x] = \mathrm{E}[\log_{D(x)}(S) \mid X = x] = 0$$

Local Polynomial Kernel Regression for SPD

Local Polynomial Kernel Regression

$$\mathrm{Log}_{D(x_0)}(D(x)) \in T_{D(x_0)} Sym^+(m)$$

$$\phi_{D(x_0)}(.) : T_{D(x_0)} Sym^+(m) \rightarrow T_{I_m} Sym^+(m)$$

$$Y(x) = \phi_{D(x_0)}(\mathrm{Log}_{D(x_0)}(D(x)))$$

$$\mathrm{Log}_{D(x_0)}(D(x))) = \phi^{-1}_{D(x_0)}(Y(x)) \approx \phi^{-1}_{D(x_0)}(Y(x_0) + \sum_{k=1}^{K} Y^{(k)}(x_0)(x - x_0)^k)$$

$$D(x) = \mathrm{Exp}_{D(x_0)}(\phi^{-1}_{D(x_0)}(Y(x))) \approx \mathrm{Exp}_{D(x_0)}(\phi^{-1}_{D(x_0)}(\sum_{k=1}^{K} Y^{(k)}(x_0)(x - x_0)^k))$$

Local Polynomial Kernel Regression for SPD

Q1: Define Intrinsic LPK Estimator

$$\hat{\alpha}_I(x_0) = \arg\min_{\alpha(x_0)} \sum_{i=1}^{n} K_h(x_i - x_0) d(S_i, \mathrm{Exp}_{D(x_0)}(\phi_{D(x_0)}^{-1}(\sum_{k=1}^{K} Y^{(k)}(x_0)(x - x_0)^k)))^2$$

K=0: **Local constant estimator;** ***K=1***: **Local linear estimator**

Cross-validation

$$\mathrm{CV} = \sum_{i=1}^{n} d(S_i, \hat{D}_I(x_i; h)^{(-i)})^2 \approx \sum_{i=1}^{n} d(S_i, \hat{D}_I(x_i; h))^2 + 2p_n(h)$$

Asymptotic average mean squared error (AMSE)

$$\mathrm{AMSE}(\log(\hat{D}_{IR}(x_0; h, k_0))) = E(tr[\{\log(\hat{D}_{IR}(x_0; h, k_0)) - \log(D(x_0))\}^2] \mid x)$$

Asymptotic average mean integrated squared error (AMISE)

$$\mathrm{AMISE}(\log(\hat{D}_{IR}(h, k_0))) = \int \mathrm{AMISE}(\log(\hat{D}_{IR}(x; h, k_0)))\omega(x)dx$$

Local Polynomial Kernel Regression for SPD

Log-Euclidean metric

$$<< U_{D(x)}, V_{D(x)} >>_{D(x),L} = \mathrm{tr}(R_{D(x)}(U_{D(x)})R_{D(x)}(V_{D(x)}))$$

Let

$$Y(x) = \phi_{D(x_0),L}(\mathrm{Log}_{D(x_0)}(D(x))) = \log(D(x)) - \log(D(x_0))$$

$$D(x) = \exp(\log(D(x_0)) + Y(x))$$

Intrinsic Mean

$$E[\log(S) - \log(D(x)) \mid X = x] = O_m$$

Geodesic Distance

$$d(D(x),S)^2 = \mathrm{tr}\{[\log(D(x)) - \log(S)]^2\}$$

Local Polynomial Kernel Regression for SPD

Q2: Intrinsic local linear is better than intrinsic local constant

$$\mathrm{AMSE}(\log(\hat{D}_{IL}(x_0;h,0))) = h^4 u_2^2 \mathrm{tr}([\mathrm{vecs}\{0.5\times\log(\mathrm{D}(x_0))^{(2)} + f_X^{(1)}(x_0) f_X(x_0)^{-1}\log(\mathrm{D}(x_0))^{(1)}\}]^{\otimes 2})$$
$$+ v_0 (nhf_X(x_0))^{-1}\mathrm{tr}(\Sigma_{\varepsilon\mathrm{D}}(x_0)) + o(h^4 + (nh)^{-1})$$

Optimal bandwidth

$$h_{opt,L}(x_0;0) = \left[\frac{n^{-1} v_0 f_X^{-1}(x_0)\mathrm{tr}(\Sigma_{\varepsilon\mathrm{D}}(x_0))}{4u_2^2\mathrm{tr}([\mathrm{vecs}\{0.5\log(\mathrm{D}(x_0))^{(2)} + f_X^{(1)}(x_0) f_X(x_0)^{-1}\log(\mathrm{D}(x_0))^{(1)}\}]^{\otimes 2})}\right]^{1/5}$$

$$\mathrm{AMSE}(\log(\hat{D}_{IL}(x_0;h,1))) = h^4 u_2^2 \mathrm{tr}([\mathrm{vecs}\{0.5\times\log(\mathrm{D}(x_0))^{(2)}\}]^{\otimes 2}) + v_0(nhf_X(x_0))^{-1}\mathrm{tr}(\Sigma_{\varepsilon\mathrm{D}}(x_0))$$
$$+ o(h^4 + (nh)^{-1})$$

Optimal bandwidth

$$h_{opt,L}(x_0;1) = \left[\frac{n^{-1} v_0 f_X^{-1}(x_0)\mathrm{tr}(\Sigma_{\varepsilon\mathrm{D}}(x_0))}{u_2^2\mathrm{tr}([\mathrm{vecs}\{\log(\mathrm{D}(x_0))^{(2)}\}]^{\otimes 2})}\right]^{1/5}$$

Ratio of AMSEs

$$\frac{\mathrm{AMSE}(\log(\hat{D}_{IL}(x_0;h,0)))}{\mathrm{AMSE}(\log(\hat{D}_{IL}(x_0;h,1)))} = \frac{\mathrm{tr}([\mathrm{vecs}\{0.5\log(\mathrm{D}(x_0))^{(2)} + f_X^{(1)}(x_0) f_X(x_0)^{-1}\log(\mathrm{D}(x_0))^{(1)}\}]^{\otimes 2})}{\mathrm{tr}([\mathrm{vecs}\{0.5\log(\mathrm{D}(x_0))^{(2)}\}]^{\otimes 2})}$$

Local Polynomial Kernel Regression for SPD

Affine invariant metric

$$<< U_{D(x)}, V_{D(x)} >>_{D(x),R} = \mathrm{tr}(U_{D(x)} D(x)^{-1} V_{D(x)} D(x)^{-1})$$

Let $D(x) = G(x)G(x)^T$

$$Y(x) = \phi_{D(x_0),R}(\mathrm{Log}_{D(x_0)}(D(x))) = \log(G(x_0)^{-1} D(x) G(x_0)^{-1})$$

$$D(x) = G(x_0)\exp(Y(x))G(x_0)^T$$

Intrinsic Mean

$$E[\log(G(x)^{-1} S G(x)^{-1}) \mid X = x] = O_m$$

Geodesic distance

$$d(D(x), S)^2 = \mathrm{tr}\{\log^2(G(x)^{-1} S G(x)^{-1})\}$$

Local Polynomial Kernel Regression for SPD

Q2: Intrinsic local linear is better than intrinsic local constant

$$AMSE(\log(\hat{D}_{IR}(x_0;h,0))) = h^4 u_2^2 \mathrm{tr}([\mathrm{G_D}(x_0)^T \mathrm{vecs}\{\mathrm{G}^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5\mathrm{G}^{(2)}(x_0)\}]^{\otimes 2})$$
$$+ (nh)^{-1} \mathrm{tr}(\mathrm{G_D}(x_0)^{\otimes 2} \Omega_0(x_0)) + o(h^4 + (nh)^{-1})$$

Optimal bandwidth $h_{opt,R}(x_0;0) = \left[\dfrac{n^{-1}\mathrm{tr}(G_{\mathrm{D}}(x_0)^{\otimes 2}\Omega_0(x_0))}{4u_2^2 \mathrm{tr}([\mathrm{G_D}(x_0)^T \mathrm{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5G^{(2)}(x_0)\}]^{\otimes 2})} \right]^{1/5}$

$$AMSE(\log(\hat{D}_{IR}(x_0;h,1))) = 0.25h^4 u_2^2 \mathrm{tr}([\mathrm{G_D}(x_0)^T \Psi_1(x_0)^{-1}\Psi_2(x_0)^T \mathrm{vecs}\{\mathrm{Y}^{(2)}(x_0)\}]^{\otimes 2})$$
$$+ (nh)^{-1} \mathrm{tr}(\mathrm{G_D}(x_0)^{\otimes 2} \Omega_0(x_0)) + o(h^4 + (nh)^{-1})$$

Optimal bandwidth $h_{opt,R}(x_0;1) = \left[\dfrac{n^{-1}\mathrm{tr}(G_{\mathrm{D}}(x_0)^{\otimes 2}\Omega_0(x_0))}{4u_2^2 \mathrm{tr}([\mathrm{G_D}(x_0)^T \Psi_1(x_0)^{-1}\Psi_2(x_0)^T \mathrm{vecs}\{Y^{(2)}(x_0)\}]^{\otimes 2})} \right]^{1/5}$

Ratio of AMSEs

$$\frac{\mathrm{AMSE}(\log(\hat{D}_{IR}(x_0;h,0)))}{\mathrm{AMSE}(\log(\hat{D}_{IR}(x_0;h,1)))} = \frac{\mathrm{tr}([G_D(x_0)^T \mathrm{vecs}\{G^{(1)}(x_0) f_X^{(1)}(x_0) f_X(x_0)^{-1} + 0.5G^{(2)}(x_0)\}]^{\otimes 2})}{\mathrm{tr}([G_D(x_0)^T \Psi_1(x_0)^{-1}\Psi_2(x_0)^T \mathrm{vecs}\{Y^{(2)}(x_0)\}]^{\otimes 2})}$$

Local Polynomial Kernel Regression for SPD

Q3: Affine Invariant Metric versus Log-Euclidean Metric

$$\mathrm{rMSE}(R,L;0)=\frac{\mathrm{AMSE}(\log(\hat{D}_{IR}(x_0;h,0)))}{\mathrm{AMSE}(\log(\hat{D}_{IL}(x_0;h,0)))}=\left[\frac{\mathrm{tr}\{G_D(x_0)^{\otimes 2}\Psi_1(x_0)^{-1}\Psi_{11}(x_0)\Psi_1(x_0)^{-1}\}}{tr\{\Sigma_{eD}(x_0)\}}\right]^{4/5}$$

$$\times\left[\frac{\mathrm{tr}([G_D(x_0)^T\,\mathrm{vecs}\{f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0)+0.5G^{(2)}(x_0)\}]^{\otimes 2})}{\mathrm{tr}([\mathrm{vecs}\{0.5\log(\mathrm{D}(x_0))^{(2)}+f_X^{(1)}(x_0)f_X(x_0)^{-1}\log(\mathrm{D}(x_0))^{(1)}\}]^{\otimes 2})}\right]^{1/5}$$

Let m=1, $D(x_0)=G(x_0)^2$ with $G(x_0)>0$

$$\mathrm{rMSE}(R,L;0)=\left[\frac{\mathrm{tr}([f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0)+0.5G^{(2)}(x_0)]^{\otimes 2})}{\mathrm{tr}([f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0)+0.5G^{(2)}(x_0)-0.5G^{(1)}(x_0)^2G(x_0)^{-1}]^{\otimes 2})}\right]^{1/5}$$

$$f_X^{(1)}(x_0)f_X(x_0)^{-1}G^{(1)}(x_0)+0.5G^{(2)}(x_0)>0.25G^{(1)}(x_0)^2G(x_0)^{-1}\Leftrightarrow \mathrm{rMSE}(R,L;0)>1$$

It depends on both design density and $D(x_0)=G(x_0)^2$.

Local Polynomial Kernel Regression for SPD

Q3: Affine Invariant Metric versus Log-Euclidean Metric

$$\mathrm{rMSE}(R,L;1) = \left[\frac{tr\{G_D(x_0)^{\otimes 2}\Omega_0(x_0)\}}{tr\{\Sigma_{eD}(x_0)\}} \right]^{4/5}$$

$$\times \left[\frac{\mathrm{tr}([G_D(x_0)^T \Psi_1(x_0)^{-1}\Psi_2(x_0)^T \,\mathrm{vecs}(\mathrm{Y}^{(2)}(x_0))\}^{\otimes 2})}{\mathrm{tr}(\mathrm{vecs}\{\log(\mathrm{D}(x_0))^{(2)}\}^{\otimes 2})} \right]^{1/5}$$

Let m=1, $D(x_0) = G(x_0)^2 \quad \text{with} \quad G(x_0) > 0$

$$\mathrm{rMSE}(R,L;1) = \frac{\mathrm{AMSE}(\log(\hat{D}_{IR}(x_0;h,1)))}{\mathrm{AMSE}(\log(\hat{D}_{IL}(x_0;h,1)))} = 1$$

Regression Models for SPD

Simulation Studies

Data model

$$S_i = C(x_i)\exp(E_i)C(x_i), \qquad E_i \sim MN(0,\Omega)$$

$$D(x) = C(x)^2 \qquad x_i \sim N(0,0.25)$$

$$C(x) = \begin{pmatrix} -0.1x & 0.2x & \sin(x) \\ 0.2x & 0.6x & -0.4x \\ \sin(x) & -0.4x & 0.5x \end{pmatrix}$$

Covariance

$$\Sigma_1 = \begin{pmatrix} 0.3 & 0.049 & 0.052 \\ 0.049 & 0.2 & 0.0424 \\ 0.052 & 0.0424 & 0.1 \end{pmatrix} \qquad \Sigma_2 = 2\Sigma_1, \quad \Sigma_3 = 4\Sigma_1, \quad \Sigma_4 = 8\Sigma_1$$

Data $\{(x_i, S_i) : i = 1, \cdots, n\}$ for $n = 50$ or 100

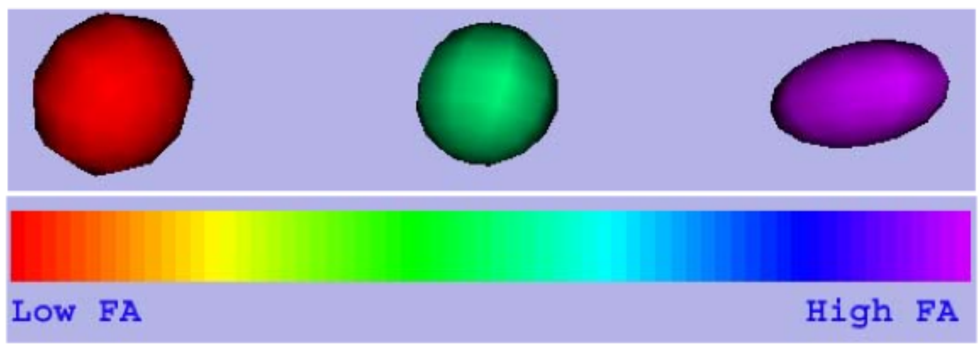
Local Polynomial Kernel Regression for SPD

$$3 \times 3 \quad SPD \qquad D(x) > 0 \qquad \lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$$

$$FA = \sqrt{\frac{3\{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2\}}{2(\lambda_1{}^2 + \lambda_2{}^2 + \lambda_3{}^2)}}$$

FA: a scalar quantity derived from diffusion tensor (SPD matrix).

- Low FA values: isotropic diffusion.
- High FA values: highly directional diffusion.



Local Polynomial Kernel Regression for SPD

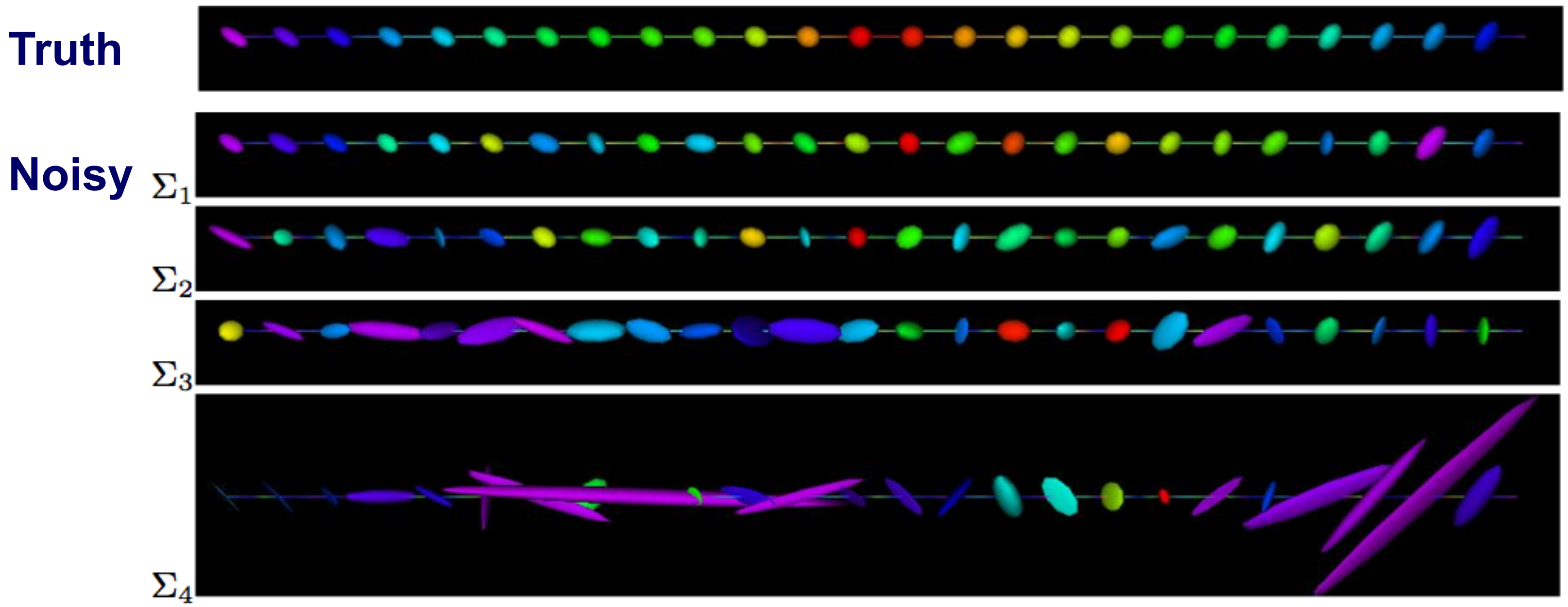


Fig. 1. Ellipsoidal representations of the true (the first row) and simulated SPD matrix data along the design points under the four different noise distributions (the second to the fifth rows: Σ_1-Σ_4) colored with FA values.

Local Polynomial Kernel Regression for SPD

Truth

Log-Euclidean Local Linear

Riemannian Local Linear

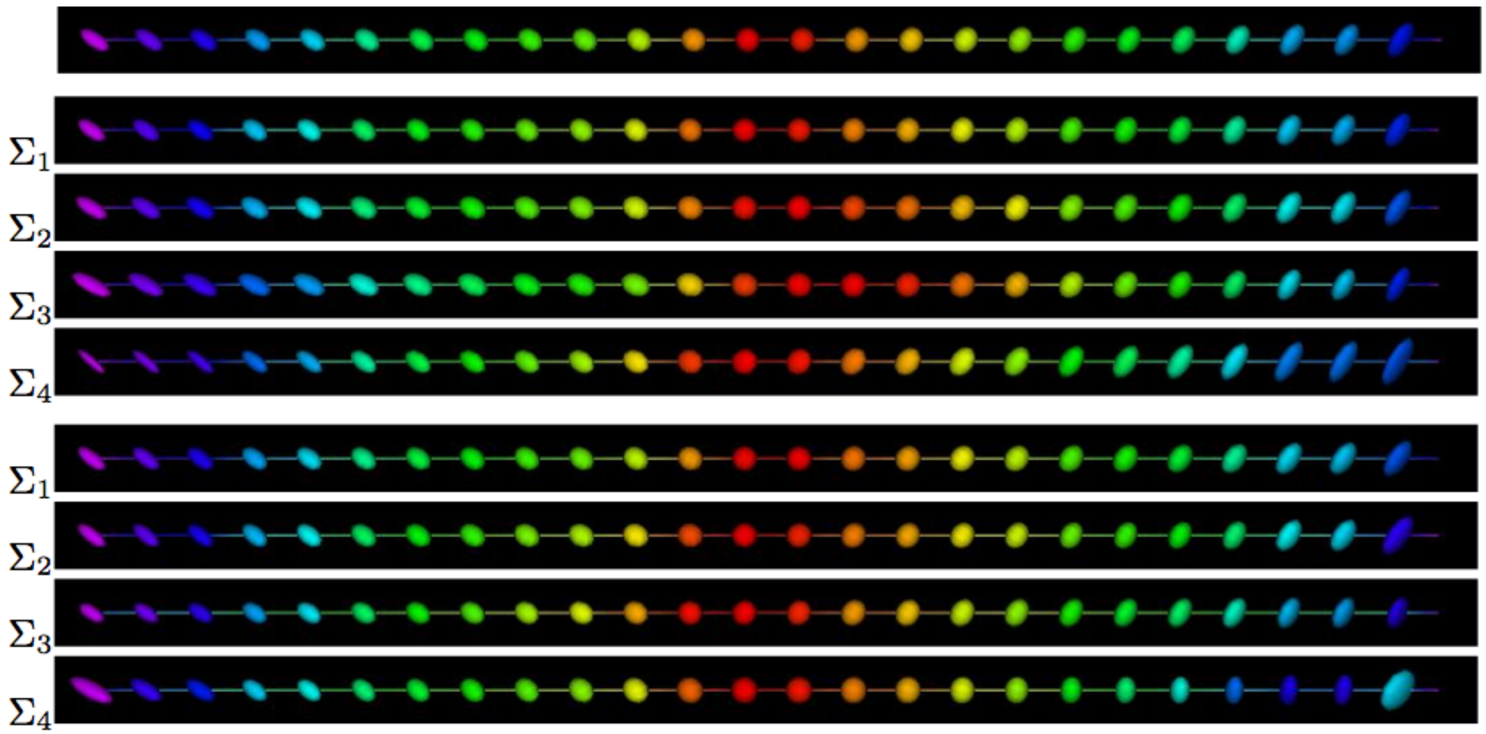


Fig. 2. Ellipsoidal representations of the true (the first row) and estimated SPD matrix data along the design points under the four different noise levels colored with FA values. The second to the fifth rows (Log-Euclidean metric): Σ_1-Σ_4, the sixth to the ninth rows (the Riemannian metric): Σ_1-Σ_4.

Local Polynomial Kernel Regression for SPD

Simulation 1.

- **Compare the performance of the local linear with the local constant**

- **Assess the performance using the Average Geodesic Distance (AGD) for each replication j=1, … , N with N as the number of replications, denoted by**

$$\text{AGD} = (nN)^{-1}\sum_{j=1}^{N}\sum_{i=1}^{n} d(\hat{D}_j(x_i), D(x_i))$$

where $\hat{D}_j(x_i)$ and $D(x_i)$ **are, respectively, the estimated and true diffusion tensors at** x_i

Local Polynomial Kernel Regression for SPD

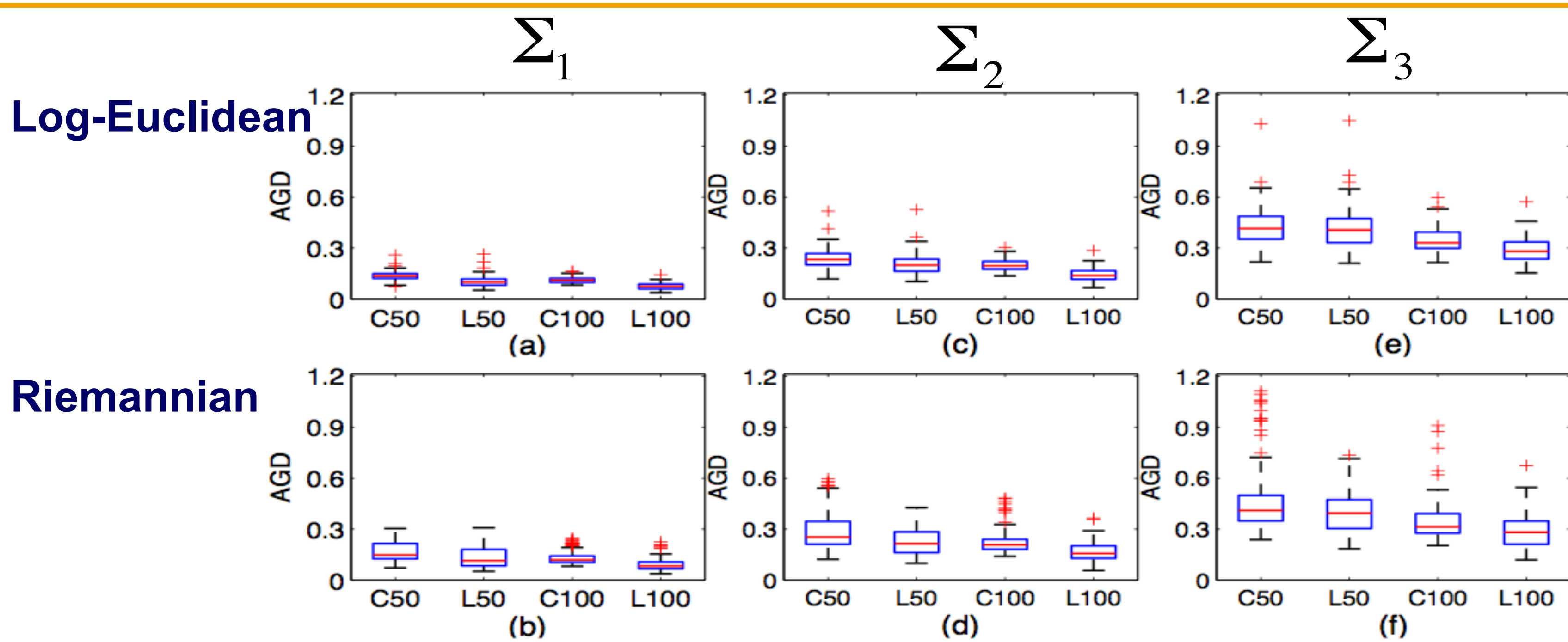


Fig. 3. Boxplots of the AGD using the intrinsic local constant and linear estimators under the log-Euclidean (the first row) and Riemannian (the second row) metrics based on 100 replications under the three covariance matrices (a)-(b) Σ_1, (c)-(d) Σ_2, and (e)-(f) Σ_3. C50 and C100 represent the intrinsic local constant estimators at sample sizes 50 and 100, respectively. L50 and L100 represent the intrinsic local linear estimators at sample sizes 50 and 100, respectively.

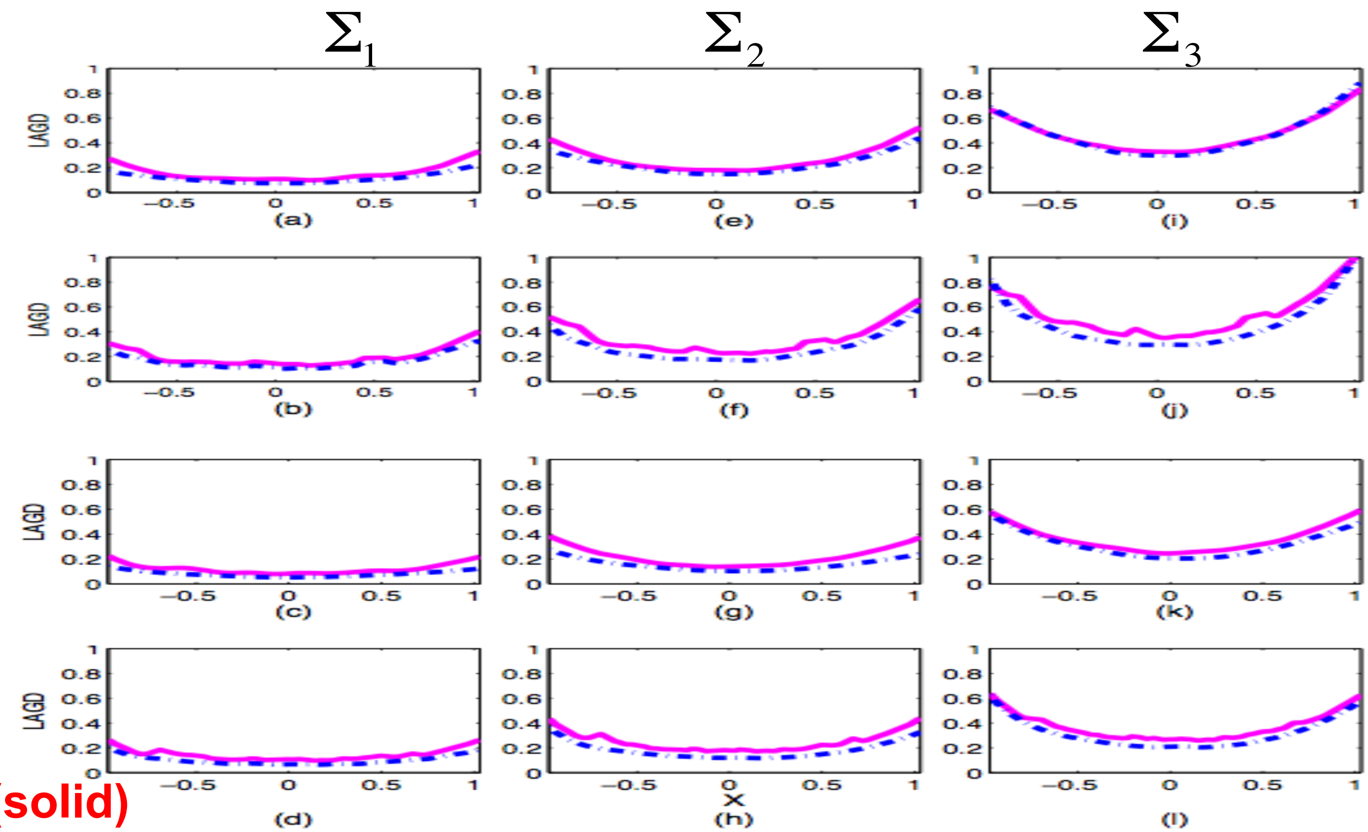
Local Polynomial Kernel Regression for SPD

Log-Euclidean

Riemannian

Log-Euclidean

Riemannian



Local constant (solid)

Local linear (dashed)

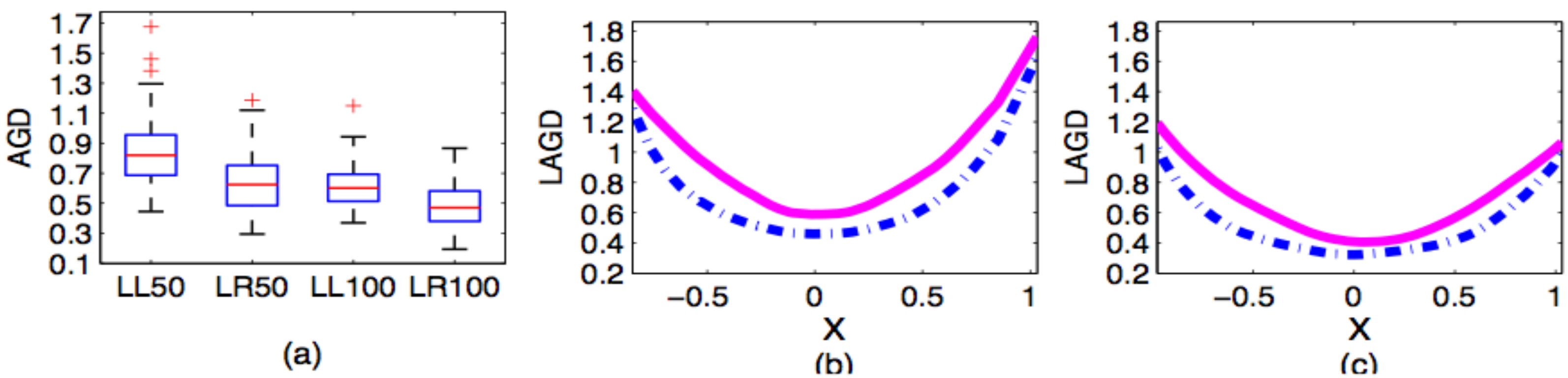
Fig. 4. The LAGD curves at each sample point using the intrinsic local constant (solid line) and linear (dash-dotted line) estimators under the three covariance matrices (a)-(d) Σ_1, (e)-(h) Σ_2, (i)-(l) Σ_3 for sample sizes 50 (the top two rows) and 100 (the bottom two rows). The first and third rows correspond to the log-Euclidean metric while the second and fourth rows correspond to the Riemannian metric.

Local Polynomial Kernel Regression for SPD

Simulation 2. High noisy level

Compare the performance of the local linear under two metrics

Σ_4



Riemannian (solid) Log-Euclidean (dashed)

Fig. 5. (a) Boxplots of the AGD's using the linear regressions based on 100 replications under the covariance matrix Σ_4, under the Log-Euclidean and Riemannian metrics, respectively. (b) and (c) LAGD curves at each sample point using the local linear regressions under the affine invariant (dash-dotted line) and Log-Euclidean (solid line) metrics under the the covariance matrix Σ_4 at sample size 50 (b) and 100 (c), respectively. LL50 (LR50) and LL100 (LR100), respectively, represent the local linear regressions under Log-Euclidean (Riemannian) metrics at sample sizes 50 and 100.

Local Polynomial Kernel Regression for SPD

Simulation 3.

- **Value of developing the LPK smoothing method**

- **Two different methods for smoothing FA values**

M1. Calculate FA values from `noisy' SPDs and then use the local linear method to smooth the FA values

M2. Use the local linear method to smooth SPDs and then calculate FA values from the smoothed SPDs

- **Calculate the Mean Absolute Deviation Error (MADE):**

$$\text{MADE} = (nN)^{-1} \sum_{j=1}^{N} \sum_{i=1}^{n} | \hat{\text{FA}}_j(x_i) - \text{FA}_j(x_i) |$$

Local Polynomial Kernel Regression for SPD

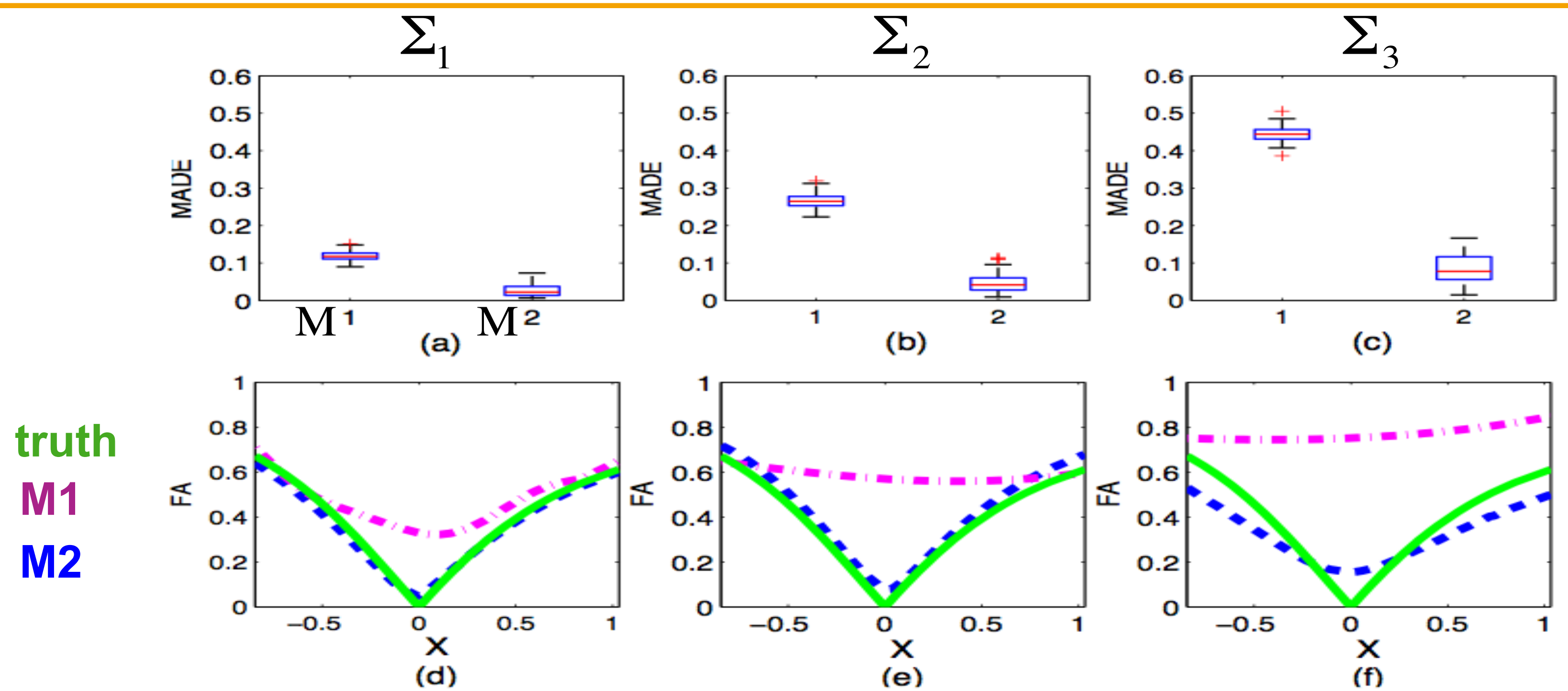


Fig. 6. Boxplot of the MADE's using the two smoothing methods based on 100 replications under the covariance matrices (a) Σ_1, (b) Σ_2, and (c)Σ_3 at sample size 50. Smoothed FA curves for the realizations with median MADE under the covariance matrices: (d) Σ_1, (e) Σ_2, and (f) Σ_3. The true FA curve (the solid line), the estimated FA curve using the first method (the dash-dotted line) and the estimated FA curve using the second method (the dashed line). This shows that the more intrinsic approach is much better.

Local Polynomial Kernel Regression for SPD

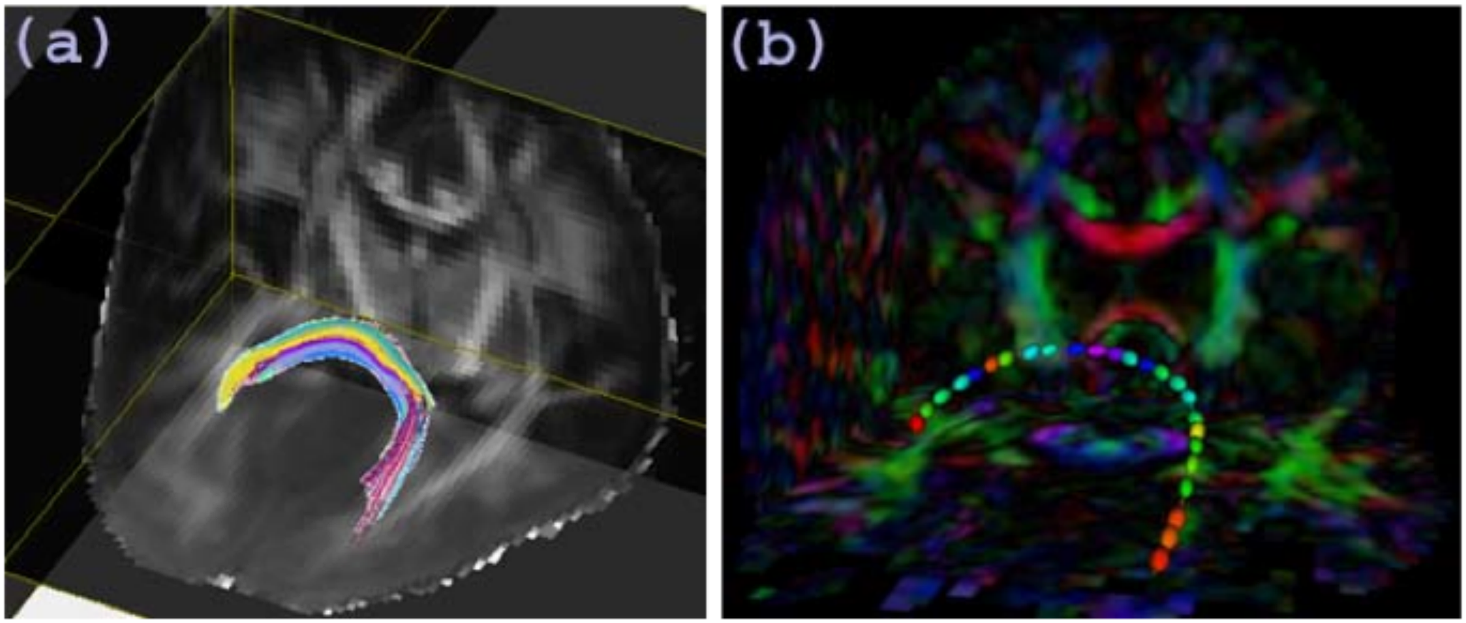


Fig. 7. (a)The splenium of the corpus callosum in the analysis of HIV DTI data. (b)The ellipsoidal representation of full tensors on the fiber tract from a selected subject.

Local Polynomial Kernel Regression for SPD

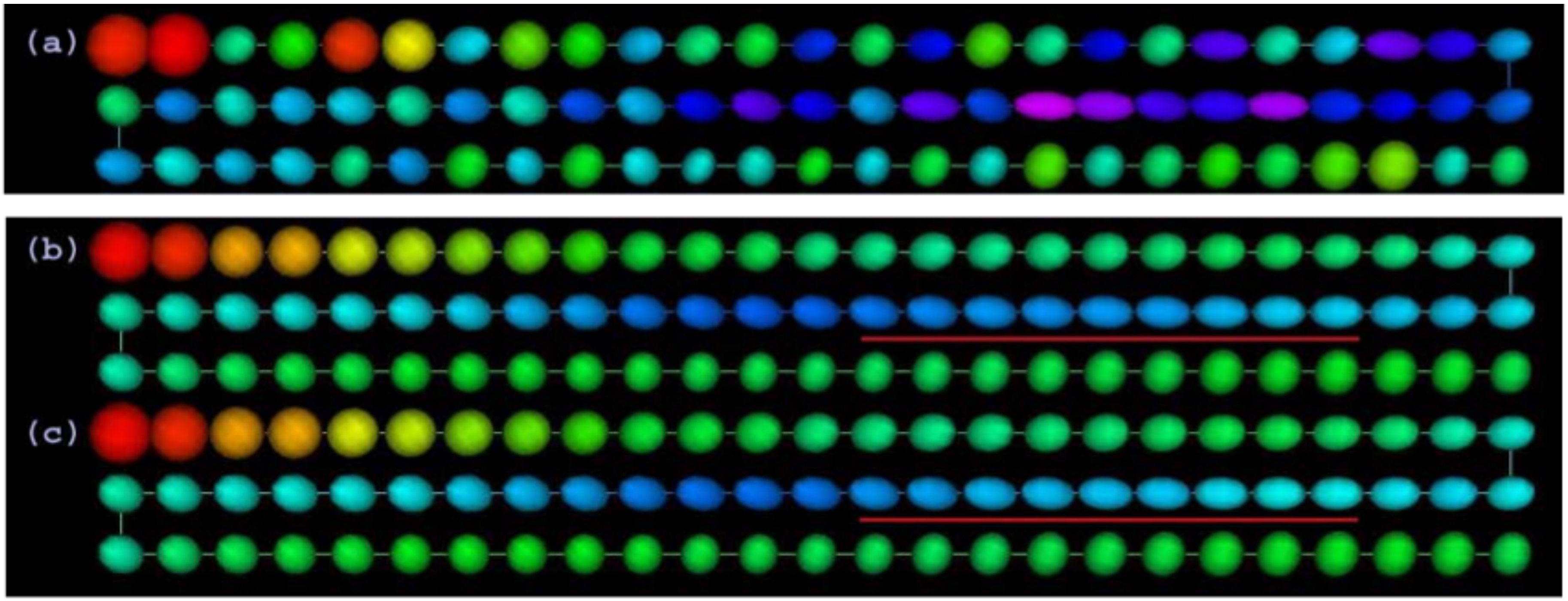


Fig. 8. (a) Ellipsoidal representations of the diffusion tensor data and estimated tensors using the intrinsic local linear regression under the (b)log-Euclidean and (c) Riemannian metrics along the fiber tract f1 colored with FA values. The estimated tensors in the middle right part (highlighted in the red line) are more anisotropic using the method under the Log-Euclidean metric.

Local Polynomial Kernel Regression for SPD

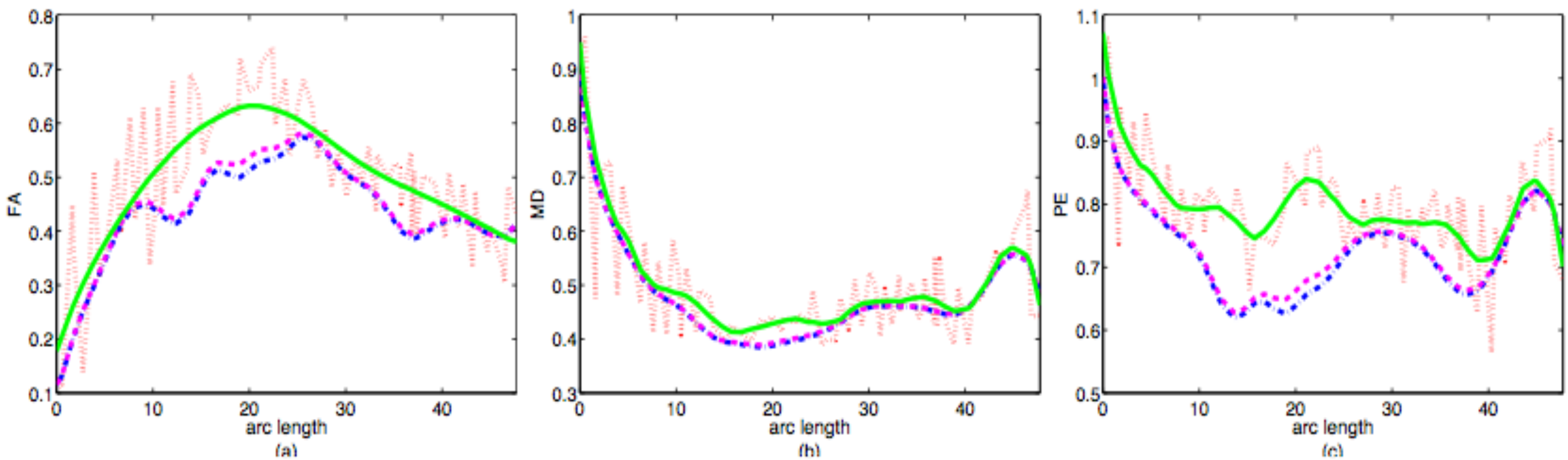


Fig. 9. (a) FA's , (b) MD's and (c) PE's derived from the raw tensor data (dot line) and estimated tensors using the intrinsic local linear regression under the Riemannian (dash-dot line) and log-Euclidean (dash line) metrics as the function of arc-length along the tract f1. Estimated FA function along the fiber tract f1 by using the standard local linear regression for scalars (solid line).

Local Polynomial Kernel Regression for SPD

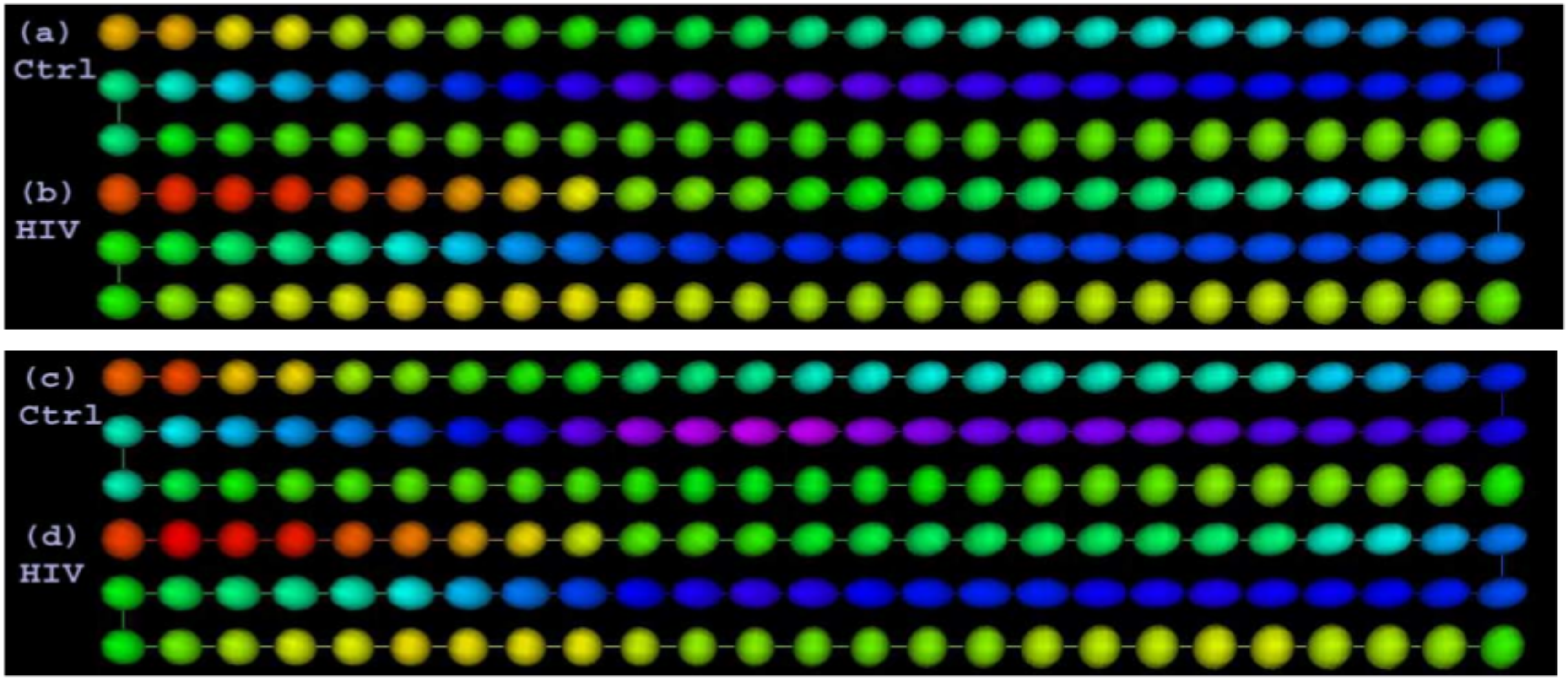


Fig. 10. Ellipsoidal representations of estimated mean tensors along the fiber tract f1 for the control and HIV groups using the intrinsic local linear regression under the log-Euclidean ((a) and (b)) and Riemannian ((c) and (d)) metrics colored with FA values.

Local Polynomial Kernel Regression for SPD

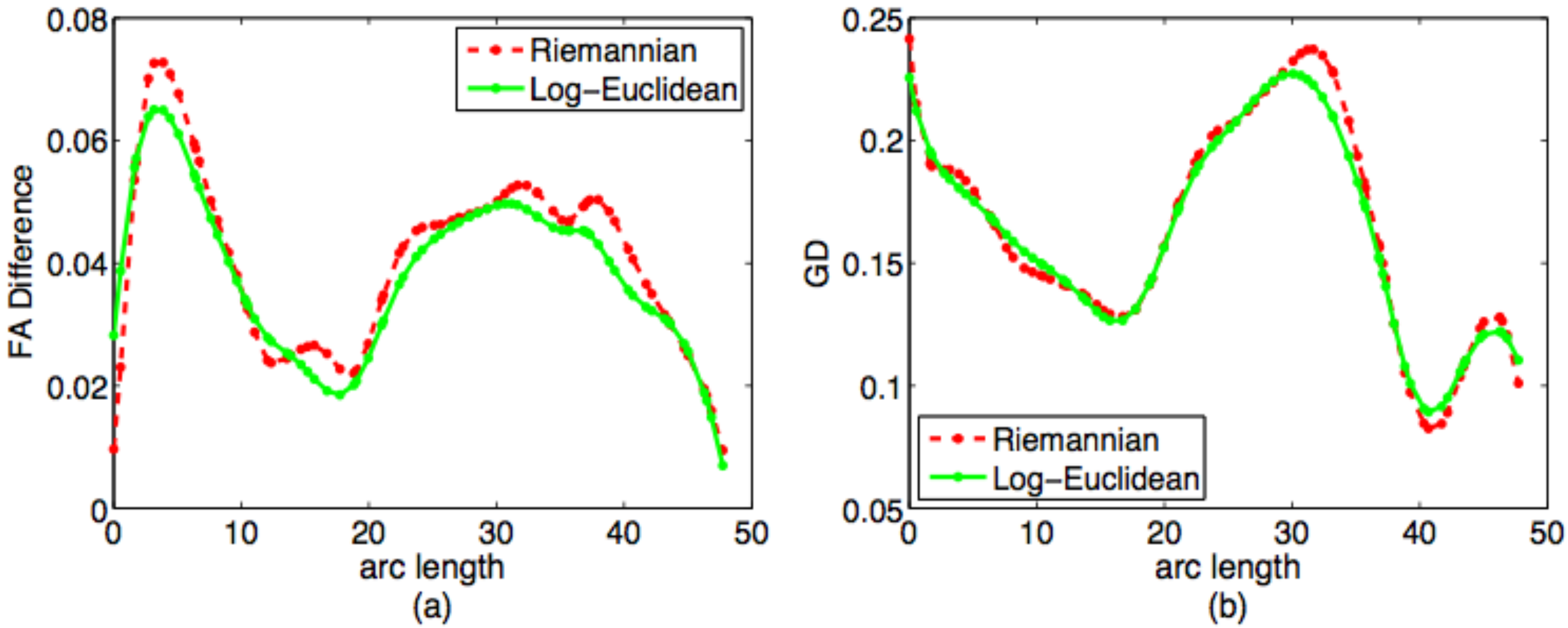


Fig. 11. (a) FA differences and (b) geodesic distances between pairs of mean diffusion tensors of HIV and control groups along the fiber f1 under the Log-Euclidean (the solid line) and Riemannian (the dashed line) metrics.

Manifold Data

- How to characterize the `variation' in the manifold data and use such information for statistical analysis?

- For a specific type of manifold data, is it possible to define an optimal geometrical structure that can capture the most important target information from the data at hand?

- How to efficiently analyze discrete time and continuous time manifold data?

Future Work

- Minimax efficiency for the local linear method for SPD
- Statistical models for Lie group
- Statistical models of correlated Manifold-valued data
- Statistical models for deformation field
- Statistical models for spherical needlets

References

Zhu, H.T., Cheng, Y.S., Ibrahim, J.G., Li, Y. M., Colin, P., Lin, W.L. (2009). Regression models for positive definite matrices, *Journal of the American Statistical Association,* 104, 1203-1212, 2009.

Yuan, Y., Zhu, H.T., Lin, W, L., and J. S. Marron (2010). Local polynomial regression for SPD matrices. In submission.

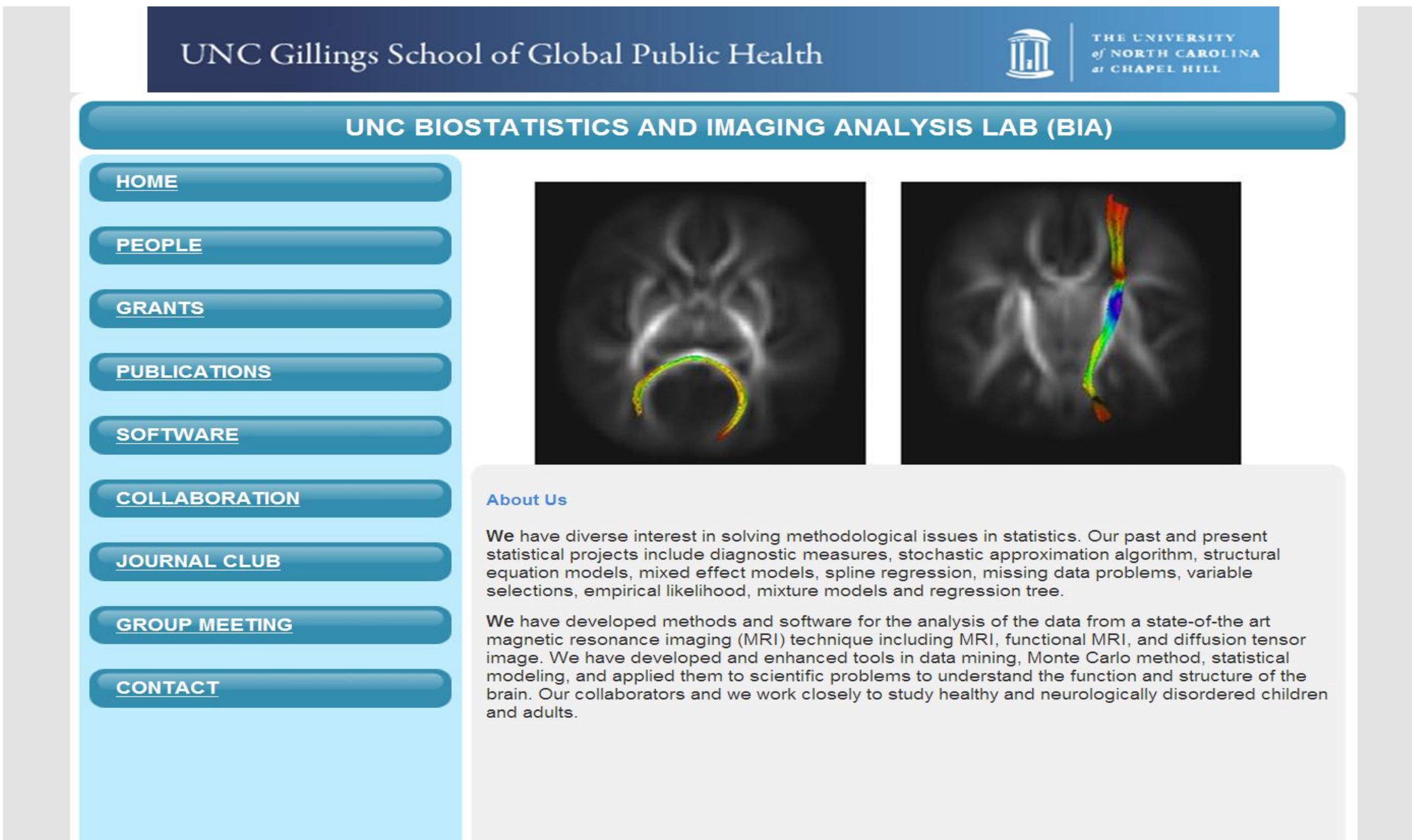
Yuan, Y., Zhu, H.T., Gilmore, J. H., Styner, M., and J. S. Marron (2010). Multivariate varying coefficient model for SPD matrices: log-euclidean metric. In submission.

Yuan, Y., Zhu, H.T., Gilmore, J. H., Styner, M., and J. S. Marron (2011). Multivariate varying coefficient model for SPD matrices: Riemannian metric. In preparation.

Yuan, Y. (2011). Statistical Analysis of SPD matrics. Ph.D. dissertation, UNC-Statistics.

UNC Biostatistics and Imaging Analysis Lab (BIAS)

http://www.bios.unc.edu/research/bia



Acknowledgements