Hongtu Zhu, Ph.D and Professor
Departments of Biostatistics, Statistics, Computer Science, and Genetics, Biomedical Research Imaging Center, University of North Carolina at Chapel Hill

Acknowledgement: Multiple slides were copied from Hans Muller, Horvath, and Kokoszka.
Reading Materials

3. Pietrosanu, Matthew; Shu, Haoxu; Jiang, Bei; Kong, Linglong; Heo, Giseon; He, Qianchuan; Gilmore, John; Zhu, Hongtu Estimation for the bivariate quantile varying coefficient model with application to diffusion tensor imaging data analysis. *Biostatistics*, in press, 2021.
Key Features
Key Features

- Infinite Dimension
- Spatial Smoothness
- Spatial Correlation
- Spatial Heterogeneity
**Infinite Dimensional Image**

**Mathematics.**

Image is the point or set of points in the range corresponding to a designated point in the domain of a given function.

- $\Omega$ is a compact set. $\tilde{x} \in \Omega \subseteq \mathbb{R}^k$

$\quad f(\tilde{x}) \in M \subseteq \mathbb{R}^m$

$f : \Omega \rightarrow M \subseteq \mathbb{R}^m$

$\star \quad \int \|f(\tilde{x})\|^k d\tilde{x} < \infty \text{ for some } k > 0$

*The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL*
Spatial Smoothness

Cartoon Model

• Disjoint Partition

\[ D = \bigcup_{l=1}^{L} D_l \quad \text{and} \quad D_l \cap D_{l'} = \emptyset \]

• Piecewise Smoothness: Lipschitz condition

• Smoothed Boundary

• Local Patch

• Degree of Jumps
$f \in F^{\text{cartoon}}$ is a “cartoon image” if it is a piecewise smooth (Hölder-$\alpha$, $\alpha \geq 1$) image with discontinuities along smooth hypersurfaces.\textsuperscript{2}

$$f(x) = 1_{\{x \in \Omega\}} f_\Omega(x) + 1_{\{x \in \Omega^c\}} f_{\Omega^c}(x),$$
Spatial Correlation

Long-range Correlation

Short-range Correlation

“Unmodelled effects”

“Signal Processing”
Random Elements and Covariance Operator
Functional Data Design

- Fully observed functions without noise at arbitrarily dense grid
  Measurements $Y_{it} = X_i(t)$ available for all $t \in T$, $i = 1, \ldots, n$:
  Often unrealistic but mathematically convenient

- Dense design with noisy measurements
  Measurements $Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij}$, where $T_{ij}$ are recorded on a regular grid, $T_{i1}, \ldots, T_{iN_i}$, and $N_i \to \infty$:
  Applies to typical functional data

- Sparse design with noisy measurements = Longitudinal data
  Measurements $Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij}$, where $T_{ij}$ are random times and their number $N_i$ per subject is random and finite.
Smoothed Functional Data

Covariates (e.g., age, gender, diagnostic)
DTI Fiber Tract Data

- Diffusion properties (e.g., FA, RA)
  \[ Y_i(s_j) = (y_{i,1}(s_j), \cdots, y_{i,m}(s_j))^T \]
- Grids \( \{s_1, \cdots, s_{n_G}\} \)
- Covariates (e.g., age, gender, diagnostic)
  \( x_1, \cdots, x_n \)
The Space $L^2([0,1])$

The space $L^2 = L^2([0, 1])$ is the set of measurable real-valued functions $x$ defined on $[0, 1]$ satisfying $\int_0^1 x^2(t)dt < \infty$. The space $L^2$ is a separable Hilbert space with the inner product

$$\langle x, y \rangle = \int x(t)y(t)dt.$$ 

An important class of operators in $L^2$ are the integral operators defined by

$$\Psi(x)(t) = \int \psi(t, s)x(s)ds, \quad x \in L^2,$$

Such operators are Hilbert–Schmidt if and only if

$$\|\Psi\|_S^2 = \iint \psi^2(t, s)dt\,ds < \infty$$

If $\psi(s, t) = \psi(t, s)$ and $\iint \psi(t, s)x(t)x(s)dt\,ds \geq 0$,

$$\psi(t, s) = \sum_{j=1}^{\infty} \lambda_j v_j(t)v_j(s) \quad \text{in } L^2([0, 1] \times [0, 1]).$$
Random Elements

A random curve $X = \{X(t), \ t \in [0, 1]\}$ as a random element of $L^2$

$X$ is integrable if $E\|X\| = E[\int X^2(t)dt]^{1/2} < \infty$

$E \langle y, X \rangle = \langle y, \mu \rangle$ for any $y \in L^2$

Mean Function  

$\mu(t) = E[X(t)]$

Covariance Operator  

$C(y) = E[\langle X, y \rangle X], \ y \in L^2.$

$EX = 0 \quad E\|X\|^2 = E \int X^2(t)dt < \infty$

$C(y)(t) = \int c(t, s)y(s)ds, \quad \text{where} \ c(t, s) = E[X(t)X(s)]$

$\iint c(t, s)y(t)y(s)dt\, ds = \iint E[X(t)X(s)]y(t)y(s)dt\, ds = E \left[ \left( \int X(t)y(t)dt \right)^2 \right] \geq 0$
A Necessary and Sufficient Condition

\( C \in \mathcal{L}(L^2) \) is a covariance operator if and only if it is symmetric positive–definite and its eigenvalues satisfy \( \sum_{j=1}^{\infty} \lambda_j < \infty \)

\[
\lambda_j = \langle C v_j, v_j \rangle = \langle E[(X, v_j) X], v_j \rangle = E \left[ \langle X, v_j \rangle^2 \right]
\]

\[
\sum_{j=1}^{\infty} \lambda_j = \sum_{j=1}^{\infty} E \left[ \langle X, v_j \rangle^2 \right] = E \|X\|^2 < \infty.
\]

A counter example: \( \Psi(x) = \sum_{j} \langle x, e_j \rangle j^{-1} e_j. \)
Two Fundamental Theorems

Theorem 2.1. Suppose \( \{X_n, \ n \geq 1\} \) is a sequence of iid mean zero random elements in a separable Hilbert space such that \( E \|X_i\|^2 < \infty \). Then

\[
N^{-1/2} \sum_{n=1}^{N} X_n \xrightarrow{d} Z,
\]

where \( Z \) is a Gaussian random element with the covariance operator

\[
C(x) = E[\langle Z, x \rangle Z] = E[\langle X_1, x \rangle X_1].
\]

Theorem 2.2. Suppose \( \{X_n, \ n \geq 1\} \) is a sequence of iid random elements in a separable Hilbert space such that \( E \|X_i\|^2 < \infty \). Then \( \mu = EX_i \) is uniquely defined by \( \langle \mu, x \rangle = E \langle X, x \rangle \), and

\[
N^{-1} \sum_{n=1}^{N} X_n \xrightarrow{a.s.} \mu.
\]
**Estimation of Mean and Covariance**

**Assumption 2.1.** The observations $X_1, X_2, \ldots X_N$ are iid in $L^2$, and have the same distribution as $X$, which is assumed to be square integrable.

\[
\begin{align*}
\mu(t) &= E[X(t)] \quad \text{(mean function)}; \\
c(t, s) &= E[((X(t) - \mu(t))(X(s) - \mu(s)))] \quad \text{(covariance function)}; \\
C &= E[(X - \mu), \cdot (X - \mu)] \quad \text{(covariance operator)}.
\end{align*}
\]

\[
\begin{align*}
\hat{\mu}(t) &= N^{-1} \sum_{i=1}^{N} X_i(t) \\
\hat{c}(t, s) &= N^{-1} \sum_{i=1}^{N} (X_i(t) - \hat{\mu}(t))(X_i(s) - \hat{\mu}(s)) \, . \\
\hat{C}(x) &= N^{-1} \sum_{i=1}^{N} \langle X_i - \hat{\mu}, x \rangle (X_i - \hat{\mu}) , \quad x \in L^2.
\end{align*}
\]
Asymptotic Properties

**Theorem 2.3.** If Assumption 2.1 holds, then $E \hat{\mu} = \mu$ and $E \| \hat{\mu} - \mu \|^2 = O(N^{-1})$.

**Theorem 2.5.** If $E \| X \|^4 < \infty$, $EX = 0$, and Assumption 2.1 holds, then

$$E \| \hat{C} - C \|^2_S \leq N^{-1} E \| X \|^4.$$

$$Z_N(t, s) = N^{1/2} (\hat{c}(t, s) - c(t, s)),$$

**Theorem 2.9.** If Assumption 2.1 holds with $EX(t) = 0$ and $E \| X \|^4 < \infty$, then $Z_N(t, s)$ converges weakly in $L^2([0, 1] \times [0, 1])$ to a Gaussian process $\Gamma(t, s)$ with $E \Gamma(t, s) = 0$ and

$$E[\Gamma(t, s)\Gamma(t', s')] = E[X(t)X(s)X(t')X(s')] - c(t, s)c(t', s').$$
Functional Principal Components

KARHUNEN-LOÈVE REPRESENTATION USING FPCs

\[ X(t) = \mu(t) + \sum_{k=1}^{\infty} A_k \phi_k(t), \]

where \( A_k = \int_0^T \{X(t) - \mu(t)\} \phi_k(t) dt \), are uncorrelated r.v. with \( EA_k = 0, \ E\lambda_k = \lambda_k \), the functional principal components.

- Parsimonious description of longitudinal/functional data as it is the unique linear representation which explains the highest fraction of variance in the data with a given number of components.
- For modeling functional regression: Functions \( f(X) \) have an equivalent function \( g(A_1, A_2, \ldots) \) so that

\[ f(X) \equiv g(A_1, A_2, \ldots) \]
Derivatives of Curves

\[ X_i^{(\nu)}(t) = \mu^{(\nu)}(t) + \sum_{k=1}^{\infty} A_{ik} \phi_k^{(\nu)}(t), \quad \nu = 0, 1, \ldots \]

- Obtain estimated random effects \( A_{ik} \) by conditioning as before.
- Estimate \( \mu^{(\nu)}(t) \) by known nonparametric 1-d differentiation, applied to pooled scatterplots.
- How to obtain \( \phi_k^{(\nu)} \)? Observe

\[
\frac{d^\nu}{dt^\nu} \int_{\mathcal{T}} G(t, s) \phi_k(s) ds = \lambda_k \frac{d^\nu}{dt^\nu} \phi_k(t),
\]

implying

\[
\phi_k^{(\nu)}(t) = \frac{1}{\lambda_k} \int_{\mathcal{T}} \frac{\partial^\nu}{\partial t^\nu} G(t, s) \phi_k(s) ds.
\]
Decomposition:

\[ y_{i,k}(s) = x_i^T B_k(s) + i,k(s) + i,k(s) \]

**Coefficients**

\[ x_1, \cdots, x_n \]

**Long-range Correlation**

\[ i,k(\cdot) \sim \text{SP}(0, \cdot) \]

**Short-range Correlation**

\[ i,k(\cdot) \sim \text{SP}(0, \cdot), \]

**Covariance operator:**

\[ y(s, s') = (s, s') + (s, s') \]

\[ \sqrt{n} \{ \text{vec}(\hat{B}(d) B(d) 0.5O(H^2)) : d \} \xrightarrow{L} \text{G}(0, B(d, d')) \]

Zhu, Li, and Kong (2012). AOS
\[
\min_{B_k(s)} \sum_{i=1}^{n} \sum_{j=1}^{n_G} K_h(s, s_j)[y_{i,k}(s_j) x_i^T B_k(s_j)]^2
\]

\[
\sqrt{n} \{\text{vec}(\hat{B}(s) - B(s) - 0.5O(H^2)) : s \in [0,L_0]\} \xrightarrow{L} G(0,\Sigma_\eta(s,s') \otimes \Omega_X^{-1})
\]

Key Advantage \rightarrow Low Frequency Signal
Smooth individual functions

\[ \min_{i,k(s)} \prod_{j=1}^{n_G} K_h(s_j, s)[y_{i,k}(s_j) - x_i \hat{B}_k(s_j)]^2 \]

Estimated covariance operator

\[ \hat{S}(s,t) = \sum_{i=1}^{n} \hat{h}_i(s) \hat{h}_i(t)^T \]

Estimated eigenfunctions

\[ \{ (\hat{\ell}_{k,l}, \hat{\ell}_{k,l}(s)) : l = 1, \ldots \} \]

Functional Principal Component Analysis

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Testing Linear Hypotheses

\[ H_0 : Cvec(B(s)) = b_0(s) \quad \text{versus} \quad H_1 : Cvec(B(s)) \neq b_0(s) \]

\[ S_n(s_j) = nd(s_j)^T [C(\Sigma_\eta(s_j,s_j) \otimes \Omega^{-1}_x)C^T]^{-1} d(s_j) \]

\[ S_n = n \int_0^{L_0} d(s)^T [C(\Sigma_\eta(s,s) \otimes \Omega^{-1}_x)C^T]^{-1} d(s)ds \]

Local Test Statistics

\[ S_n(s_j) \overset{2}{\sim} (m) \quad \text{and} \quad S_n \overset{\sum_{k=1}^{K} w_k}{\sim} 2(1), \]

Global Test Statistics

MVCM

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Asymptotics

\[ \sqrt{n} [\hat{b}_{k,l}(s) - b_{k,l}(s) - \text{bias}(\hat{b}_{k,l}(s))] \quad G_{k,l}(\cdot) \]

Critical point

\[ P(\sup_{s \in [0,L_0]} |G_{k,l}(s)| > C_{k,l}(\cdot)) = 1 - \]

Confidence band

\[ (\hat{b}_{k,l}(s) - \frac{C_{k,l}(\cdot)}{\sqrt{n}}, \hat{b}_{k,l}(s) + \frac{C_{k,l}(\cdot)}{\sqrt{n}}) \]
FADTTS

Multivariate varying coefficient model

Weighted least square estimation

Functional principal component analysis

Resampling methods

Confidence bands

Hypothesis test

Resampling methods

\[ y_{i,k}(s) = \mathbf{x}_i^T B_k(s) + \eta_{i,k}(s) + \epsilon_{i,k}(s) \]

\[ B_k(s_j) = B_k(s) + \dot{B}_k(s)(s_j - s) \]

\[ \Sigma_{\eta,k,k}(s,t) = \sum_{l=1}^{\hat{\lambda}_k,\hat{\lambda}_k} \hat{\lambda}_k,\hat{\lambda}_k(s) \psi_{k,l}(t) \psi_{k,l}(t) \]

\[ H_0 : \text{Cvec}(B(s)) = b_0(s) \]

\[ H_1 : \text{Cvec}(B(s)) \neq b_0(s) \]

\[ S_n = n \int_0^{L_0} d(s)^T [C(\hat{\Sigma}_n(s,s) \otimes \hat{\Omega}_n^{-1})C^T]^{-1} d(s) ds \]

\[ p = G^{-1} \sum_{g=1}^{G} 1(S_n^{(g)} \geq S_n) \]
Real Data

- PI: Dr. John H. Gilmore from Dept of Psychiatry at UNC-CH
- 128 healthy full-term infants: 75 males and 53 females
- Mean gestational age: $298 \pm 17.6$ days, range: 262 - 433 days
- 5 diffusive outcomes: FA, MD, $\lambda_1, \lambda_2, \lambda_3$
- Internal Capsule
Real Data

Diffusion properties = Gender + Gestational age
Real Data

Global p-value: Age (<0.001), Gender (0.341)
FPCA

(a) Eigenvalues
(b) FA
(c) MD

(d) 1st eigenvector
(e) 2nd eigenvector
(f) 3rd eigenvector

1
2
3
FMEM

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Motivation

Diffusion Tensor Tract Statistics

<table>
<thead>
<tr>
<th></th>
<th>FA</th>
<th>Tensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1</td>
<td>2 week</td>
<td></td>
</tr>
<tr>
<td>a2</td>
<td>1 year</td>
<td></td>
</tr>
<tr>
<td>a3</td>
<td>2 year</td>
<td></td>
</tr>
<tr>
<td>b1</td>
<td>2 week</td>
<td></td>
</tr>
<tr>
<td>b2</td>
<td>1 year</td>
<td></td>
</tr>
<tr>
<td>b3</td>
<td>2 year</td>
<td></td>
</tr>
</tbody>
</table>
Longitudinal Extensions

Longitudinal Data

Spatial-temporal Process

$y_i(s, t) = x_i(t)^T B(s) + z_i(t)^T i(s) + i(s, t) + i(s, t)$

Objectives:
Dynamic functional effects of covariates of interest on functional response.
Decomposition:

\[ y_i(d,t) = x_i(t)^T B(d) + z_i(t)^T i(d) + i(d,t) + \text{Global Noise Components} \]

\[ \sim SP(0,\cdot), \quad \sim SP(0,\cdot) \]

\[ \text{Local Correlated Noise} \]

\[ i(\cdot) \sim SP(0,\cdot), \]

\[ \sqrt{n}\{\text{vec}(\hat{B}(d) - B(d) - 0.5O(H^2)) : d \} \xrightarrow{L} G(0, B(d,d')) \]

Ying et al. (2014). NeuroImage.
## Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>Data format</th>
<th>Estimation</th>
<th>Inference tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear mixed effects models</td>
<td>Standard longitudinal data $y_{ij}$</td>
<td>Fixed effects</td>
<td>Test statistics</td>
</tr>
<tr>
<td>Guo (2002)’s method</td>
<td>One-time-measured curves $y_i(s)$</td>
<td></td>
<td>Covariance</td>
</tr>
<tr>
<td>Greven et al. (2010)’s method</td>
<td>Multiple-time-measured curves $y_{ij}(s)$</td>
<td>Fixed effect functions</td>
<td>Test statistics</td>
</tr>
<tr>
<td>FMEM</td>
<td>Multiple-time-measured curves $y_{ij}(s)$</td>
<td>Fixed effect functions</td>
<td>Covariance functions</td>
</tr>
</tbody>
</table>

*The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL*
Functional mixed effects model

\[ y_{ij}(s) = x_{ij}^T B(s) + z_{ij}^T \xi(s) + \eta_{ij}(s) + \epsilon_{ij}(s) \]

Initial estimator
Refined estimator

Local constant and functional principal component analysis

\[ \tilde{\Sigma}_\xi(s,t) = \sum_{l=1}^{n}\tilde{\lambda}_{i}^{l}\tilde{\psi}_{i}^{l}(s)\tilde{\psi}_{i}^{l}(t)' \]

\[ \tilde{\Sigma}_\eta(s,t) = \sum_{l=1}^{n}\tilde{\lambda}_{i}^{n}\tilde{\psi}_{i}^{n}(s)\tilde{\psi}_{i}^{n}(t) \]

Hypothesis test

\[ H_0 : CB(s) = b_0(s) \]
\[ H_1 : CB(s) \neq b_0(s) \]
### Real Data

- **Gender:** Male/Female 83/54
- **Gestational age at birth (weeks):** 38.67 ± 1.74
- **Age at scan 1 (days):** 297.89 ± 13.90
- **Age at scan 2 (days):** 655.34 ± 24.00
- **Age at scan 3 (days):** 1021.70 ± 28.26
- **Number of Gradient directions:**
  - dir6/dir42 at scan 1: 80/24
  - dir6/dir42 at scan 2: 59/44
  - dir6/dir42 at scan 3: 42/49

### DTImaging parameters:

- **TR/TE = 5200/73 ms**
- **Slice thickness = 2mm**
- **In-plane resolution = 2x2 mm^2**
- **b = 1000 s/mm^2**
- **One reference scan b = 0 s/mm^2**
- **Repeated 5 times when 6 gradient directions applied.**
Real Data Analysis Results
Real Data Analysis Results

FA
- Intercept
  - Dir
  - Gender
  - Age$_1$
  - Age$_2$

RD
- Intercept
  - Dir
  - Gender
  - Age$_1$
  - Age$_2$

AD
- Intercept
  - Dir
  - Gender
  - Age$_1$
  - Age$_2$
FNMEM

Functional Nonlinear Mixed Effects Model

Decomposition:

\[ y_{i,j}(s) = f(i(s), x_{i,j}) + e_{i,j}(s), \]
\[ i(s) = (s) + b_i(s) \]

Asymptotic Normality:

\[ \sqrt{n}\{\text{vec}(\hat{\beta}(s) - \beta(s) - O(h^2)) : d \in D\} \xrightarrow{L} G(0, \Sigma_\beta(s, s')) \]

Estimation Procedure

MLE for each grid point $s_m$:

$$y_{i,j}(s_m) = f(i(s_m), x_{i,j}) + i,j(s_m) \rightarrow \hat{(s_m)}$$

Smoothing:

$$\tilde{\beta}(s) = \sum_{m=1}^{M} \tilde{K}_h(s_m - s) \hat{\beta}(s_m) \quad \text{with kernel function}$$

$$\tilde{K}_h(s_m - s) = K_h(s_m - s) / \sum_{m=1}^{M} K_h(s_m - s) \hat{\beta}(s_m),$$

$$K_h(\cdot) = K(\cdot \mid h) / h$$
Inference Procedure

Hypothesis Test:

\[ H_0 : R\beta(s) = b_0(s) \text{ for all } s \quad \text{vs.} \quad H_1 : R\beta(s) \neq b_0(s) \]

Simultaneous Confidence Bands:

\[ P\left( \hat{\beta}_{l}^{L,\alpha}(s) < \beta_l(s) < \hat{\beta}_{l}^{U,\alpha}(s) \text{ for all } s \right) = 1 - \alpha \]
Plots of power curves. Rejection rates based on score bootstrap method are calculated using FNMEM and NMEM, with sample size 50 and 100 at significant levels 5% and 1%.
Simulations

Typical 95% and 99% simultaneous confidence bands. The black solid, green solid, and red dash curves are, respectively.
### Simulations

<table>
<thead>
<tr>
<th>M</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha = 0.05$</td>
<td></td>
<td></td>
<td>$\alpha = 0.01$</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.935</td>
<td>0.925</td>
<td>0.975</td>
<td>0.975</td>
</tr>
<tr>
<td>50</td>
<td>0.935</td>
<td>0.930</td>
<td>0.980</td>
<td>0.980</td>
</tr>
<tr>
<td>75</td>
<td>0.950</td>
<td>0.945</td>
<td>0.985</td>
<td>0.990</td>
</tr>
</tbody>
</table>

Empirical coverage probabilities of $1 - \alpha$ simultaneous confidence bands for all components of $\beta$ based on 200 simulated data sets.
We analyzed a data set taken from a national database for autism research (NDAR) (http://ndar.nih.gov/), an NIH-funded research data repository, that aims to accelerate progress in autism spectrum disorders (ASD) research through data sharing, data harmonization, and the reporting of research results. 416 high quality MRI scans are available for 253 children (126 males and 127 females) with 45 grid points.
Real Data

Table 2. Demographic information for participants.

<table>
<thead>
<tr>
<th>Visit</th>
<th>Number of subjects</th>
<th>Age(years)</th>
<th>Range(years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>58</td>
<td>10.53(5.96)</td>
<td>[0, 18]</td>
</tr>
<tr>
<td>2</td>
<td>148</td>
<td>12.25(4.62)</td>
<td>[0, 21]</td>
</tr>
<tr>
<td>3</td>
<td>160</td>
<td>12.29(5.14)</td>
<td>[1, 22]</td>
</tr>
<tr>
<td>4</td>
<td>19</td>
<td>1.84(1.42)</td>
<td>[1, 6]</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>1.57(0.79)</td>
<td>[1, 3]</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>2.70(0.67)</td>
<td>[2, 4]</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3.17(0.75)</td>
<td>[2, 4]</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>3.40(1.14)</td>
<td>[2, 5]</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>3.67(1.15)</td>
<td>[3, 5]</td>
</tr>
</tbody>
</table>

Gender | Male/Female | 126/127

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Gompertz Function:

\[ y = \text{asymptote} \cdot \exp(-\text{delay} \cdot \exp(-\text{speed} \cdot t)) \]

used to characterize longitudinal white matter development during early childhood.

Functional Version:

\[ y_{i,j}(s) = \phi_{1i}(s) \cdot \exp(-\phi_{2i}(s) \cdot \phi_{3i}(s)^{t_{i,j}}) + \varepsilon_{i,j}(s) \]
Real Data Analysis Results

Tract (red solid lines) varying as a function of age for grid points from 25 to 40, the black dash curves are estimated curves.
The 100(1 − α)% simultaneous confidence bands of parameters for α = 0.05 (the first row) and α = 0.01 (the second row). The green solid and red dash curves are, respectively, the estimated curves and their corresponding 95% and 99% simultaneous confidence bands.
Functional Structural Equation Models


\[ y_{ij}(d) = x_{ij}^T (d) + \sqrt{0.51(DZ)}a_{ij}(d) + [1(MZ) + \sqrt{0.51(DZ)}]a_i(d) + c_i(d) + e_{ij}(d) \]

- \( a_{ij}(d) \sim GP(0, a) \)
- \( a_i(d) \sim GP(0, a) \)
- \( c_i(d) \sim GP(0, c) \)
- \( e_{ij}(d) \sim GP(0, e) \)

\[ Y(d, d') = \begin{bmatrix} a + c + e & a + c \\ a + c & a + c + e \end{bmatrix} (d, d') \]

\[ Y(d, d') = \begin{bmatrix} a + c + e & 0.5 & a + c \\ 0.5 & a + c & a + c + e \end{bmatrix} (d, d') \]
Three-stage Estimation Method

- **Mean structure**
  \[ Y_{ij}(d) = x_{ij}^T(d) + \hat{\epsilon}_{ij}(d) \{ \hat{(d;h)} : d \in D \} \]

- **Variance structure (Weighted likelihood)**
  \[ \{ Y_{ij}(d) x_{ij}^T(d;h) \}^2 = z_{ij}^T(d) + \hat{\epsilon}_{ij}(d) \{ \hat{(d;h)} : d \in D \} \]

- **Estimate covariance operators**
  \[ a(d,d') \text{ and } c(d,d') \]
Test Procedure

Two Key Test Statistics

• **Test marginal genetic and environmental effects**

  \[ H_0^A(d) : a(d,d) = 0 \text{ versus } H_1^A(d) : a(d,d) > 0 \]

  \[ H_0^C(d) : c(d,d) = 0 \text{ versus } H_1^C(d) : c(d,d) > 0 \]

• **Test global genetic and environmental effects**

  \[ H_0^A : a(d,d)m(d) = 0 \text{ versus } H_1^A(d) : a(d,d)m(d) > 0 \]

  \[ H_0^C : c(d,d)m(d) = 0 \text{ versus } H_1^C(d) : c(d,d)m(d) > 0 \]
Simulations

True

$\mathbf{S}_a(d,d')$

Estimated

$\mathbf{b}(d)$

$\mathbf{S}_b(d,d')$

$\mathbf{c}(d,d')$
Simulations

Histograms of 100 genetic variance estimates when genetic variance = 0 via LR

(a)

Histograms of 100 genetic variance estimates when genetic variance = 0.09 via LR

(c)

Histograms of 100 genetic variance estimates when genetic variance = 0 via WLR

(b)

Histograms of 100 genetic variance estimates when genetic variance = 0.09 via WLR

(d)
Simulations

Type I

Power
UNC Early Brain Development Studies

To track changes in behavior with brain structure, connectivity, and function, in order to characterize the progression from primary changes to subsequent clinical presentation, and to identify predictors of divergence from the typical trajectory.

PIs: Dr. John H. Gilmore

- Singletons, twins, high risk
- A longitudinal prospective study
- 900 young children aged 0 to 6 years
- Recruited prenatally
  - Exclusion: ultrasound abnormality, significant fetal/maternal medical problem, substance abuse
- 3T MRI (Siemens Allegra)
  - T1, T2, DTI, resting state fMRI
- Scanned during normal sleep (no meds)
- Ear protection, head in vac-fix device
- Success rate: 87% @ 2 weeks, 71% @ 1 year, 62% at 2 years
Quantitative tract-based white matter heritability in twin neonates

- The data set consists of 356 healthy twin neonates with 190 males and 166 females from the neonatal project as part of the UNC Early Brain Development Studies.
- There are 129 twin pairs (48 MZ twin pairs and 81 DZ twin pairs) and 98 unrelated "singleton" twins - a single unpaired twin subject in which a usable scan was not obtained from the co-twin.
- The gestational ages of these infants range from 257 to 401 days, and their mean gestational age is 289 days with standard deviation 18 days.

Question of interest:
- comprehensive heritability data on white matter microstructure fractional anisotropy (FA), radial diffusion (RD), and axial diffusion (AD) along 47 major fiber bundles.
Twin Functional Data
Coefficient Functions

(a) FA: intercept

(b) FA: gender

(c) FA: age

(d) MD: intercept

(e) MD: gender

(f) MD: age
Genetic and Environmental COs

(a) FA: Estimated Genetic Covariance

(b) FA: Estimated Common Environmental Covariance

(c) MD: Estimated Genetic Covariance

(d) MD: Estimated Common Environmental Covariance
Genetic Effects

FA: Estimated genetic variance along genu fiber tract

MD: Estimated genetic variance along genu fiber tract

FA: p values along genu fiber tract

MD: p values along genu fiber tract
Extension to High-dimensional Functions

Derive alternative estimators. Let $\Sigma_a = Z_aZ'_a$ and $\Sigma_c = Z_cZ'_c$.

$$J^{PSD} = \arg\min_{Z_a \in \mathbb{R}^{V \times d_a}, Z_c \in \mathbb{R}^{V \times d_c}}$$

(slef) $$\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{2} \sum_{v,v'} \sum_{v_0,v'_0} \left\{ S_{0ijv_0v'_0} - z^a_v z^a_{v'} - z^c_v z^c_{v'} \right\}^2 K_h(v_0,v)K_h(v'_0,v')$$

(MZ) $$+ \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{v,v'} \sum_{v_0,v'_0} \left\{ S_{1iv_0v'_0} - z^a_v z^a_{v'} - z^c_v z^c_{v'} \right\}^2 K_h(v_0,v)K_h(v'_0,v')$$

(DZ) $$+ \frac{1}{n_2} \sum_{i=1}^{n_2} \sum_{v,v'} \sum_{v_0,v'_0} \left\{ S_{2iv_0v'_0} - 0.5z^a_v z^a_{v'} - z^c_v z^c_{v'} \right\}^2 K_h(v_0,v)K_h(v'_0,v')$$
Mean Squared Errors

Figure: Mean squared error of covariance estimates.
HCP Cortical Thickness

- Preprocessed data from HCP [Glasser et al., 2013]: cortical thickness estimated using FreeSurfer. 556 subjects: 328 females, 228 males.
- No Smoothing.
- Age: 28.9 ± 3.55.
- 90 MZ pairs, 88 DZ pairs, 200 singletons.
- Assessed covariates: gender, age, handedness, height, weight, BMI, ICV.
- Kept age and ICV.
Heritability Estimates
Big Missing Data Problem

- Missing imaging data caused by design
  - Add new imaging techniques in the middle of studies (ADNI),
  - Drop out,

- Missing imaging data caused by acquisition
  - Head motion, Physiological fluctuations, Artifact-induced problem, Susceptibility artifact.

- Predicting neural activity or brain development longitudinally
Data Structure

Across modalities

Across subjects

Local spatial smoothness
SGPP Model

\[ y_{i,j}(d) = \mathbf{x}_i^T \beta_j(d) + \eta_{i,j}(d) + \epsilon_{i,j}(d) \text{ for } i = 1, \ldots, n; \ j = 1, \ldots, J, \quad (1) \]

where

\[ \beta_j(d) = (\beta_{j1}(d), \ldots, \beta_{jp}(d))^T \]

is a \( p \times 1 \) vector of regression coefficients at voxel \( d \),

\[ \eta_{i,j}(d) \]

characterizes both individual image variations from \( \mathbf{x}_i^T \beta_j(d) \) and the medium-to-long-range dependence of imaging data between \( y_{i,j}(d) \) and \( y_{i,j}(d') \) for any \( d \neq d' \),

\[ \epsilon_{i,j}(d) \]

are spatially correlated errors that capture the local (or short-range) dependence of imaging data.


SGPP. NeuroImage
Combining a functional principal component model and a multivariate simultaneous autoregressive model, we obtain an approximation of model (1) given by

\[
y_{i,j}(d) \approx x_i^T \beta_j(d) + \sum_{l=1}^{L_0} \xi_{ij,l} \psi_{j,l}(d) \\
+ \rho \frac{1}{|N(d)|} \sum_{d' \in N(d)} \left( y_{i,j}(d') - x_i^T \beta_j(d') - \sum_{l=1}^{L_0} \xi_{ij,l} \psi_{j,l}(d') \right) \\
+ e_{i,j}(d).
\]
Estimation Procedure

Spatial Gaussian predictive process model

\[ y_{i,j}(d) = x_i^T \beta_j(d) + \eta_{i,j}(d) + \epsilon_{i,j}(d) \]

Least squares estimation

\[ \hat{\beta}_j(d) \]

\[ y_{i,j}(d) - x_i^T \hat{\beta}_j(d) \]

Functional principal component analysis

\[ \hat{\Sigma}_{\eta,j}(d,d') = \sum_{l=1}^{L} \hat{\lambda}_{j,l} \hat{\psi}_{j,l}(d) \hat{\psi}_{j,l}(d') \]

\[ \hat{\epsilon}_{i,j}(d) = y_{i,j}(d) - x_i^T \hat{\beta}_j(d) - \sum_{l=1}^{L} \hat{\xi}_{i,l} \hat{\psi}_{j,l}(d) \]

Nonparametric regression

\[ \hat{\eta}_{i,j}(d) \]

Multivariate spatial autoregressive model

\[ \epsilon_i \sim N(0, \Psi(\rho, \theta)) \]

\[ \Psi(\rho, \theta) = (I_m - \rho W \otimes I_j)^{-1} \Sigma(\theta) (I_m - \rho W \otimes I_j)^{-1} \]

Restricted maximum likelihood estimation

\[ \hat{\rho}, \hat{\theta} \]

Figure: The first stage of the estimation procedure is the least squares estimation of the regression coefficients \( \beta(d) = [\beta_1(d), \ldots, \beta_J(d)] \), the second stage is the nonparametric estimation of \( \Sigma_\eta \) and its associated eigenvalues and eigenfunctions, and the third stage is the restricted maximum likelihood estimation of all the parameters in the spatial autoregressive model.
Simulation

We simulated data at all 900 pixels on a $30 \times 30$ phantom image for $n = 50$ subjects. At a given pixel $d_m = (d_{m1}, d_{m2})^T$, the data were generated from a bivariate spatial Gaussian process model according to

\[ y_{i,j}(d_m) = \beta_{j1}(d_m) + x_{i2}\beta_{j2}(d_m) + \sum_{l=1}^{2} \xi_{ij,l}\psi_{j,l}(d_m) + \epsilon_{i,j}(d_m), \]

\[ i = 1, \ldots, 50; \quad j = 1, 2. \]
Simulation Results

Figure: (a) true $\beta_{11}(d)$; (b) true $\beta_{12}(d)$; (c) true $\beta_{21}(d)$; (d) true $\beta_{22}(d)$; (e) $\hat{\beta}_{11}(d)$; (f) $\hat{\beta}_{12}(d)$; (g) $\hat{\beta}_{21}(d)$; (h) $\hat{\beta}_{22}(d)$.
Simulation Results

Figure: (a) true $\psi_{1,1}(d)$; (b) true $\psi_{1,2}(d)$; (c) true $\psi_{2,1}(d)$; (d) true $\psi_{2,2}(d)$; (e) $\hat{\psi}_{1,1}(d)$; (f) $\hat{\psi}_{1,2}(d)$; (g) $\hat{\psi}_{2,1}(d)$; and (h) $\hat{\psi}_{2,2}(d)$. 
Ventricle Surface Data

- The surface data set of the left lateral ventricle consists of 43 infants (23 males and 20 females).
- The gestational ages of the 43 infants range from 234 to 295 days and their mean gestational age is 263 days with standard deviation 12.8 days.
- The left lateral ventricle surface of each infant is represented by 1002 location vectors with each location vector consisting of the spatial x, y, and z coordinates of the corresponding vertex on the SPHARM-PDM surface.
- We randomly splitted the data set into a training set (70%) and a test set (30%).
- We fitted the SGPP model to the training set and predicted the multiple measurements at the hold-out voxels, based on the measurements at other voxels and the fitted model, for each subject in the test set.
Ventricle Surface Data

Figure: Results from the surface data of the left lateral ventricle: (a) $\hat{\beta}_{11}(d)$; (b) $\hat{\beta}_{12}(d)$; (c) $\hat{\beta}_{13}(d)$; (d) $\hat{\beta}_{21}(d)$; (e) $\hat{\beta}_{22}(d)$; (f) $\hat{\beta}_{23}(d)$; (g) $\hat{\beta}_{31}(d)$; (h) $\hat{\beta}_{32}(d)$; and (i) $\hat{\beta}_{33}(d)$. 
Ventricle Surface Data

Figure: Results from the surface data of the left lateral ventricle: (a) $\psi_{1,1}(d)$; (b) $\psi_{1,2}(d)$; (c) $\psi_{1,3}(d)$; (d) $\psi_{2,1}(d)$; (e) $\psi_{2,2}(d)$; (f) $\psi_{2,3}(d)$; (g) $\psi_{3,1}(d)$; (h) $\psi_{3,2}(d)$; and (i) $\psi_{3,3}(d)$.
## Ventricle Surface Data

Table: rtMSPE for the surface data of the left lateral ventricle

<table>
<thead>
<tr>
<th>Missingness</th>
<th>x-coordinate</th>
<th>y-coordinate</th>
<th>z-coordinate</th>
<th>x-coordinate</th>
<th>y-coordinate</th>
<th>z-coordinate</th>
<th>x-coordinate</th>
<th>y-coordinate</th>
<th>z-coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>1.9272</td>
<td>0.9810</td>
<td>0.0738</td>
<td>1.9337</td>
<td>1.0197</td>
<td>0.1156</td>
<td>1.9263</td>
<td>1.0294</td>
<td>0.1615</td>
</tr>
<tr>
<td></td>
<td>2.2448</td>
<td>1.3455</td>
<td>0.1067</td>
<td>2.2655</td>
<td>1.3827</td>
<td>0.1657</td>
<td>2.2012</td>
<td>1.3471</td>
<td>0.2204</td>
</tr>
<tr>
<td></td>
<td>2.1554</td>
<td>1.1753</td>
<td>0.0926</td>
<td>2.1906</td>
<td>1.2069</td>
<td>0.1446</td>
<td>2.1862</td>
<td>1.1830</td>
<td>0.1924</td>
</tr>
<tr>
<td>30%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Structure

T1w

2 weeks

1 year

2 years

T2w
Longitudinal Neuroimaging Data

\[ y_i(d, t) = \mu(d, x_i(t)) + \eta_i(d, t) + \epsilon_i(d, t) \text{ for } i = 1, \ldots, n, \] (2)

where

Across subjects & time \( \mu(d, x_i(t)) \) is the fixed main effect, which depends semi-parametrically on the covariates \( x_i(t) = (x_{i,1}(t), \ldots, x_{i,p}(t))^T \),

Across Modality & time \( \eta_i(d, t) \) characterizes both individual image variations from \( \mu(d, x_i(t)) \) and the medium-to-long-range dependence of imaging data between \( y_i(d, t) \) and \( y_i(d', t') \) for any \( (d, t) \neq (d', t') \),

Local spatial-temporal smoothness \( \epsilon_i(d, t) \) are spatially and temporally correlated errors that capture the local (or short-range) dependence of imaging data,

\( \eta_i(d, t) \) and \( \epsilon_i(d, t) \) are, respectively, independent and identical copies of \( \text{GP}(0, \Sigma_\eta) \) and \( \text{GP}(0, \Sigma_\epsilon) \) and mutually independent.

LSGPP Model

We add a subscript $k$ to denote the functional cluster to which voxel $d$ belongs ($k = 1, \ldots, K$). Combining a functional principal component model and a spatial-temporal model, we obtain an approximation of model (2) given by

$$y_{ik}(d, t) \approx \mu(d, x_i(t)) + \sum_{l=1}^{L_0} \xi_{i,l} \psi_l(d, t) + \epsilon_{ik}(d, t),$$

where

$$\text{Cov}(\epsilon_{ik}(d, t), \epsilon_{ik'}(d', t')) = \begin{cases} 
\sum_{\epsilon}((d, t), (d', t'); \theta(k)) & \text{if } k = k', \\
0 & \text{otherwise},
\end{cases}$$

where $\theta(k)$ is a vector of unknown parameters in $\sum_{\epsilon}((d, t), (d', t'); \theta(k))$. 
ADNI PET Data

- Data were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database.
- We consider PET scans obtained at baseline, 6 months, and 12 months.
- Subjects are classified as having mild cognitive impairment (MCI), as AD patients, or as Normal Controls (NC).

<table>
<thead>
<tr>
<th>Diagnostic status</th>
<th>age (years)</th>
<th>N male</th>
<th>N female</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>75.9± 6.9</td>
<td>34</td>
<td>17</td>
</tr>
<tr>
<td>MCI</td>
<td>76.3± 7.3</td>
<td>33</td>
<td>25</td>
</tr>
<tr>
<td>NC</td>
<td>77.0± 4.2</td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>

- We randomly chose 80 subjects for the training set to develop the prediction model.
- We predicted the PET scans at month 12, based on the baseline and 6-month scans for 79 subjects in the test set.
- We used gender, diagnostic status (MCI, AD, NC), and age (55-90 years) as covariates for the semi-parametric model.
We used a nonseparable space-time covariance function proposed by Gneiting (2002) for the spatial-temporal model, which is given by

\[
C(h, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{3/2}} \exp\left(-\frac{c \|h\|}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right), (h, u) \in \mathbb{R}^3 \times \mathbb{R},
\]

where \( a, c > 0 \) are scaling parameters of time and space, respectively; \( \alpha \in (0, 1] \) is the smoothness parameter; \( \beta \in [0, 1] \) is the space-time interaction parameter, and \( \sigma^2 \) is the variance of the spatio-temporal process.
ADNI PET Data

Figure: Results from the ADNI PET data: (a) $\hat{\psi}_1(d, t)$; (b) $\hat{\psi}_2(d, t)$; (c) $\hat{\psi}_3(d, t)$. One selected slice is shown.
Figure: Observed (upper panel) and predicted (bottom panel) PET images at month 12 for (a) an AD patient, (b) an MCI subject, and (c) a NC subject. One selected slice is shown.
Figure: Observed (upper panel) and predicted (bottom panel) PET images at month 12 for (a) an AD patient, (b) an MCI subject, and (c) a NC subject. One selected slice is shown.

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
**ADNI PET Data**

**Figure**: rtMSPE maps for prediction of ADNI PET images at month 12 for 79 test subjects. Selected slices are shown for (a) Semi-parametric model; (b) Semi-parametric model+FPCA; (c) Semi-parametric model+FPCA+Spatial-temporal model.
ADNI PET Data

Table: rtMSPE for ADNI PET images

<table>
<thead>
<tr>
<th>Model</th>
<th>rtMSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semi-parametric model</td>
<td>0.0692</td>
</tr>
<tr>
<td>Semi-parametric model + FPCA</td>
<td>0.0550</td>
</tr>
<tr>
<td>Semi-parametric model + FPCA + Spatial-temporal model</td>
<td>0.0354</td>
</tr>
</tbody>
</table>
Conclusion

- The proposed SGPP can accurately approximate the unstructured variance-covariance matrix of ultra-high dimensional data by explicitly modeling the long-to-medium-to-short-range spatial dependence.
- For standard hypothesis testing, our GLM based parameter estimation can handle many scientific questions in neuroimaging applications.
- SGPP and our prediction method can be used to directly solve missing data problems in neuroimaging studies.
- The current spatial model can be extended to a spatial-temporal model to analyze longitudinal neuroimaging data.
Image is the point or set of points in the range corresponding to a designated point in the domain of a given function.

\[ \Omega \text{ is a compact set. } \tilde{x} \in \Omega \subseteq R^k \]

\[ f(\tilde{x}) \in M \subseteq R^m \quad f : \Omega \rightarrow M \subseteq R^m \]

\[ \int \| f(\tilde{x}) \|^k d\tilde{x} < \infty \text{ for some } k > 0 \]
Neuroimaging Data with Discontinuity

Noisy Piecewise Smooth Function with Unknown Jumps and Edges

Subject1  Subject2

Covariates (e.g., age, gender, diagnostic, stimulus)
Decomposition:

\[ y_i(d) = f(x_i, B(d) + \eta_i(d)) + \epsilon_i(d), d \in D \]

**Piecewise Smooth Varying Coefficients**

\[ B(d) \in L^K \]

**Long-range Correlation**

\[ \eta_{ij}(\bullet) \sim SP(0, \Sigma_\eta) \]

**Short-range Correlation**

\[ \epsilon_{ij}(\ ) \sim SP(0, S) \]

**Covariance operator:**

\[ y(d, d') = (d, d') + (d, d) \]

3D volume/2D surface

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
SVCM

Cartoon Model

- **Disjoint Partition**
  \[ D = \bigcup_{l=1}^{L} D_l \text{ and } D_l \cap D_{l'} = \emptyset \]

- **Piecewise Smoothness: Lipschitz condition**

- **Smoothed Boundary**

- **Local Patch**

- **Degree of Jumps**
Kernel-based Smoothing Methods

Observed image $y$ = Underlying scene $f$ + Noise $\varepsilon$

$y = f + \varepsilon$; $\varepsilon$ uncorrelated, mean=0, var=$\sigma^2$

Estimate $f_i$ as a weighted average of the noisy pixels:

$$\hat{f}_i = \sum_j w_{i,j} y_j$$

Arias-Casto, Salmon, Willett (2011)

- Local constant/linear
- Yaroslavsky/Bilateral Filter
- Nonlocal Means
- PS
Smoothing Methods

MARM

At each voxel $d$

- Increasing Bandwidth
- Adaptive Weights
- Adaptive Estimates

\[
0 < h_0 < h_1 < \cdots < h_S = r_0
\]

\[
(d, d'; h_1) \quad (d, d'; h_2)
\]

\[
\hat{m}(d; h_0) \quad \hat{m}(d; h_1)
\]

\[
(d', d'; h_2)
\]

\[
\hat{m}(d; h_S)
\]

Stopping Rule

\[
(d, d'; h_s) = K_{loc}(\|d - d'\|/h_s) K_{st}(D(d, d'; h_{s-1}) / C_n)
\]

\[
D(d, d'; h_{s-1}) = (\hat{m}(d; h_{s-1}), \hat{m}(d'; h_{s-1}))
\]
Adaptively Smooth coefficients

$$\beta_j(d; h_1) = \sum_{d' \in B(d, h_1)} w(d, d'; h_1) \beta_j(d; h_0) \sum_{d' \in B(d, h_1)} w(d, d'; h_1)$$

Estimated covariance operator

$$\hat{\Sigma}(d, d') = \sum_{i=1}^{n} \hat{\eta}_i(d) \hat{\eta}_i(d')^T$$

Estimated eigenfunctions

$$\{(\hat{\lambda}_{kl}, \hat{\psi}_{kl}(d)) : l = 1, \ldots, \infty\}$$

Functional Principal Component Analysis
Simulation

True Image

SVCM

Initial Estimate in SVCM

Estimate with LF and r=0

Estimate with LF and r=1

Estimate with LF and r=2
Real Data

- Attention deficit hyperactivity disorder (ADHD) is a developmental disorder.
- ADHD is the most commonly studied and diagnosed psychiatric disorder in children.
- It affects about 3 to 5 percent of children globally and diagnosed in about 2 to 16 percent of school aged children.
- It directly cost about $36 billion per year in US.
- ADHD-200 Global Competition is a grassroots initiative event to accelerate the understanding of ADHD.
Real Data

- 174 subjects, 99 normal, and 75 ADHD-combined
- Response: White Matter, original $256 \times 256 \times 198$, downsize to $128 \times 128 \times 99$
- Covariate variables: age, gender, group (diagnosis status), and whole brain volume
- Goal: look at the group effects, including interaction effects with age and gender
Interaction effect estimates

$h_0$

Age $\times$ Diagnostic Status

$h_{10}$

Gender $\times$ Diagnostic status

$0$ $0$ $h$

$0$ $h$

$0$ $h$

$10$ $h$

$10$ $h$

$10$ $h$

$0.4$ $0.4$

$0.4$ $0.4$

$0.5$ $0.5$

$0.5$ $0.5$

$3$ $3$ $L$

$3$ $L$

$2$ $2$ $L$

$2$ $L$

$3$ $3$

$3$

$3$

$L$
First four eigenfunctions
Significant regions overlaid on template

Age × Diagnostic Status

Gender × Diagnostic status
Table 3: The first two largest significant regions of the first three largest significant blocks for hypothesis tests $H_0 : \beta_6(d) = 0$ and $H_0 : \beta_7(d) = 0$ with block and region voxel sizes. WM, L and R, respectively represent white matter, left and right.

<table>
<thead>
<tr>
<th>block</th>
<th>size</th>
<th>1st largest ROI</th>
<th>2nd largest ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ROI label</td>
<td>size</td>
</tr>
<tr>
<td>A × D</td>
<td>1</td>
<td>frontal lobe WM L</td>
<td>1567</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>frontal lobe WM R</td>
<td>900</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>nucleus accumbens L</td>
<td>1019</td>
</tr>
<tr>
<td>G × D</td>
<td>1</td>
<td>temporal lobe WM L</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>frontal lobe WM R</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>temporal lobe WM R</td>
<td>66</td>
</tr>
</tbody>
</table>
Reading materials:

Matrix-valued Data
Low-rank Linear Regression Models
Background: Emerging neuroimaging and genetic data

THE CANCER GENOME ATLAS
National Cancer Institute
National Human Genome Research Institute
Omnis characterizations

HUMAN Connectome PROJECT

Structural MRI
Resting-state MRI
Task MRI
Diffusion MRI

Figure credit: Van Essen, et al. (2013)

TCGA
HCP
ADNI

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Background: Imaging Genetics

I: Imaging

G: Genetics/Genomics

D: Disease
Hippocampus Surface Data

Data format: $100 \times 150$ matrix for each part of hippocampus (left and right)

Each element of the matrix represents radial distance of the corresponding vertex

Projected onto the hippocampus surface
Genetics Data

- Single nucleotide polymorphism (SNPs), the most common type of genetic variation among people.
- Each SNP three different genotypes, coded as 0, 1, 2.
- 6,087,205 SNPs.
Previous Work

- Voxel-wise genome-wide association analysis: fit a univariate model to each voxel and SNP pair. (approximately $2 \times 10^{10}$ pairs)

- Fit a gigantic model accommodating all genetic variation and imaging measurements using penalized regression such as Lasso. (Out of memory)

- Main disadvantage: computationally expensive / infeasible.
Low-rank Linear Regression Models for High-dimensional Matrix Responses

- \( \{(x_i, Y_i), 1 \leq i \leq n\} \) i.i.d. observations.
- \( x_i = (x_{i1}, \ldots, x_{is})^T \), \( s \times 1 \) vector of scalar covariates.
- \( Y_i \), a \( p \times q \) response matrix.
- \( x_{il} \) and \( \{Y_i, 1 \leq i \leq n\} \) standardized.

Our model is

\[
Y_i = \sum_{l=1}^{s} x_{il} \ast B_l + E_i.
\]
Two-step Screening and Estimation Procedure

- First step: Dimension reduction for the SNP covariates.
- Develop a sure independence screening procedure (Fan and Lv, 2008) for our matrix regression model.

- Second step: Refined estimation for those significant covariates.
- Develop a low rank estimation procedure using nuclear norm penalization.
Screening Procedure

- Screen covariates based on the estimated marginal OLS coefficient matrix $\hat{\mathbf{B}}_i^M = n^{-1} \sum_{i=1}^{n} x_{il} \cdot \mathbf{Y}_i$ for $l = 1, \ldots, s_n$.

- Some Summary Statistics using $\hat{\mathbf{B}}_i^M$.

- Rank the Summary Statistics.
Screening Procedure

- We use $\|\hat{\mathbf{B}}_l^M\|_{op}$ to do screening.
- More robust to noise and small effective regions of interest, rank-one screening.
- Define the selected submodel as

$$\hat{\mathcal{M}}_{\gamma_n} = \{1 \leq l \leq s_n : \|\hat{\mathbf{B}}_l^M\|_{op} \geq \gamma_n\},$$
Suppose selected set is $\hat{\mathcal{M}}$ after screening.

- Estimation of $\mathbf{B} = [\mathbf{B}_l, l \in \hat{\mathcal{M}}] \in \mathbb{R}^{p \times q|\hat{\mathcal{M}}}$.

- Nuclear norm penalization: $\min Q(\mathbf{B})$:

$$Q(\mathbf{B}) = \frac{1}{2n} \sum_{i=1}^{n} \left\| \mathbf{Y}_i - \sum_{l \in \hat{\mathcal{M}}} x_{il} \ast \mathbf{B}_l \right\|_F^2 + \lambda \sum_{l \in \hat{\mathcal{M}}} \| \mathbf{B}_l \|_*,$$

where $\| \mathbf{B}_l \|_* = \sum_k \sigma_k(\mathbf{B}_l)$.

- If $\lambda = 0$, Multivariate Ordinary Least Squares.
If $\gamma_n = \alpha C_1 n^{-\kappa}$ with $0 < \alpha < 1$, and under some other regularity conditions:

- Sure Independence Screening Property: $P(\mathcal{M} \subset \hat{\mathcal{M}}_{\gamma_n}) \to 1$.
- Vanishing False Selection Rate: $|\hat{\mathcal{M}}_{\gamma_n}|$ is at the polynomial order of $n$ with probability goes to 1.
If $\hat{\mathcal{M}} = \mathcal{M}$, when $p$ and $q$ are fixed, under some regularity conditions:

- **Estimation Consistency:**
  When $n^{1/2} \lambda \to \rho \in [0, \infty)$, the rate is $\hat{\mathbf{B}}_l - \mathbf{B}_{l0} = O_p(n^{-1/2})$, and when $\lambda \to 0$ and $n^{1/2} \lambda \to \infty$, the rate is $\hat{\mathbf{B}}_l - \mathbf{B}_{l0} = O_p(\lambda)$.

- **Rank Consistency:**
  When $\lambda \to 0$ and $n^{1/2} \lambda \to \infty$, $P(rank(\hat{\mathbf{B}}_l) = rank(\mathbf{B}_{l0})) \to 1$ for all $1 \leq l \leq s$. 
Unified theories for Two-Step Estimator

If $\hat{\mathcal{M}}$ may not be the same as $\mathcal{M}$, (account for the randomness of the first-step screening)

When $\min(p, q) \to \infty$ and $\max(p, q) = o(n)$, under some regularity conditions, when $\lambda \geq 4 C_5 n^{\tau-1/2}(p^{1/2} + q^{1/2})$, there exist positive constants $c_1, c_2, c_3, c_4, c_5$ such that with probability at least

$$1 - c_1 n^{2\kappa+\tau} \exp\{-c_2(p+q)\} - c_3 n^{2\kappa+\tau} \exp(-n) - c_4 \exp(-c_5 n^{1-2\kappa})$$

we have

$$\|\hat{\mathbf{B}}^{\hat{\mathcal{M}}} - \mathbf{B}_0^{\hat{\mathcal{M}}}\|_F^2 \leq C \left( \sum_{l \in \mathcal{M}} r_l \right) \lambda^2 \nu_L^{-2}.$$ 

Implies $\|\hat{\mathbf{B}} - \mathbf{B}_0\|_F^2 = O_P(n^{2\tau-1}(p + q)).$
ADNI Application

- 735 subjects, each with two $100 \times 150$ matrices representing the surface data from left and right hippocampus.

- Clinical covariates: Age and Gender.

- Genetic covariates: 6,087,205 SNPs.

- Adjust for Age and Gender.

- Adjust for the population stratification, the top 5 principal components of the whole chromosome data.
### Left Hippocampus

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Chromosome</th>
<th>SNP name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>rs429358</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3:143999928:C_CTG</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>rs4681527</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>rs34151823</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>rs742030</td>
</tr>
<tr>
<td>6</td>
<td>22</td>
<td>rs57622409</td>
</tr>
</tbody>
</table>

### Right Hippocampus

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Chromosome</th>
<th>SNP name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>rs429358</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>rs71517814</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>rs17196760</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>rs4074702</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>rs13265414</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>rs13264935</td>
</tr>
</tbody>
</table>
The most significant SNP for both left and right hippocampus: \textit{rs}429358 on the 19th chromosome.

The SNP \textit{rs}429358: one of the two SNPs relevant to the APOE genotype, a well-known genetic marker for the Alzheimer’s disease.

The SNP \textit{rs}429358 is strongly related to the hippocampal atrophy (Potkin et al. 2009).
ADNI data: Panel (a) is the plot for the matrix coefficient estimates for the rs429358 SNP on Chromosome 19 for both left and right hippocampus. The top part corresponds to the right hippocampus, and the bottom part corresponds to the left hippocampus. Panel (b) is hippocampal subfields mapped onto a representative hippocampal surface.
ADNI Application: Estimation Result

- The rs429358 SNP has some negative effects on CA1 as well as some parts of CA2 and CA3 of both hippocampi.

- SNP Genotype 2, more shrinkage on CA1, CA2 and CA3 subregions.
Matrix Treatment
A Causal Analysis of Hippocampal Atrophy on Behavioral Deficits in Alzheimer's Studies
A hypothetical model of AD

Figure 1: A hypothetical model of AD pathogenesis by Jack Jr et al. (2010). The red arrow denotes the causal effect we are interested in estimating.
Potential Outcome for 2D Treatment

- $Y(z)$, potential outcome under exposure level $z \in \mathcal{Z} \subseteq \mathbb{R}^{p \times q}$, 2D continuous treatment (hippocampus surface matrix).

- $Y_i = Y_i(Z_i)$, the potential outcome under $Z_i$, the level of exposure observed

Aim: To identify effect surface $\mu(z) = E[Y(z)]$. 
Assumptions

- STUVA.

- (Ignorability): $Y(z) \perp Z \mid X$.

- (Positivity): $\pi(z \mid x) \geq \pi_{\text{min}} > 0$. 
Outcome Generating Model

\[ X_i = (x_{i1}, \ldots, x_{is})^T, \text{ } s \times 1 \text{ vector of standardized scalar covariates, } s \gg n. \]

\[ Y_i = \sum_{l=1}^{S} x_{il} \beta_l + \langle Z_i, B \rangle + \epsilon_i. \]

\( B \), the effect of the image on the outcome, a \( p \times q \) matrix.

The *effect surface* can be identified as

\[
\mu(z) = E_x E[Y \mid Z = z, X = x] = \langle z, B \rangle
\]

*The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL*
Treatment Generating Model: L2RM

\[ Z_i = \sum_{l=1}^{S} x_{il} \times C_l + E_i. \]

- \( C_l \) is a \( p \times q \) matrix, denoting the effect of \( l \)th covariate on the image.

- \( E_i \), a \( p \times q \) error matrix.
Covariates Selection in Causal Inference

Goal: select confounders $X^C$ and precision variables $X^P$

$C = \{ l \in \{1, \ldots, s\} | \beta_l \neq 0 \text{ and } C_l \neq 0 \}$

$P = \{ l \in \{1, \ldots, s\} | \beta_l \neq 0 \text{ and } C_l = 0 \}$

$I = \{ l \in \{1, \ldots, s\} | \beta_l = 0 \text{ and } C_l \neq 0 \}$

$N = \{ l \in \{1, \ldots, s\} | \beta_l = 0 \text{ and } C_l = 0 \}$. 
Marginal Screening Based on Both Models

\[ \hat{\mathcal{M}}_1^* = \{ 1 \leq l \leq s : |\hat{\beta}_l^M| \geq \gamma_{1,n} \}, \]

\[ \hat{\mathcal{M}}_2 = \{ 1 \leq l \leq s : \|\hat{C}_l^M\|_{op} \geq \gamma_{2,n} \}. \]

Combine outcome and treatment model

Selected submodel $\hat{\mathcal{M}}$: Union $\hat{\mathcal{M}}_1^* \cup \hat{\mathcal{M}}_2$. 

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Refined Estimation

- $\hat{\mathcal{M}}$ include true confounders, precision and instrumental variables.

- Minimize

$$\frac{1}{2} \sum_{i=1}^{n} \left( Y_i - \langle Z_i, B \rangle - \sum_{l \in \hat{\mathcal{M}}} X_{il} \beta_l \right)^2 + \lambda_{1,n} \sum_{l \in \hat{\mathcal{M}}} |\beta_l| + \lambda_{2,n} \| B \|_*$$

where $\| B \|_* = \sum_k \sigma_k(B)$.

- $\ell_1$ penalty, exclude the instrumental variables.

- Nuclear penalty, low-rank estimation of $B$.

- Estimated effect surface $\hat{\mu}(z) = \langle z, \hat{B} \rangle$. 

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
584 subjects with complete data information

Perform analysis on left and right hippocampus separately.

Age, Gender and Education length, adjusted

The population stratification, the top 5 principal components of the whole chromosome data, adjusted.
Panel (a) plots the $\hat{B}$ corresponding to the left hippocampus (left part) and $\hat{B}$ corresponding to the right hippocampus (right part). The estimated effect surface is $\hat{\mu}(z) = \langle \hat{B}, z \rangle$. Panel (b) plots the hippocampal subfield.
The atrophy of CA1 and subiculum may cause more severe behavioral deficits compared to CA2 and CA3 subregions.

Existing literature (Schoenheit et al., 2004, Apostolova et al., 2010) has found that as Alzheimer’s disease progresses, it first affects CA1 and subiculum subregions and later CA2 and CA3 subregions.
Optimal Minimax Variable Selection for Large-Scale Matrix Linear Regression Model

- \( \{(x_i, Y_i), 1 \leq i \leq n\} \) i.i.d. observations.
- \( x_i = (x_{i1}, \ldots, x_{is})^T \), \( s \times 1 \) vector of scalar covariates.
- \( Y_i \), a \( p \times q \) response matrix.
- \( x_{il} \) and \( \{Y_i, 1 \leq i \leq n\} \) standardized.

Our model is

\[
Y_i = \sum_{l=1}^{s} x_{il} * B_l + E_i.
\]
Two Selection Procedures

- First procedure: covariates selection and coefficient estimation
  - Screen out unimportant variables and yield consistent estimate of the coefficient matrix simultaneously
  - Smooth estimators

- Second procedure: subregions detection
  - Find out areas of the matrix response affected by the predictors
Covariates Selection and Coefficient Estimation

Sparsity-restricted least squares

- Estimation of $\mathbb{B} = (B_1^T, B_2^T, \ldots, B_s^T)^T$, denoted as $\hat{\mathbb{B}}$.

- Hard-thresholding penalization: $\min_{\mathbb{B}} Q(\mathbb{B})$,

$$Q(\mathbb{B}) = \min_{\mathbb{B}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \| Y_i - \sum_{j=1}^{s} x_{ij} B_j \|_F^2 \right\} \text{ subject to } \sum_{j=1}^{s} I(\| B_j \|_F \neq 0) \leq \tau,$$

where $I(\cdot)$ is an indicator function.

- Screening based on $\hat{\mathbb{B}}$. 
The selected submodel: $\hat{M}_\tau = \{1 \leq j \leq s : \|\hat{B}_{j\tau}\|_F \neq 0\}$.

Criterion for the final model size: $\min \text{EBIC}(\hat{M}_\tau)$,

$$\text{EBIC}(\hat{M}_\tau) = \log \left\{ \frac{1}{n} \|Y - X\hat{B}_\tau\|_F^2 \right\} + \tau \frac{c_n \log(n)}{n}.$$

Joint effects of the variables rather than the marginal correlation.

Sum of all the signals rather than the largest signal.
Strong sure screening property:

\[ pr(\mathcal{M}^* = \hat{\mathcal{M}}_\tau) \to 1 \quad \text{as} \quad n \to \infty. \]

Optimal minimax rate:

\[ \| \mathbb{B}^{[l]} - \mathbb{B}^* \|_F = O_p(\{\tau \log(s/\tau)/n\}^{1/2}). \]
ADNI Application

- 735 subjects, each with two $100 \times 150$ matrices representing the surface data from left and right hippocampus.

- Genetic covariates: 2000 SNPs on 19th chromosome.

- Clinical covariates: Age and Gender.

- Adjust for the population stratification, the top 5 principal components of the whole chromosome data.
## ADNI Application: Screening Result

<table>
<thead>
<tr>
<th>Hippocampal surface</th>
<th>SNP</th>
<th>SRLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>left</td>
<td>rs8105522</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs3119815</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>rs8103479</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>rs266875</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>rs12610273</td>
<td>✓</td>
</tr>
<tr>
<td>right</td>
<td>rs8105522</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs12974560</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>rs3119815</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>rs11667541</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>rs10415851</td>
<td></td>
</tr>
</tbody>
</table>
ADNI Application: Estimating Result

Figure 1: Alzheimer’s Disease Neuroimaging Initiative data: The panels (a) and (b) show the plots for our proposed estimates corresponding to the 3 selected SNPs associated with the left and right hippocampal surfaces, respectively.
ADNI Application: Conclusion

- The SNP rs3119815 is on gene LSM14A, which is a co-expressed gene in the incipient Alzheimer’s disease samples (Chowriappa et al., 2013).

- The SNP rs12974560 is on gene LAIR1, which may also be a key in the progression of Alzheimer’s disease (Wirz et al., 2013; Sadigh-Eteghad et al., 2015).

- The SNP rs12974560 is on gene ARHGEF18, which is identified as a significantly differentially expressed gene in the lung cancer, malignant glioblastomas and Alzheimer’s disease (Sánchez – Valle, 2017).
Smoothing Estimate

- Estimate: $\hat{B}_j = (\hat{b}_{j,sk})$.

- Neighbourhood of point $(s, k)$: $\mathcal{N}_{sk} = \{(s', k'): \| (s', k') - (s, k) \|_2 < r \}$.

- Kernel smoothing estimator: $\tilde{b}_{j,sk}$,

$$\tilde{b}_{j,sk} = \sum_{(s', k') \in \mathcal{N}_{sk}} w(\hat{b}_{j,s'k'}, \hat{b}_{j,sk}; r, h) \hat{b}_{j,s'k'},$$

where

$$w(v, u; r, h) = \frac{\mathcal{K}_1(\| (s', k') - (s, k) \|_2 / r) \mathcal{K}_2(\| v - u \| / h)}{\sum_{s', k' \in \mathcal{N}_{sk}} \mathcal{K}_1(\| (s', k') - (s, k) \|_2 / r) \mathcal{K}_h(\| v - u \| / h)}.$$
**Inactive Entry** $Y_{i,sk}$:

\[ E(Y_{i,sk}|X_i) = 0 \text{ almost surely.} \]

- Inactive set: $\Psi^c = \{(s,k) : E(Y_{i,sk}|X_i) = 0 \text{ almost surely}\}$.

- Active set: $\Psi$.

- Detected subregions:

\[
\bar{\Psi}_\varsigma = \left\{ (s,k) : \frac{1}{n} \sum_{i=1}^{n} \bar{Y}_{i,sk}^2 > \varsigma \right\},
\]

where $\bar{Y}_{i,sk} = \sum_{j=1}^{r} x_{ij} \bar{b}_{j,sk}$.

- Final response area: $\min \text{ EBIC}$. 

*The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL*
Figure 3: Heat maps of the nonzeros identified out of 100 replications. White indicates 100 1s identified out of 100 replications; black indicates 100 0s identified out of 100 replications.
Conclusion

- Contributions:
  - Novel models for imaging genetics analysis with 2D imaging response
  - Fast low-rank screening step
  - Applications in causal inference
Decomposition-based Canonical Correlation Analysis
Background: Emerging multi-platform/modal biomedical data

Image credit: Van Essen, et al. (2013)
Background: Emerging multi-platform/modal biomedical data

Big Data Integration
Genomic, Imaging, Clinical data, etc.

Screening
High/Low Risk

Diagnosis
DZ/NC Degree

Treatment
Planning

Prognosis
Response

Understand the connection between disease process and data

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Background: Multi-type datasets for the same set of objects

\[ Y_1 = \begin{cases} \end{cases} \]

\[ Y_k \in \mathbb{R}^{p_k \times n} \]

\[ k = 1, \ldots, K \]

Example:

- mRNA expression data for \( p_1 \) genes measured on \( n \) tumor samples

- DNA methylation data for \( p_K \) probes measured on \( n \) tumor samples
Background: A popular decomposition model

- Only \( \{Y_k\}_{k=1}^K \) are observable.
- \( Y_k = X_k + E_k \) is low-rank + noise.
- Decompose \( X_k = C_k + D_k \).
Background: A popular decomposition model

▪ "Low-rank plus noise" model:
  \[ Y_k = X_k + E_k = C_k + D_k + E_k, \quad k = 1, \ldots, K \]

▪ All common-source matrices
  \( \{C_k\}_{k=1}^K \): from the shared underlying mechanism (latent factors) of the \( K \) datasets.

▪ \( k \)-th distinctive-source matrix
  \( D_k \): from the individual underlying mechanism (latent factors) of \( k \)-th dataset.

Mechanism: the underlying causes of variation in the data
Previous Work

- The decomposition model:

\[ Y_k = X_k + E_k = C_k + D_k + E_k, \quad k = 1, \ldots, K \]

- Existing methods: different in defining \( \{C_k, D_k\}_{k=1}^K \).

<table>
<thead>
<tr>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>JIVE</td>
<td>Lock et al. (2013)</td>
</tr>
<tr>
<td>R.JIVE</td>
<td>O’Connell &amp; Lock (2016)</td>
</tr>
<tr>
<td>AJIVE</td>
<td>Feng et al. (2018)</td>
</tr>
<tr>
<td>OnPLS</td>
<td>Löfstedt &amp; Trygg (2011)</td>
</tr>
<tr>
<td>DISCO-SCA</td>
<td>Schouteden et al. (2014)</td>
</tr>
<tr>
<td>COBE</td>
<td>Zhou et al. (2016)</td>
</tr>
</tbody>
</table>
Previous Work: Two Drawbacks

Assume the columns of $Y_k = X_k + E_k = C_k + D_k + E_k \in \mathbb{R}^{p_k \times n}$ are $n$ i.i.d. copies of mean-zero $y_k = x_k + e_k = c_k + d_k + e_k \in \mathbb{R}^{p_k}$.

**Drawback 1**

- Define $\{C_k, D_k\}_{k=1}^K$ from $(\mathbb{R}^n, \cdot)$ not $(L_0^2, \text{Cov})$. Their $\{C_k, D_k\}_{k=1}^K$ in $(\mathbb{R}^n, \cdot)$ are approximations of those in $(L_0^2, \text{Cov})$.

$(\mathbb{R}^n, \cdot): n$-dim Euclidean space

$(L_0^2, \text{Cov})$: the set of all random variables with zero mean and finite variance

$C_k D_k \geq 0$

$\gamma$

$\hat{\text{Corr}}(c_k, d_k) = 0$

approx.

$Corr(c_k, d_k) = 0$
Previous Work: Two Drawbacks

Assume the columns of \( Y_k = X_k + E_k = C_k + D_k + E_k \in \mathbb{R}^{p_k \times n} \) are \( n \) i.i.d. copies of mean-zero \( y_k = x_k + e_k = c_k + d_k + e_k \in \mathbb{R}^{p_k} \).

### Drawback 2

- Focus on \( C_k \perp D_k \) not well consider the orthogonality among \( \{D_k\}_{k=1}^K \).

E.g., for \( K=2 \) datasets,

\[
D_1 \perp D_2 \text{ in } (\mathbb{R}^n, \cdot ) \quad \text{approx.} \quad \text{Corr}(d_1, d_2) = 0
\]

to avoid "common" latent factors between \( d_1 \) and \( d_2 \).
Decomposition-based canonical correlation analysis (D-CCA) for $K=2$ datasets

Assume the columns of $\mathbf{Y}_k = \mathbf{X}_k + \mathbf{E}_k = \mathbf{C}_k + \mathbf{D}_k + \mathbf{E}_k \in \mathbb{R}^{p_k \times n}$ are $n$ i.i.d. copies of mean-zero $\mathbf{y}_k = \mathbf{x}_k + \mathbf{e}_k = \mathbf{c}_k + \mathbf{d}_k + \mathbf{e}_k \in \mathbb{R}^{p_k}$.

**D-CCA**

For $\mathbf{x}_k = \mathbf{c}_k + \mathbf{d}_k$ ($k = 1,2$), define $\{\mathbf{c}_k, \mathbf{d}_k\}_{k=1}^K$ from $(\mathcal{L}_0^2, \text{Cov})$ s.t.

- **Same common latent factors:**
  \[
  \text{span}(\mathbf{c}_1^T) = \text{span}(\mathbf{c}_2^T)
  \]
  Notation: $\text{span}(\nu^T) = \text{span}(\{\nu_i\}_{i=1}^p)$

- **Uncorrelated distinctive latent factors:**
  \[
  \text{span}(\mathbf{d}_1^T) \perp \text{span}(\mathbf{d}_2^T), \text{i.e., Corr}(\mathbf{d}_1, \mathbf{d}_2)=0
  \]

- **Parsimonious representation:**
  \[
  \text{span}([\mathbf{x}_1^T, \mathbf{x}_2^T]) = \text{span}([\mathbf{c}_1^T, \mathbf{c}_2^T, \mathbf{d}_1^T, \mathbf{d}_2^T])
  \]
D-CCA for $K=2$ datasets

Latent Structure

- $\text{span}(c_1^T) = \text{span}(c_2^T)$
- $\text{span}(d_1^T) \perp \text{span}(d_2^T)$
- $\text{span}([x_1^T, x_2^T]) = \text{span}([c_1^T, c_2^T, d_1^T, d_2^T])$
D-CCA for $K=2$ datasets with dimensions $p_1 = p_2 = 1$

Decompose 2 standardized (mean-0 & var-1) variables $z_1$ and $z_2$ with correlation $\rho \geq 0$ by $z_k = c + d_k$, $k = 1, 2$

- The six competing methods: under their condition $c \perp \{d_1, d_2\}$
  (i) $c = 0$ if $z_1 \neq z_2$; $c = z_k$ if $z_1 = z_2$.
  (ii) $c \perp \{d_1, d_2\}$ but $d_1 = -d_2$
D-CCA for $K=2$ datasets with dimensions $p_1 = p_2 = 1$

Decompose 2 standardized (mean-0 & var-1) variables $z_1$ and $z_2$ with correlation $\rho \geq 0$ by $z_k = c + d_k$, $k = 1,2$

- **Our D-CCA:**
  - $c$ is the closest to both $z_1$ and $z_2$:
    \[ c \propto \arg\max_{w \in \mathcal{L}_0^2} \{ \text{Corr}^2(z_1, w) + \text{Corr}^2(z_2, w) \} \]
  - $d_1$ and $d_2$ are uncorrelated.
  - Larger $\|c\|$ shows stronger $\rho$: $\|c\| \uparrow$ as $\theta \downarrow$ (i.e., as $\rho \uparrow$).
D-CCA for $K=2$ datasets (any dims)

### Canonical Correlation Analysis (CCA)

<table>
<thead>
<tr>
<th>Canonical Correlation Analysis (CCA)</th>
<th>Canonical variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CCA ($r_1 \leq r_2$)</strong></td>
<td><strong>Canonical variables</strong></td>
</tr>
<tr>
<td><strong>Step 1:</strong></td>
<td>$z_1^{(1)}$, $z_2^{(1)}$</td>
</tr>
<tr>
<td></td>
<td>$z_1^{(r_1)}$, $z_2^{(r_1)}$</td>
</tr>
<tr>
<td><strong>Step $r_1$:</strong></td>
<td>$z_1^{(r_1)}$, $z_2^{(r_1)}$</td>
</tr>
<tr>
<td><strong>Step $(r_1+1)$:</strong></td>
<td>$0$, $z_2^{(r_1+1)}$</td>
</tr>
<tr>
<td></td>
<td>$0$, $z_2^{(r_2)}$</td>
</tr>
</tbody>
</table>

- For each $k = 1, 2$, $z_k^{(1)}, \ldots, z_k^{(r_k)}$ form an orthogonal basis of $\text{span}(x_k^\top)$.

So, $x_k = \sum_{l=1}^{r_2} \beta_{k,l}^{(l)} z_k^{(l)}$

Signal subspace: $\text{span}(x_1^\top)$, $\text{span}(x_2^\top)$

$r_k = \dim\{\text{span}(x_k^\top)\}$ for $k = 1, 2$
## Canonical Correlation Analysis (CCA)

<table>
<thead>
<tr>
<th>CCA ($r_1 \leq r_2$)</th>
<th>Canonical variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1:</strong></td>
<td>$z_1^{(1)}, z_2^{(1)}$</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td><strong>Step $r_1$:</strong></td>
<td>$z_1^{(r_1)}, z_2^{(r_1)}$</td>
</tr>
<tr>
<td><strong>Step $(r_1+1)$:</strong></td>
<td>$0, z_2^{(r_1+1)}$</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td><strong>Step $r_2$:</strong></td>
<td>$0, z_2^{(r_2)}$</td>
</tr>
</tbody>
</table>

- **Cross-orthogonality:**
  \[
  \{z_1^{(\ell)}, z_2^{(\ell)}\} \perp \{z_1^{(m)}, z_2^{(m)}\} \\
  \text{for } \ell \neq m
  \]

- **Signal subspace:**
  \[
  \text{span}(x_1^T), \text{span}(x_2^T)
  \]

- **Dimensionality:**
  \[
  r_k = \text{dim}\{\text{span}(x_k^T)\} \text{ for } k = 1, 2
  \]
### Canonical Correlation Analysis (CCA)

<table>
<thead>
<tr>
<th>CCA ($r_1 \leq r_2$)</th>
<th>Canonical variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1:</td>
<td>$z_1^{(1)}$, $z_2^{(1)}$</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>Step $r_1$:</td>
<td>$z_1^{(r_1)}$, $z_2^{(r_1)}$</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>Step $(r_1+1)$:</td>
<td>0, $z_2^{(r_1+1)}$</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>Step $r_2$:</td>
<td>0, $z_2^{(r_2)}$</td>
</tr>
</tbody>
</table>

Within each pair

- Signal subspace: $\text{span}(x_1^T)$, $\text{span}(x_2^T)$
- $r_k = \text{dim}\{\text{span}(x_k^T)\}$ for $k = 1, 2$
D-CCA for $K=2$ datasets (any dims)

D-CCA: for $k = 1, 2$,

$$x_k = \sum_{l=1}^{r_2} \beta_k^{(l)} z_k^{(l)} = \sum_{l=1}^{r_2} \beta_k^{(l)} (c^{(l)} + d_k^{(l)}) = c_k + d_k$$

By the cross-orthogonality that $\{z_1^{(l)}, z_2^{(l)}\} \perp \{z_1^{(m)}, z_2^{(m)}\}$ for $l \neq m$, D-CCA satisfies

- Same common latent factors:
  $$\text{span}(c_1^T) = \text{span}(c_2^T)$$

- Uncorrelated distinctive latent factors:
  $$\text{span}(d_1^T) \perp \text{span}(d_2^T), \text{i.e.}, \text{Corr}(d_1, d_2) = 0$$

- Parsimonious representation:
  $$\text{span}([x_1^T, x_2^T]) = \text{span}([c_1^T, c_2^T, d_1^T, d_2^T])$$
D-CCA for $K \geq 2$ datasets

- The ideal orthogonality structure:
  At least one pair in $\{\text{span}(d_k^T)\}_{k=1}^{K}$ are orthogonal.

- Method:

  ![Step Table](image)

  1. CCA (Hotelling, 1936) Generalized CCA (Carroll, 1968)

  2. Our decomp. Within each $l$-th set of canonical variables:

     $z_k^{(l)} = c^{(l)} + d_k^{(l)}, \quad k = 1, \ldots, K$
Matrix Estimation

Recall: the columns of $Y_k = X_k + E_k = C_k + D_k + E_k \in \mathbb{R}^{p_k \times n}$ are $n$ i.i.d. copies of mean-zero $y_k = x_k + e_k = c_k + d_k + e_k \in \mathbb{R}^{p_k}$.

- **Difficulty:**
  Only high-dim $Y_k, k = 1, \ldots, K$ are observable;
  $\text{Cov}(x_j, x_k)$ may not be consistently est’ed by sample cov.

- **Assumption:**
  $\text{span}(x_k^T)$ is low-dim (i.e., $\text{Cov}(x_k)$ is low-rank, so $X_k$ is low-rank)
D-CCA for $K \geq 2$ datasets

Matrix Estimation

- Define $\hat{X}_k$ by the soft-thresholded SVD (Wang & Fan, 2017) of $Y_k$;
- From $\hat{X}_k$ using the sample (generalized) CCA and our D-CCA construction yields $\hat{C}_k$ and $\hat{D}_k$.

**Theorem.** Under certain assumptions, we have

$$\frac{\|\hat{X}_k - X_k\|^2}{\|X_k\|^2} = O_P\left(\frac{p_k \log p_k}{nR_k}\right),$$

$$\max\left\{\frac{\|\hat{C}_k - C_k\|^2}{\|X_k\|^2}, \frac{\|\hat{D}_k - D_k\|^2}{\|X_k\|^2}\right\} = O_P\left(\sum_{k=1}^{K} \sqrt{\frac{p_k \log p_k}{nR_k}}\right),$$

where $R_k = \frac{\lambda_{r_k}\{\text{Cov}(x_k)\}}{\lambda_{\max}\{\text{Cov}(e_k)\}},$ and $\|\cdot\|$ is valid for both the Frobenius and spectral norms.
D-CCA Simulation Example

\[ x_k = \sum_{l=1}^{r_k} \beta_k^{(l)} z_k^{(l)} = \sum_{l=1}^{r_k} \beta_k^{(l)} (c^{(l)} + d_k^{(l)}) = c_k + d_k \text{ for } k = 1,2 \]

Setting:

- Data size: \( p_1 = 900, p_2 = n = 300 \)
- \( r_k = \dim\{\text{span}(x_k^T)\}: r_1 = 3, r_2 = 5 \)
- 1\text{st} canonical pair: \( \theta \left( z_1^{(1)}, z_2^{(1)} \right) = 45^\circ \), i.e., \( \text{Corr} \left( z_1^{(1)}, z_2^{(1)} \right) = 0.707 \)
- Other canonical pairs: \( \{z_1^{(l)}\}_{l=2}^3, \{z_2^{(l)}\}_{l=2}^5 \sim iid \mathcal{N}(0,1) \)
- Entries of noises \( \{e_k\}_{k=1}^2 \sim iid \mathcal{N}(0,1) \)
- Signal-to-noise eigen-ratio: \( R_k = \frac{\lambda_{r_k}\{\text{Cov}(x_k)\}}{\lambda_{\max}\{\text{Cov}(e_k)\}} = 100 \)
- Generated data: \( y_k = x_k + e_k \)
D-CCA Simulation Example Result

(a) Ground truth

(b) D-CCA estimates

(c) $\mathbf{g}_k$ by other methods
Two Real-Data Examples

TCGA breast cancer genomic data

HCP motor-task fMRI

The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
TCGA Breast Cancer Data

- 683 samples (4 intrinsic subtypes):
  - 118 Basal-like
  - 57 HER2-enriched
  - 341 Luminal A
  - 167 Luminal B
- 3 genomic datasets:
  - mRNA expression (2930 genes)
  - miRNA expression (526 miRNAs)
  - DNA methylation (3067 probes)
- 3 data matrices:
  \[ \mathbf{Y}_{\text{mRNA}} \in \mathbb{R}^{2930 \times 683}, \mathbf{Y}_{\text{miRNA}} \in \mathbb{R}^{526 \times 683}, \mathbf{Y}_{\text{DNA}} \in \mathbb{R}^{3067 \times 683} \]
TCGA Breast Cancer Data

- **Aim:** 4 intrinsic cancer subtypes
  - Common/distinctive mechanisms underlying the 3 datasets (mRNA, miRNA, DNA)
- **Model:** \( Y_k = X_k + E_k = C_k + D_k + E_k, \ k \in \{\text{mRNA, miRNA, DNA}\} \)
- **Method:** Investigate each matrix’s ability to separate the cancer subtypes.
- **Hypothesis:** For each \( k \), matrix \( C_k \) has the best subtype separation.
- **Metric:** \( R^2 \) score (Cabanski et al., 2010)
  \[
  R^2 (A) = 1 - \frac{\sum_{i=1}^{p} \sum_{j=1}^{n} (A_{ij} - \bar{A}_{i,.})^2}{\sum_{i=1}^{p} \sum_{j=1}^{n} (A_{ij} - \bar{A}_i)^2} = 1 - \frac{SSW}{SST} \in [0,1]
  \]
  A higher score indicates better subtype separation.
TCGA Data: compare $Y_k$ and $\hat{X}_k$

- $R^2(\hat{X}_k) > R^2(Y_k)$ by all methods
TCGA Data: compare $\hat{C}_k$, $\hat{X}_k$ and $\hat{D}_k$

- $R^2(\hat{C}_k) > R^2(\hat{X}_k) > R^2(\hat{D}_k)$ by all except AJIVE and COBE with $\hat{C}_k = 0$:
  - The 4 intrinsic cancer subtypes are more likely a feature of the common underlying mechanism.
TCGA Data: compare methods by $\hat{C}_k$

- D-CCA has the highest $R^2(\hat{C}_k)$ for each $k$.

All $R^2$ scores in the above comparisons are significantly different with p-value <0.001 using the permutation test of Cabanski et al. (2010).
TCGA Breast Cancer Data: Results

Separation of the 4 intrinsic cancer subtypes
(A higher $R^2$ score indicates better separation):

For each $k \in \{\text{mRNA, miRNA, DNA}\}$,

- $R^2(\hat{X}_k) > R^2(Y_k)$ by all methods;
- $R^2(\hat{C}_k) > R^2(\hat{X}_k) > R^2(\hat{D}_k)$ by all except AJIVE and COBE with $\hat{C}_k = 0$:
  4 intrinsic cancer subtypes $\leftrightarrow$ common mechanism of mRNA, miRNA & DNA
- Our D-CCA has the highest $R^2(\hat{C}_k)$ for each $k$. 
HCP moto-task fMRI data

- n=1080 young healthy adults
- Each subject has $K=3$ z-statistic maps
  - (A map: a task vs. fixation baseline)
    - Left-hand task
    - Right-hand task
    - Overall motor task
      - average contrast of left/right-hand, left/right-foot, and tongue
- $p=91,282$ grayordinates:
  - 59,412 cortical surface vertices + 31,870 subcortical gray matter voxels
- $K=3$ data matrices:
  $\mathbf{Y}_{\text{left-hand}}, \mathbf{Y}_{\text{right-hand}}, \mathbf{Y}_{\text{overall}} \in \mathbb{R}^{91,282 \times 1080}$
HCP moto-task fMRI data

- **Aim:**
  - brain regions
  - Common mechanism of left-hand, right-hand & overall motor tasks

- **Model:** the columns of $\mathbf{Y}_k = \mathbf{X}_k + \mathbf{E}_k = \mathbf{C}_k + \mathbf{D}_k + \mathbf{E}_k \in \mathbb{R}^{p \times n}$ are $n$ iid obs. of mean-zero $\mathbf{y}_k = \mathbf{x}_k + \mathbf{e}_k = \mathbf{c}_k + \mathbf{d}_k + \mathbf{e}_k \in \mathbb{R}^p$.

- **Metric:** local variance ratio

$$VR(\mathbf{c}_k[i]) = \frac{\text{Var}(\mathbf{c}_k[i])}{\text{Var}(\mathbf{x}_k[i])} \approx \hat{VR}(\mathbf{c}_k[i]) = \frac{\left\| \hat{\mathbf{c}}_k[i,:] \right\|_F^2}{\left\| \hat{\mathbf{X}}_k[i,:] \right\|_F^2}$$

to assess the influence of the common mechanism on $i$-th brain voxel in $k$-th dataset.

- $\mathbf{c}_k[i]$ : $i$-th entry of $\mathbf{c}_k$
- $\hat{\mathbf{c}}_k[i,:]$ : $i$-th row of $\hat{\mathbf{C}}_k$
HCP moto-task fMRI data: D-CCA’s VR map for left-hand task
(see right somatomotor cortex and left cerebellum)
HCP moto-task fMRI data: D-CCA’s VR map for right-hand task
(see left somatomotor cortex and right cerebellum)
HCP moto-task fMRI data:
D-CCA’s VR map for overall-motor task
(see somatomotor cortex and cerebellum)
HCP moto-task fMRI: results by D-CCA

Table 1: Regions most affected by the common mechanism of the 3 motor tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Somatomotor cortex</th>
<th>Cerebellum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left-hand</td>
<td>Right</td>
<td>Left</td>
</tr>
<tr>
<td>Right-hand</td>
<td>Left</td>
<td>Right</td>
</tr>
<tr>
<td>Overall-motor</td>
<td>Two-side</td>
<td>Two-side</td>
</tr>
</tbody>
</table>

- The **contralateral** change in the somatomotor cortex and the cerebellum is supported by their intrinsic functional connectivity shown in Buckner et al. (2011, *J. Neurophysiology*)
HCP moto-task fMRI: compare methods

- Run on a single computing node with two 10-core Intel Xeon E5-2690v2 3.0GHz CPUs, 62GB memory for 24hrs wall time limit.

<table>
<thead>
<tr>
<th>Method</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-CCA</td>
<td>18.0 s</td>
</tr>
<tr>
<td>COBE</td>
<td>25.3 s, $\hat{C}_k = 0$</td>
</tr>
<tr>
<td>AJIVE</td>
<td>180.5 s, $\hat{C}_k = 0$</td>
</tr>
<tr>
<td>JIVE</td>
<td>5.47 hrs</td>
</tr>
<tr>
<td>R.JIVE</td>
<td>17.4 hrs</td>
</tr>
<tr>
<td>DISCO-SCA</td>
<td>out of 24 hrs</td>
</tr>
<tr>
<td>OnPLS</td>
<td>out of memory</td>
</tr>
</tbody>
</table>
Conclusion

- Contributions:
  - Well-defined the common and distinctive-source matrices.
  - Proved the asymptotic convergence of matrix estimators.
  - Showed D-CCA’s better performance in simulations and real-data analysis.
Thanks