Functional Analysis in Data Modelling

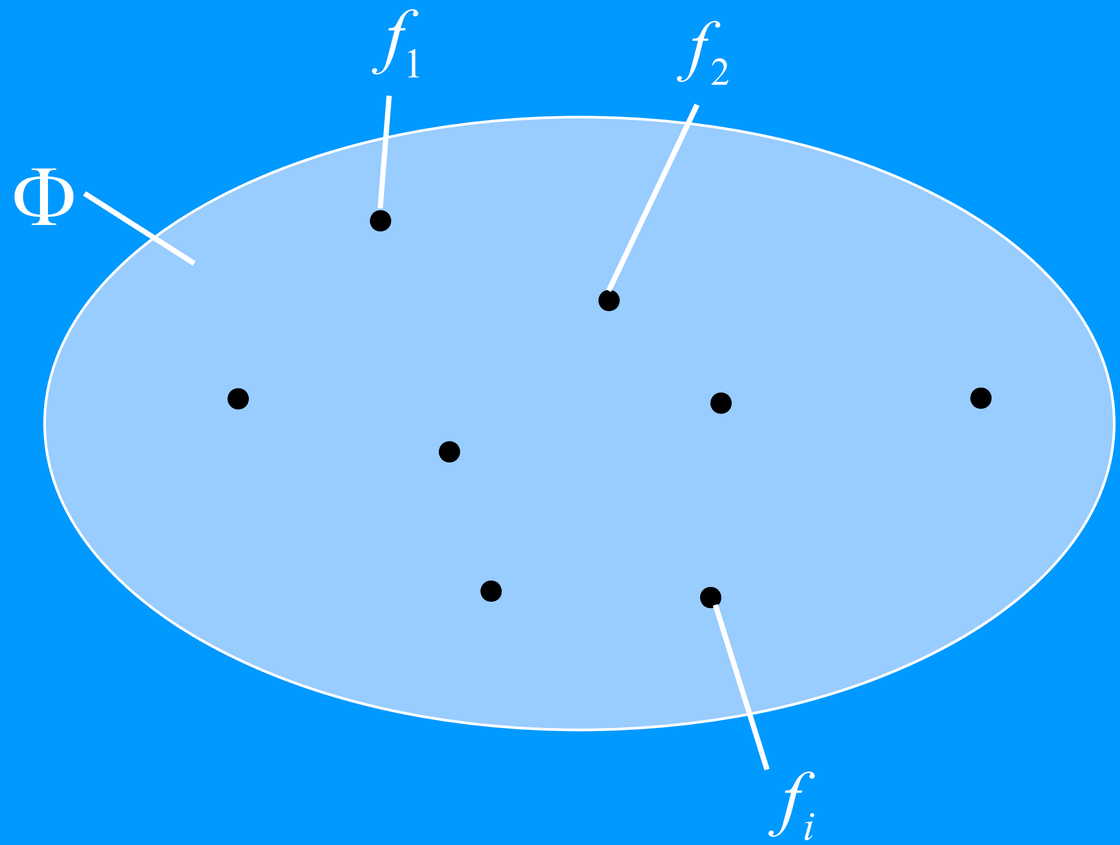
Tony Dodd

t.j.dodd@shef.ac.uk

Overview

- Metric spaces
- Linear spaces
- Normed, inner product and Hilbert spaces
- Best approximation
- Reproducing kernel Hilbert spaces
- Approximation vs estimation

Spaces



Spaces

Define some space Φ

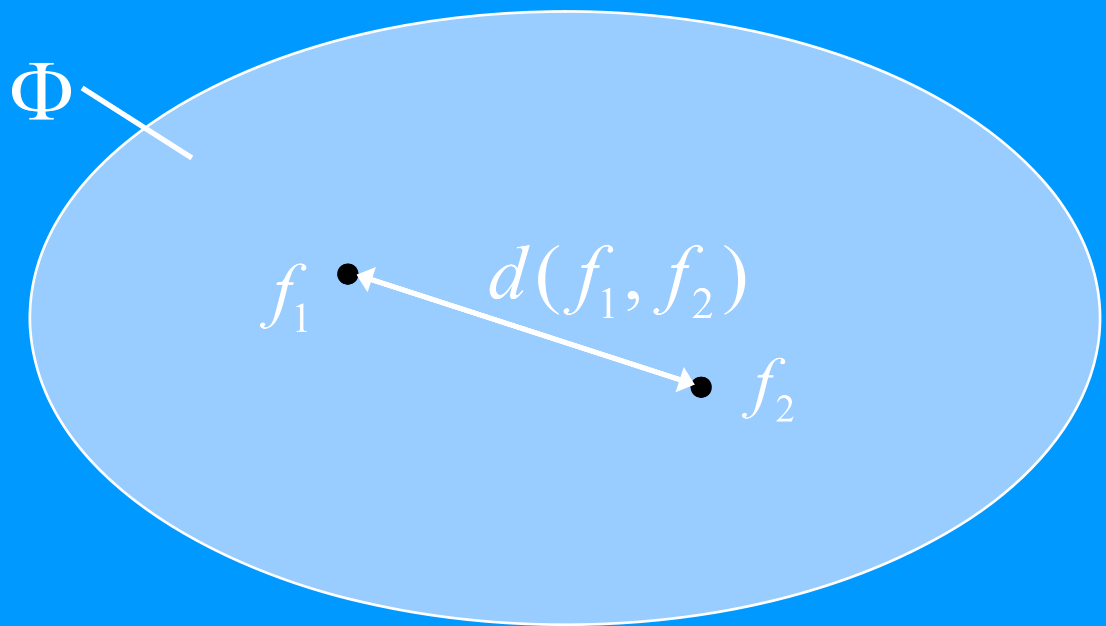
Points/elements in Φ denoted f_i

What is Φ ?

- Euclidean space
- Space of sines and cosines (Fourier)
- Space of bandlimited functions (Paley-Wiener)
- L_2
- Can be a nonlinear space

Metric spaces

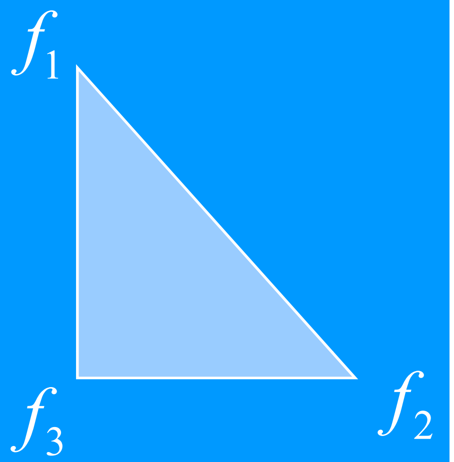
- Put some structure on our space Φ



Metric spaces

Define the distance $d : \mathrm{F} \times \mathrm{F} \to \mathbb{R}$

- $d(f_1, f_2) \geq 0$
- $d(f_1, f_2) = 0$ if and and only if $f_1 = f_2$
- $d(f_1, f_2) = d(f_2, f_1)$
- $d(f_1, f_2) \leq d(f_1, f_3) + d(f_3, f_2)$



Why are metric spaces important?

- Allow us to define the distance between functions
- Can then talk about best approximations
- Completeness – no holes in the space
- We want to look at spaces that are very similar to Euclidean space

Linear spaces

Need two algebraic operations:

- Addition $f_3 = f_1 + f_2$:
  - $f_1 + f_2 = f_2 + f_1$ and $f_1 + (f_2 + f_3) = (f_1 + f_2) + f_3$
- Multiplication by scalars $f_2 = \alpha f_1$:
  - $\alpha(f_1 + f_2) = \alpha f_1 + \alpha f_2$
  - $(\alpha + \beta) f_1 = \alpha f_1 + \beta f_1$
  - $\alpha(\beta f_1) = (\alpha\beta) f_1$
  - $1 \cdot f_1 = f_1$; $0 \cdot f_1 = 0$

Linear spaces and basis

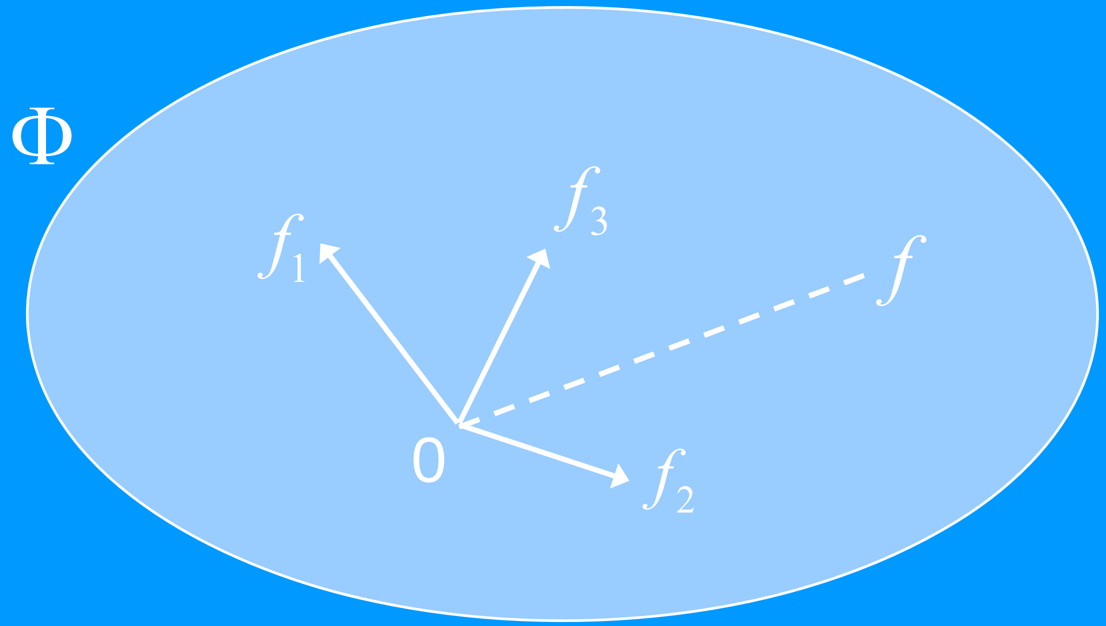
- Set of $f_1, f_2, \ldots, f_n$ is linearly independent if

$$\alpha_1 f_1 + \alpha_2 f_2 + \ldots + \alpha_n f_n = 0$$

- Holds only if each $\alpha_i = 0$.
- Finite dimensional if only n linearly independent elements
- Linear manifold: $\alpha f_1 + \beta f_2$ in Φ
- Basis: can express every f in Φ in the form

$$f = \alpha_1 f_1 + \alpha_2 f_2 + \ldots + \alpha_n f_n$$

Basis



Normed spaces

- Define the notion of the size of an element in Φ
- Norm $\|f\|$

$$\|f\| \geq 0, \ \|f\| = 0 \Leftrightarrow f = 0$$

$$\|\alpha f\| = |\alpha| \|f\|$$

$$\|f_1 + f_2\| \leq \|f_1\| + \|f_2\|$$

- Defines a metric $d(f_1, f_2) = \|f_1 - f_2\|$

Algebraic and geometric

- By defining the metric based on the norm we have combined the algebraic (metric) and geometric (linear, norm) properties.
- Algebraic – the technical bits that we need but are difficult!
- Geometric – the intuitive bits.

Inner product spaces

- Linear space with an inner product

$$\langle \cdot,\cdot \rangle : F \times F \to \mathbb{R}$$

$$1.\langle f_1, f_2 \rangle = \langle f_2, f_1 \rangle$$

$$2.\langle \alpha_1 f_1 + \alpha_2 f_2, f_3 \rangle = \alpha_1 \langle f_1, f \rangle_3 + \alpha_2 \langle f_2, f \rangle_3$$

$3.\langle f, f \rangle \geq 0$ with equality iff $f = 0$

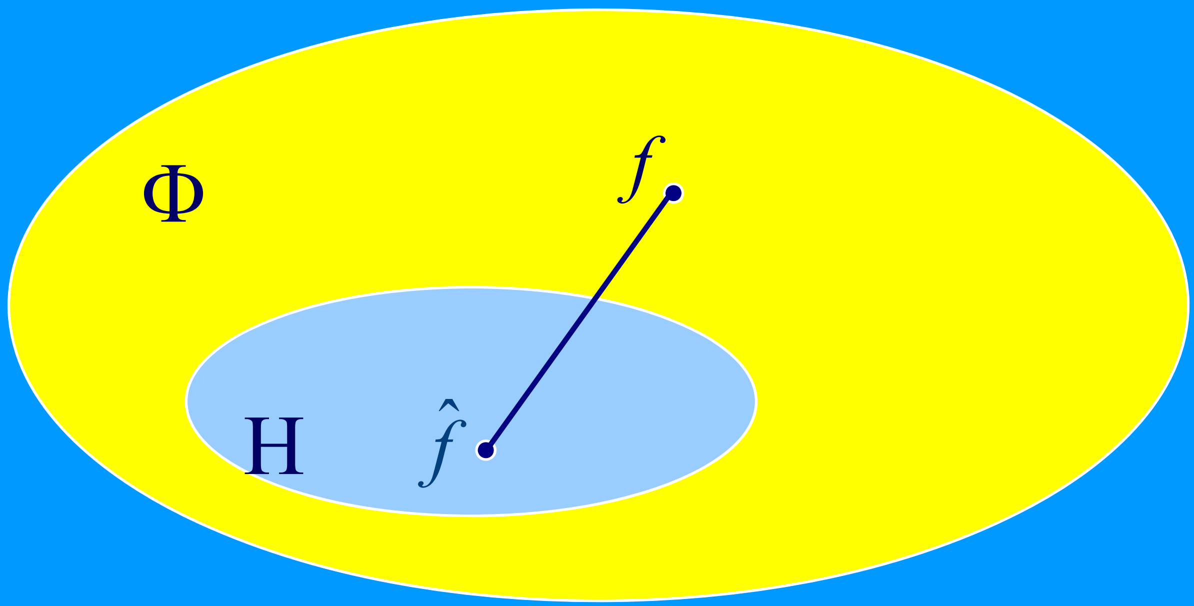
- Define the norm as

$$\|f\| = \sqrt{\langle f, f \rangle}$$

Hilbert spaces

- These “look like” Euclidean space.
- *An inner product space which is complete with respect to the metric defined from the inner product is called a Hilbert space.*
- In simple terms they have all the nice mathematical properties we need – so don’t worry about the complicated bits.

Best approximation



Best approximation

- For any given f in a Hilbert space Φ and a closed subspace $H \subset F$ there exists a unique best approximation $\hat{f}$ to f out of H.
- In fact $F = H \oplus H^{\perp}$
- i.e. $f - \hat{f} \perp H$

Best approximation

- Assume H is finite dimensional with basis

$$\{k_1, k_2, \ldots, k_m\}$$

- i.e. $\hat{f} = \alpha_1 k_1 + \alpha_2 k_2 + \cdots + \alpha_m k_m$
- $f - \hat{f} \perp H$ gives m conditions ($i=1,\ldots,m$)

$$\langle k_i, f - \alpha_1 k_1 - \alpha_2 k_2 - \cdots - \alpha_m k_m \rangle = 0 \quad \text{or}$$

$$\langle k_i, f \rangle - \alpha_1 \langle k_i, k_1 \rangle - \alpha_2 \langle k_i, k_2 \rangle - \cdots - \alpha_m \langle k_i, k_m \rangle = 0$$

Reproducing kernel Hilbert spaces (RKHS)

- A particular class of Hilbert space very important in machine learning.
- Splines, kernel machines, support vector machines, neural networks, Gaussian processes, time series analysis, bandlimited signals…

RKHS

- Hilbert space with even more structure.
- Not worry about technical details here.
- Main properties:
  - Observations: $y_i = \langle k_i, f \rangle$
  - $\left\langle k_i, k_j \right\rangle = k\left(x_i, x_j\right)$
  - k are positive definite functions

RKHS approximation

m conditions become:

$$y_i - \alpha_1 k\left(x_i, x_1\right) - \alpha_2 k\left(x_i, x_2\right) - \cdots - \alpha_m k\left(x_i, x_m\right) = 0$$

Can then estimate the parameters using:

$$\alpha = K^{-1} y$$

In practise can be ill-conditioned/noise on the data so minimise:

$$\sum_{i=1}^{m}\left(\hat{f}(x_i) - z_i\right)^2 + \lambda\left\|f\right\|^2, \lambda \geq 0$$

$$\alpha = \left(K + \lambda I\right)^{-1} y$$

Approximation vs estimation