# Cell-Well Data Objects & Fisher Rao Curve Warping

Xiaosun Lu

UNC Chapel Hill

October 16, 2012

# Outline

Two topics in cell culture biology

(1) Object Oriented Data Analysis (OODA)

- Motivation: Analysis of cell images
- How the choice of data objects orient further analyses

# Outline

Two topics in cell culture biology

(1) Object Oriented Data Analysis (OODA)

- Motivation: Analysis of cell images
- How the choice of data objects orient further analyses

(2) Functional Data Analysis

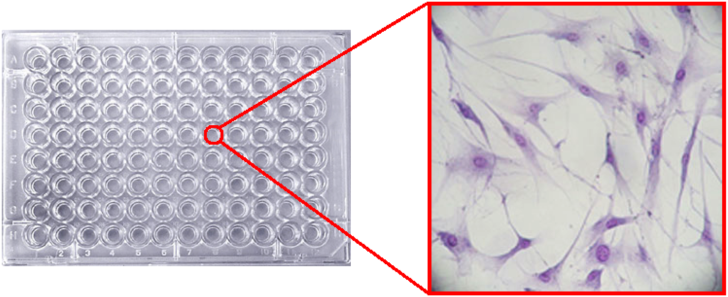
- Motivation: Analysis of cell growth curves
- Decompose horizontal and vertical variabilities

# Object Oriented Data Analysis

- Proposed by Wang & Marron, 2007
- Data objects: Atoms of statistical analysis
  - Numbers
  - Vectors (Multivariate analysis)
  - Curves (Functional data analysis)
- More complex objects
  - Trees
  - Images
  - Shapes

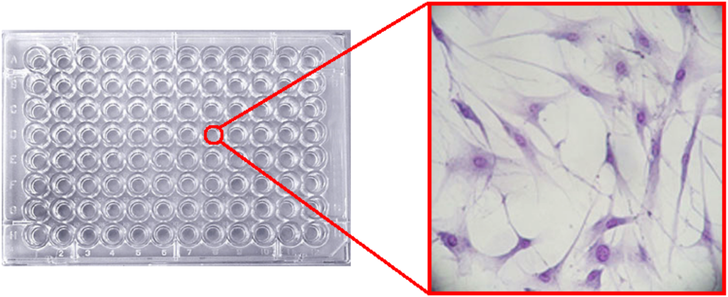
# Background

Goal: Media development for cell culture



# Data Objects

- Wells?
- Cells?



# How OODA works

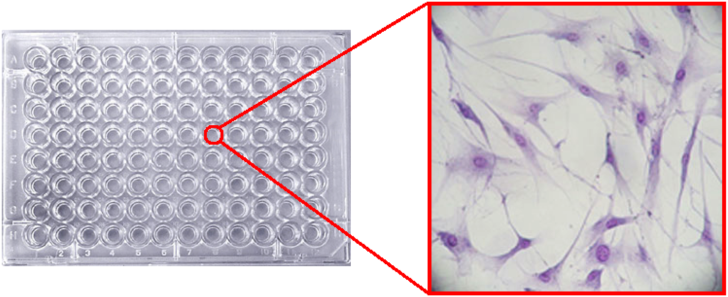
(1) Fully understand data structure

(2) Choose appropriate data objects

(3) Come up with an “appropriate” analysis

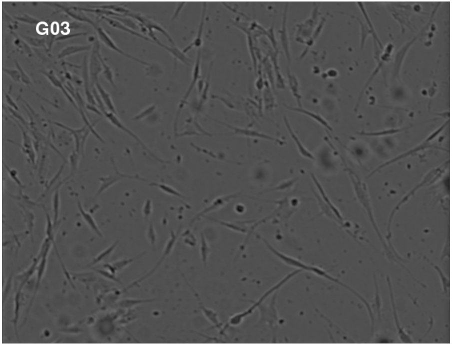
# Motivation

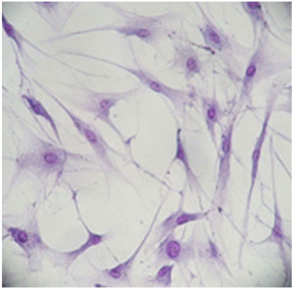
- Confluence: Percent of environment used by cells
- Passage cell culture based on confluence level
- Image a well $\rightarrow$ Estimate confluence level $\rightarrow$ Passage



# The Challenge: Bright Field Imaging

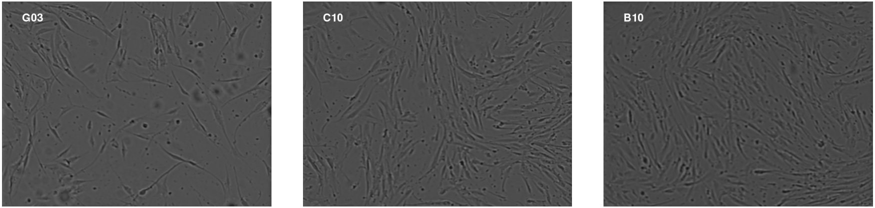
- Defocused image of cell shadows
- Difficult to estimate confluence using BF images



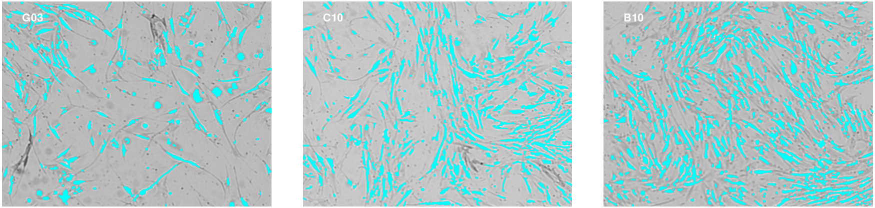



# The Challenge: Bright Field Imaging

## Cell shadows




## IPLab cell identification



# How to Estimate Confluence

(1) Counting the cells (cyan objects)

(2) Biologists' manual estimation

(3) Objective statistical estimation

- Improves over the counting approach
- Support automated passaging

# How to Estimate Confluence

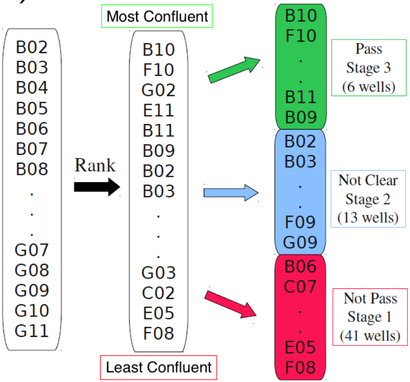
(1) Counting the cells (cyan objects)

(2) Biologists' manual estimation

- Varies among people
- Consensus estimation

  *bio-rank*

  *bio-classification*



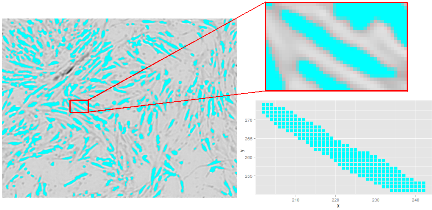

(3) Objective statistical estimation

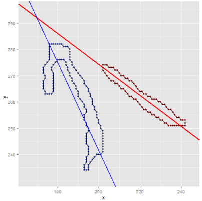
- Improves over the counting approach
- Support automated passaging

# Image Feature Extraction

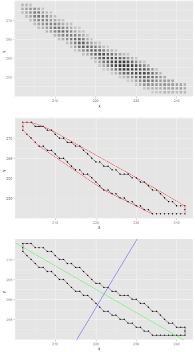
- Image preprocessing
  - Remove uneven background shading
  - Remove granular noise
  - Intensity normalization
- Two types of confluence-related features
  - Features of An Individual Cell (32)
  - Additional Entire-Well Features (13)

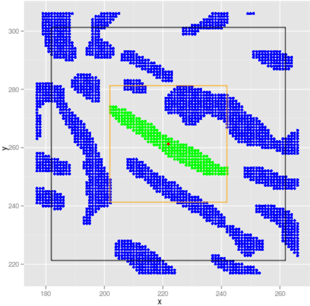
# Features of An Individual Cell





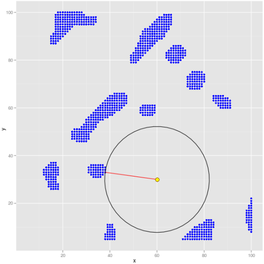
- Intensity
- Shape & size
- Local density
- Cell orientation

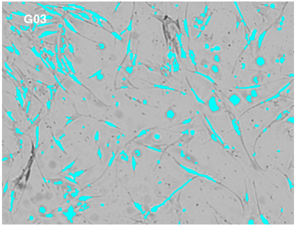


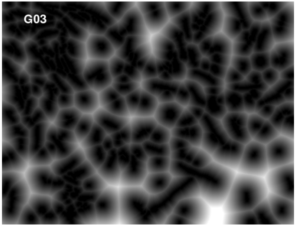


# Additional Entire-Well Features

- Cell number
- Cell gap size/intensity



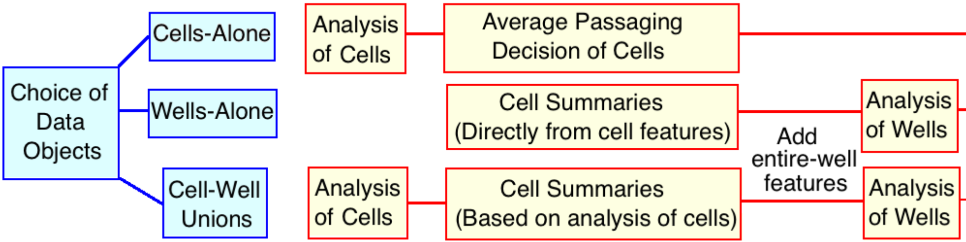



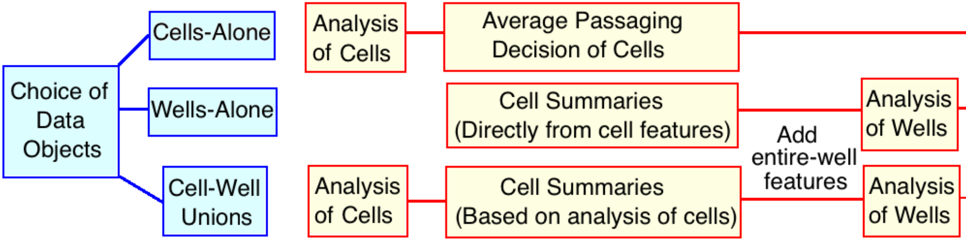
# The Choice of Data Objects

- Two data sets
  - Cell data (features of each individual cell);
  - Well data (additional entire-well features)
- Different choices of data objects
  - Cells-alone
  - Wells-alone
  - A new type of data objects: Wells $\cup$ Cells

# The Choice of Data Objects



# The Choice of Data Objects




- Cells-Alone: Ignore additional well data
- Wells-Alone vs. C-W Unions
  How to summarize the cells?
  E.g. feature-wise summaries, PC summaries

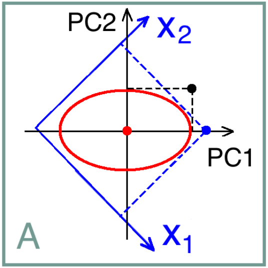
# Compare Data Objects

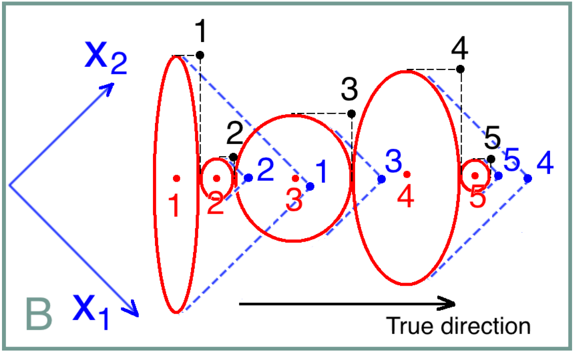
- DWD of passaging groups
- % of false passaging decision
  - Cells-Alone: 25%
  - Wells-Alone: 8.6%
  - Cell-Well Unions: 5.2%

- Cells-alone are not a good choice
- Further study of the wells-alone and the unions...

# Cell Summarization

Can either impair or preserve the bio-pattern in cell data



# Cell Summarization

- How well the bio-pattern is pressved depends on
  - Choice of statistics
  - Variability of cell distributions across wells
  - Rotation of cell data before summaring
- OODA is independent of and suggests analysis method
- C-W unions are a good choice for such data structure

# Outline

Two topics in cell culture biology

(1) Object Oriented Data Analysis (OODA)

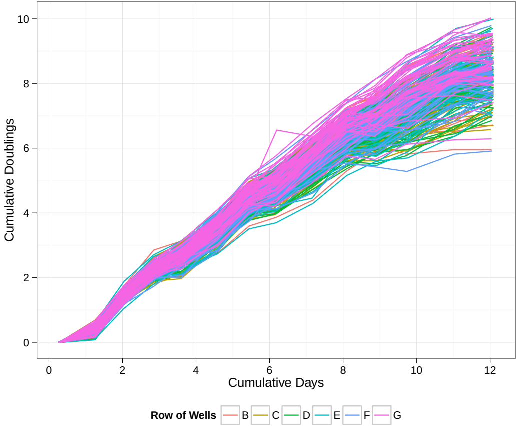
- Motivation: Analysis of cell images
- How the choice of data objects orient further analyses

(2) Functional Data Analysis

- Motivation: Analysis of cell growth curves
- Decompose horizontal and vertical variabilities

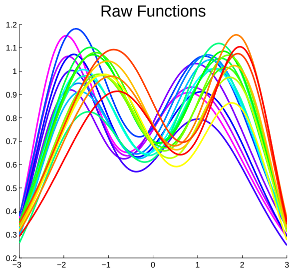
# Motivation

- Cell growth curves: Media effect, Batch effect, etc.
- Analysis of variabilities among curves



# How to Understand the Variability?

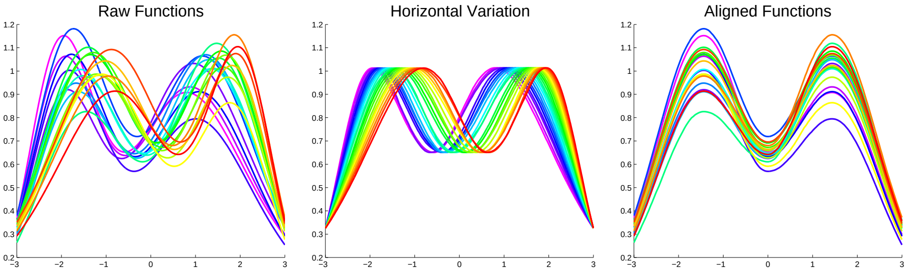
- Toy Example to develop appropriate approaches



# How to Understand the Variability?

- Toy Example to develop appropriate approaches
- Insightful decomposition

$\rightarrow$ Horizontal var + Vertical var

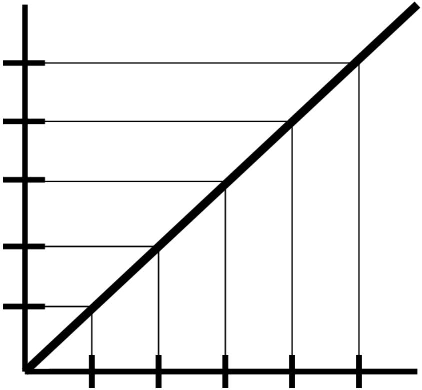


# Curve Registration

- Consider domain warping $\gamma : [0, 1] \rightarrow [0, 1]$
- $\gamma(x)$ is a diffeomophism (smooth)

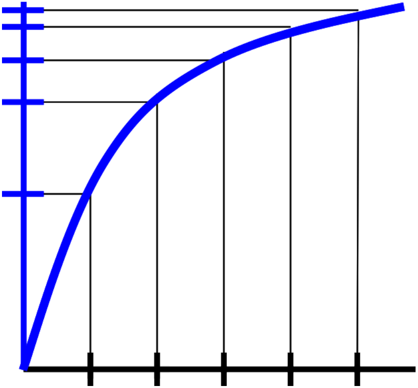
# Curve Registration

- Consider domain warping $\gamma : [0,1] \rightarrow [0,1]$
- $\gamma(x)$ is a diffeomophism (smooth)



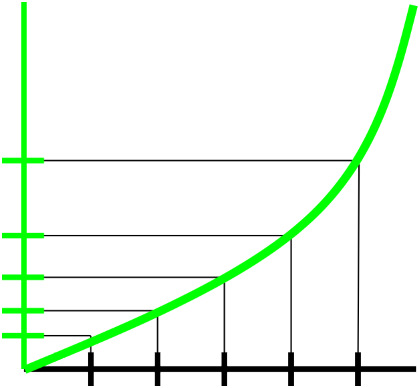
# Curve Registration

- Consider domain warping $\gamma : [0,1] \to [0,1]$
- $\gamma(x)$ is a diffeomophism (smooth)



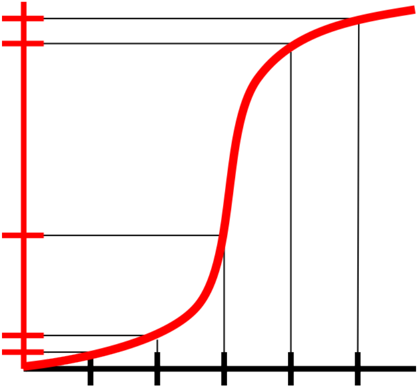
# Curve Registration

- Consider domain warping $\gamma : [0, 1] \to [0, 1]$
- $\gamma(x)$ is a diffeomophism (smooth)



# Curve Registration

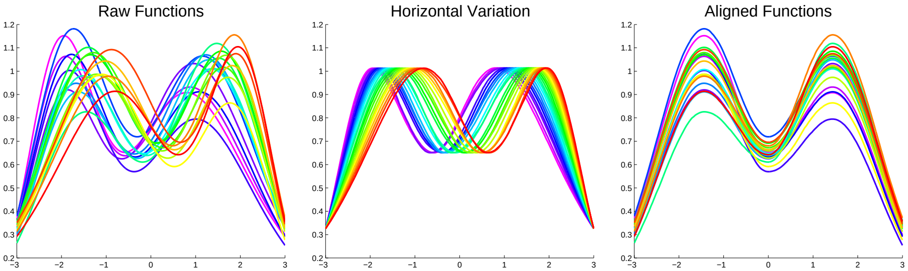
- Consider domain warping $\gamma : [0,1] \to [0,1]$
- $\gamma(x)$ is a diffeomophism (smooth)



# How to Understand the Variability?

- Toy Example to develop appropriate approaches
- Insightful decomposition

$\rightarrow$ Horizontal var + Vertical var

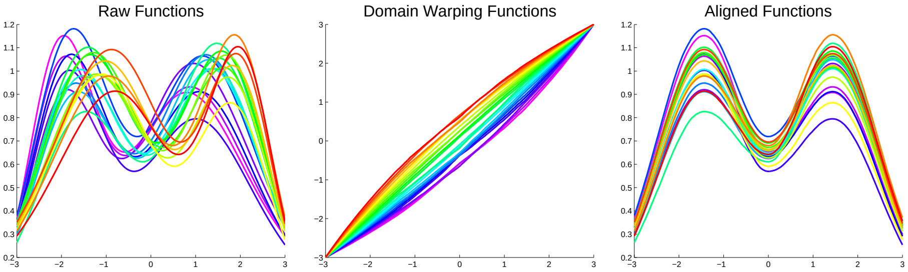


# How to Understand the Variability?

- Toy Example to develop appropriate approaches
- Curve registration: $f(\gamma(x)) = \tilde{f}(x)$

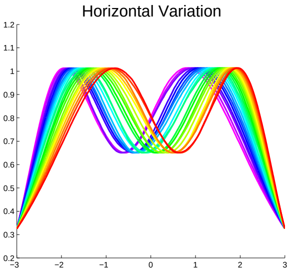
  $\rightarrow$ Warping functions + Aligned functions

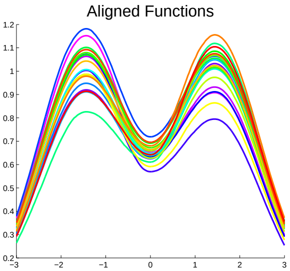
  $\rightarrow$ Horizontal var + Vertical var



# What Are the Data Objects?

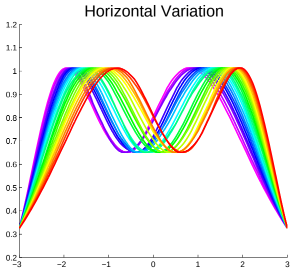
- "Equivalence" of two curves: $f_1 \sim f_2$
- $\exists\ \gamma$ so that $f_1 \circ \gamma = f_2$

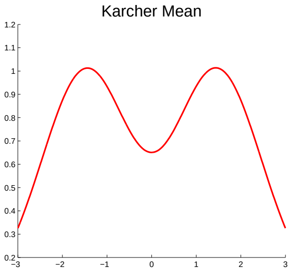


# What Are the Data Objects?

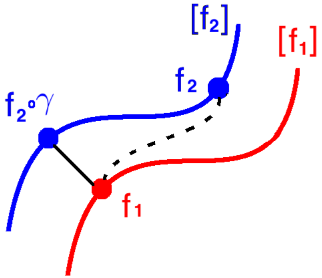
- Data object = Equivalence group of curves
- A representer of the group: $f$
- Notation of a data object: $[f]$
- Orbit, Quotient space



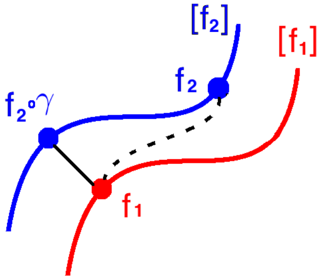
# Curve Registration

- Align $f_2$ to $f_1$
- Find a "good" representer of $[f_2]$, i.e. $f_2 \circ \gamma$
- $\inf_{\gamma \in \Gamma} d(f_1 - f_2 \circ \gamma)$



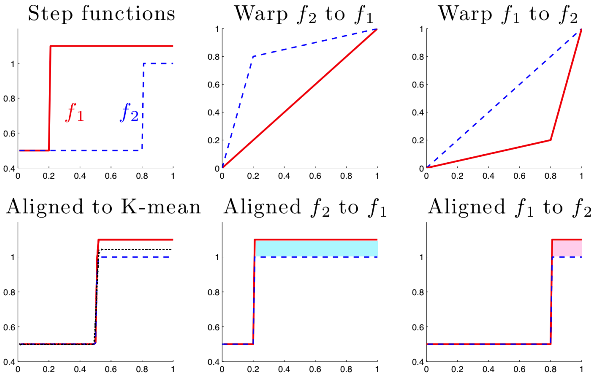
# Metrics in Curve Space

- What is the appropriate metric $d$?
- Traditional choice: $\|.\|$
- $\inf_{\gamma \in \Gamma} d(f_1 - f_2 \circ \gamma)$



# Metrics in Curve Space

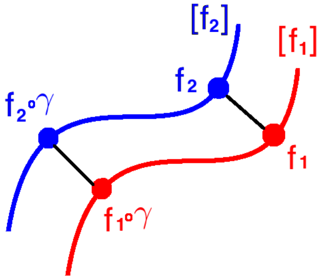
- Issues in $\mathcal{L}^2$ Metric
- $\inf_{\gamma \in \Gamma} \|f_1 - (f_2 \circ \gamma)\| \neq \inf_{\gamma \in \Gamma} \|(f_1 \circ \gamma) - f_2\|$



# Metrics in Curve Space

- Solution: Warping-invariant metric

$$d(f_1, f_2) = d(f_1 \circ \gamma, f_2 \circ \gamma)$$

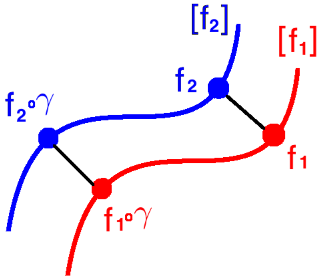


# Metrics in Curve Space

- Fisher Rao Metric (C. R. Rao, 1945)
- It is the unique solution (Cencov, 1982)

$$d_{FR}(f_1, f_2) = d_{FR}(f_1 \circ \gamma, f_2 \circ \gamma)$$

- Challenge: Complicated
  Sample statistics are not clear



# Metrics in Curve Space

- Square Root Velocity Function

$$q_f(t) = \frac{\dot{f}(t)}{\sqrt{|\ \dot{f}(t)\ |}}$$

$$f(t) = f(0) + \int_0^t q_f(s)|q_f(s)|ds$$

- Simplifies FR framework (Srivastava et al, 2010)

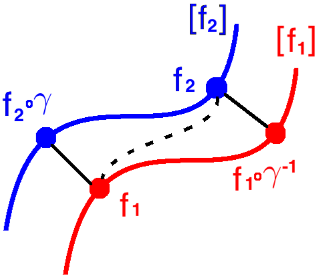
$$d_{FR}(f_1, f_2) = \|q_{f_1} - q_{f_2}\|$$

# Metrics in Quotient Space

- Distance between equivalence groups

$$d_Q([f_1],[f_2]) = \inf_{\gamma \in \Gamma} d_{FR}(f_1, f_2 \circ \gamma) = \inf_{\gamma \in \Gamma} \|q_{f_1} - q_{f_2 \circ \gamma}\|$$

- Independent of the choice of $f_1$, $f_2$



# Mean in Quotient Space

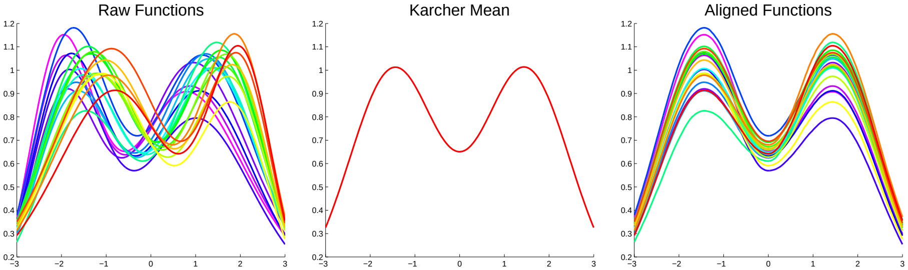
- Consider equivalence groups

$$[f_1], [f_2], \ldots, [f_n]$$

- Karcher mean $[\mu] = argmin_{[f]} \Sigma_{i=1}^{n} d_Q([f], [f_i])^2$

- Choose "best" representer of $[\mu]$ so that the mean of warping functions $=$ Identity

# Mean in Quotient Space

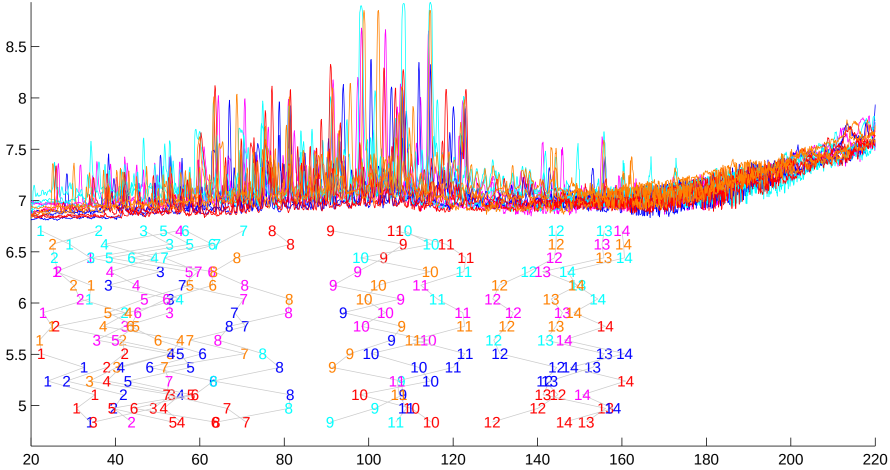


# Proteomics Data

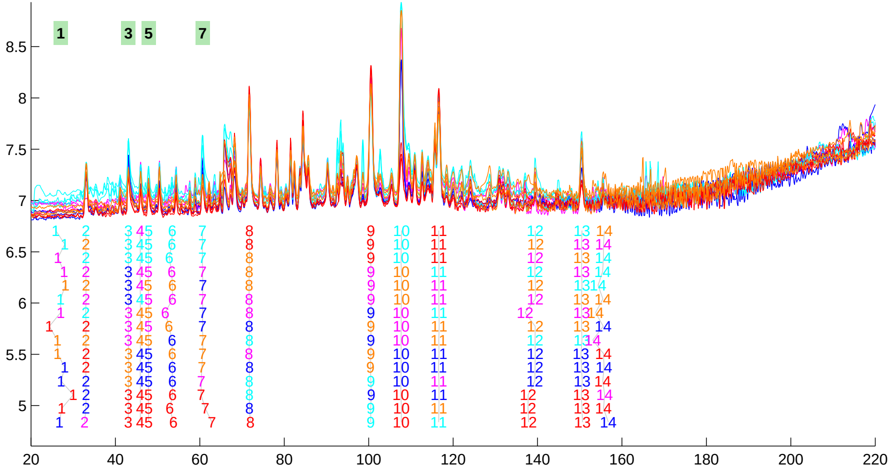
- Measurements: TIC (Total Ion Count) Chromatograms
  Modern type of chemical spectra
- Intensity as a function of time
- 15 functions
  - Samples: A, B, C, X, Y
  - Runs: 1, 2, 3

  Functions are colored by sample
- 14 features are marked, including “spiked in” features (1, 3, 5, 7)
- Goal: Warp the functions to line up the features

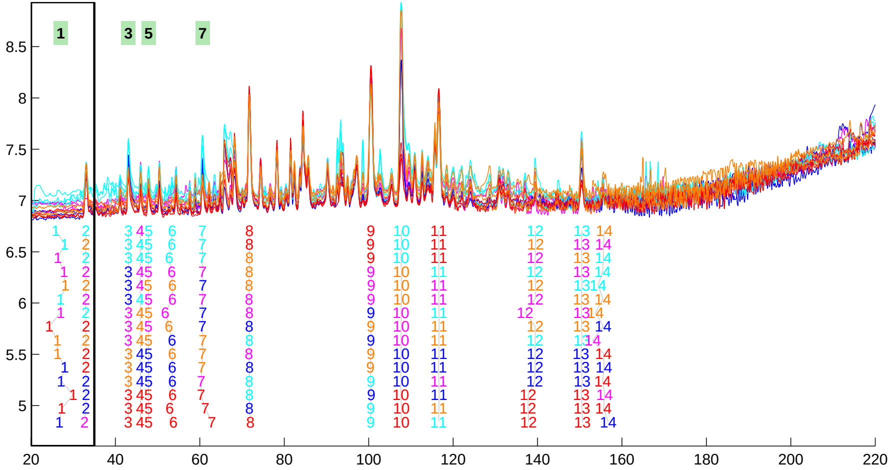
# Unaligned Functions



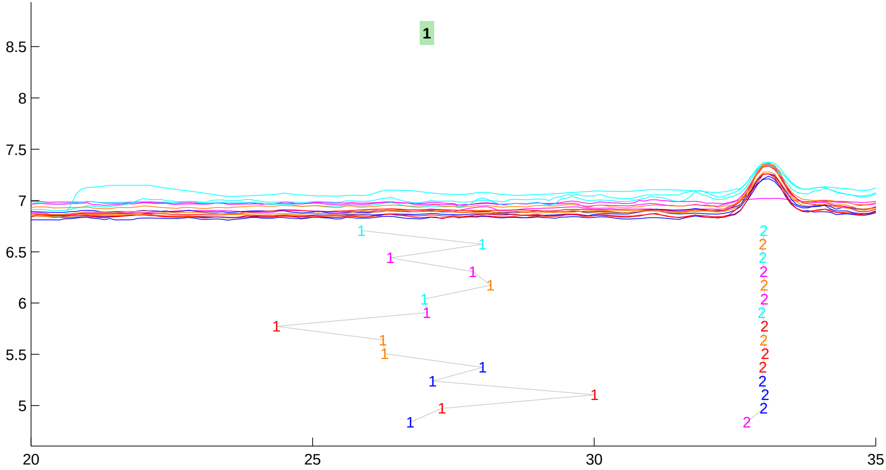
# Aligned Functions



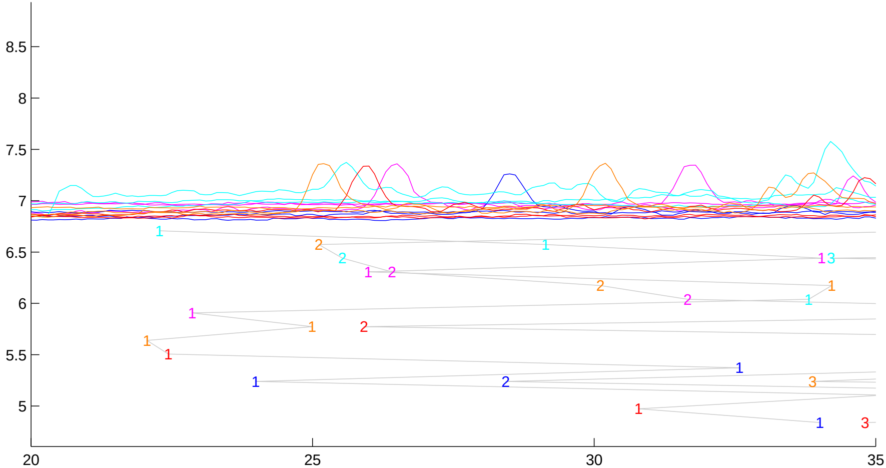
# Zoom-in



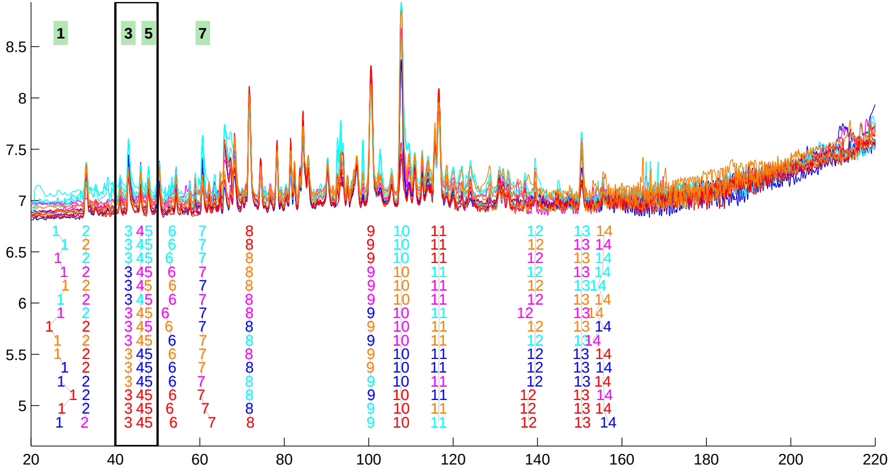
# Zoom-in: Aligned Functions



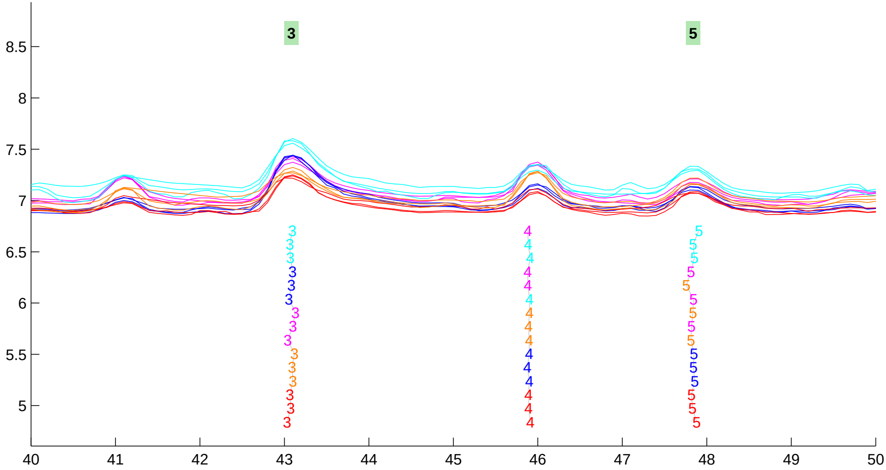
# Zoom-in: Unaligned Functions



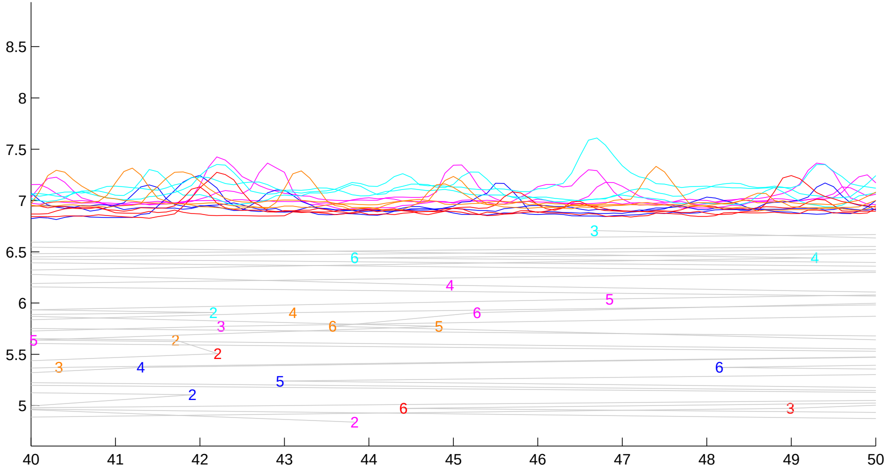
# Zoom-in



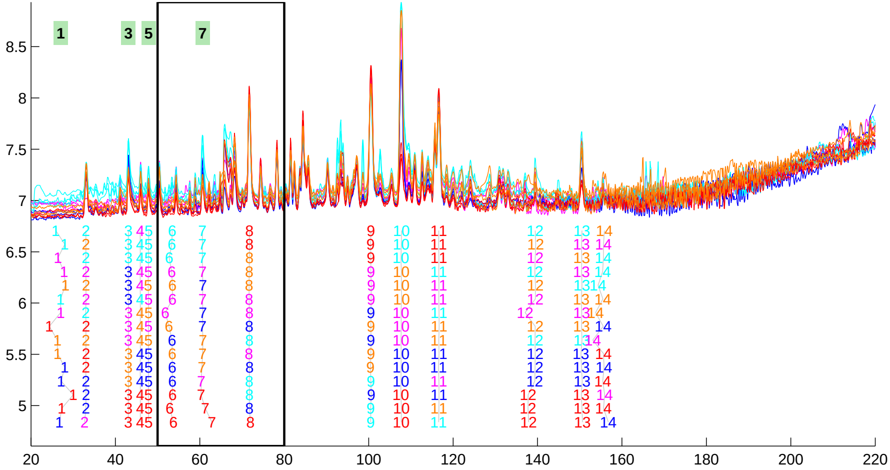
# Zoom-in: Aligned Functions



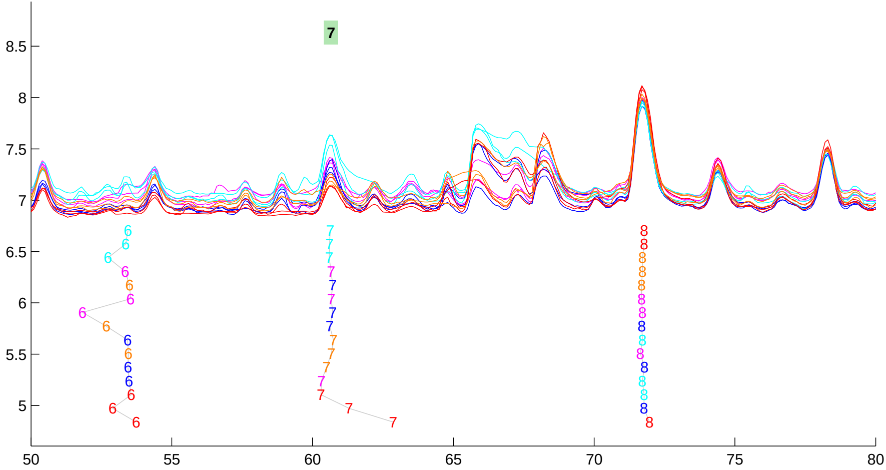
# Zoom-in: Unaligned Functions



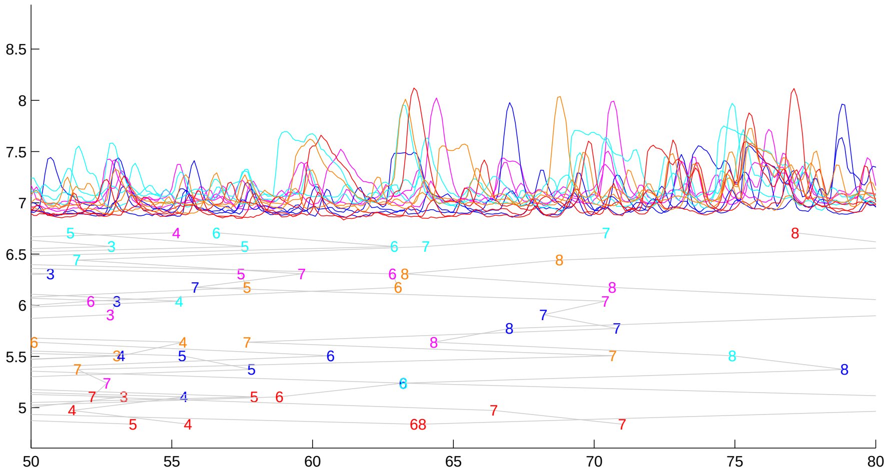
# Zoom-in



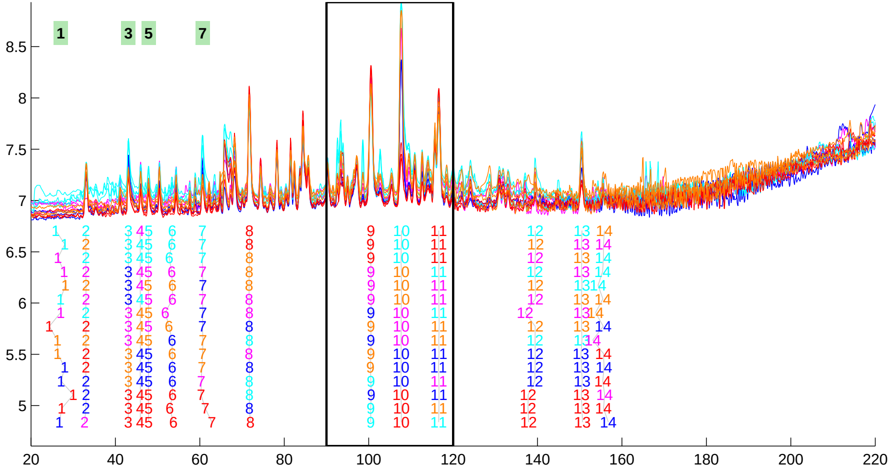
# Zoom-in: Aligned Functions



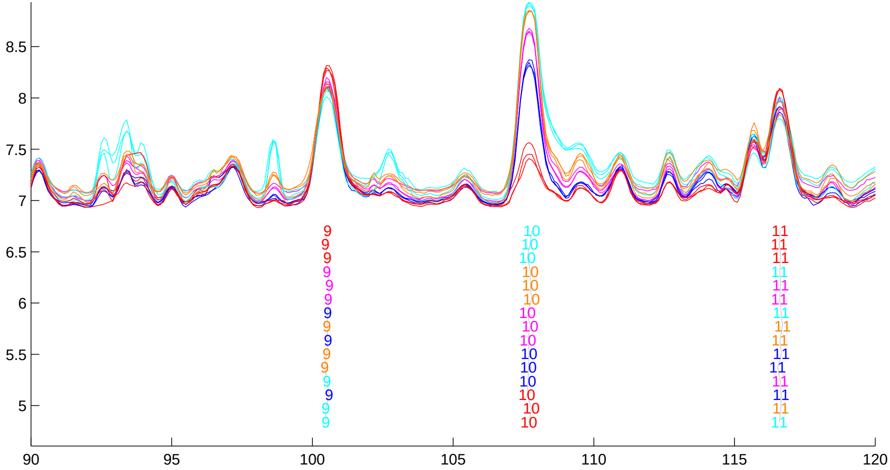
# Zoom-in: Unaligned Functions



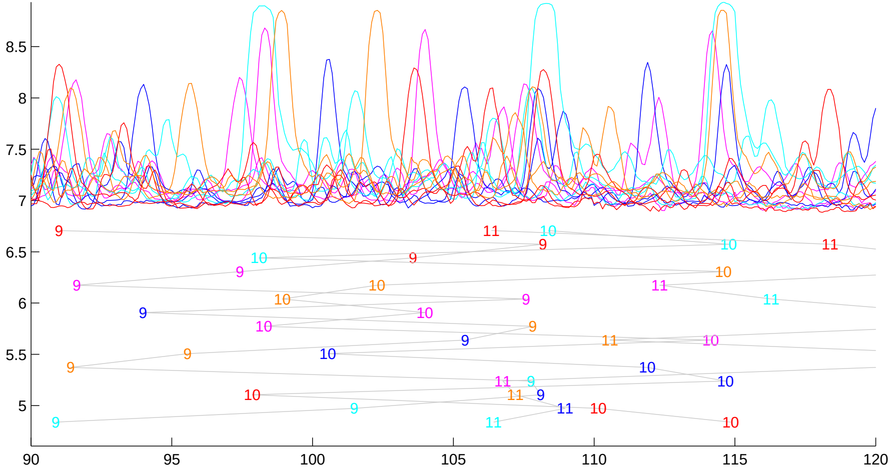
# Zoom-in



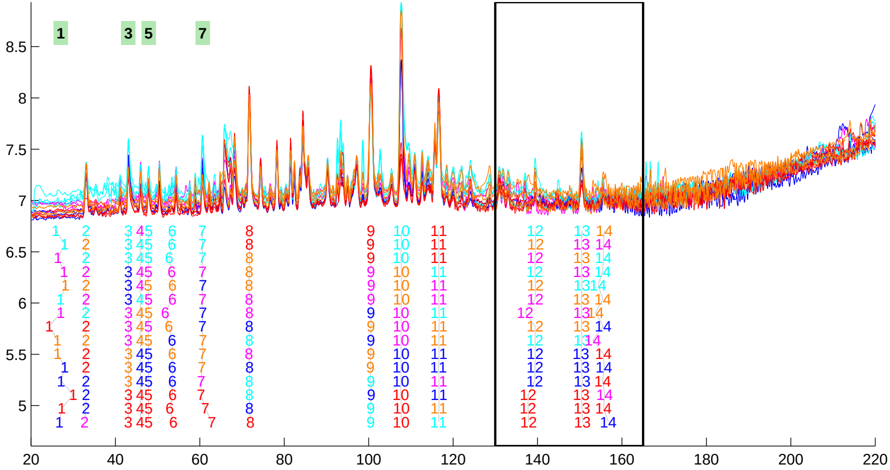
# Zoom-in: Aligned Functions



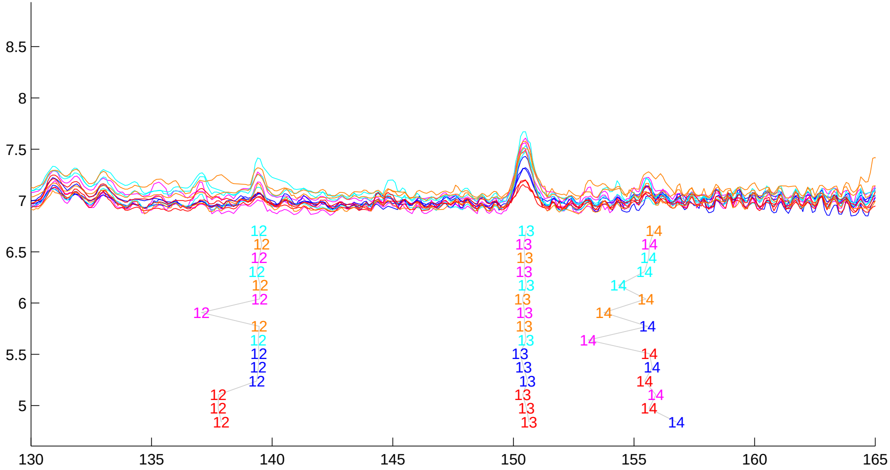
# Zoom-in: Unaligned Functions



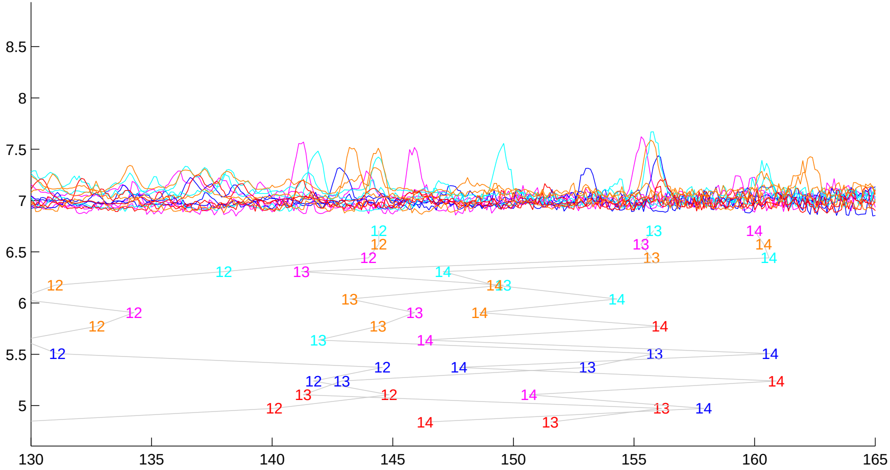
# Zoom-in



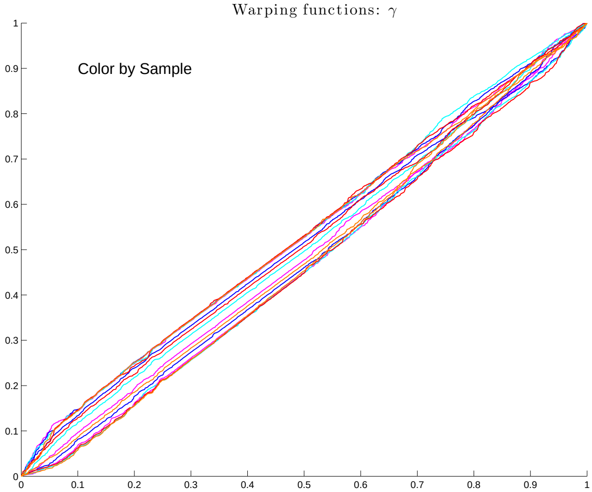
# Zoom-in: Aligned Functions



# Zoom-in: Unaligned Functions



# Warping Functions



Additional Information

# Appendix: Fisher Rao Metric

- Define Fisher Rao metric as

$$\ll v_1, v_2 \gg_f = \frac{1}{4}\int_0^1 \dot{v}_1(t)\dot{v}_2(t)\frac{1}{\mid \dot{f}(t) \mid}dt$$

where $f \in \mathcal{F}$ and $v_1, v_2 \in T_f(\mathcal{F})$

- Define Fisher Rao distance as

$$d_{FR}(f_1, f_2) = \inf_{\alpha:[0,1]\to\mathcal{F},\alpha(0)=f_1,\alpha(1)=f_2} L[\alpha]$$

where $\alpha(\tau)$ is a differentiable path connecting $f_1$ and $f_2$ in $\mathcal{F}$

$$L[\alpha] = \int_0^1 (\ll \dot{\alpha}(\tau), \dot{\alpha}(\tau) \gg_{\alpha(\tau)})^{1/2} d\tau$$

# Cell Image Data Visualization