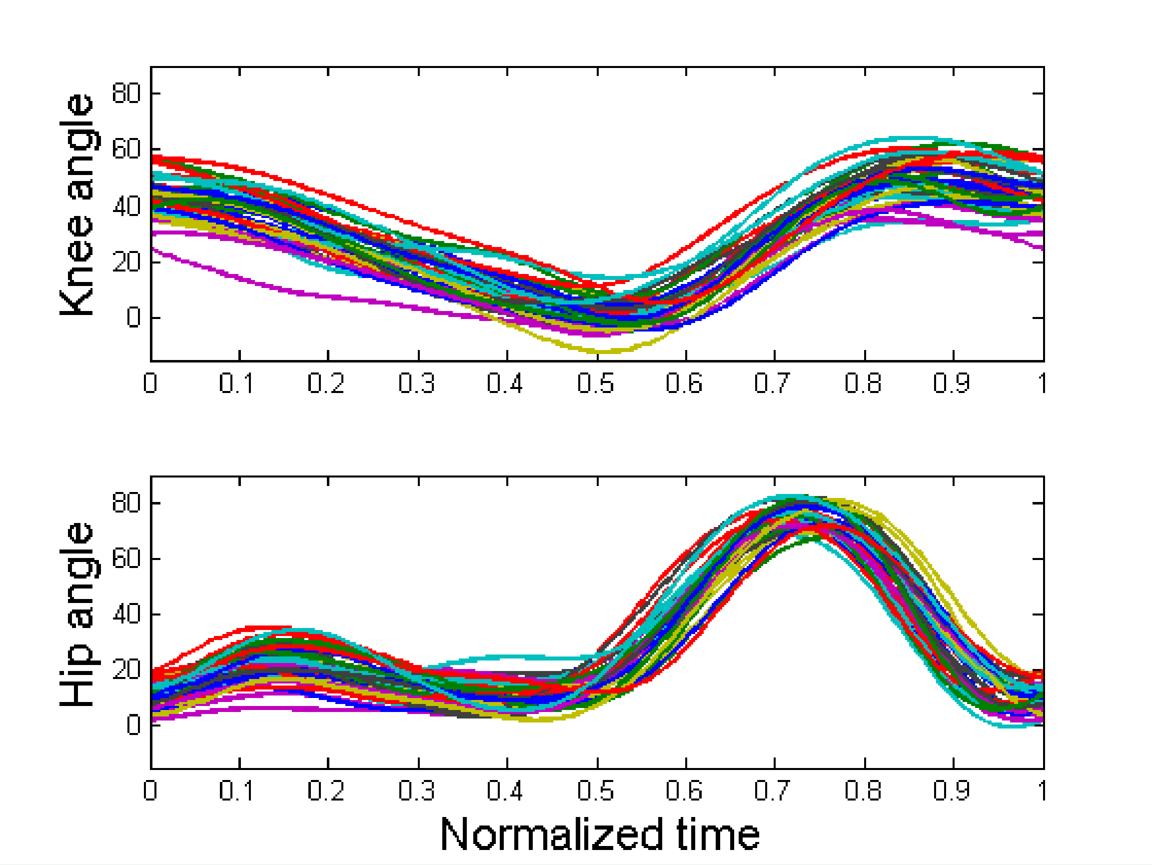
Canonical correlation analysis

Exploring correlations between two sets of functions.

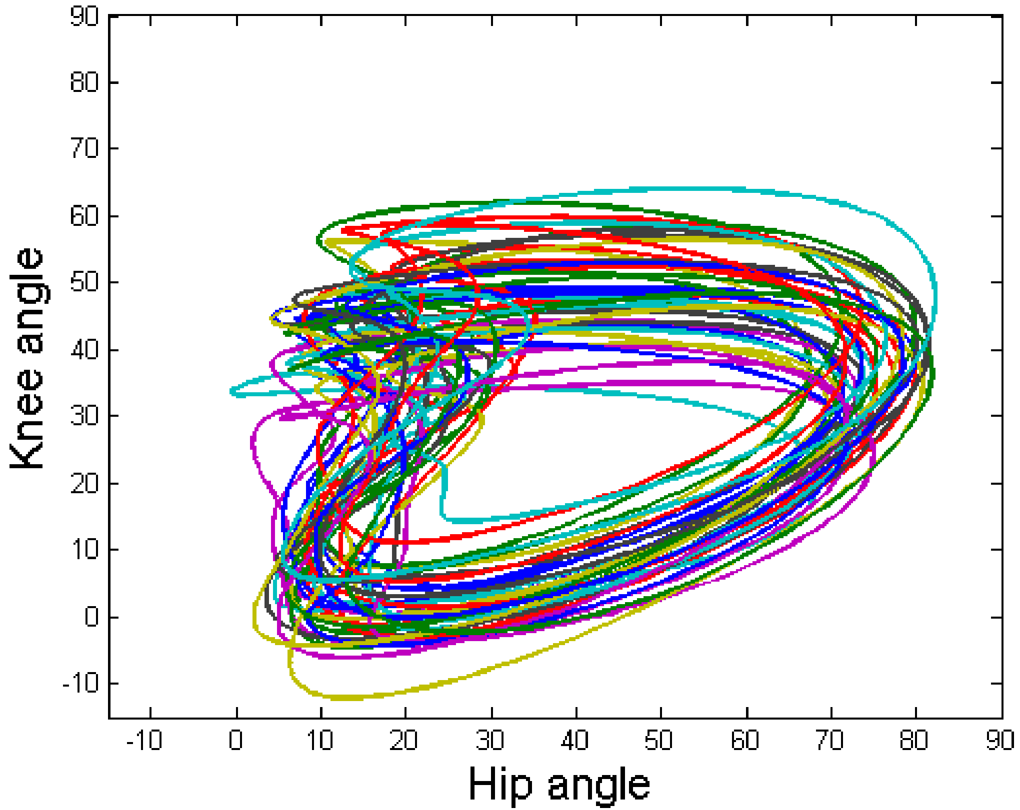
1. Overview

- Let the two sets of functions be $x_i(t)$ and $y_i(t), i = 1, \ldots, N$.
- The correlation surface $r(s, t)$ offers correlations for each pair of values $[x_i(s), y_i(t)]$. This may be more information than we can deal with.
- We might prefer to only look at the *dominant* modes of correlation between two sets of functions.
- We will explore the types of correlation between knee and hip angles for the gait data.

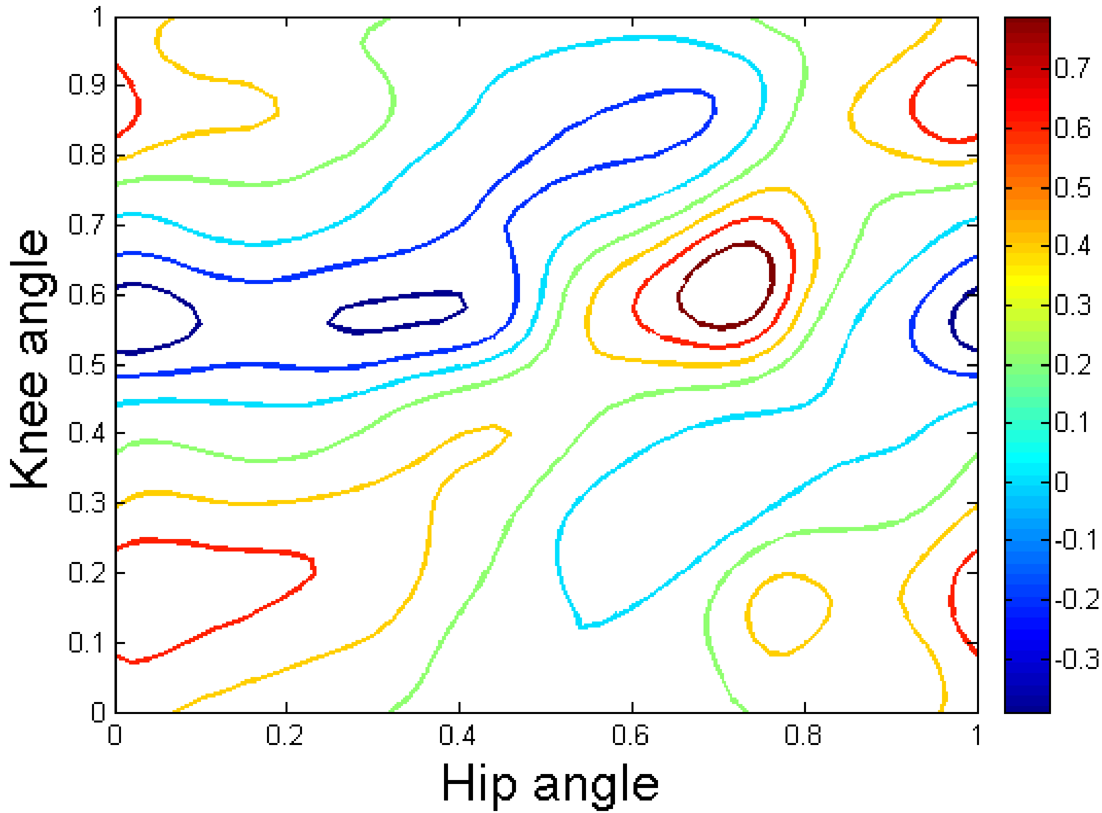
Knee and hip angles for the gait data



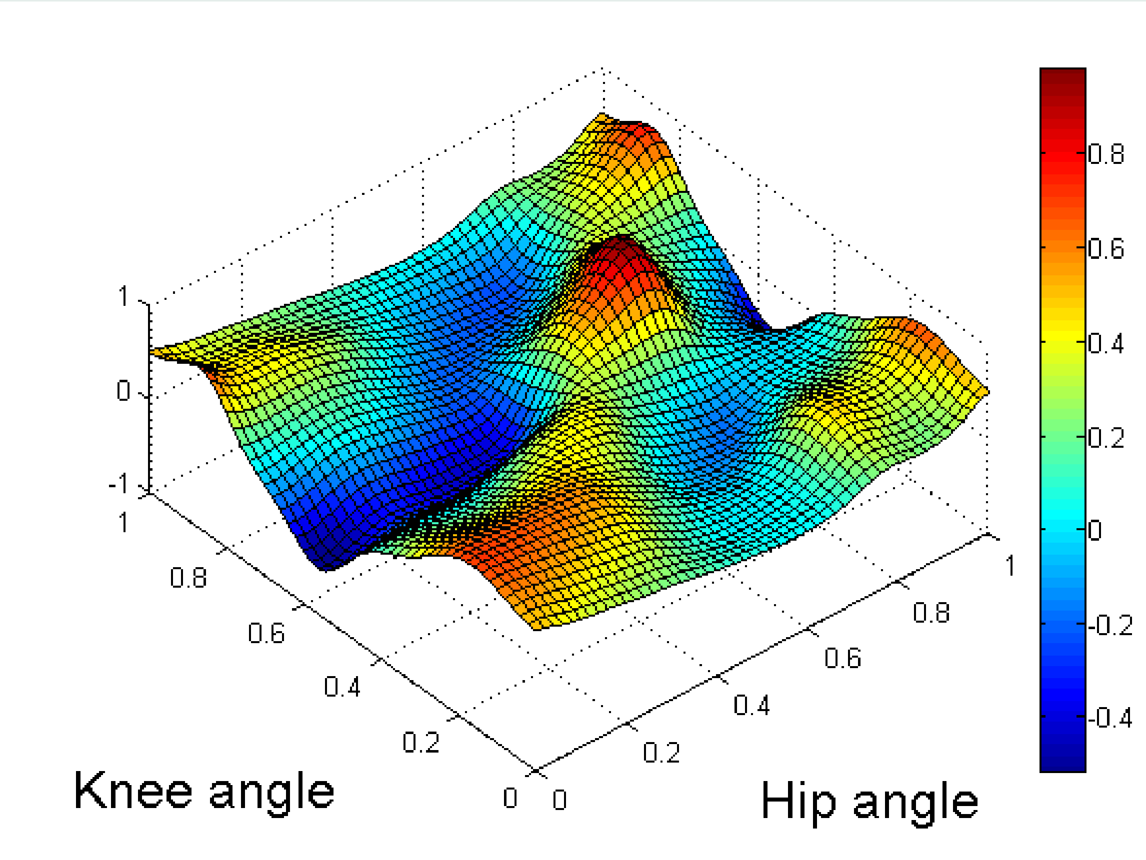
Knee and hip angles for the gait data



Contour plot of correlations between knee and hip angles



Surface plot of correlations between knee and hip angles



2. Functional canonical correlation analysis defined

Probes for modes of correlation

- Let the two sets of functions be $x_i(t)$ and $y_i(t), i = 1, \ldots, N$.
- Let a probe for the x functions be defined by

$$\int \xi(t) x_i(i)\, dt.$$

- Let a probe for the y functions be defined by

$$\int \eta(t) y_i(i)\, dt.$$

- Given a pair of probe weighting functions (ξ, η), how strongly are the two sets of probe values correlated?

The first canonical correlation ρ_1

- We can re-phrase this question: What pair of probe weighting functions would give the highest possible correlation between probe values?
- The correlation that results is called the first *canonical correlation*, and indicated by ρ_1.
- Let the maximizing probe weighting functions be indicated by ξ_1 and η_1. These are called the first pair of *canonical weighting functions*.
- The probe scores

$$f_{i1} = \int \xi_1(t) x_i(t)\, dt \text{ and } g_{i1} = \int \eta_1(t) y_i(t)\, dt$$

are called the first pair of *canonical variables*.

Subsequent canonical correlations ρ_j

- We can repeat this process. Let a new pair of canonical weight functions ξ_2 and η_2 be defined, and let them satisfy the two restrictions

$$\int \xi_1(t)\xi_2(t)\,dt = 0 \text{ and } \int \eta_1(t)\eta_2(t)\,dt = 0.$$

- We can again ask for the pair of ξ_2 and η_2 that maximize the correlation ρ_2 between the pairs of canonical variable values (f_{i2}, g_{i2}) defined by these canonical weight functions.
- And so on, until there are no more interesting canonical correlations to be found.

Smoothing canonical weight functions.

- If you already know about the multivariate version of canonical correlation, you know that the method tends to produce a rather large number of large correlations, and that these are often difficult to interpret.
- There is the same problem in the functional environment. In fact, even worse. Subsequent canonical weight functions tend to emphasize higher and higher frequency modes of variation that are usually impossible to interpret, or else are not very interesting.

- The solution here as in principal components analysis and smoothing, is to *regularize* the canonical weights.
- We measure the roughness of each canonical weight function by the two penalties

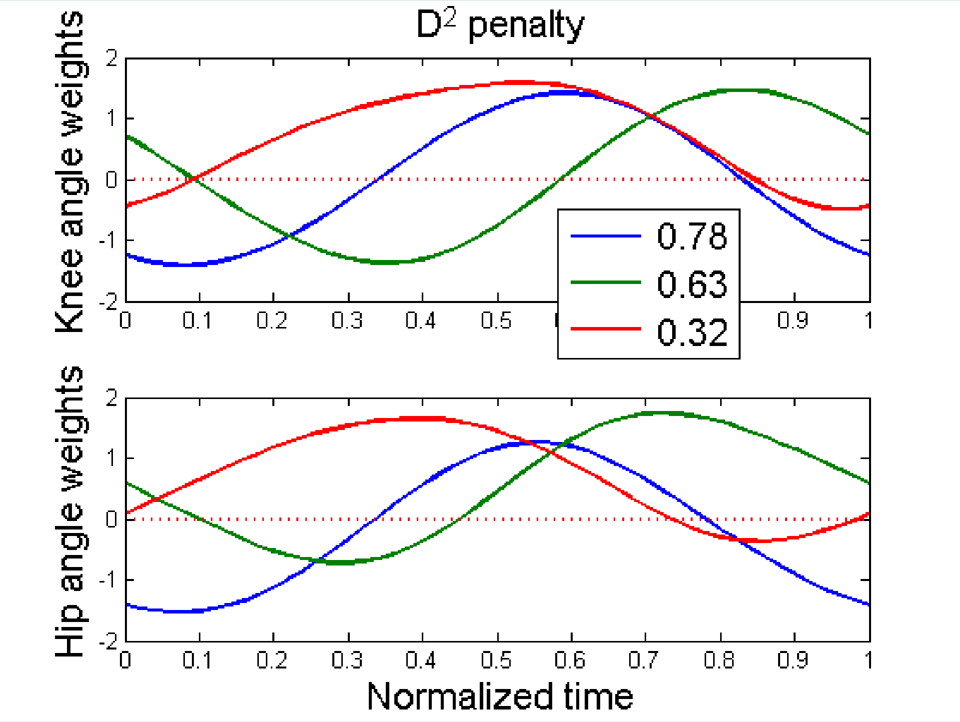
$$\int [L\xi_j(t)]^2\,dt \quad \text{and} \quad \int [M\eta_j(t)]^2\,dt$$

where L and M are suitably chosen linear differential operators. $L = M = D^2$ will often be reasonable.
- Each of these penalties is multiplied by a smoothing parameter λ.

3. Canonical correlation analysis for the gait data

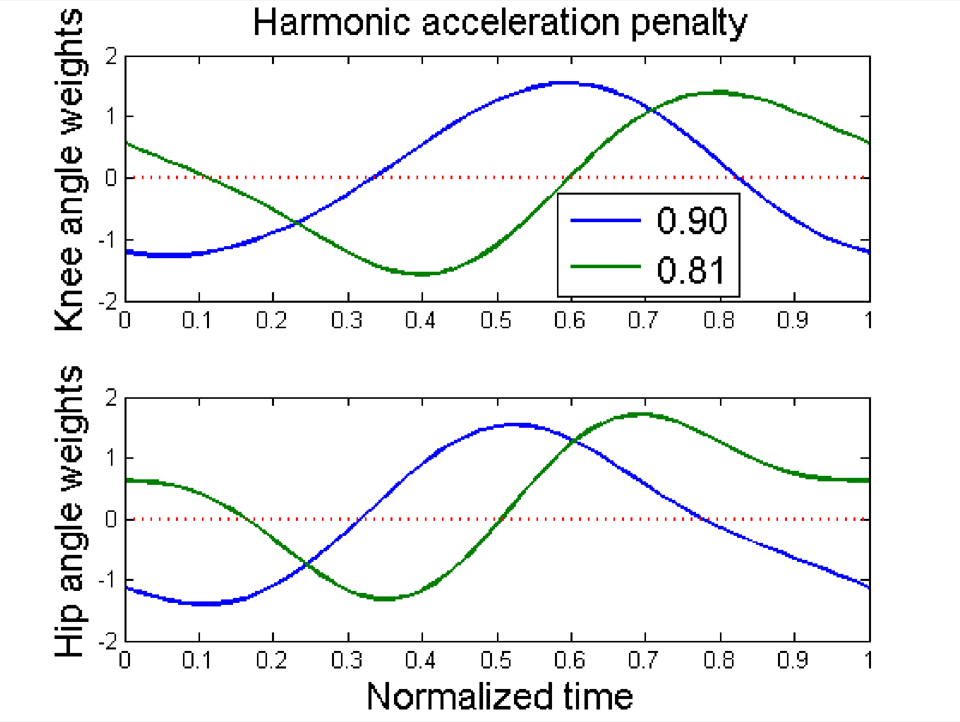
- In order to explore the correlations between knee and hip angles for the gait data, we first used D^2 to define both penalties on the roughness of each pair of weighting functions.
- The weights placed on these two penalties were $\lambda = 0.001$.
- The first three canonical correlations were 0.73, 0.58 and 0.31. The next was 0.05.
- It seemed that there were only three important modes of correlation between knee and hip angle.

The first two canonical weight functions for knee and hip angles using D^2 smoothing.

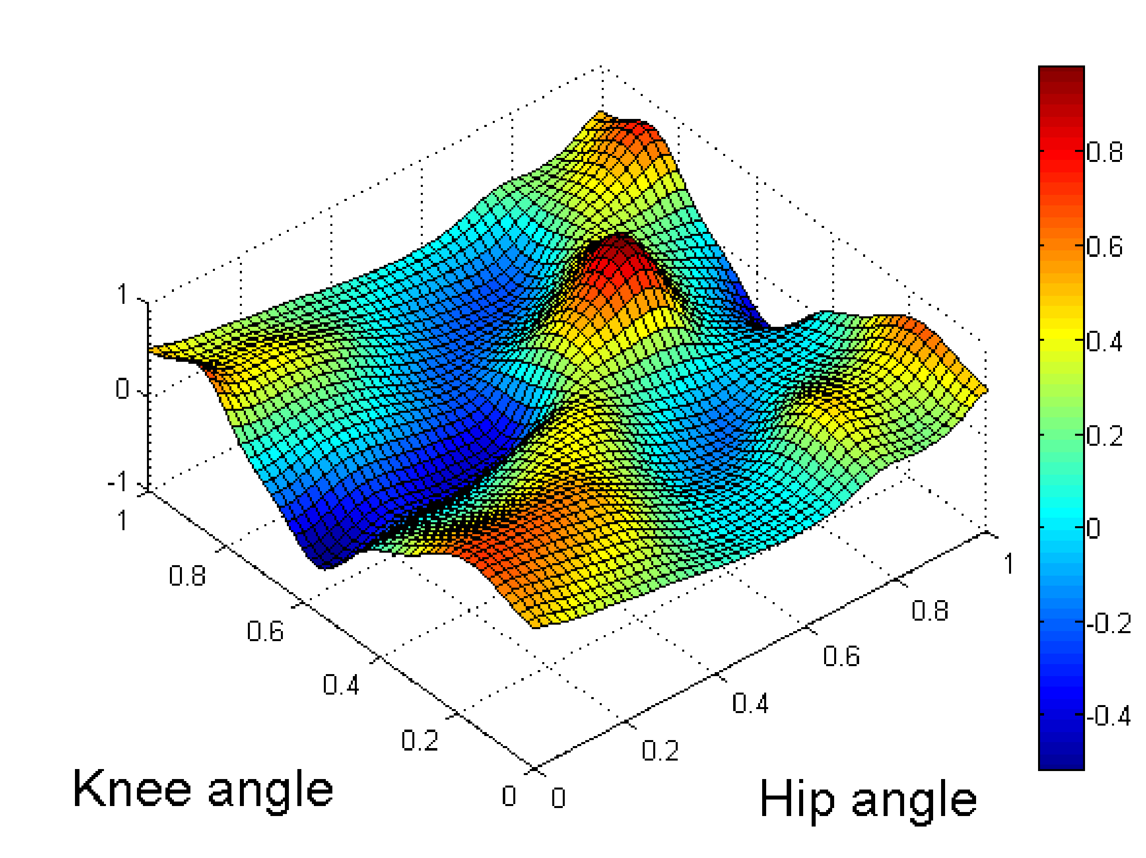


- But since the functions are essentially periodic, we decided that the harmonic acceleration operator would be a better choice of L and M.
- Using $\lambda = 10^{-6}$, we found the first four canonical correlations to be 0.90, 0.81, 0.39 and 0.26.
- This analysis highlighted only two dominant modes of variation.

The first two canonical weight functions for knee and hip angles using harmonic acceleration smoothing.



Surface plot of correlations between knee and hip angles



4. Summary

- Canonical correlation allows us to simplify the study of the ways in which two sets of functions are correlated.
- But we need to impose smoothness on the canonical weighting functions in order to keep the number of modes of variation with high correlations small.