STAT 597: Functional Data Analysis
Lecture 4 – Chapter 2 – Further topics

Instructor: Matthew Reimherr

Homework

First homework is going up soon. Remember, it must be prepared using sweave, markdown, or knitr (or some equivalent software). You will turn the homework in through canvas.

Project

You may work in groups of up to 3. Timeline:

- Midterm report: Oct 20th.
- Presentation: Nov 29, Dec 1, Dec 6. (time depends on groups)
- Final Report: Dec 13th (Tuesday of finals)

Outline

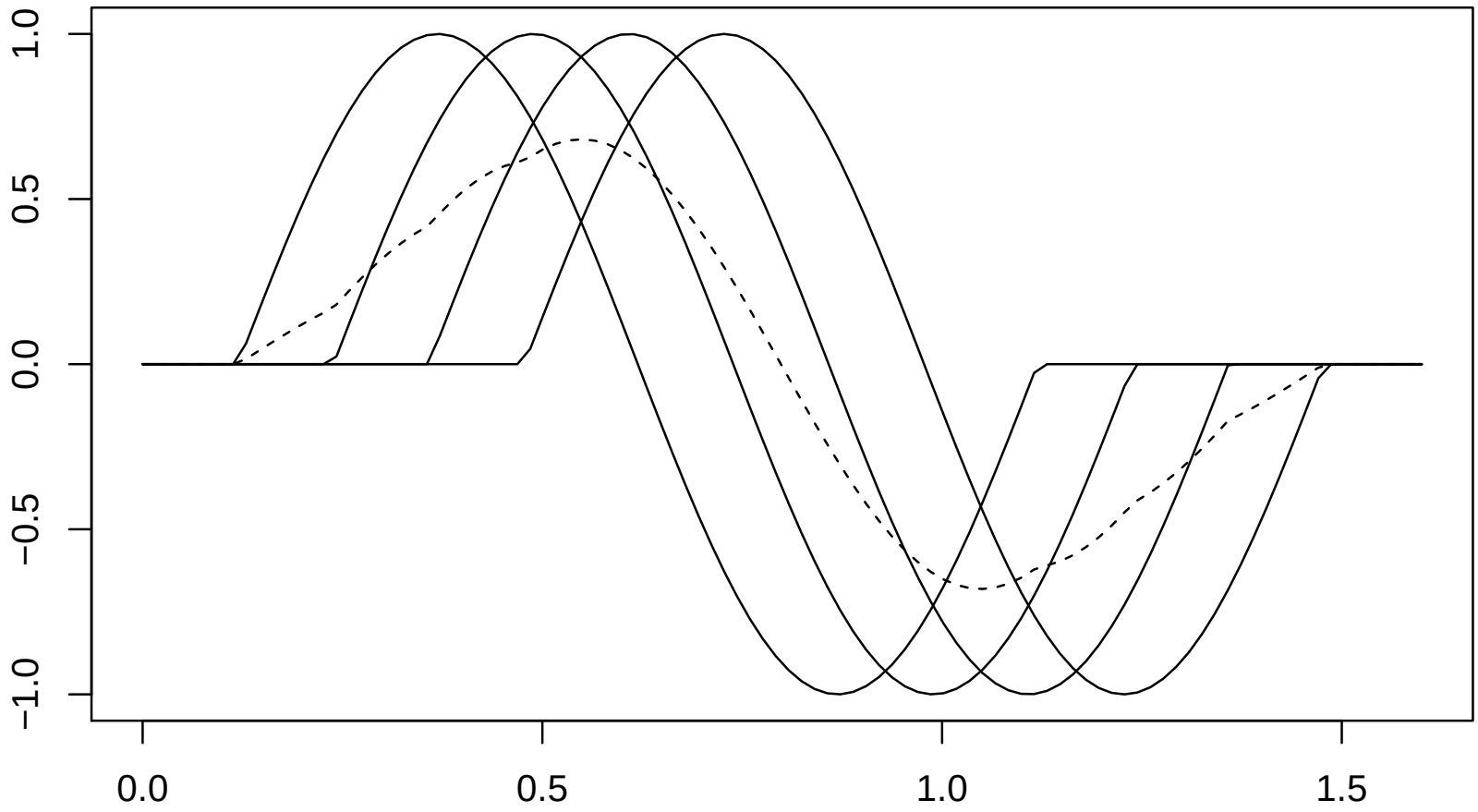
We continue Chapter 2 - Further topics.

- Curve Alignment and examples.

Curve alignment

Often samples of curves may be "out of alignment", and inference can be more accurate/interpretable if the curves are proprely aligned. The most common reason for being out of alignment is that each unit has its own "time scale". For example, subjects will hit puberty/growth spurts at slightly different ages, but we might want these events to be aligned across subjects.

Example



Modes of variation

A sample of curves can be decomposed into two sources of variation:

- amplitude variation - variation between curves in terms of height/amplitude,
- phase variation - variation between curves in terms of the domain.

Often one wishes to eliminate the phase variation and focus exlusively on the amplitude variation (but this is not always the case).

Modeling phase variation

We model phase variation by introduciont *warping functions*:

$$X_n(t) = X_n^*(h_n(t)).$$

Here X_n is the observed (unaligned) curve, X_n^* is the uobserved aligned curve, and h_n is a subject specific warping function that is also unobserved. Variation between $X_n(t)$ and $X_n^*(t)$ describes the phase variation, while variation between different $X_n^*(t)$ describes the phase variation.

Warping functions

There are a few properties that make sense for the warping functions $h_n(t)$:

- $h(0) = 0$,
- $h(T) = T$,
- $h(t)$ is monotonically increasing.

Any concerns? If we can estimate the warping functions $h_n(t)$ then the aligned curves will be

$$X_n^*(t) = X_n(h_n^{-1}(t)).$$

Landmark Registration

The first approach to alignment is called landmark registration. The idea is to choose landmarks from each function to align. For example, the maximum, minimum, and so on. Often this involves manually choosing the location of the landmarks through some point and click function, so it can be very time consuming.

Let $t_{n1}, \ldots, t_{nJ}$ be the selected time points, and suppose $t_1, \ldots, t_n$ are the selection made off of the mean function (or some function from the sample). Then the goal is to find h_n such

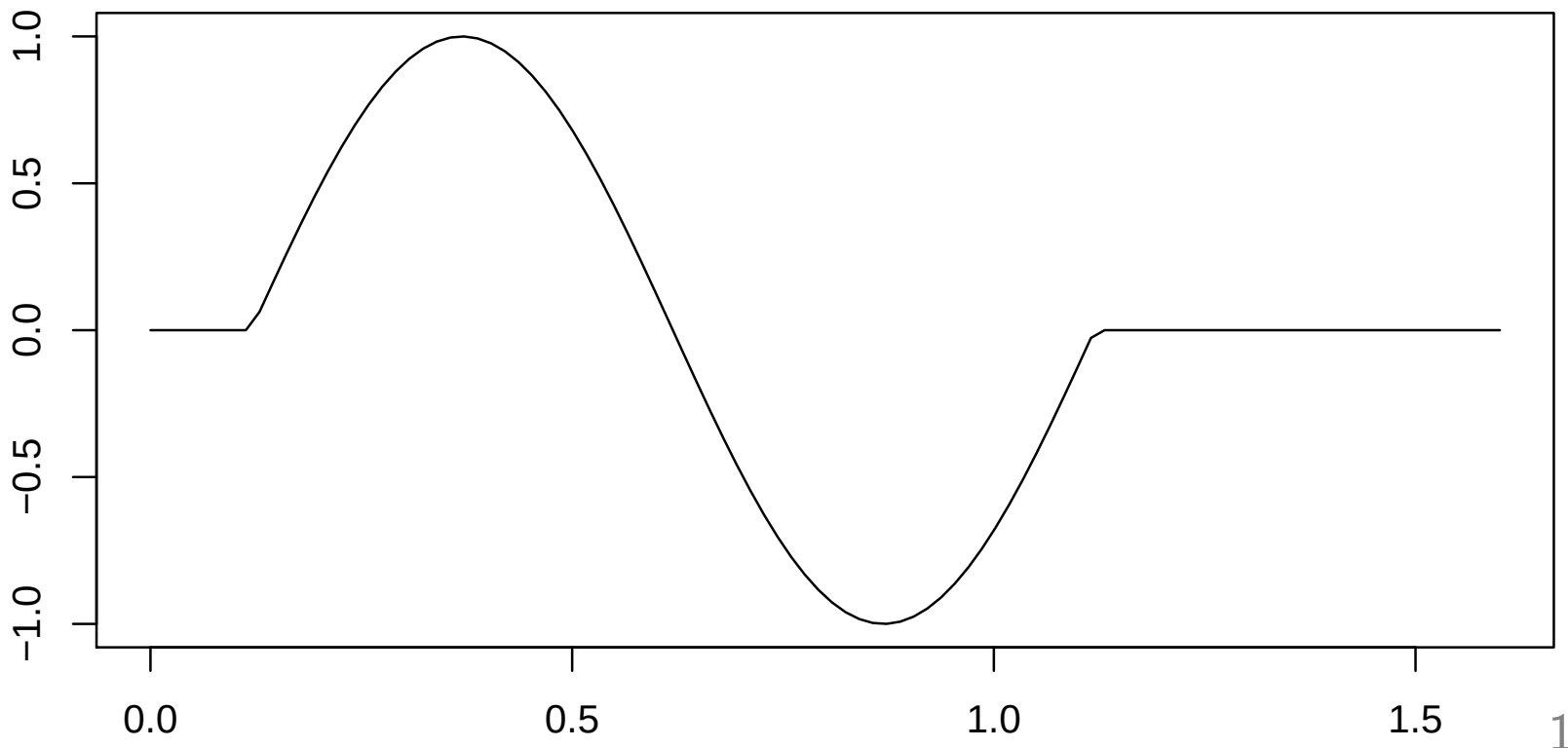
$$h_n(t_j) = t_{nj},$$

which can be done using fitting a polynomial or some basis expansion.

Landmark - R

Using the `locator` function, you can click on plots to find the coordinates of landmarks. You can then use the `landmarkreg` function to align the functions.

```
plot(times,x1,type="l",ylab="",xlab="")
num_loc<-2
t_loc<-locator(num_loc)
```



Continuous Registration

Continuous registration is much easier to automate. The idea is that you choose a specific curve (usually the mean function) and align "all" time points to that curve. So basically, one would choose h_n such that

$$\int (X(h_n^{-1}(t)) - \bar{X}(t))^2 dt,$$

is as small as possible. Typically the h_n^{-1} is expanded using some basis, while imposing our other discussed constraints. This can be carried out using the `register.fd` function in `R`.

Example - Berkeley

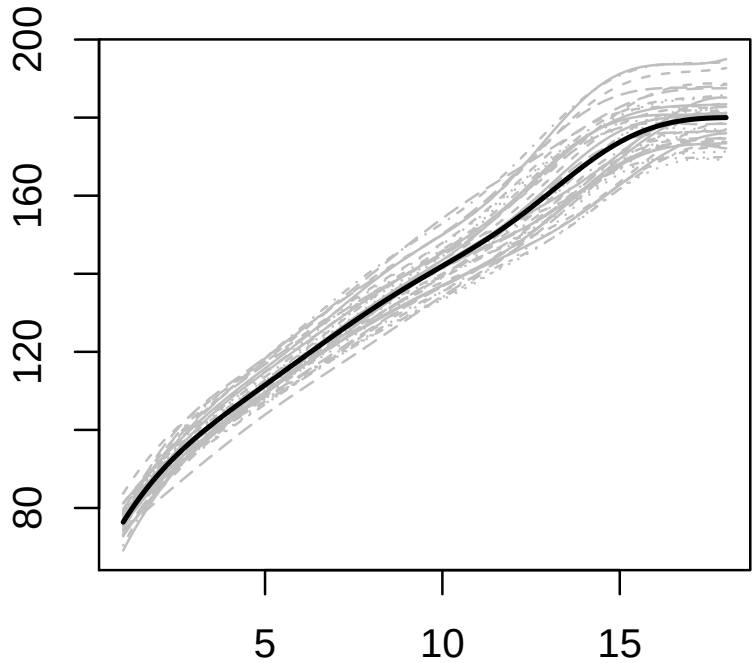
Let's return to the Berkeley growth data, and try out the alignment. Why does registering such data make sense?

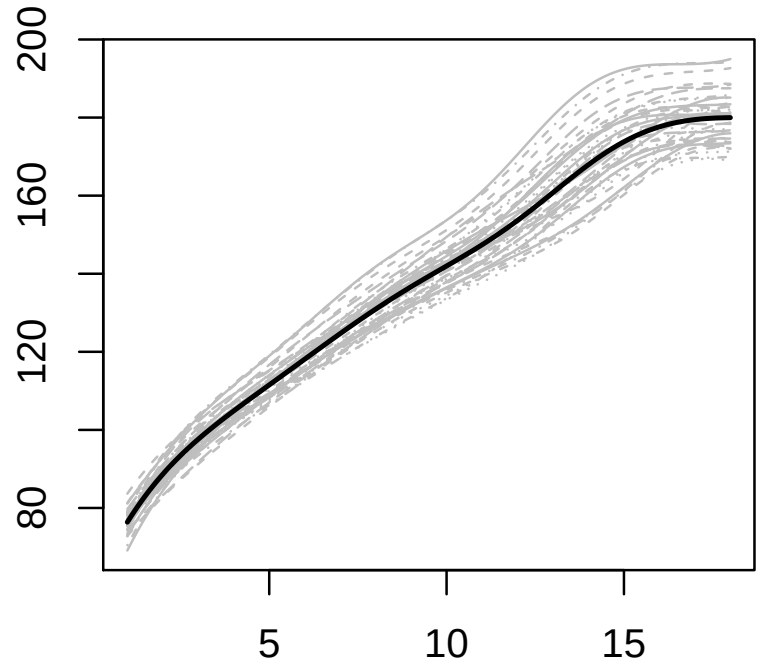
```
hgtbasis <- create.bspline.basis(c(1,18), norder = 6, breal
growfdPar <- fdPar(hgtbasis, 4, 10^(-.5))
X.f<-smooth.basis(growth$age,growth$hgtm,growfdPar)$fd
X.reg<-register.fd(X.f)
```

```
names(X.reg)

## [1] "regfd"  "warpfd" "Wfd"    "shift"  "y0fd"
## [6] "yfd"
```

Example - Berkeley

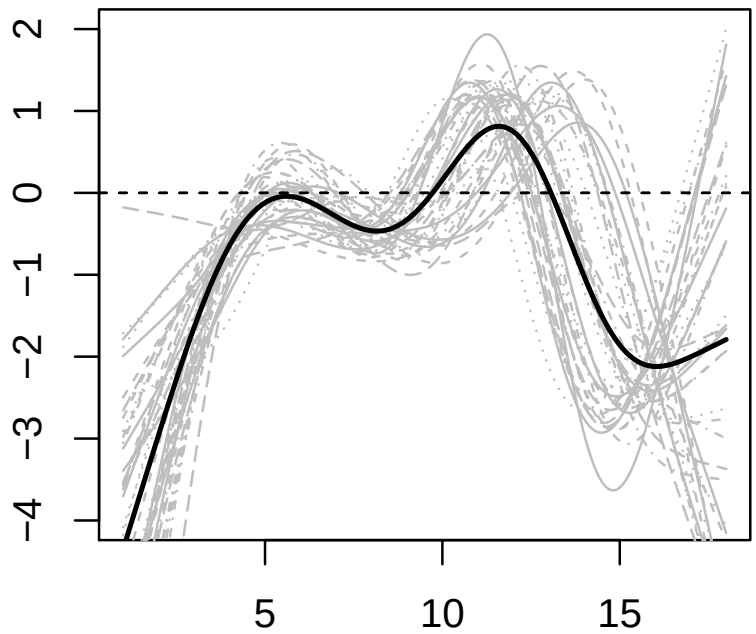
Let's return to the Berkeley growth data, and try out the alignment. Why does registering such data make sense?

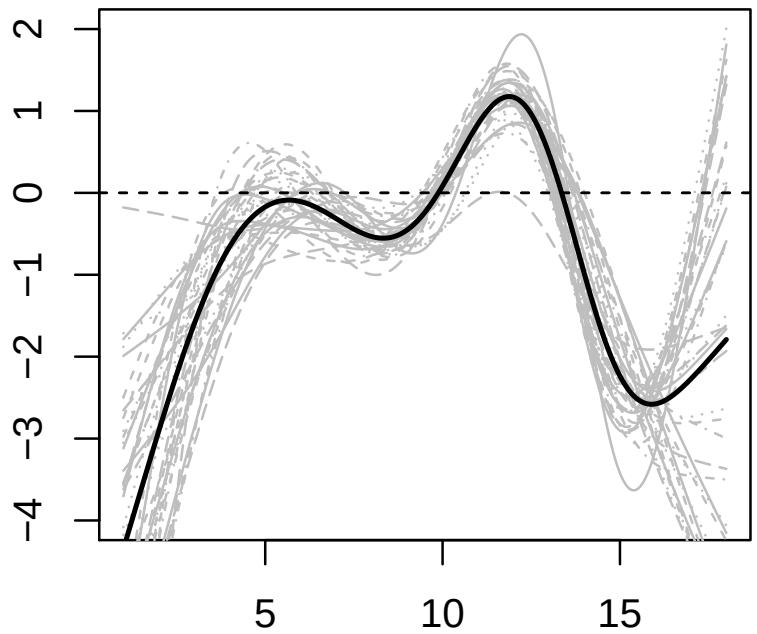




Example - Berkeley

Let's take a look at the 2nd derivatives. First we take the derivatives, then we align them.





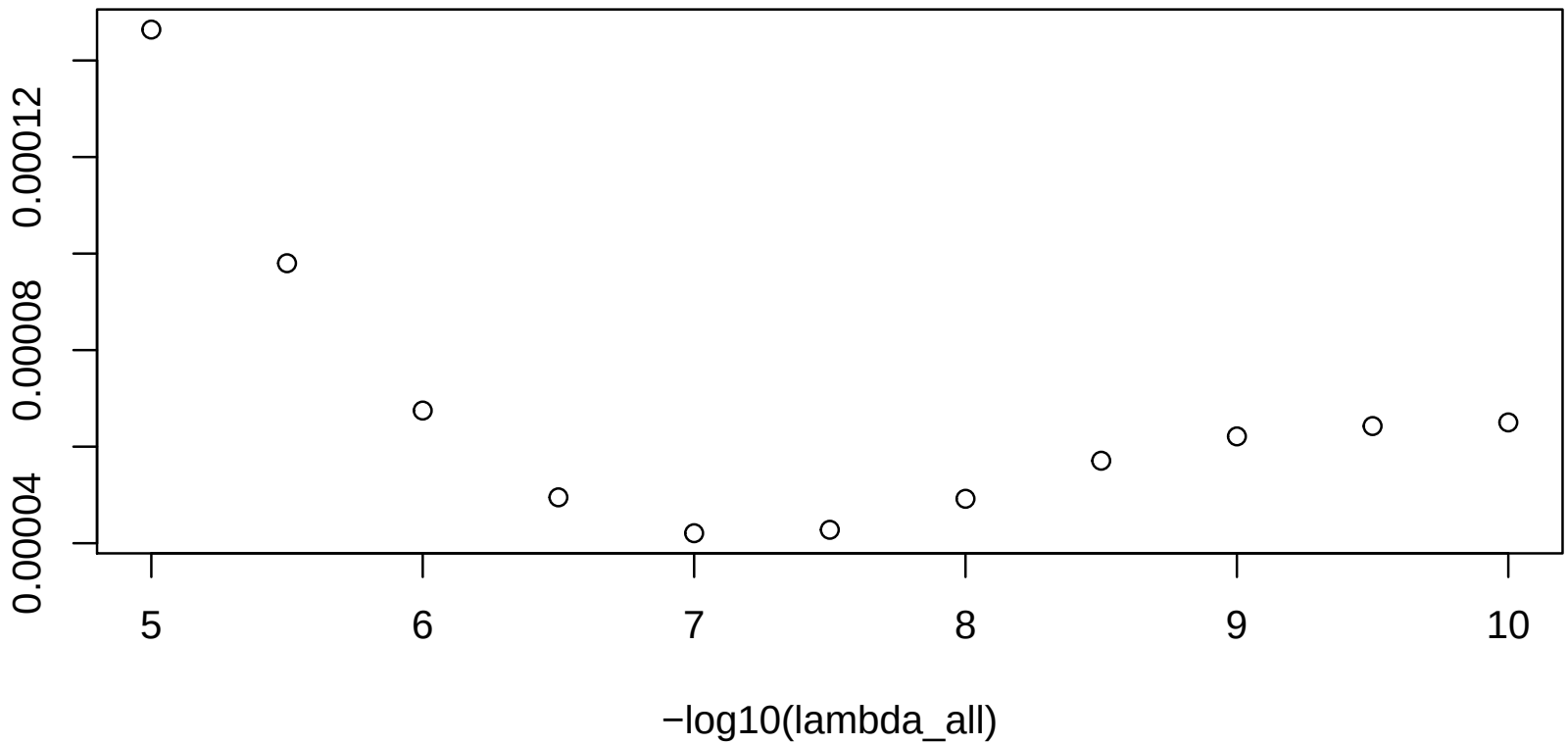
Example - DTI

Lets head back to the DTI data. Let's more carefully smooth.

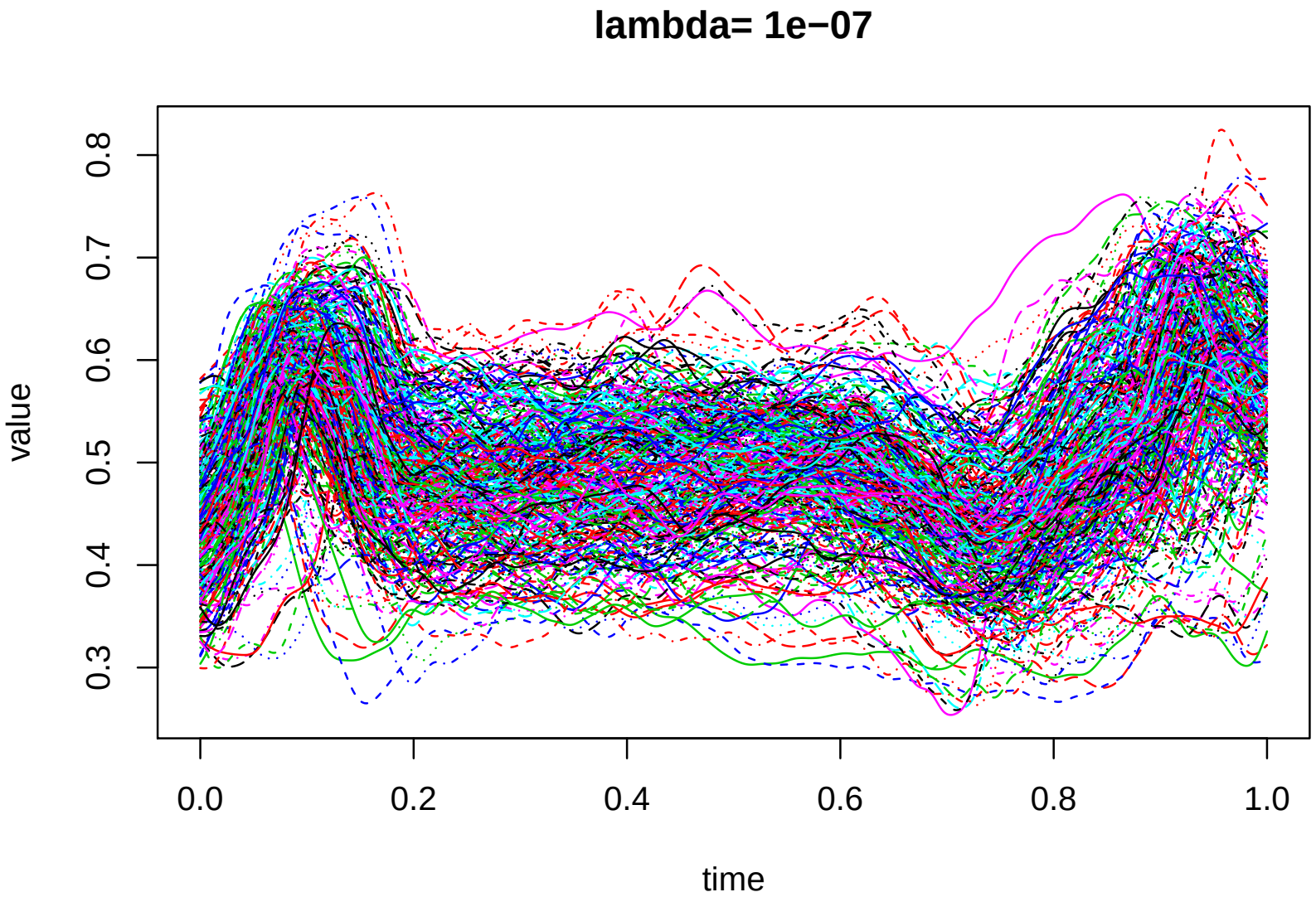
```
Corp<-DTI$cca
drop<-unique(which(is.na(Corp),arr.ind=TRUE)[,1])
Corp<-Corp[-drop,] # Missing value
pts<-seq(0,1,length=93)
my_basis<-create.bspline.basis(c(0,1),
                nbasis=100,norder=6)
lambda_all<-10^(-(10:20)/2)
gcv_all<-numeric(0)
for(lambda in lambda_all){
myPar<-fdPar(my_basis,2,lambda)
Corp.F<-smooth.basis(pts,t(Corp),myPar)
gcv_all<-c(gcv_all,mean(Corp.F$gcv))}
```

Example - DTI

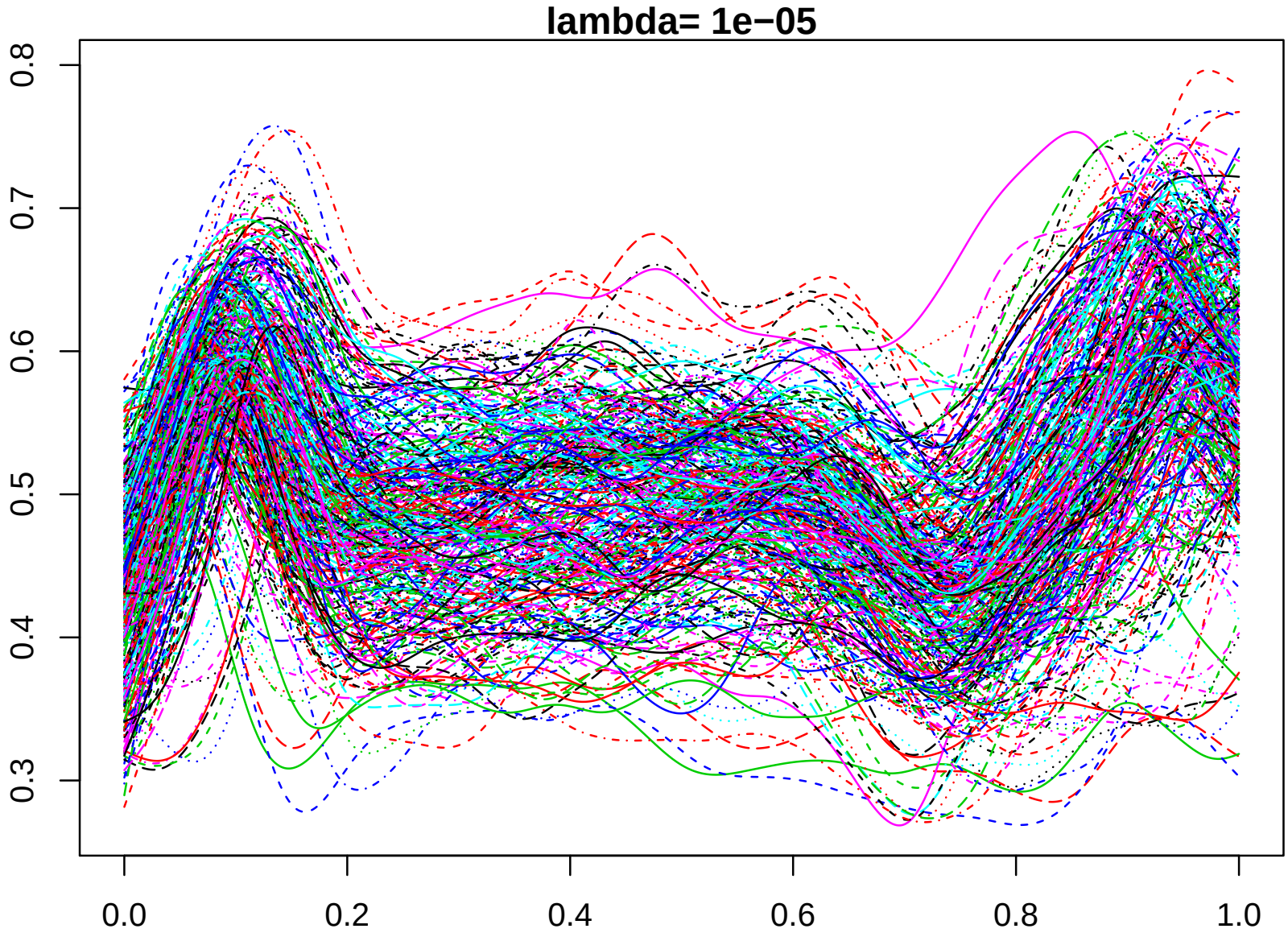
```
lambda_all[which.min(gcv_all)]

## [1] 1e-07

plot(-log10(lambda_all),gcv_all)
```



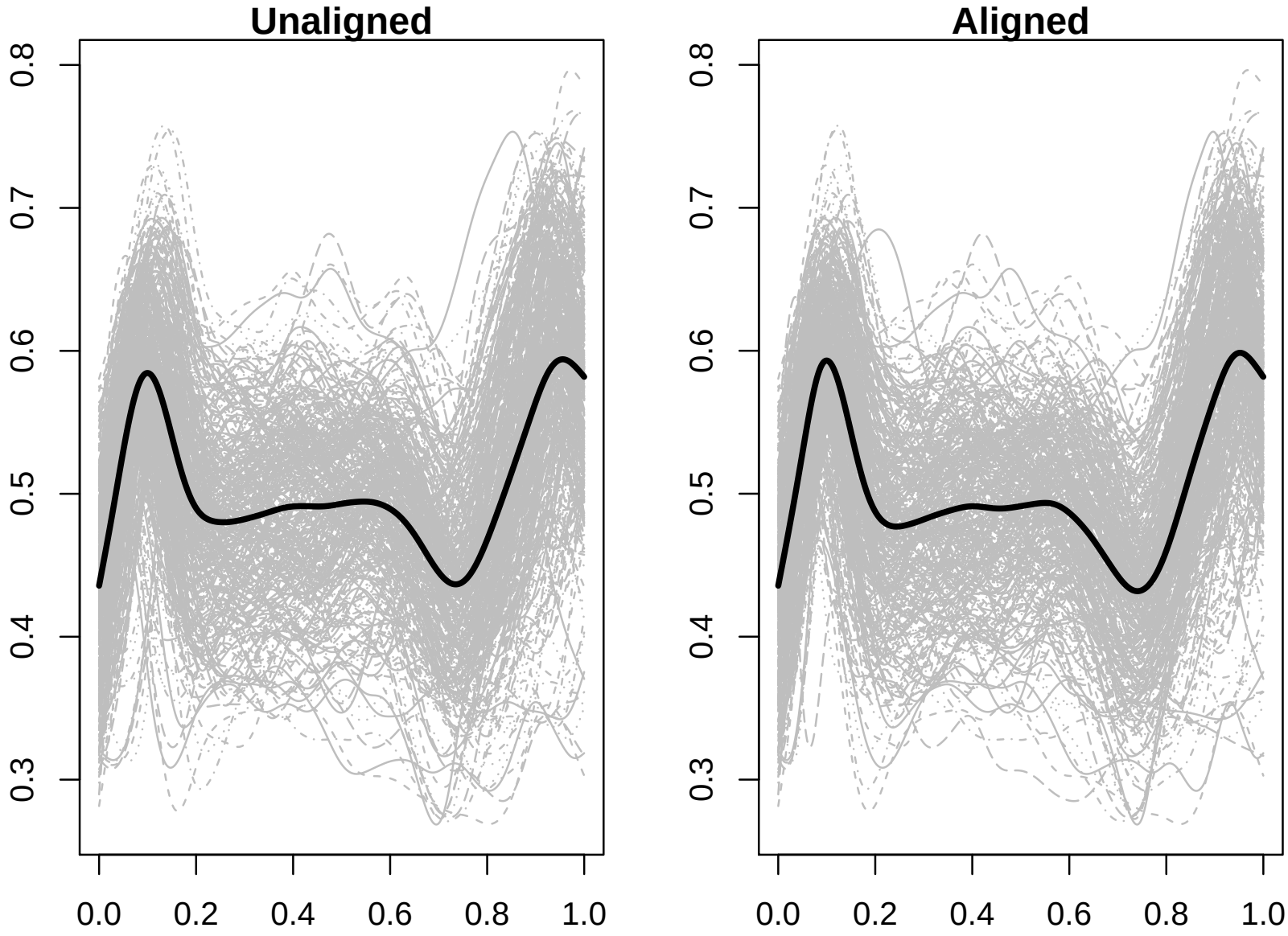
Example - DTI



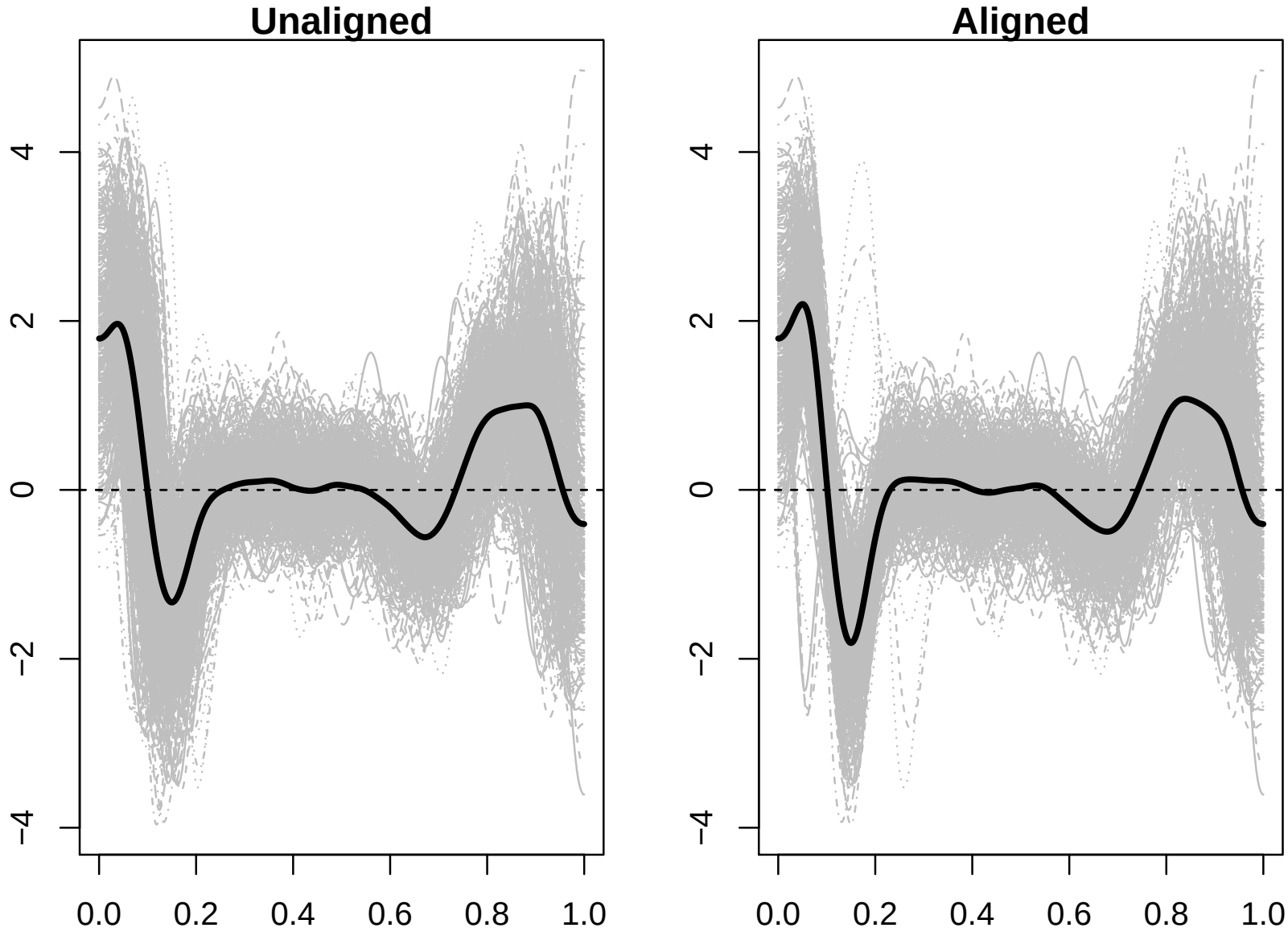
Example - DTI



Example - DTI



Example - DTI



Mathematical Framework

Next week we will move into the theoretical framework for FDA. This will be covered in Chapters 11, 12, and 13. We will skip Chapter 3 as it is designed as a lower level replacement for Chapters 11 and 12:

- Chapter 10 - Hilbert spaces,
- Chapter 11 - Random Functions,
- Chapter 12 - Statistical inference.