Functional Data Analysis for Medical Imaging Data

Jane-Ling Wang
University of California, Davis

2015 SAMSI CCNS Summer School

Research supported by NSF and Gates Foundation
Outline

1. Introduction
2. Mean and Covariance Functions
3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?
4. Inverse Problem in Functional Correlations and Regression
5. Stringing high-dim data to functional data
6. Next Generation Functional Data
Disclaimer

- This course has limited coverage of FDA applications to medical imaging data.
  - Some (brain) images were downloaded from the web.
This course has limited coverage of FDA applications to medical imaging data.
- Some (brain) images were downloaded from the web.

But you’ll hopefully see why FDA is relevant to neuroimaging data analysis (NDA).
This course has limited coverage of FDA applications to medical imaging data.

- Some (brain) images were downloaded from the web.

But you’ll hopefully see why FDA is relevant to neuroimaging data analysis (NDA).

Plenty of FDA approaches have been applied to neuroimaging data, e.g. by Aston, Caffo, Crainiceanu, Lindquist, Morris, Zhu, among others.

- But there’s much more that FDA can contribute!
Disclaimer

- This course is not comprehensive for FDA.
  e.g. We will not cover functional classification and clustering.
Disclaimer

- This course is not comprehensive for FDA.
  
  e.g. We will not cover functional classification and clustering.

- It mainly represents our views of FDA.
  
  - The emphasis is on concepts rather than details, but it does include some theory.
1. Introduction
2. Mean and Covariance Functions
3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?
4. Inverse Problem in Functional Correlations and Regression
5. Stringing high-dim data to functional data
6. Next Generation Functional Data
What is Functional Data?

- Functional Data: A sample of random functions, with one function per subject.
  - These functions can be curves (1D), images (2D or 3D), or higher dimension object data.
What is Functional Data?

- Functional Data: A sample of random functions, with one function per subject.
  - These functions can be curves (1D), images (2D or 3D), or higher dimension object data.

- Characteristics of functional data:
  (i) The atom of functional data is a “function”.
  (ii) They are $\infty$—dimensional data.
What is Functional Data?

- Functional Data: A sample of random functions, with one function per subject.
  - These functions can be curves (1D), images (2D or 3D), or higher dimension object data.

- Characteristics of functional data:
  (i) The atom of functional data is a “function”.
  (ii) They are $\infty$—dimensional data.

- Example: Curve data
  - real-valued functions defined on an interval $I \in \mathbb{R}$.
  - one curve per subject
  - These curves are usually considered realizations of a stochastic process $X(t)$ in a **Hilbert space**, e.g. in $L^2(I)$, ($\int_I [X(t)]^2 dt < \infty$).
Examples of Functional Data

- fMRI data at a particular voxel of 20 subjects $\rightarrow n = 20$. 
Examples of Functional Data

- Spectrum data for meat content - here the function is over the spectrum channels of n pieces of meat.
Additional Examples of Functional Data

- 10 minutes EKG recordings of 100 patients.
  \[ n = 100 \]

- Daily temperature recording in January at 240 locations.
  \[ n = 240 \]

- Daily reproduction (\# of eggs) of 1000 female medflies (Mediterranean fruit flies) till death.
  \[ n = 1000 \]
In reality, functional data are recorded at a regular and dense time grid \(\Rightarrow\) high-dimensional data.

- The fMRI data were recorded every two seconds for about 10 minutes (300 time points) \(\Rightarrow\) 300 dimensional data.
Real/Observed Functional Data

- The spectrum data were recorded at 100 frequency channels (hence 100-dim) and smoothed individually, i.e. pre-smoothed.

Is there a curse of high dimensionality?
Longitudinal Data as Functional Data

- Longitudinal Data - Irregularly sampled functional data.

- They are often sparse, as in medical follow-up or social studies.
Longitudinal AIDS Data

CD4 counts of 369 patients.

\[ n_i = \# \text{ of repeated measurements for subject } i, \]

varies with subject with an average of 6.44 measurement per subject.
CD4 counts of 369 patients.

\[ n_i = \# \text{ of repeated measurements for subject } i, \]

varies with subject with an average of 6.44 measurement per subject.

This results in longitudinal data with uneven \# of measurements at irregular time-points.
CD4 Counts of First 25 Patients
Because of the dense recording, functional data have been customarily modeled with a nonparametric approach.

- Often smoothness of the functions is assumed.
Because of the dense recording, functional data have been customarily modeled with a nonparametric approach.

- Often smoothness of the functions is assumed.

Longitudinal data have traditionally been modeled by a parametric approach, such as a linear mixed-effects model.

- However, it may not be easy to spot the pattern due to sparsity of and noise in the longitudinal data.
This motivates a data oriented nonparametric approach, which is often considered a divide between LDA and FDA.
The analysis of functional data is termed “Functional Data Analysis” (FDA).

Let the observed data for subject $i$ might be

$$Y_{ij} = X_i(t_{ij}) + e_{ij}; j = 1;\ldots;n_i;$$

where $X_i(t)$ is a smooth random function and $e_{ij}$ are independent $\varepsilon$ $i;j$.

A strength of the FDA approach is its ability to handle noise.
The analysis of functional data is termed “Functional Data Analysis” (FDA).

Both longitudinal and functional data may be observed with noise (measurement errors).

\[ Y_{ij} = X_i(t_{ij}) + e_{ij}, j = 1, \ldots, n_i, \]

where \( X_i(t) \) is a smooth random function and \( e_{ij} \) are independent \( \forall i, j \).
Longitudinal vs Functional data

- The analysis of functional data is termed “Functional Data Analysis” (FDA).

- Both longitudinal and functional data may be observed with noise (measurement errors).

  \[ Y_{ij} = X_i(t_{ij}) + e_{ij}, j = 1, \ldots, n_i, \]

  where \( X_i(t) \) is a smooth random function and \( e_{ij} \) are independent \( \forall i, j \).

- A strength of the FDA approach is its ability to handle noise.
There are the three types of functional data:

(i) stochastic processes $\rightarrow \infty$-dimensional data
(ii) dense functional data $\rightarrow$ high dimensional data
(iii) sparse functional/longitudinal data $\rightarrow$ irregular dim. data.
There are the three types of functional data:

(i) stochastic processes \( \xrightarrow{} \) \( \infty \)-dimensional data
(ii) dense functional data \( \xrightarrow{} \) high dimensional data
(iii) sparse functional/longitudinal data \( \xrightarrow{} \) irregular dim. data.

The stochastic process \( X_i(t) \) is assumed to be a continuous function.
(Both dense and sparse functional data may contain noise, a.k.a. measurement error.)
There are the three types of functional data:

(i) stochastic processes $\implies \infty$-dimensional data
(ii) dense functional data $\implies$ high dimensional data
(iii) sparse functional/longitudinal data $\implies$ irregular dim. data.

The stochastic process $X_i(t)$ is assumed to be a continuous function.
(Both dense and sparse functional data may contain noise, a.k.a. measurement error.)

Nonparametric approaches are typically employed to all three types of functional data.
The term “functional data” was coined by Ramsey (1982) and Ramsey and Dalzell (1991) but the term “curve data” was often used as well, e.g. by Gasser et al. (1984) for growth curves.
The term “functional data” was coined by Ramsey (1982) and Ramsey and Dalzell (1991) but the term “curve data” was often used as well, e.g. by Gasser et al. (1984) for growth curves.

Growth curves and the first implementation of functional principal component analysis (FPCA) were attributed to Rao (1958), where the growth data were recorded as multivariate data.
The term “functional data” was coined by Ramsey (1982) and Ramsey and Dalzell (1991) but the term “curve data” was often used as well, e.g. by Gasser et al. (1984) for growth curves.

Growth curves and the first implementation of functional principal component analysis (FPCA) were attributed to Rao (1958), where the growth data were recorded as multivariate data.

The analysis of stochastic processes went even further back to Grenander (1950), Karhunen (1946), Loève (1946), and Dauxois and Poussé (1976).
The handling of longitudinal data as sparse functional data was the focus in Yao et al. (2005), but nonparametric approaches for longitudinal data had already been employed by Shi et al. (1996), Staniswalis and Lee (1998), James et al. (2000), and Rice and Wu (2001).
The handling of longitudinal data as sparse functional data was the focus in Yao et al. (2005), but nonparametric approaches for longitudinal data had already been employed by Shi et al. (1996), Staniswalis and Lee (1998), James et al. (2000), and Rice and Wu (2001).

First generation functional data typically consist of a random sample, $X_1(t), \ldots, X_n(t)$, of independent real-valued functions.
Next generation functional data refers to functional data that are part of complex data objects, possibly multivariate, correlated, or involve images or shapes.


Next generation functional data refers to functional data that are part of complex data objects, possibly multivariate, correlated, or involve images or shapes.

- See a review by W. et al. (2015, arXiv)

and page 23 of the report

Brain and neuroimaging data are examples of next generation functional data.
- See an upcoming review by Aston (2015).
What is the difference between functional and time-series data?
Questions

- What is the difference between functional and time-series data?

- What is the benefit of functional data over multivariate data?
Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- What are the three types of functional data?
- What are the pros and cons for each type?
Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- What are the three types of functional data?
- What are the pros and cons for each type?
  - Which one is the easiest or hardest to handle?
    (We’ll discuss how to handle them in the next section.)
- Is there a curse of dimensionality for dense functional data?
Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- What are the three types of functional data?
- What are the pros and cons for each type?
  - Which one is the easiest or hardest to handle? (We’ll discuss how to handle them in the next section.)
- Is there a curse of dimensionality for dense functional data?
  - What does “dense” mean?
End of Introduction
Outline

1. Introduction

2. Mean and Covariance Functions

3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?

4. Inverse Problem in Functional Correlations and Regression

5. Stringing high-dim data to functional data

6. Next Generation Functional Data
Mean and Covariance Functions

Data ($\{X_1, \ldots, X_n\}$) are i.i.d. copies of a random function $X(t)$:

- **Mean function:** $\mu(t) = E(X(t))$

- **Covariance function:** $\Sigma(s, t) = \text{cov}(X(s), X(t))$, where $s \& t \in$ interval $I$. 

Regular functional data - All subjects were measured at the same time grid, $t_1; \ldots; t_m$, often equally spaced = multivariate data.

Irregular functional data - The measurement schedule for subject $i$ is $t_i_1; \ldots; t_i_i = \text{longitudinal data.}$
Mean and Covariance Functions

Data \(\{X_1, \ldots, X_n\}\) are i.i.d. copies of a random function \(X(t)\):

- Mean function: \(\mu(t) = E(X(t))\)

- Covariance function: \(\Sigma(s, t) = \text{cov}(X(s), X(t))\),
  where \(s, t \in \text{interval } I\).

- Regular functional data - All subjects were measured at the same time grid, \(t_1, \ldots, t_m\), often equally spaced \(\implies\) multivariate data.

- Irregular functional data - The measurement schedule for subject \(i\) is \(t_{i1}, \ldots, t_{in_i}\) \(\implies\) longitudinal data.
Estimation of Mean and Covariance Functions:
Dense and Regular Functional Data

- For regular or fully observed functional data, the cross-sectional (sample) mean at the observed time $t$ provides a $\sqrt{n}$-consistent (pointwise) estimate of the mean $\mu(t)$ even in the presence of measurement errors.  

(WHY?)
Estimation of Mean and Covariance Functions: 
Dense and Regular Functional Data

- For regular or fully observed functional data, the cross-sectional (sample) mean at the observed time $t$ provides a $\sqrt{n}$-consistent (pointwise) estimate of the mean $\mu(t)$ even in the presence of measurement errors. (WHY?)

- The cross-section mean can further be smoothed to obtain a smooth mean estimate (This requires a dense measurement schedule).
  - Think how you should smooth to retain the $\sqrt{n}$-consistent of the smooth estimate.
For regular or fully observed functional data, the cross-sectional (sample) mean at the observed time $t$ provides a $\sqrt{n}$-consistent (pointwise) estimate of the mean $\mu(t)$ even in the presence of measurement errors.

The cross-section mean can further be smoothed to obtain a smooth mean estimate (This requires a dense measurement schedule).

- **Think how you should smooth to retain the $\sqrt{n}$-consistent of the smooth estimate**.

Likewise, the sample covariance matrix is also $\sqrt{n}$-consistent and can be smoothed to retain the $\sqrt{n}$-rate of consistency.
The cross-sectional approaches no longer work for irregular data.
Estimation of Mean and Covariance Functions: Irregular Functional Data

- The cross-sectional approaches no longer work for irregular data.
  - Smoothing is needed to estimate the mean function.
Estimation of Mean and Covariance Functions: Irregular Functional Data

- The cross-sectional approaches no longer work for irregular data.

  - Smoothing is needed to estimate the mean function.

  - Such a mean estimate could be $\sqrt{n}$-consistent (pointwise) for "dense" functional data but will have nonparametric rates otherwise. (More about the theory and formal definition of dense functional data later.)
Estimation of Mean and Covariance Functions: Irregular Functional Data

- The cross-sectional approaches no longer work for irregular data.

- Smoothing is needed to estimate the mean function.

- Such a mean estimate could be $\sqrt{n}$-consistent (pointwise) for “dense” functional data but will have nonparametric rates otherwise. (More about the theory and formal definition of dense functional data later.)

- Likewise, 2D smoothing is needed to estimate the covariance function and $\sqrt{n}$-consistency can be achieved for dense functional data.
CD4 Counts of First 25 Patients
CD4 Counts of first 25 Patients
Mean Curve: CD4 Counts of first 25 Patients
Mean Curve: CD4 counts of all patients
If we employ the local linear smoother, the estimate for the mean function is:

\[ \hat{\mu}(t) = \hat{\beta}_0, \quad \text{where} \]

\[ (\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[ Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) \right]^2 K_{h_\mu}(T_{ij} - t). \]
Remarks

- In addition to the local linear (or polynomial) smoother, any smoothing method, such as penalized splines, B-splines, Wavelets, and Fourier filtering, can be employed.
Remarks

- In addition to the local linear (or polynomial) smoother, any smoothing method, such as penalized splines, B-splines, Wavelets, and Fourier filtering, can be employed.

- These smoothing methods (scatter plot smoothers) can also be applied to dense data, whether regular or not, so a unified approach is feasible.
In addition to the local linear (or polynomial) smoother, any smoothing method, such as penalized splines, B-splines, Wavelets, and Fourier filtering, can be employed.

These smoothing methods (scatter plot smoothers) can also be applied to dense data, whether regular or not, so a unified approach is feasible.

Another common approach for dense data is to pre-smooth the data from each subject separately, then take the cross-section mean at each time point $t$. 
Remarks

This is the case for the spectrum data.
Remarks

Such a pre-smoothing typically aims at:

(i) getting rid of the noise in the data,
(ii) revealing the latent smooth curve for each subject.

Remarks

Such a pre-smoothing typically aims at:

(i) getting rid of the noise in the data,
(ii) revealing the latent smooth curve for each subject.


However, both depend on the amount of smoothing and design of the dense data, so whether the mission has been accomplished is unclear.

Because of this uncertainty and because pre-smoothing alters the data, we prefer not to adopt a pre-smoothing approach.
End of Estimation of Mean Function
Our target is $\Sigma(s, t) = \text{cov}(X(s), X(t))$, but we do not observe $X$. 

Estimation of Covariance Function

- Our target is $\Sigma(s, t) = \text{cov}(X(s), X(t))$, but we do not observe $X$.

- Observe $Y_{ij} = Y_i(t_{ij}) = X_{ij} + e_{ij}$, where $\text{var}(e_{ij}) = \sigma^2(t_{ij})$.

  \[ \Rightarrow \text{cov}(Y(s), Y(t)) = \text{cov}(X(s), X(t)), \text{ when } s \neq t, \]

  \[ \text{var}(Y(t)) = \text{cov}(Y(t), Y(t)) = \text{cov}(X(t), X(t)) + \sigma^2(t). \]
Estimation of Covariance Function

- Our target is $\Sigma(s, t) = \text{cov}(X(s), X(t))$, but we do not observe $X$.

- Observe $Y_{ij} = Y_i(t_{ij}) = X_{ij} + e_{ij}$, where $\text{var}(e_{ij}) = \sigma^2(t_{ij})$.

\[
\Rightarrow \text{cov}(Y(s), Y(t)) = \text{cov}(X(s), X(t)), \text{ when } s \neq t,
\]

\[
\text{var}(Y(t)) = \text{cov}(Y(t), Y(t)) = \text{cov}(X(t), X(t)) + \sigma^2(t).
\]

- This means we need to handle the diagonal part of $\text{cov}(Y)$, which corresponds to the variance of $Y$, differently from the rest of $\text{cov}(Y)$. 
Raw Covariance Plot:

\[ \{ Y(t_{ij}) - \hat{\mu}(t_{ij}) \} \{ Y(t_{ik}) - \hat{\mu}(t_{ik}) \}, \quad \forall \ i, j, k \]
Design Plot for Covariance: \((t_{ij}, t_{ik}), \forall i, j, k\)
Design Plot for One Subject: \((t_{ij}, t_{ik}), \forall j, k\)
Design Plot for Two Subjects: \((t_{ij}, t_{ik}), \forall j, k\)
Design Plot for All Subjects: $(t_{ij}, t_{ik}), \forall i, j, k$
Raw Covariance Plot: Diagonal Data in Black
Raw Covariance with Diagonal Data Removed
Estimated Covariance Surface of $X(t)$
**Estimated Covariance Surface of** \( X(t) \)

Let \( D_{ijl} = (Y_{ij} - \hat{\mu}(t_{ij}))(Y_{il} - \hat{\mu}(t_{il})) \), be the raw covariances.

Employing a local linear smoother, the estimate for \( \Sigma(s, t) \) is:

\[
\hat{\Sigma}(s, t) = \hat{\beta}_0, \quad \text{where}
\]

\[
(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^{n} \sum_{1 \leq j \neq l \leq N_i} \left[ D_{ijl} - \beta_0 - \beta_1(t_{ij} - s) - \beta_2(t_{il} - t) \right]^2 K_{h_S}(t_{ij} - s)K_{h_S}(t_{il} - t).
\]
Raw Variance Plot: \[ Y(t_{ij}) - \mu(t_{ij}) \]^2, \ \forall \ i, j \]
Estimated Variance function of $Y(t)$
Estimated Covariance & Variance of $Y(t)$
Estimates of $\sigma^2(t)$: Variance of Measurement Errors

\[ \sigma^2(t) = \text{var}(Y(t)) - \text{var}(X(t)). \]
Estimates of $\sigma^2(t)$: Variance of Measurement Errors

$\sigma^2(t) = \text{var}(Y(t)) - \text{var}(X(t))$.

When $\sigma^2(t) = \sigma^2$ for all $t$, one can estimate $\sigma^2$ by $\int_I \hat{\sigma}^2(t)dt$.

Due to boundary effects, PACE replaces $I$ by a sub-interval.
Theory: Mean and Covariance Estimation
(Zhang and W., 2014)

- We adopt a unified approach that works for both dense and sparse functional data.

- Use scatter plot smoothers to estimate the mean and covariance functions that remove the measurement errors.
  - This is the method of PACE (a Matlab package for functional data), first proposed in Yao, M. and W. (2005, JASA).
We adopt a unified approach that works for both dense and sparse functional data.

Use scatter plot smoothers to estimate the mean and covariance functions that remove the measurement errors.
- This is the method of PACE (a Matlab package for functional data), first proposed in Yao, M. and W. (2005, JASA).

Next, we present asymptotic properties of these estimators and define what “dense” functional data means.

We also show another way to smooth the data by Li and Hsing (2010) and compare it with the scatter plot smoother in PACE.
Estimation of Mean Function

- Equal weight per observation (Yao M. and W., 2005)

\[
\hat{\mu}_{obs}(t) = \hat{\beta}_0 \text{ where }
\]

\[
(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[ Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) \right]^2 K_{h\mu}(T_{ij} - t)
\]
Estimation of Mean Function

- Equal weight per observation (Yao M. and W., 2005)

\[ \hat{\mu}_{\text{obs}}(t) = \hat{\beta}_0 \quad \text{where} \]

\[ (\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left[ Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) \right]^2 K_{h_\mu}(T_{ij} - t) \]

- Equal weight per subject (Li and Hsing, 2010)

\[ \hat{\mu}_{\text{sub}}(t) = \hat{\beta}_0 \quad \text{where} \]

\[ (\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \frac{1}{N_i} \sum_{j=1}^{N_i} \left[ Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) \right]^2 K_{h_\mu}(T_{ij} - t) \]
Important notation

\[ n = \text{Sample size}, \quad N_i = \# \text{observations of subject } i. \]
Important notation

$n = \text{Sample size, } N_i = \# \text{ observations of subject } i.$

- OBS: arithmetic mean of $N_i$ and of $N_i^2$,

$$\bar{N} = \frac{1}{n} \sum_{i=1}^{n} N_i, \quad \bar{N}_{S2} = \frac{1}{n} \sum_{i=1}^{n} N_i^2.$$
Important notation

\( n = \) Sample size, \( N_i = \) # observations of subject \( i \).

- **OBS**: arithmetic mean of \( N_i \) and of \( N_i^2 \),

\[
\overline{N} = \frac{1}{n} \sum_{i=1}^{n} N_i, \quad \overline{N}_{S2} = \frac{1}{n} \sum_{i=1}^{n} N_i^2.
\]

- **SUB**: harmonic mean of \( N_i \),

\[
\overline{N}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{N_i}}.
\]
$L^2$ Convergence: Non-random $N_i$

Equal weight per observation:

$$\| \hat{\mu}_{obs} - \mu \|_2 = O_p \left( h_{\mu}^2 + \sqrt{ \left( 1 + \frac{1}{\bar{N} h_{\mu}} \right) \frac{1}{n}} \right).$$
**Convergence: Non-random $\mathcal{N}_i$**

- Equal weight per observation:

$$\| \hat{\mu}_{obs} - \mu \|_2 = O_p \left( h^2_\mu + \sqrt{\left( 1 + \frac{1}{\bar{N}h_\mu} \right) \frac{1}{n}} \right).$$

- Equal weight per subject:

$$\| \hat{\mu}_{sub} - \mu \|_2 = O_p \left( h^2_\mu + \sqrt{\left( 1 + \frac{1}{\bar{N}_H h_\mu} \right) \frac{1}{n}} \right).$$
Uniform Convergence

- Equal weight per observation:

\[
\sup_{t \in [0,1]} |\hat{\mu}_{obs}(t) - \mu(t)| = O \left( h_{\mu}^2 + \sqrt{ \left( 1 + \frac{1}{\bar{N} h_{\mu}} \right) \frac{\log(n)}{n} } \right) \quad \text{a.s.}
\]
Uniform Convergence

- Equal weight per observation:

\[
\sup_{t \in [0,1]} |\hat{\mu}_{obs}(t) - \mu(t)| = O \left( h^2_{\mu} + \sqrt{\left(1 + \frac{1}{N_{\mu}}\right) \frac{\log(n)}{n}} \right) \quad a.s.
\]

- Equal weight per subject:

\[
\sup_{t \in [0,1]} |\hat{\mu}_{sub}(t) - \mu(t)| = O \left( h^2_{\mu} + \sqrt{\left(1 + \frac{1}{N_H h_{\mu}}\right) \frac{\log(n)}{n}} \right) \quad a.s.
\]
Asymptotic Normality

- OBS: Assume \( \limsup_n (\bar{N} S^2)/(\bar{N})^2 < \infty \)

\[
\left[ \Gamma_{obs}(t) \right]^{-1/2} \left\{ \hat{\mu}_{obs}(t) - \mu(t) - \frac{1}{2} h^2 \sigma^2 K \mu^{(2)}(t) \right\} \overset{d}{\to} \mathcal{N}(0, 1),
\]

where

\[
\Gamma_{obs}(t) = \|K\|^2 \frac{\Sigma(t, t) + \sigma^2}{n \bar{N} h_{\mu} f(t)} + \frac{(\bar{N} S^2 - \bar{N})}{n (\bar{N})^2} \Sigma(t, t).
\]
Asymptotic Normality

• OBS: Assume \( \lim_{n \to \infty} \sup_n (\bar{N}_{S2})/(\bar{N})^2 < \infty \)

\[
[\Gamma_{obs}(t)]^{-1/2} \{ \hat{\mu}_{obs}(t) - \mu(t) - \frac{1}{2} h_\mu^2 \sigma_K^2 \mu^{(2)}(t) \} \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{where}
\]

\[
\Gamma_{obs}(t) = \| K \|^2 \frac{\sum(t, t) + \sigma^2}{n \bar{N} h_\mu f(t)} + \frac{(\bar{N}_{S2} - \bar{N})}{n(\bar{N})^2} \Sigma(t, t).
\]

• SUB

\[
[\Gamma_{sub}(t)]^{-1/2} \{ \hat{\mu}_{sub}(t) - \mu(t) - \frac{1}{2} h_\mu^2 \sigma_K^2 \mu^{(2)}(t) \} \xrightarrow{d} \mathcal{N}(0, 1), \quad \text{where}
\]

\[
\Gamma_{sub}(t) = \| K \|^2 \frac{\sum(t, t) + \sigma^2}{n \bar{N}_H h_\mu f(t)} + \frac{1}{n} \left( 1 - \frac{1}{\bar{N}_H} \right) \Sigma(t, t)
\]

\{z\} asymptotic variance

\{z\} asymptotic bias
Rates of Convergence

- Non-dense: Slower than $\sqrt{n}$-rate.
  Sparse data $= \text{finite } N_i$, is a special case.
Rates of Convergence

- Non-dense: Slower than $\sqrt{n}$-rate.
  Sparse data $= \text{finite } N_i$, is a special case.

- Dense: $\sqrt{n}$-rate, with asymptotic bias
  (between non- and parametric paradigm)
Rates of Convergence

- Non-dense: Slower than $\sqrt{n}$-rate.
  Sparse data $=$ finite $N_i$, is a special case.

- Dense: $\sqrt{n}$-rate, with asymptotic bias
  (between non- and parametric paradigm)

- Ultra-dense: $\sqrt{n}$-rate, no asymptotic bias
  (Parametric paradigm)
Partition of Functional Data: OBS

- **Non-Dense data:** When $\N/n^{1/4} \to 0$ and $h_\mu \asymp (n\N)^{-1/5}$,

  $\sqrt{n\N} h_\mu [\hat{\mu}_{obs}(t) - \mu(t) - \frac{1}{2} h_\mu^2 \sigma_K^2(t)] \overset{d}{\to} \mathcal{N}\left(0, \|K\|^2 \frac{\Sigma(t, t) + \sigma^2}{f(t)}\right).$

- **Dense data:** When $\N/n^{1/4} \to C$ and $h_\mu/n^{-1/4} \to C_1$ where $0 < C, C_1 < \infty$,

  $\sqrt{n}[\hat{\mu}_{obs}(t) - \mu(t) - \frac{1}{2} h_\mu^2 \sigma_K^2(t)] \overset{d}{\to} \mathcal{N}\left(0, \|K\|^2 \frac{\Sigma(t, t) + \sigma^2}{f(t)C \cdot C_1} + \frac{\N S^2}{(\N)^2} \Sigma(t, t)\right).$

- **Ultra-Dense data:** When $\N/n^{1/4} \to \infty$, $h_\mu = o(n^{-1/4})$, and $h_\mu \N \to \infty$,

  $\sqrt{n}[\hat{\mu}_{obs}(t) - \mu(t)] \overset{d}{\to} \mathcal{N}\left(0, \frac{\N S^2}{(\N)^2} \Sigma(t, t)\right).$
**Non-Dense data:** When $\bar{N}_H/n^{1/4} \to 0$ and $h_\mu \asymp (n\bar{N}_H)^{-1/5}$,

$$\sqrt{n\bar{N}_H h_\mu} [\hat{\mu}_{\text{sub}}(t) - \mu(t) - \frac{1}{2} h_\mu^2 \sigma_K^2 \mu^{(2)}(t)] \xrightarrow{d} \mathcal{N} \left( 0, \|K\|_2^2 \frac{\Sigma(t,t) + \sigma^2}{f(t)} \right).$$

**Dense data:** When $\bar{N}_H/n^{1/4} \to C$ and $h_\mu/n^{-1/4} \to C_1$ where $0 < C, C_1 < \infty$,

$$\sqrt{n} [\hat{\mu}_{\text{sub}}(t) - \mu(t) - \frac{1}{2} h_\mu^2 \sigma_K^2 \mu^{(2)}(t)] \xrightarrow{d} \mathcal{N} \left( 0, \|K\|_2^2 \frac{\Sigma(t,t) + \sigma^2}{f(t)C \cdot C_1} + \Sigma(t,t) \right).$$

**Ultra-Dense data:** When $\bar{N}_H/n^{1/4} \to \infty$, $h_\mu = o(n^{-1/4})$, and $h_\mu \bar{N}_H \to \infty$,

$$\sqrt{n} [\hat{\mu}_{\text{sub}}(t) - \mu(t)] \xrightarrow{d} \mathcal{N} \left( 0, \Sigma(t,t) \right).$$
Asymptotic bias: both $\frac{1}{2} h^2 \mu^2 \sigma^2_K \mu^{(2)}(t)$. 
Comparison of Two Schemes

1. Asymptotic bias: both \( \frac{1}{2} h^2 \mu \sigma^2 K \mu^{(2)}(t) \).

2. Asymptotic variance:
   - Non-dense data: \( \hat{\mu}_{obs} \leq \hat{\mu}_{sub} \), so OBS more efficient.
   - Ultra-dense data: \( \hat{\mu}_{obs} \geq \hat{\mu}_{sub} \), so SUB more efficient.
Equal weight per observation: For $s = t,$

\[
\Gamma_{obs}(t, t)^{-1/2} \left[ \hat{\Sigma}_{obs}(t, t) - \Sigma(t, t) - \frac{1}{2} h_S^2 \sigma_K^2 \left( \frac{\partial^2 \Sigma}{\partial s^2} (t, t) + \frac{\partial^2 \Sigma}{\partial t^2} (t, t) \right) \right] \xrightarrow{d} \mathcal{N}(0, 1),
\]

\[
\Gamma_{obs}(t, t) = \frac{1}{n h_S^2} \left[ 2 P_1(N) \| K \|^4 \frac{V_1(t, t)}{f(t)^2} + 4 h_S P_2(N) \| K \|^2 \frac{V_2(t, t)}{f(t)} + h_S^2 P_3(N) V_1(t, t) \right]
\]
Asymptotic Normality: Variance function

Equal weight per subject: Let

\[ D_{ijl} = (Y_{ij} - \hat{\mu}_{sub}(T_{ij}))(Y_{il} - \hat{\mu}_{sub}(T_{il})). \]

The weight \( 1/(N_i(N_i - 1)) \) is attached to each observation for the \( i \)th subject. Thus \( \hat{\Sigma}_{sub}(s, t) = \hat{\beta}_0 \)

where

\[
(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \arg\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^{n} \frac{1}{N_i(N_i - 1)} \sum_{1 \leq j \neq l \leq N_i} \left[ D_{ijl} - \beta_0 - \beta_1(T_{ij} - s) - \beta_2(T_{il} - t) \right]^2 K_{h \Sigma}(T_{ij} - s)K_{h \Sigma}(T_{il} - t)
\]

Asymptotic Normality: Replace \( P_1(N) \), \( P_2(N) \), \( P_3(N) \) with \( P_4(N) \), \( P_5(N) \), \( P_6(N) \).
Covariance function

Unified asymptotic normality:

- Three partitions: non-dense, dense, and ultra-dense.

- Two weighing schemes:
  \( \hat{\Sigma}_{\text{obs}} \) more efficient for non-dense data;
  \( \hat{\Sigma}_{\text{sub}} \) more efficient for ultra-dense data.

- *Discontinuity of the asymptotic variance*: asymptotic variance expressions are different between \( s = t \) and \( s \neq t \).
Discontinuity of the asymptotic variance

OBS:

- For $s \neq t$, the asymptotic variance of $\hat{\Sigma}_{obs}(s, t)$ is

$$\frac{P_1(N)\|K\|^4 V_1(s, t)}{n h^2 \Sigma f(s)f(t)} + \frac{P_2(N)\|K\|^2 [f(s)V_2(t, s) + f(t)V_2(s, t)]}{n h \Sigma f(s)f(t)} + \frac{P_3(N)V_1(s, t)}{n}.$$

- For $s = t$, the asymptotic variance of $\hat{\Sigma}_{obs}(s, t)$ is

$$2 \frac{P_1(N)\|K\|^4 V_1(t, t)}{f(t)^2} + 4 \frac{P_2(N)\|K\|^2 V_2(t, t)}{n h \Sigma f(t)} + \frac{P_3(N)V_1(t, t)}{n}.$$
Discontinuity of the variance

Explanation:

\[ E[K_h(T - t)K_h(T - s)] = 0 \] for \( s \neq t \) when \( h \to 0 \) for \( K \) on an interval;

\[ E[K_h(T - t)K_h(T - s)] = \|K\|^2f(t)/h + o(1/h) \] for \( s = t \).
Discontinuity of the variance

Explanation:

\[ E[K_h(T-t)K_h(T-s)] = 0 \] for \( s \neq t \) when \( h \to 0 \) for \( K \) on an interval;

\[ E[K_h(T-t)K_h(T-s)] = \|K\|^2 f(t)/h + o(1/h) \] for \( s = t \).

- Intuitive explanation?
End of Asymptotic Theory
References for Mean and Covariance Function

- Yao Mueller and Wang (2005, JASA)
- Li and Hsing (2010, AoS)
- Zhang and Wang (2014, Manuscript)

**Additional References:**

End of Lecture 1: Mean and Covariance Functions
What is the difference between functional and time-series data?
What is the difference between functional and time-series data?

What is the benefit of functional data over multivariate data?
Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- What are the three types of functional data?
- What are the pros and cons for each type?
  - Which one is the easiest or hardest to handle?
  (We’ll discuss how to handle them in the next section.)
Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- What are the three types of functional data?
- What are the pros and cons for each type?
  - Which one is the easiest or hardest to handle? (We’ll discuss how to handle them in the next section.)
- Is there a curse of dimensionality for dense functional data?
  - What does “dense” mean?
For regular or fully observed functional data, the cross-sectional mean at time $t$ can be used to estimate the mean function.

- Such a mean estimate is $\sqrt{n}$-consistent (pointwise) even in the presence of measurement errors.  (WHY?)

- The cross-section mean can further be smoothed to obtain a smooth mean estimate (This requires a dense measurement schedule).

- Think how to make the resulted mean estimate still $\sqrt{n}$-consistent.

Likewise, the sample covariance matrix is also $\sqrt{n}$-consistent and can be smoothed to retain the $\sqrt{n}$-rate of consistency.
Outline

1. Introduction

2. Mean and Covariance Functions

3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?

4. Inverse Problem in Functional Correlations and Regression

5. Stringing high-dim data to functional data

6. Next Generation Functional Data
Since functional data are intrinsically infinite dimension, their analyses often rely on dimension reduction methods.
Dimension Reduction for Functional Data

Since functional data are intrinsically infinite dimension, their analyses often rely on dimension reduction methods.

There are two types of dimension reduction for functional data:
- one on the data themselves and
- another on statistical modeling of such data.
Since functional data are intrinsically infinite dimension, their analyses often rely on dimension reduction methods.

There are two types of dimension reduction for functional data:
- one on the data themselves and
- another on statistical modeling of such data.

The latter is the topic of functional regression (Lecture 3).
Since functional data are intrinsically infinite dimension, their analyses often rely on dimension reduction methods.

There are two types of dimension reduction for functional data:
- one on the data themselves and
- another on statistical modeling of such data.

The latter is the topic of functional regression (Lecture 3).

In this section we focus on the first one:
  dimension reduction on the data.
Principal component analysis for multivariate \((p\text{-dim})\) data is a dimension reduction tool to transform (linearly) the data to orthogonal \((p\text{-dim})\) data so that the first few \((k)\) of them explains most of the variation.

The first eigenfunction \(\lambda_1 = \text{argmax}_{1 \leq k \leq p} \text{var}(hX_i)\) represents the direction of the data with the largest variation. The second eigenfunction \(\lambda_2 = \text{argmax}_{1 \leq k \leq p; k=1} \text{var}(h(X_i))\).
Principal component analysis for multivariate \((p\text{-dim})\) data is a dimension reduction tool to transform (linearly) the data to orthogonal \((p\text{-dim})\) data so that the first few \((k)\) of them explains most of the variation.

- The first eigenfunction \(\phi_1 = \arg\max_{\phi \in \mathbb{R}^p, \|\phi\|=1} \text{var}(\langle X - \mu, \phi \rangle)\)
  - The first principal direction \(\phi_1\), represents the direction of the data with the largest variation.
Principal component analysis for multivariate \((p\text{-dim})\) data is a dimension reduction tool to transform (linearly) the data to orthogonal \((p\text{-dim})\) data so that the first few \((k)\) of them explains most of the variation.

The first eigenfunction \(\phi_1 = \arg\max_{\phi \in \mathbb{R}^p, \|\phi\| = 1} \text{var}(\langle X - \mu, \phi \rangle)\)
- The first principal direction \(\phi_1\), represents the direction of the data with the largest variation.

The second eigenfunction \(\phi_2 = \arg\max_{\phi \in \mathbb{R}^p, \|\phi\| = 1, \phi \perp \phi_1} \text{var}(\langle (X - \mu), \phi \rangle)\)
Which one is the least squares line?
Review of PCA

- The least squares line minimizes the vertical squared distance, but the first PC line minimized the perpendicular squared distance.

- In this example the first PC explains 75% of the variations.
How would you extend PCA to functional data?

- The first eigenfunction \( \phi_1 = \arg\max_{\phi \in \mathcal{R}^p, \|\phi\| = 1} \text{var}(\langle X - \mu, \phi \rangle) \)
  
  - The first principal direction \( \phi_1 \), represents the direction of the data with the largest variation.

- The second eigenfunction \( \phi_2 = \arg\max_{\phi \perp \phi_1, \phi \in \mathcal{R}^p, \|\phi\| = 1} \text{var}(\langle X - \mu, \phi \rangle) \)
How would you extend PCA to functional data?

- The first eigenfunction \( \phi_1 = \arg\max_{\phi \in \mathbb{R}^p, \|\phi\|=1} \text{var}(\langle X - \mu, \phi \rangle) \)
  
  - The first principal direction \( \phi_1 \), represents the direction of the data with the largest variation.

- The second eigenfunction \( \phi_2 = \arg\max_{\phi \perp \phi_1, \phi \in \mathbb{R}^p, \|\phi\|=1} \text{var}(\langle X - \mu, \phi \rangle) \)

- All we have to do is to change the inner product from a vector to a function in \( L^2 \), a Hilbert space (this explains why we need a Hilbert space structure for functional data).

\[ \langle f, g \rangle = \int_I f(t)g(t)dt, \text{ for any functions } f \text{ and } g \text{ in } L^2(I). \]
Definition of FPCA

- The first eigenfunction $\phi_1$
  \[
  \phi_1 = \arg\max_{\phi \in L^2(I), \|\phi\|=1} \text{var}(\langle X - \mu, \phi \rangle)
  \]
  \[
  = \arg\max_{\phi \in L^2(I), \|\phi\|=1} \text{var}(\int [X(t) - \mu(t)]\phi(t)dt)
  \]
  The first principal direction $\phi_1$, represents the direction of the data with the largest variation.

- The second eigenfunction $\phi_2$
  \[
  \phi_2 = \arg\max_{\phi \perp \phi_1, \phi \in L^2(I), \|\phi\|=1} \text{var}(\int [X(t) - \mu(t)]\phi(t)dt)
  \]
For vector $X$, the first eigenfunction $\phi_1$

$$= \arg\max_{\phi \in \mathcal{R}^p, \|\phi\| = 1} \text{var}(\langle X - \mu, \phi \rangle)$$

is equivalent to the eigenfunction corresponding to the largest eigenvalue of the covariance matrix $\text{cov}(X)$. 
Properties of FPCA

- For vector $X$, the first eigenfunction $\phi_1$
  
  \[ = \underset{\phi \in \mathcal{R}^p, \|\phi\|=1}{\arg\max} \text{var}(\langle X - \mu, \phi \rangle) \]

  is equivalent to the eigenfunction corresponding to the largest eigenvalue of the covariance matrix $\text{cov}(X)$.

  \[ \implies \text{cov}(X)\phi_1 = \lambda_1 \phi_1 \quad \text{and} \quad \text{cov}(X) = \sum_{k=1}^{p} \lambda_k \phi_k^T \phi_k. \]

- This concept can be extended to function data but we need to define what “$\text{cov}(X)\phi_1$” mean.
This leads to the definition of a covariance operator.

Covariance function: $\Sigma(s, t) = \text{cov}(X(s), X(t))$, $s \& t \in I$.

We'll use the same notation $\Sigma$ for the covariance function and its operator and define the covariance operator $\Sigma$ (from $L^2(I)$ to $L^2(I)$) as:

$$\Sigma(f) = \int_I \Sigma(s, t)f(s)ds, \text{ for any } f \in L^2(I).$$
Properties of FPCA

- This leads to the definition of a covariance operator.

  Covariance function: \( \Sigma(s, t) = \text{cov}(X(s), X(t)), s \& t \in I. \)

  We’ll use the same notation \( \Sigma \) for the covariance function and its operator and define the covariance operator \( \Sigma \) (from \( L^2(I) \) to \( L^2(I) \)) as:

  \[
  \Sigma(f) = \int_I \Sigma(s, t)f(s)ds, \quad \text{for any } f \in L^2(I).
  \]

- FPCA = spectral decomposition of the covariance operator:

  \[
  \Sigma(\phi_k) = \lambda_k \phi_k,
  \]

  \( \lambda_k \) \& \( \phi_k \) are the eigenvalues and eigenfunctions of \( \Sigma \).
Properties of FPCA

- Mercer’s theorem (with mild assumption) implies that

\[ \Sigma(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t), \]

and the convergence above is uniform over \( s \) and \( t \).
Properties of FPCA

- Mercer’s theorem (with mild assumption) implies that

\[ \Sigma(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t), \]

and the convergence above is uniform over \( s \) and \( t \).

- This leads to the Karhunen-Loève decomposition:

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

\[ \text{var}(A_k) = \lambda_k, \text{ the } k\text{-th largest eigenvalue of } \Sigma, \]

\[ A_k = \int_I [X(t) - \mu(t)] \phi_k(t) dt, \text{ are uncorrelated PC (scores)}. \]
Steps to FPCA (Yao, Müller & Wang, 2005)

1. Estimate the mean $\mu(t)$ and covariance $\Sigma(s, t)$. (This usually involves smoothing)
Steps to FPCA (Yao, Müller & Wang, 2005)

1. Estimate the mean $\mu(t)$ and covariance $\Sigma(s, t)$. (This usually involves smoothing)

2. Estimate the eigenvalues and eigenfunctions of $\Sigma(s, t)$. 
Steps to FPCA (Yao, Müller & Wang, 2005)

1. Estimate the mean $\mu(t)$ and covariance $\Sigma(s, t)$. (This usually involves smoothing)

2. Estimate the eigenvalues and eigenfunctions of $\Sigma(s, t)$.

3. Estimate the PC scores $A_k = \int (X(t) - \mu(t))\phi_k(t)dt$. 
Steps to FPCA (Yao, Müller & Wang, 2005)

1. Estimate the PC scores $A_k = \int (X(t) - \mu(t))\phi_k(t)dt$.

For functional data that are recorded on a dense grid ($N_i \to \infty$), numerical integration method can be employed to estimate the PC score.
Steps to FPCA (Yao, Müller & Wang, 2005)

1. Estimate the PC scores $A_k = \int (X(t) - \mu(t))\phi_k(t)dt$.

- For functional data that are recorded on a dense grid ($N_i \to \infty$), numerical integration method can be employed to estimate the PC score.

- When functional data are observed at a few time points, the numerical integration method does not work.
1. Estimate the PC scores $A_k = \int (X(t) - \mu(t))\phi_k(t)dt$.

   For functional data that are recorded on a dense grid ($N_i \to \infty$), numerical integration method can be employed to estimate the PC score.

   When functional data are observed at a few time points, the numerical integration method does not work.

   Yao et al. (2005) proposed PACE to resolve this issue.

   $\hat{A}_{ik} = \hat{E}(A_{ik}|Y_i) = \hat{\lambda}_k\hat{\phi}_k^T\hat{\Sigma}_i^{-1}(Y_i - \mu_i)$
Use Karhunen Loève decomposition to recover the latent curve

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

\[
\downarrow
\]

\[
\hat{X}_{ik}(t) = \hat{\mu}(t) + \sum_{k=1}^{\infty} \hat{A}_{ik} \hat{\phi}_k(t)
\]
The mean function can be estimated at the one-dim nonparametric rate.

Perturbation theory further implies that the eigen-values and eigen-functions can be estimated at the 2-dim nonparametric rate, but this rate is suboptimal!

Hall, M. and W. (2006) showed that by undersmoothing the covariance function:

the first $K$ (finite) eigen-values can be estimated at the $\sqrt{n}$ rate,
the first $K$ (finite) eigen-functions can be estimated at the one-dim nonparametric rate.
Convergence Rates for FPCA

- The mean function can be estimated at the one-dim nonparametric rate.

- The covariance function can be estimated at the 2-dim nonparametric rate.

- Perturbation theory further implies that the eigen-values and eigen-functions can be estimated at the 2-dim nonparametric rate, but this rate is suboptimal!
Convergence Rates for FPCA

- The mean function can be estimated at the one-dim nonparametric rate.

- The covariance function can be estimated at the 2-dim nonparametric rate.
  - Perturbation theory further implies that the eigen-values and eigen-functions can be estimated at the 2-dim nonparametric rate, but this rate is suboptimal!

- Hall, M. and W. (2006) showed that by undersmoothing the covariance function:
  the first $K$ (finite) eigen-values can be estimated at the $\sqrt{n}$-rate, the first $K$ (finite) eigen-functions can be estimated at the one-dim nonparametric rate.
AIDS CD4: 6 Randomly Selected Subjects
What Does FPCA Offer?

- The Karhunen-Loève decomposition is useful to impute functional data.
What Does FPCA Offer?

- The Karhunen-Loève decomposition is useful to impute functional data.

- FPCA provides a way to project an infinite dimensional function onto a finite $K$-dimensional subspace, the space spanned by the first $K$ eigenfunctions.

  → The main information of functional data can be summarized by finitely many ($K$) PC components.
The Karhunen-Loève decomposition is useful to impute functional data.

FPCA provides a way to project an infinite dimensional function onto a finite $K$-dimensional subspace, the space spanned by the first $K$ eigenfunctions.

$\Rightarrow$ The main information of functional data can be summarized by finitely many ($K$) PC components.

(The proportion of variation explained by the first $K$ PC components is the ratio $\frac{\sum_{k=1}^{K} \lambda_k}{\sum_{k=1}^{\infty} \lambda_k}$.)

What Does FPCA Offer?
What Does FPCA Offer?

- This process transfers functional data to $K$-dim multivariate data consisting of the first $K$ PC scores, so any existing method for multivariate data can be applied to these scores providing off-the-shelf methods for functional data.

**Examples**: Clustering and classification of functional data.
What Does FPCA Offer?

While it is possible to expand functional data with any basis functions, such as B-splines, Fourier bases, wavelets, etc., FPCA provides the most parsimonious way to do so.

By virtue of its definition, FPCA requires less components than other basis functions.
While it is possible to expand functional data with any basis functions, such as B-splines, Fourier bases, wavelets, etc., FPCA provides the most parsimonious way to do so.

By virtue of its definition, FPCA requires less components than other basis functions.

However, the basis functions for FPCA needs to be estimated, which makes the theory a little harder than for preselected basis functions.
What Does FPCA Offer?

- The shape of $\phi_k$ may help us to settle for a more parsimonious or parametric model.
  **Example**: CD4 counts.
AIDS CD4: Eigenfunctions
AIDS data has been modeled as linear mixed-effects model with linear time trend and random effects on the intercepts only. However, we can increase the total variation explained to over 96% if a second PC is added. Note that even with two components, the computational effort may be less than that for a good parametric random effects model, which may need five random effects (if a fourth order polynomial is used to capture the shape of the AIDS data).
AIDS data has been modeled as linear mixed-effects model with linear time trend and random effects on the intercepts only.

The linear time trend is arguable but the random intercept is not far off as the first eigenfunction looks flat and already explains over 84% of the variation in the data.
AIDS data has been modeled as linear mixed-effects model with linear time trend and random effects on the intercepts only.

The linear time trend is arguable but the random intercept is not far off as the first eigenfunction looks flat and already explains over 84% of the variation in the data.

However, we can increase the total variation explained to over 96% if a second PC is added.

- Note that even with two components, the computational effort may be less than that for a good parametric random effects model, which may need five random effects (if a fourth order polynomial is used to capture the shape of the AIDS data).
Mean Curve: CD4 counts of all patients
What Else can FPCA Offer?

- The principal directions $\phi_k$ explain the modes of variations of functional data (Rice and Jones, 1991).
AIDS CD4: Modes of Variation

![Graph showing CD4 count over time with different modes of variation.](image-url)
AIDS CD4: Modes of Variation
AIDS CD4: Modes of Variation
The above plots remind us of Tukey’s Box plot for scalar data.

Box plots have been extended to functional data, but the extension is non trivial and still open for improvements!

**Why?**
Functional Box Plot

- The above plots remind us of Tukey’s Box plot for scalar data.

- Box plots have been extended to functional data, but the extension is non trivial and still open for improvements!

  **Why?**

- There are two R-packages for functional box plots.
  
  Hyndman and Shang (2010, JCGS) - rainbow, box, and bag plots
  Sun and Genton (2011, JCGS) - boxplots
Yao, Müller and Wang (2005, JASA)  
Methods and theory

Hall, Müller and Wang (2006, AOS)  
Improved theory on eigenfunctions and eigenvalues.
References for FPCA

- **Functional data**
  - Dauxois, Pousse & Romain (1982)
  - Rice & Silverman (1991)
  - Cardot (2000)
  - Hall & Hosseini-Nasab (2006)

- **Longitudinal data**
  - Shi, Weiss & Taylor (1996)
  - James, Sugar & Hastie (2000)
  - Rice & Wu (2001)
  - Yao, Müller & Wang (2005)
End of FPCA
Outline

1. Introduction

2. Mean and Covariance Functions

3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?

4. Inverse Problem in Functional Correlations and Regression

5. Stringing high-dim data to functional data

6. Next Generation Functional Data
Covariate Adjusted FPCA

The above FPCA assumes that data come from one population. What if we have additional information of a covariate $Z$ (a scalar) or $Z(t)$ (a functional or longitudinal covariate)?

- This is straightforward for dense functional data with scalar covariate $Z$.
  
  Chiou, Müller & Wang (2003)
  Cardot (2006)
The above FPCA assumes that data come from one population. What if we have additional information of a covariate $Z$ (a scalar) or $Z(t)$ (a functional or longitudinal covariate)?

- This is straightforward for dense functional data with scalar covariate $Z$.

  Chiou, Müller & Wang (2003)
  Cardot (2006)

- Their methods do not work for sparse functional data or longitudinal covariates.
First, pool all the data together to get the overall mean function $\mu(\bullet)$ and eigenfunctions $\phi_k(\bullet)$ of the overall covariance function.
First, pool all the data together to get the overall mean function \( \mu(\cdot) \) and eigenfunctions \( \phi_k(\cdot) \) of the overall covariance function.

Next, incorporate the covariate information through the conditional mean function:

\[
\mu(t, Z) = E(Y(t)|Z) = \mu(t) + \sum_k E(A_k|Z)\phi_k(t)
\]
First, pool all the data together to get the overall mean function $\mu(\cdot)$ and eigenfunctions $\phi_k(\cdot)$ of the overall covariance function.

Next, incorporate the covariate information through the conditional mean function:

$$
\mu(t, Z) = E(Y(t)|Z) = \mu(t) + \sum_k E(A_k|Z)\phi_k(t)
$$

However, this does not work for sparse data as PACE cannot be directly applied since in general:

$$
E(E(A_k|Y_u)|Z) \neq E(A_k|Z)
$$
Assume no measurement errors, but the mean $\mu(t, z)$ and covariance $\Sigma(s, t, z)$ functions both vary with $Z = z$, hence the eigenfunctions and PC scores also vary with $z$.

\[ Y(t, z) = \mu(t, z) + \sum_k A_k(z) \phi(t, z). \]
Assume no measurement errors, but the mean $\mu(t, z)$ and covariance $\Sigma(s, t, z)$ functions both vary with $Z = z$, hence the eigenfunctions and PC scores also vary with $z$.

$$Y(t, z) = \mu(t, z) + \sum_k A_k(z) \phi(t, z).$$

Since the whole random functions are observable, one can perform one-dimensional smoothing on $Z$ to estimate $\mu(t, z)$ (at each fixed time $t$) and $\Sigma(s, t, z)$ (at each fixed $(s, t)$).
Assume no measurement errors, but the mean $\mu(t, z)$ and covariance $\Sigma(s, t, z)$ functions both vary with $Z = z$, hence the eigenfunctions and PC scores also vary with $z$.

$$Y(t, z) = \mu(t, z) + \sum_k A_k(z) \phi(t, z).$$

Since the whole random functions are observable, one can perform one-dimensional smoothing on $Z$ to estimate $\mu(t, z)$ (at each fixed time $t$) and $\Sigma(s, t, z)$ (at each fixed $(s, t)$).

This approach does not work for sparse data or even irregular dense data.

- Proposed two ways to extend the FPCA approach to accommodate covariate information: fFPCA and mFPCA
Proposed two ways to extend the FPCA approach to accommodate covariate information: fFPCA and mFPCA

Both approaches consist of two parts: A systematic part corresponding to the mean function and a stochastic part comprising the random components.

- Proposed two ways to extend the FPCA approach to accommodate covariate information: fFPCA and mFPCA

- Both approaches consist of two parts: A systematic part corresponding to the mean function and a stochastic part comprising the random components.

- The difference between these two approaches is in the handling of the covariance structure.
Suppose the data originate from a random function $X(t, z)$ with mean $\mu(t, z)$, where $z$ is the value of a covariate $Z$. 

There are two ways to handle the covariance function:

(i) Fully adjusted FPCA (fFPCA) - the covariance function $(s; t; z)$ varies with the covariate $z$.

(ii) Mean adjusted FPCA (mFPCA) - the covariance function $(s; t)$ does not vary with the covariate.

Covariate adjusted FPCA: Longitudinal Data
Suppose the data originate from a random function $X(t, z)$ with mean $\mu(t, z)$, where $z$ is the value of a covariate $Z$.

There are two ways to handle the covariance function:

(i) Fully adjusted FPCA (fFPCA)
- the covariance function $\Sigma(s, t, z)$ varies with the covariate $z$,

(ii) Mean adjusted FPCA (mFPCA)
- the covariance function $\Sigma(s, t)$ does not vary with the covariate.
This approach assumes that the covariance function $\Sigma(s, t, z)$ varies with $z$.

$\Rightarrow$ the corresponding eigenfunctions $\phi_k(t, z)$ and eigenvalues $\lambda_k(z)$ vary with $Z$:

$$
\Sigma(s, t, z) = \sum_k \lambda_k(z) \phi_k(s, z) \phi_k(t, z)
$$
This approach assumes that the covariance function $\Sigma(s, t, z)$ varies with $z$.

$\longrightarrow$ the corresponding eigenfunctions $\phi_k(t, z)$ and eigenvalues $\lambda_k(z)$ vary with $Z$:

$$\Sigma(s, t, z) = \sum_k \lambda_k(z) \phi_k(s, z) \phi_k(t, z)$$

Karhunen-Loeve expansion implies random trajectory $X(t, z)$ can be represented as:

$$X(t, z) = \mu(t, z) + \sum_k A_k(z) \phi_k(t, z)$$
The second approach took the view that the covariate $Z$ is a random variable.

If we pool all the subjects together after centering each individual curve to zero, we would have a pooled covariance function:

$$\Sigma(s, t) = \sum_k \lambda_k^* \phi_k^*(s) \phi_k^*(t)$$
The second approach took the view that the covariate $Z$ is a random variable.

If we pool all the subjects together after centering each individual curve to zero, we would have a pooled covariance function:

$$\Sigma(s, t) = \sum_k \lambda_k^* \phi_k^*(s) \phi_k^*(t)$$

Karhunen-Loeve expansion thus implies that the random trajectory $X(t, z)$ can be represented as:

$$X(t, z) = \mu(t, z) + \sum_k A_k^*(z) \phi_k^*(t)$$
The mean function for fFPCA and mFPCA are the same and can be estimated using any two-dimensional scatter-plot smoother of $Y_{ij}$ on $(T_{ij}, Z_i)$.

Local linear estimator: $\hat{\mu}_L(t, z) = \hat{\beta}_0$, where for $\beta = (\beta_0, \beta_1, \beta_2)$

$$\hat{\beta} = \arg\min_\beta \sum_{i=1}^n \sum_{j=1}^{N_i} K_2\left( \frac{t - T_{ij}}{h_{\mu,t}}, \frac{z - Z_i}{h_{\mu,z}} \right) \times \left[ Y_{ij} - \beta_0 - \beta_1(T_{ij} - t) - \beta_2(Z_i - z) \right]^2$$
AIDS CD4: Mean Function
Estimation: Covariance Function

The covariance can also be estimated by a scatter-plot smoother of the raw covariances defined as:

\[ C_{ijk} = (Y_{ij} - \hat{\mu}(T_{ij}, Z_i))(Y_{ik} - \hat{\mu}(T_{ik}, Z_i)) \]

- fFPCA: three-dimensional smoother of \( C_{ijk} \) on \( (T_{ij}, T_{ik}, Z_i) \)
- mFPCA: two-dimensional smoother of \( C_{ijk} \) on \( (T_{ij}, T_{ik}) \)
Since:

\[
\text{cov}(Y_{ij}, Y_{ik}|T_{ij}, T_{ik}, Z_i) = \text{cov}(X(T_{ij}, Z_i), X(T_{ik}, Z_i)) + \sigma^2 \delta_{jk}
\]

where \( \delta_{jk} \) is 1 if \( j = k \), and 0 otherwise, the diagonal of the raw covariances \( C_{ijk} \) should not be included in the covariance function smoothing step.
Example of Covariance Estimates

- Linear local smoother for fFPCA:

$$\Sigma_L(t, s, z) = \hat{\beta}_0, \text{ where:}$$

$$\hat{\beta} = \text{argmin}_{\beta} \left\{ \sum_{i=1}^{n} \sum_{1 \leq j \neq k \leq N_i} K_3 \left( \frac{t - T_{ij}}{h_{G,t}}, \frac{s - T_{ik}}{h_{G,t}}, \frac{z - Z_i}{h_{G,z}} \right) \times \left[ C_{ijk} - (\beta_0 + \beta_1(T_{ij} - t) + \beta_2(T_{ik} - s) - \beta_3(Z_i - z))^2 \right] \right\}$$

- Linear local smoother for mFPCA:

$$\Sigma^*(t, s) = \hat{\beta}_0, \text{ where:}$$

$$\hat{\beta} = \text{argmin}_{\beta} \left\{ \sum_{i=1}^{n} \sum_{1 \leq j \neq k \leq N_i} K_1 \left( \frac{t - T_{ij}}{h_{G^*}} \right) K_1 \left( \frac{s - T_{ik}}{h_{G^*}} \right) \times \left[ C_{ijk} - (\beta_0 + \beta_1(T_{ij} - t) + \beta_2(T_{ik} - s))^2 \right] \right\}$$
The variance of \( Y(t) \) for a given \( z \) is:

\[
V(t, z) = \Sigma(t, t, z) + \sigma^2
\]

\[
\hat{V}(t, z) = \hat{\beta}_0, \text{ where:}
\]

\[
\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_2 \left( \frac{t - T_{ij}}{h_{V,t}}, \frac{z - Z_i}{h_{V,z}} \right)
\]

\[
\times [C_{ijj} - \beta_0 + \beta_1(T_{ij} - t) + \beta_2(Z_i - z)]^2
\]
Estimation: Variance of Measurement Errors

For stability,

\[ \hat{\sigma}^2 = \frac{2}{T} \int_{Z} \int_{T_1} \{ \hat{V}(t, z) - \hat{\Sigma}_L(t, t, z) \} dt \, dz, \]

where:

\[ T_1 = [\inf\{t : t \in T\} + |T|/4, \sup\{t : t \in T\} - |T|/4] \]
AIDS: Estimated Covariance + measurement error
Estimation: Eigenvalues and Eigenfunctions

- **fFPCA**: The solutions of the eigen-equations,

\[
\int \hat{\Sigma}_L(t, s, z) \hat{\phi}_k(s, z) ds = \hat{\lambda}_k(z) \hat{\phi}_k(t, z),
\]

where the \( \hat{\phi}_k(t, z) \) satisfies \( \int \hat{\phi}_k^2(t, z) dt = 1 \) and \( \int \hat{\phi}_k(t, z) \hat{\phi}_m(t, z) dt = 0 \) for \( m < k \).

- **mFPCA**: The solutions of the eigen-equations,

\[
\int \hat{\Sigma}^*_L(t, s) \hat{\phi}^*_k(s) ds = \hat{\lambda}^*_k \hat{\phi}_k(t),
\]

where the \( \hat{\phi}^*_k(t) \) satisfies \( \int (\hat{\phi}^*_k(t))^2 dt = 1 \) and \( \int \hat{\phi}^*_k(t) \hat{\phi}^*_m(t) dt = 0 \) for \( m < k \).
Estimation: Principal Component Scores

- **fFPCA:**
  
  Use the conditional expectation (PACE) \( E(A_{ik}(Z_i | \tilde{Y}_i)) \) to estimate the principal component scores, where \( \tilde{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})^T \)

- Under the assumption that \( \tilde{Y}_i \) is multivariate normal:

\[
\hat{A}_{ik}(Z_i) = \hat{\lambda}\hat{\phi}_{ik}^T\Sigma_{\tilde{Y}_i}^{-1}(\tilde{Y}_i - \hat{\mu}_i)
\]

where

\[
\hat{\mu}_i = (\hat{\mu}(T_{i1}, Z_i), \ldots, \hat{\mu}(T_{iN_i}, Z_i))^T
\]

\[
(\hat{\Sigma}_{\tilde{Y}_i})_{j,k} = \hat{\Sigma}_L(T_{ij}, T_{ik}, Z_i) + \hat{\sigma}^2 \delta_{jk}
\]

\[
\hat{\phi}_{ik} = (\hat{\phi}_k(T_{ij}, Z_i), \ldots, \hat{\phi}_k(T_{iN_i}, Z_i))^T
\]
The prediction of principal component scores in mFPCA is similar.
Rate of Convergence

- If $E(N) < \infty$ the rate of convergence for the 2D mean and covariance function is $n^{1/3}$.
  
  This is the optimal rate of convergence for 2D smoothers with independent data.

- If $E(N) \to \infty$, the rate of convergence can be as close to $n^{2/5}$ as possible but not equal to $n^{2/5}$.

- If $N_i \to \infty$, the convergence rate is $\sqrt{n}$. 

Rates of Convergence

- If $E(N) < \infty$ the rate of convergence for the 3D mean and covariance function is $n^{2/7}$.
  
  This is the optimal rate of convergence for 3D smoothers with independent data.

- If $E(N) \to \infty$, the rate of convergence can be as close to $n^{2/5}$ as possible but not equal to $n^{2/5}$.

- If $N_i \to \infty$, the convergence rate should be $\sqrt{n}$.
Optimal Rates of Convergence

- The first $k$ eigenfunctions can be estimates at the same optimal rate as a 1- or 2-dim nonparametric regression function.

- The largest $k$ eigenvalues can be estimated at the $\sqrt{n}$ rate.
Bandwidth Selection

- Mean function $\mu(t, z)$ and covariance $\Gamma^*(s, t)$:
  Leave one subject out cross-validation

- Covariance Function $\Gamma(s, t, z)$: $k$-fold cross-validation
  Suppose that the subjects are randomly assigned to $k$ sets ($S_1, S_2, \ldots, S_k$).

\[
h = \arg\min \sum_{l=1}^{k} \sum_{i \in S_l} \sum_{1 \leq j \neq m \leq N_i} \{C_{ijm} - \hat{\Gamma}^{-S_l}(T_{ij}, T_{im}, z_i)\}^2
\]

where $\hat{\Gamma}^{-S_l}(T_{ij}, T_{im}, z_i)$ is the estimated covariance function at $(T_{ij}, T_{im}, z_i)$ when the subjects in $S_l$ are not used to estimate $\Gamma(t, s, z)$. 
Number of Eigenfunctions

Three methods:

- AIC
- BIC
- FVE: minimum number of eigen-components needed to explained at least a specified total fraction of the variation.
End of Covariate Adjusted FPCA
Outline

1. Introduction
2. Mean and Covariance Functions
3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?
4. Inverse Problem in Functional Correlations and Regression
5. Stringing high-dim data to functional data
6. Next Generation Functional Data
Multidimensional Covariates

- Although both fPFCA and mFPCA can accommodate several covariates (including longitudinal covariates) through multivariate smoothing, the computation escalates fast so dimension reduction models are called for to overcome this nonparametric curse of dimensionality.
Multidimensional Covariates

- Although both fPFCA and mFPCA can accommodate several covariates (including longitudinal covariates) through multivariate smoothing, the computation escalates fast so dimension reduction models are called for to overcome this nonparametric curse of dimensionality.

- Assume that $Z \in \mathbb{R}^p$, and for simplicity only the mean function depends on $Z$ (i.e. mFPCA).
Multidimensional Covariates

- Although both fPFCA and mFPCA can accommodate several covariates (including longitudinal covariates) through multivariate smoothing, the computation escalates fast so dimension reduction models are called for to overcome this nonparametric curse of dimensionality.

- Assume that $Z \in \mathbb{R}^p$, and for simplicity only the mean function depends on $Z$ (i.e. mFPCA).

$$
\implies \mu(t, z) = \mu(t, \beta^T z) \rightarrow \text{single index}
$$

or

$$
\mu(t, z) = \mu(t, \beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z), k < p
$$

\downarrow

multiple indices
There are many ways to estimate the indices for independent data, i.e. when there is no $t$, including SIR (Li, 1991) and MAVE (Xia and Li, 2002).

$$Y = \mu(\beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon.$$
Dimension Reduction Models

- There are many ways to estimate the indices for independent data, i.e. when there is no \( t \), including SIR (Li, 1991) and MAVE (Xia and Li, 2002).
  \[
  Y = \mu(\beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon.
  \]

- A few have been extended to functional or longitudinal data, but none for the model:
  \[
  Y(t) = \mu(t, \beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon(t).
  \]
There are many ways to estimate the indices for independent data, i.e. when there is no $t$, including SIR (Li, 1991) and MAVE (Xia and Li, 2002).

$$Y = \mu(\beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon.$$ 

A few have been extended to functional or longitudinal data, but none for the model:

$$Y(t) = \mu(t, \beta_1^T z, \beta_2^T z, \ldots, \beta_k^T z) + \epsilon(t).$$


AIDS CD4: Estimated Mean
AIDS: Estimated Covariance + measurement error
End of Multidimensional Covariates
Outline

1. Introduction

2. Mean and Covariance Functions

3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?

4. Inverse Problem in Functional Correlations and Regression

5. Stringing high-dim data to functional data

6. Next Generation Functional Data
FPCA can be the end product, or it can be further used

- to explore the covariate effects,
- to recover the trajectories of each subject,
- to explore the modes of variation
- etc.
What’s Next After FPCA?

- FPCA can be the end product, or it can be further used
  - to explore the covariate effects,
  - to recover the trajectories of each subject,
  - to explore the modes of variation
  - etc.

- FPCA can help to find a more parsimonious model.
  - We have illustrated this already when no covariates are involved.
  - Next we explore model building when covariates are present.
AIDS CD4: Estimated Mean with Covariates adjusted
This suggests the possibility of a more parsimonious model with additive or multiplicative covariate effects.

\[ Y(t) = \mu(t) + \psi(\beta^T z) + e(t) \rightarrow \mu(t) \] could be parametric, eg. a polynomial.
AIDS CD4 Data

- This suggests the possibility of a more parsimonious model with additive or multiplicative covariate effects.

\[ Y(t) = \mu(t) + \psi(\beta^T z) + e(t) \rightarrow \mu(t) \] could be parametric, eg. a polynomial.

- Common marginal models for longitudinal data take the additive form, and employ parametric models for both the mean and covariance function.

Both parametric forms are difficult to detect for sparse and noisy longitudinal data.
AIDS CD4: Estimated Covariance
AIDS CD4: Estimated Eigenfunctions

<table>
<thead>
<tr>
<th>FVE</th>
<th>AIC (BIC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>K</td>
</tr>
<tr>
<td>0.1154</td>
<td>1</td>
</tr>
</tbody>
</table>
Adding Random Effects

Help to identify the form of the random effects.

\[ Y(t) = \mu(t)\psi(\beta^T z) = a + bt + e(t) \]

\[ \downarrow \quad \downarrow \]

random effects
If the first eigenfunction is proportional to the population mean function $\mu(t, Z)$ and explains almost all the variations of the data, we can discard the remaining eigenfunctions and arrive at the following multiplicative random effect model:

$$Y(t) = \mu(t, z) + A\mu(t, z) + e(t) + b\mu(t, z) + e(t) + e(t)$$

Random effects
If the first eigenfunction is proportional to the population mean function $\mu(t, Z)$ and explains almost all the variations of the data, we can discard the remaining eigenfunctions and arrive at the following multiplicative random effect model:

$$Y(t) = \mu(t, z) + A\mu(t, z) + e(t)$$

Random effects

Examples of such multiplicative random effects models are bountiful and includes the PET data in Jiang, Aston and W. (2009) and the PBC data (bilirubin) in Ding and W. (2008).
References

- Jiang and Wang (2010, AoS)
  covariate adjusted FPCA

- Jiang and Wang (2011, AoS)
  dimension reduction (Semi-parametric index) model

- Jiang, Aston and Wang (2009, NeuroImage)
  multiplicative random effects model for PET data

- Ding and Wang (2008)
  multiplicative random effects model for PBC data
End of What’s Next After FPCA?
End of Lecture 2: FPCA
Outline

1. Introduction
2. Mean and Covariance Functions
3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?
4. Inverse Problem in Functional Correlations and Regression
5. Stringing high-dim data to functional data
6. Next Generation Functional Data
A major challenge in FDA is the inverse problem, which stems from the inversion of the covariance operator.

- The covariance operator is compact, hence its inverse is not a bounded operator, whence the complication.
Inverse Problem with the Covariance Operator

- A major challenge in FDA is the inverse problem, which stems from the inversion of the covariance operator.
  - The covariance operator is compact, hence its inverse is not a bounded operator, whence the complication.

- We illustrate this through two examples where such an inverse problem occurs.
A major challenge in FDA is the inverse problem, which stems from the inversion of the covariance operator.

- The covariance operator is compact, hence its inverse is not a bounded operator, whence the complication.

We illustrate this through two examples where such an inverse problem occurs.

(i) Functional canonical correlation analysis (FCCA)
(ii) Functional linear models.
Let \((X(t), Y(t))\) be a pair of functional data.

How do we extend the concept of correlation to functional data?
Let \((X(t), Y(t))\) be a pair of functional data.

How do we extend the concept of correlation to functional data?

- The first attempt by Leurgans et al. (1993) is to extend the canonical correlation for multivariate data to functional data.
Let \((X(t), Y(t))\) be a pair of functional data.

How do we extend the concept of correlation to functional data?

- The first attempt by Leurgans et al. (1993) is to extend the canonical correlation for multivariate data to functional data.

- For \(p\)-dimensional multivariate data \((X, Y)\), there are \(p\) canonical correlations.
The first canonical correlation $\rho_1$ is the maximal Pearson correlation between any two linear directions $\langle \alpha, X \rangle$ and $\langle \beta, Y \rangle$, 
The first canonical correlation $\rho_1$ is the maximal Pearson correlation between any two linear directions $\langle \alpha, X \rangle$ and $\langle \beta, Y \rangle$, 

$$\rho_1 = \sup_{\alpha, \beta \in \mathbb{R}^p} \text{cov}(\langle \alpha, X \rangle, \langle \beta, Y \rangle) = \text{cov}(\langle \alpha_1, X \rangle, \langle \beta_1, Y \rangle),$$

subject to $\text{var}(\langle \alpha, X \rangle) = 1$ and $\text{var}(\langle \beta, Y \rangle) = 1$. 

The first canonical correlation $\rho_1$ and its associated weights $(\alpha_1, \beta_1)$ are defined as follows:
The \( k \)-th (\( k > 1 \)) canonical correlation \( \rho_k \) and its associated weights can be defined similarly as:

\[
\rho_k = \sup_{\alpha, \beta \in \mathbb{R}^p} \text{cov}(\langle \alpha, X \rangle, \langle \beta, Y \rangle) = \text{cov}(\langle \alpha_k, X \rangle, \langle \beta_k, Y \rangle),
\]

subject to \( \text{var}(\langle \alpha, X \rangle) = 1, \text{var}(\langle \beta, Y \rangle) = 1 \), and

\[(U_k, V_k) = (\langle \alpha_k, X \rangle, \langle \beta_k, Y \rangle)\]

is uncorrelated to all previous pairs

\[(U_j, V_j) = (\langle \alpha_j, X \rangle, \langle \beta_j, Y \rangle), \text{ for } j = 1, \ldots, k - 1.\]
Definition of Functional Canonical Correlation

Replacing the inner product in Euclidean space with the $L^2$-inner product, $\langle f, g \rangle = \int_I f(t)g(t)dt$, we arrive at functional canonical correlations with a series of functional canonical components

$$(\rho_k, \alpha_k, \beta_k, U_k, V_k), \ k \geq 1,$$

where $U_k$ and $V_k$ are maximally correlated and $(U_k, V_k)$ are uncorrelated across $k$. 
Definition of Functional Canonical Correlation

Replacing the inner product in Euclidean space with the $L^2$-inner product, $\langle f, g \rangle = \int_I f(t)g(t)dt$, we arrive at functional canonical correlations with a series of functional canonical components

$$\left(\rho_k, \alpha_k, \beta_k, U_k, V_k, k \geq 1,\right)$$

where $U_k$ and $V_k$ are maximally correlated and $(U_k, V_k)$ are uncorrelated across $k$.

It can be shown (as for multivariate data) that functional canonical correlation analysis (FCCA) corresponds to eigenanalysis of the operator $R = \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2}$, where $\Sigma_{XX}$ and $\Sigma_{YY}$ are the corresponding covariance operator for $X$ and $Y$ respectively, and $\Sigma_{XY}$ is the cross-covariance operator.
Existence of the canonical components is guaranteed if $R$ is a bounded operator.
Inverse Problem of FCCA

- Existence of the canonical components is guaranteed if $R$ is a bounded operator.

- While this is feasible for multivariate data, the inverse of a covariance operator is NEVER bounded because a covariance operator is a compact operator.
Existence of the canonical components is guaranteed if $R$ is a bounded operator.

While this is feasible for multivariate data, the inverse of a covariance operator is NEVER bounded because a covariance operator is a compact operator.

- A remedy was provided in He, M. and W. (2003), which defined a generalized inverse under strong conditions that requires the eigenvalues decay fast to zero.
Although it is possible to resolve the inverse problem by imposing strong conditions, CCA often has an overfitting problem in that the first canonical correlation tends to be very large and hard to interpret.
Although it is possible to resolve the inverse problem by imposing strong conditions, CCA often has an overfitting problem in that the first canonical correlation tends to be very large and hard to interpret.

- This overfitting problem already exists for multivariate data when the dimension is relatively high (but still finite).

- It is caused by the large degree of freedoms in choosing the $p$-dim canonical weights.
More Problems with FCCA

- This overfitting problem is magnified for functional data as theoretically the d.f. is $\infty$ because the weight functions are infinitely dimensional.

  In practice, the weight functions are obtained on a dense grid but it still involves a large d.f.
More Problems with FCCA

- This overfitting problem is magnified for functional data as theoretically the d.f. is $\infty$ because the weight functions are infinitely dimensional.

In practice, the weight functions are obtained on a dense grid but it still involves a large d.f.

- Another challenge with FCCA is the theory. The ill-posed nature of FCCA triggers theoretical challenges.
  - e.g. $\sqrt{n}$-rate of convergence is not feasible for estimates of the canonical correlations $\rho_k$ within the $L^2$ paradigm.
For these reasons, FCCA must be used with caution.
For these reasons, FCCA must be used with caution.

Alternative functional correlations are called for.

One of them is the dynamic correlation (DC) proposed by Dubin and M. (2005), which is a functional version of Pearson correlation.
For these reasons, FCCA must be used with caution.

Alternative functional correlations are called for.

One of them is the dynamic correlation (DC) proposed by Dubin and M. (2005), which is a functional version of Pearson correlation.

DC avoids involving the entire covariance operator to overcome both the inverse and overfitting problem intrinsic to FCCA.

- We may explore this in a later section with neuroimaging applications, if time permits.
End of FCCA
Intuitively, any regression method/model involves an inverse problem, but not all inverse problems are ill-posed.
Intuitively, any regression method/model involves an inverse problem, but not all inverse problems are ill-posed.

Any procedure that involves inverting a covariance operator is ill-posed.
Intuitively, any regression method/model involves an inverse problem, but not all inverse problems are ill-posed.

Any procedure that involves inverting a covariance operator is ill-posed.

There are three scenarios for functional regressions:
(i) scalar (or vector) response with functional and possibly additional vector covariates
(ii) Functional response with scalar covariates
(iii) Functional response with functional and possibly additional vector covariates.
How to construct a functional linear model?
How to construct a functional linear model?

**Answer**: Always think of the vector case first, then replace the inner product!

Linear model: \( Y = \beta_0 + \langle \beta_1, X \rangle + \text{error}. \)
Functional Linear Regression:

Scalar Response $Y$ & Functional Covariate $X(t)$

- How to construct a functional linear model?
  
  **Answer**: Always think of the vector case first, then replace the inner product!

  Linear model: $Y = \beta_0 + \langle \beta_1, X \rangle + \text{error}$. 

- Functional linear model: $Y = \beta_0 + \int_I \beta_1(t)X(t)dt + \text{error}$. 
Functional Linear Regression:

Scalar Response $Y$ & Functional Covariate $X(t)$

- How to construct a functional linear model?

**Answer:** Always think of the vector case first, then replace the inner product!

Linear model: $Y = \beta_0 + \langle \beta_1, X \rangle + \text{error}$.

- Functional linear model: $Y = \beta_0 + \int_I \beta_1(t)X(t)dt + \text{error}$.

- Is this an ill-posed problem?
Functional Linear Regression:

Scalar Response $Y$ & Functional Covariate $X(t)$

How to construct a functional linear model?

**Answer**: Always think of the vector case first, then replace the inner product!

Linear model: $Y = \beta_0 + \langle \beta_1, X \rangle + \text{error}.$

Functional linear model: $Y = \beta_0 + \int_I \beta_1(t) X(t) dt + \text{error}.$

Is this an ill-posed problem?

**YES!** $\beta_1 = \Sigma_{XX}^{-1} \text{cov} (Y, X).$
How to construct a functional linear model?
How to construct a functional linear model?

**Answer:** Replace the inner product in the vector (Y) case!

Multivariate Linear model: \[ Y = \beta_0 + \beta_1 X + \text{error}. \]
- Here \( Y \) and \( \beta_0 \) are \( q \)-dim vectors, but \( \beta_1 \) is a \( q \times p \) matrix.
How to construct a functional linear model?

**Answer:** Replace the inner product in the vector \( Y \) case!

Multivariate Linear model: \[ Y = \beta_0 + \beta_1 X + \text{error}. \]
Here \( Y \) and \( \beta_0 \) are \( q \)-dim vectors, but \( \beta_1 \) is a \( q \times p \) matrix.

Functional linear model: \[ Y(t) = \beta_0(t) + \langle \beta_1(t), X \rangle + b(t) + \text{error}, \]
where \( \beta_1(t) \) consists of \( p \) columns of functions and \( b(t) \) is a random function. \( \rightarrow \) “varying-coefficient” model.
Functional Linear Regression:

Functional Response \( Y(t) \) & \( p \)-dim Vector Covariate \( X \)

- How to construct a functional linear model?

**Answer:** Replace the inner product in the vector (\( Y \)) case!

Multivariate Linear model: \( Y = \beta_0 + \beta_1 X + \text{error}. \)
- Here \( Y \) and \( \beta_0 \) are \( q \)-dim vectors, but \( \beta_1 \) is a \( q \times p \) matrix.

- Functional linear model: \( Y(t) = \beta_0(t) + \langle \beta_1(t), X \rangle + b(t) + \text{error}, \)
where \( \beta_1(t) \) consists of \( p \) columns of functions and \( b(t) \) is a random function. \( \rightarrow \) “varying-coefficient” model.

- Is this an ill-posed problem?
Functional Linear Regression:

**Functional Response Y(t) & p-dim Vector Covariate X**

- How to construct a functional linear model?
  - **Answer:** Replace the inner product in the vector (Y) case!

  **Multivariate Linear model:** \( Y = \beta_0 + \beta_1 X + \text{error} \).
  - Here \( Y \) and \( \beta_0 \) are \( q \)-dim vectors, but \( \beta_1 \) is a \( q \times p \) matrix.

- Functional linear model: \( Y(t) = \beta_0(t) + \langle \beta_1(t), X \rangle + b(t) + \text{error} \), where \( \beta_1(t) \) consists of \( p \) columns of functions and \( b(t) \) is a random function. \( \rightarrow \) “varying-coefficient” model.

- Is this an ill-posed problem?
  - No! \( \beta_1(t) = \Sigma_{XX}^{-1} \text{cov} (Y(t), X) \).
How to construct a functional linear model?

\[
Y(t) = \theta(t) + \int R(s; t) X(s) \, ds + b(t) + \text{error.}
\]

Is this an ill-posed problem?

\[1 \text{ cov}(Y(t); X(t)) = X \text{ cov } XX^\top:\]
How to construct a functional linear model?

**Answer:** Replace the inner product in

Multivariate Linear model: \( Y = \beta_0 + \langle \beta_1, X \rangle + \text{error}, \)

where \( X \) and \( Y \) are vectors and \( \beta_1 \) is a matrix.
Functional Linear Regression:

Functional Response & and Functional Covariate

• How to construct a functional linear model?

**Answer:** Replace the inner product in

Multivariate Linear model: \( Y = \beta_0 + \langle \beta_1, X \rangle + \text{error}, \)
where \( X \) and \( Y \) are vectors and \( \beta_1 \) is a matrix.

• Functional linear model:
\[
Y(t) = \beta_0(t) + \int_{I_X} \beta_1(s, t)X(s)ds + b(t) + \text{error}.
\]
Functional Linear Regression:

Functional Response & and Functional Covariate

How to construct a functional linear model?

**Answer**: Replace the inner product in

Multivariate Linear model: \( Y = \beta_0 + \langle \beta_1, X \rangle + \text{error} \), where \( X \) and \( Y \) are vectors and \( \beta_1 \) is a matrix.

Functional linear model:
\[
Y(t) = \beta_0(t) + \int_{I_X} \beta_1(s, t)X(s)ds + b(t) + \text{error}.
\]

Is this an ill-posed problem?
Functional Linear Regression:

Functional Response & Functional Covariate

How to construct a functional linear model?

**Answer:** Replace the inner product in

Multivariate Linear model: \( Y = \beta_0 + \langle \beta_1, X \rangle + \text{error}, \) where \( X \) and \( Y \) are vectors and \( \beta_1 \) is a matrix.

Functional linear model:

\[
Y(t) = \beta_0(t) + \int_{I_X} \beta_1(s, t)X(s)ds + b(t) + \text{error}.
\]

Is this an ill-posed problem?

**YES!** \( \beta_1(t) = \Sigma_{XX}^{-1} \text{cov} (Y(t), X(t)). \)
There is another functional linear model:

$$Y(t) = \beta_0(t) + \beta_1(t)X(t) + b(t) \text{ error.}$$

- This is a “concurrent” varying-coefficient model as only the current value of $X$ is associated with the current $Y$ value at time $t$. 

- This model is NOT ill-posed.

- How about $Y(t) = \beta_0(t) + \int_0^t \beta(s,t)X(s)ds + b(t) + \text{error}$?

- This is called a “historical model” as the entire history of $X$ up to time $t$ is used to predict $Y$ at time $t$.

- This model is ill-posed.
There is another functional linear model:

\[ Y(t) = \beta_0(t) + \beta_1(t)X(t) + b(t) \text{ error.} \]

- This is a “concurrent” varying-coefficient model as only the current value of \( X \) is associated with the current \( Y \) value at time \( t \).
- This model is NOT ill-posed.
Functional Linear Regression:

There is another functional linear model:

\[ Y(t) = \beta_0(t) + \beta_1(t)X(t) + b(t) \text{ error}. \]

- This is a “concurrent” varying-coefficient model as only the current value of \( X \) is associated with the current \( Y \) value at time \( t \).
- This model is NOT ill-posed.

How about \( Y(t) = \beta_0(t) + \int_0^t \beta_1(s, t)X(s)ds + b(t) \text{ error} \)?

- This is called a “historical model” as the entire history of \( X \) up to time \( t \) is used to predict \( Y \) at time \( t \).
There is another functional linear model:

\[ Y(t) = \beta_0(t) + \beta_1(t)X(t) + b(t) \text{ error.} \]

- This is a "concurrent" varying-coefficient model as only the current value of \( X \) is associated with the current \( Y \) value at time \( t \).
- This model is NOT ill-posed.

How about \( Y(t) = \beta_0(t) + \int_0^t \beta_1(s, t)X(s)ds + b(t) + \text{ error} \)?
- This is called a "historical model" as the entire history of \( X \) up to time \( t \) is used to predict \( Y \) at time \( t \).
- This model is ill-posed.
Comparison of Functional Linear Models vs Linear Mixed-effects Model

- Functional varying-coefficient model:
  \[ Y(t) = \beta_0(t) + \beta_1(t)X(t) + b(t) + \text{error}. \]

- Historical Functional linear model:
  \[ Y(t) = \beta_0 + \int_0^t \beta_1(s, t)X(s)ds + b(t) + \text{error} \]

- Functional linear model using the entire \( X \) trajectory:
  \[ Y(t) = \beta_0 + \int_{I_X} \beta_1(s, t)X(s)ds + b(t) + \text{error} \]
  - Note here that \( X \) and \( Y \) need not be measured at the same time period.
Comparison of Functional Linear Models vs Linear Mixed-effects Model

- Functional varying-coefficient model:
  \[ Y(t) = \beta_0(t) + \beta_1(t)X(t) + b(t) + \text{error}. \]

- Historical Functional linear model:
  \[ Y(t) = \beta_0 + \int_0^t \beta_1(s,t)X(s)ds + b(t) + \text{error} \]

- Functional linear model using the entire \( X \) trajectory:
  \[ Y(t) = \beta_0 + \int_{I_X} \beta_1(s,t)X(s)ds + b(t) + \text{error} \]
- Note here that \( X \) and \( Y \) need not be measured at the same time period.

- Linear Mixed-effects Model: \( Y(t) = \beta_0 + \beta_1X(t) + bZ(t) + \text{error} \), where the coefficient \( \beta_1 \) is time-invariant, so is the random effect \( b \).
All the aforementioned functional linear models can be extended to functional generalized linear model by adding a known link function and a prespecified variance function.
All the aforementioned functional linear models can be extended to functional generalized linear model by adding a known link function and a prespecified variance function.

They can also be extended to an unknown link function, termed “functional single index model”.

\[ Y = \beta_0 + g(\int_I \beta_1(t)X(t)) + \text{error}, \]

where \( g \) is an unknown function.

(X needs to be a continuous process)
All the aforementioned functional linear models can be extended to functional generalized linear model by adding a known link function and a prespecified variance function.

They can also be extended to an unknown link function, termed “functional single index model”.

e.g. \( Y = \beta_0 + g(\int I \beta_1(t)X(t)) + \text{error}, \)

where \( g \) is an unknown function.

(X needs to be a continuous process)

Additional continuous time-dependent and time-independent covariates \( Z \) can be added to the model (How?)
More about Functional Regression Models

- If some of these covariates are discrete, a partial linear single-index model will be needed to model the discrete covariates linearly.

\[ Y = \beta_0 + \theta_1 Z_1 + g(\theta_2 Z_2 + \int_I \beta_q(t) X(t) dt) + \text{error}. \]
More about Functional Regression Models

- If some of these covariates are discrete, a partial linear single-index model will be needed to model the discrete covariates linearly.
  
  e.g. \( Y = \beta_0 + \theta_1 Z_1 + g(\theta_2 Z_2 + \int_I \beta_q(t) X(t) dt) + \text{error}. \)

- The single-index can be extended to multiple indices etc.
  
  e.g. \( Y = \beta_0 + g(\int_I \beta_1(t) X(t) dt + \int_I \beta_2(t) X(t) dt) + \text{error}. \)
More about Functional Regression Models

- If some of these covariates are discrete, a partial linear single-index model will be needed to model the discrete covariates linearly.
  
  e.g. \[ Y = \beta_0 + \theta_1 Z_1 + g(\theta_2 Z_2 + \int_I \beta_q(t)X(t)dt) + \text{error}. \]

- The single-index can be extended to multiple indices etc.
  
  e.g. \[ Y = \beta_0 + g(\int_I \beta_1(t)X(t)dt + \int_I \beta_2(t)X(t)dt) + \text{error}. \]

- All these index based models are “dimension reduction” models. Another type of dimension reduction model is the “additive model”,
  
  e.g. \[ Y(t) = \beta_0(t) + \sum \phi_k(X_k) + b(t) + \text{error}. \]
The additive model, \( Y(t) = \beta_0(t) + \sum \phi_k(X_k) + b(t) + \text{error} \), can be extended to allow for time-varying coefficients.

\[ Y(t) = \beta_0(t) + \sum \beta_k(t)\phi_k(X_k) = b(t) + \text{error}. \]

“Time-varying additive model”
The additive model, $Y(t) = \beta_0(t) + \sum \phi_k(X_k) + b(t) + \text{error}$, can be extended to allow for time-varying coefficients.

$Y(t) = \beta_0(t) + \sum \beta_k(t) \phi_k(X_k) = b(t) + \text{error.}$

“Time-varying additive model”

There are lots of models that one can construct!

More later by Hans.
End of Functional Regression
End of Inverse Problem: Just hang in there!
Outline

1 Introduction

2 Mean and Covariance Functions

3 Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?

4 Inverse Problem in Functional Correlations and Regression

5 Stringing high-dim data to functional data

6 Next Generation Functional Data
We have mentioned the advantages of a natural ordering for functional data.
We have mentioned the advantages of a natural ordering for functional data.

- What are the advantages?
Stringing High-dim Data to Functional Data

- We have mentioned the advantages of a natural ordering for functional data.
  
  - What are the advantages?
    
    - Easy to visualize the data.
    - Information can be borrowed from neighboring data to overcome the curve of high dimensionality.
We have mentioned the advantages of a natural ordering for functional data.

- What are the advantages?
  - Easy to visualize the data.
  - Information can be borrowed from neighboring data to overcome the curve of high dimensionality.

What if the data are really high-dim without a natural ordering in time?
Stringing High-dim Data to Functional Data

- We have mentioned the advantages of a natural ordering for functional data.

- What are the advantages?

  - Easy to visualize the data.
  - Information can be borrowed from neighboring data to overcome the curve of high dimensionality.

- What if the data are really high-dim without a natural ordering in time?

  - We reorder the data to make it smooth!
Stringing - Use a similarity measure between feature variables $X_j$ & $X_k$ to map each feature to a location on $[0,1]$ by MDS (multi-dimensional scaling).
Stringing - Use a similarity measure between feature variables $X_j$ & $X_k$ to map each feature to a location on $[0, 1]$ by MDS (multi-dimensional scaling).

0++++++++++++++++++++++++++++++++++++++++++++++++++1
Stringing - Use a similarity measure between feature variables \(X_j\) & \(X_k\) to map each feature to a location on \([0, 1]\) by MDS (multi-dimensional scaling).

Convert the high-dim data to FD by permuting the order of the high-dim data to the new *stringed* location on \([0, 1]\) for each subject.
Stringing High-dim Data to FD

- Assumption: There is a latent order which corresponds to a smooth process.
The latent order was scrambled.
Stringing High-dim Data to FD

- Stringing restores the latent order which corresponds to a smooth process.
Molecular classification problem to diagnose type of leukaemia based on gene expression array data (Golub et al. 1999).

The gene expression data were obtained from Affymetrix chips, containing 7129 genes, reduced to 50 most significant genes by $t$-tests.

Data for 38 patients, 27 with acute lymphoblastic leukaemia (ALL) with 11 with acute myeloid leukaemia (AML) cases.
Stringing Gene Expression

Leukaemia gene expressions for 50 significant genes. Left: Original order. Right: Stringed order. Double arrow: True separation between ALL and AML patients.
For the gene expression data, stringing separates the leukaemia types.
Stringing: Tree Ring Example

Top: Observed series of tree ring width data. Bottom: Ordered tree ring series, obtained by Stringing, actually tracks precipitation.
For the tree ring data, the reordering reveals a natural physical ordering, precipitation, of the data.
For the tree ring data, the reordering reveals a natural physical ordering, precipitation, of the data.

For the gene expression data, stringing separates the leukaemia types.
Conclusion on Stringing

This approach trades the assumptions of sparsity and low-correlation on high-dim data for the existence of a latent smooth ordering of features.
Conclusion on Stringing

- This approach trades the assumptions of sparsity and low-correlation on high-dim data for the existence of a latent smooth ordering of features.

- The latent smooth ordering assumption might be a useful assumption.
Conclusion on Stringing

- This approach trades the assumptions of *sparsity and low-correlation* on high-dim data for the existence of a *latent smooth ordering of features*.

- The latent smooth ordering assumption might be a useful assumption.

- Applications to Gene data and tree ring data demonstrate this.
Conclusion on Stringing

- This approach trades the assumptions of *sparsity and low-correlation* on high-dim data for the existence of a *latent smooth ordering of features*.

- The latent smooth ordering assumption might be a useful assumption.

- Applications to Gene data and tree ring data demonstrate this.

- Stringing could be used as a visualization tool to view high-Dim data.
Leukaemia gene expressions for 50 significant genes. Left: Original order. Right: Stringed order. Double arrow: True separation between ALL and AML patients.
Stringing: Tree Ring Example

Top: Observed series of tree ring width data. Bottom: Ordered tree ring series, obtained by Stringing, actually tracks precipitation.
End of Stringing
Outline

1. Introduction
2. Mean and Covariance Functions
3. Principal Component Analysis
   - Covariate adjusted FPCA
   - Multidimensional Covariates
   - What’s Next After FPCA?
4. Inverse Problem in Functional Correlations and Regression
5. Stringing high-dim data to functional data
6. Next Generation Functional Data
Next Generation FD

- So far, the functional data are 1D independent curve data.
- These curve data can be dependent
  - e.g. Functional time series and spatio-temporal data.
Next Generation FD

- So far, the functional data are 1D independent curve data.

- These curve data can be dependent
e.g. Functional time series and spatio-temporal data.

- Functional data can be 2D, 3D images, or even 4D data.

- Functional data can also include objects, shapes, trees, networks, etc.
2D Functional Data:
http://www.pnas.org/content/97/11/6150/F1.expansion.html
3D Functional Data:
http://www.musicianbrain.com/images/fmri.jpg
What is the Dimension of fMRI data?

- For a single subject:
  
  Spatially (3D) correlated temporal (1D) data
  Temporally correlated 3D data
  → Atom of Longitudinal 3D functional data
What is the Dimension of fMRI data?

- For a single subject:
  - Spatially (3D) correlated temporal (1D) data
  - Temporally correlated 3D data
  → Atom of Longitudinal 3D functional data

- For multi subjects:
  - Independent Sample of 4D functional data
  - multi-level data if there are multiple scans
Neuroimaging data is intrinsically functional, often multifaceted.
FDA for Neuroimaging Data

- Neuroimaging data is intrinsically functional, often multifaceted.
- Neuroimaging data is a gold mine for FDA.
Neuroimaging data is intrinsically functional, often multifaceted.

Neuroimaging data is a gold mine for FDA.

The most prevailing FDA method for NDA has been Functional PCA.
FDA for Neuroimaging Data

- Neuroimaging data is intrinsically functional, often multifaceted.

- Neuroimaging data is a gold mine for FDA.

- The most prevailing FDA method for NDA has been Functional PCA.

- Other functional approaches are emerging daily!
Summary of Next Generation Functional Data

- Correlated functional data
  - e.g. functional time series, spatio-temporal data
- Independent k-D data
- Correlated k-D data
- Longitudinal functional data
- Multi-level functional data
- ...... any others?
Next Generation Functional Data

- **Object data**
  
e.g. shapes, trees, networks, etc.
Next Generation Functional Data

- *Object data*
  - e.g. shapes, trees, networks, etc.

- All seems complex - Some non-Euclidean!
Next Generation Functional Data

- **Object data**
  
e.g. shapes, trees, networks, etc.

- All seems complex - Some non-Euclidean!

- Many are BIG!
  
e.g. Neuroimaging data
Next Generation Functional Data

- **Object data**
  
e.g. shapes, trees, networks, etc.

- All seems complex - Some non-Euclidean!

- Many are BIG!
  
e.g. Neuroimaging data

- They are challenging, but represent opportunities!
List of Review Papers


- Aston (2015) also has a review paper on FDA for NDA.


Aston (2015) also has a review paper on FDA for NDA.


Additional review papers:
Müller (2005, Scan J. Stat.) - regression and classification
List of Books on Functional Data

- Ramsay and Silverman (2005)
- Ferraty and Vieu (2006)
- Wu and Zhang (2006)
- Horváth and Kokoska (2012)
- Li and Eubank (2015)
Thank You!