Math Camp 1: Functional analysis

Sayan Mukherjee + Alessandro Verri

About the primer

Goal To briefly review concepts in functional analysis that will be used throughout the course.* The following concepts will be described

1. Function spaces

2. Metric spaces

3. Convergence

4. Measure

5. Dense subsets

*The definitions and concepts come primarily from "Introductory Real Analysis" by Kolmogorov and Fomin (highly recommended).

6. Separable spaces
7. Complete metric spaces
8. Compact metric spaces
9. Linear spaces
10. Linear functionals
11. Norms and semi-norms of linear spaces
12. Convergence revisited
13. Euclidean spaces
14. Orthogonality and bases
15. Hilbert spaces

16. Delta functions

17. Fourier transform

18. Functional derivatives

19. Expectations

20. Law of large numbers

Function space

A **function space** is a space made of functions. Each function in the space can be thought of as a point. Examples:

1. $C[a, b]$, the set of all real-valued *continuous* functions in the interval $[a, b]$;

2. $L_1[a, b]$, the set of all real-valued functions whose absolute value is integrable in the interval $[a, b]$;

3. $L_2[a, b]$, the set of all real-valued functions square integrable in the interval $[a, b]$

Note that the functions in 2 and 3 are not necessarily continuous!

Metric space

By a **metric space** is meant a pair (X, ρ) consisting of a space X and a distance ρ, a single-valued, nonnegative, real function $\rho(x, y)$ defined for all $x, y \in X$ which has the following three properties:

1. $\rho(x, y) = 0$ *iff* $x = y$;

2. $\rho(x, y) = \rho(y, x)$;

3. Triangle inequality: $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$

Examples

1. The set of all real numbers with distance

$$\rho(x, y) = |x - y|$$

is the metric space $\mathbb{R}^1$.

2. The set of all ordered n-tuples

$$x = (x_1, ..., x_n)$$

of real numbers with distance

$$\rho(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

is the metric space $\mathbb{R}^n$.

3. The set of all functions satisfying the criteria

$$\int f^2(x)dx < \infty$$

with distance

$$\rho(f_1(x), f_2(x)) = \sqrt{\int (f_1(x) - f_2(x))^2 dx}$$

is the metric space $L_2(\mathbb{R})$.

4. The set of all probability densities with Kullback-Leibler *divergence*

$$\rho(p_1(x), p_2(x)) = \int \ln \frac{p_1(x)}{p_2(x)} p_1(x)dx$$

is not a metric space. The divergence is not symmetric

$$\rho(p_1(x), p_2(x)) \neq \rho(p_2(x), p_1(x)).$$

Convergence

An *open/closed sphere* in a metric space S is the set of points $x \in \mathbb{R}$ for which

$$\rho(x_0, x) < r \quad \text{o}pen$$
$$\rho(x_0, x) \le r \quad \text{c}losed.$$

An open sphere of radius ϵ with center x_0 will be called an ϵ-*neighborhood* of x_0, denoted $O_\epsilon(x_0)$.

A sequence of points $\{x_n\} = x_1, x_2, ..., x_n, ...$ in a metric space S **converges** to a point $x \in S$ if every neighborhood $O_\epsilon(x)$ of x contains all points x_n starting from a certain integer. Given any $\epsilon > 0$ there is an integer N_ϵ such that $O_\epsilon(x)$ contains all points x_n with $n > N_\epsilon$. $\{x_n\}$ converges to x *iff*

$$\lim_{n\to\infty} \rho(x, x_n) = 0.$$

Measure

Throughout the course we will see integrals of the form

$$\int V(f(x), y) d\nu(x) \rightarrow \int V(f(x), y)\, p(x) dx$$

$\nu(x)$ is the measure.

The concept of the measure $\nu(E)$ of a set E is a natural extension of the concept

1) The length $l(\Delta)$ of a line segment Δ
2) The volume $V(G)$ of a space G
3) The integral of a nonnegative function of a region in space.

Lebesgue measure

Let f be a ν-measurable function (it has finite measure) taking no more than countably many distinct values

$$y_1, y_2, \ldots, y_n, \ldots$$

Then by the Lebesgue integral of f over the set A denoted

$$\int_A f(x)\, d\nu,$$

we mean the quantity

$$\sum_n y_n \nu(A_n)$$

where

$$A_n = \{x : x \in A, f(x) = y_n\},$$

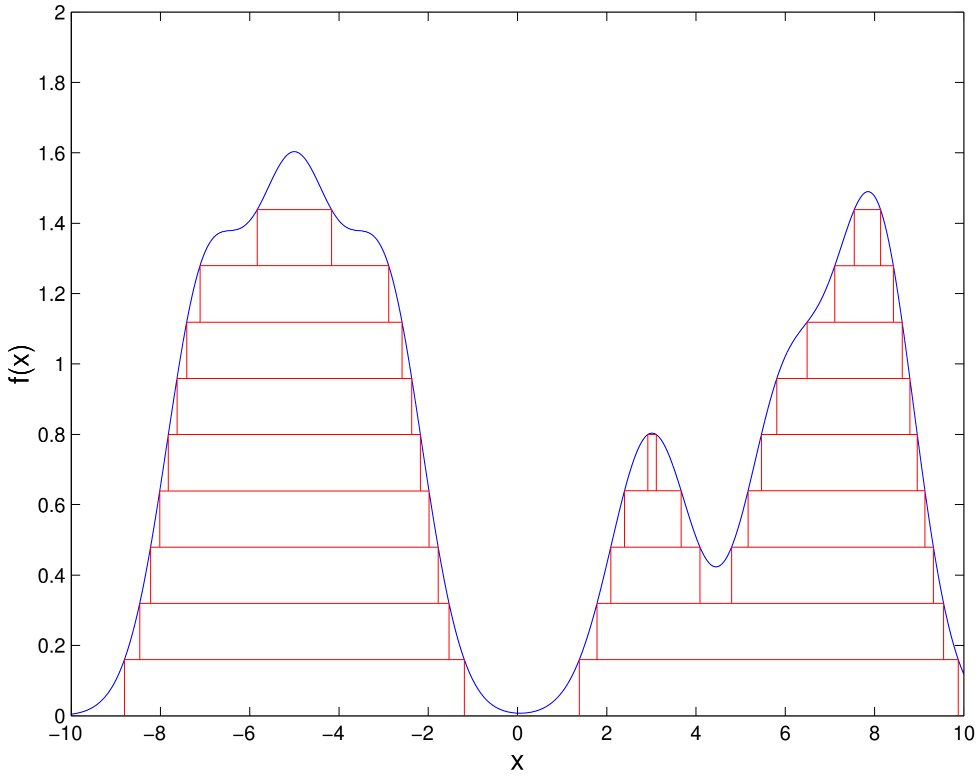
provided the series is absolutely convergent. The measure ν is the Lebesgue measure.

Lebesgue integral

We can compute the integral

$$\int f(x)dx$$

by adding up the area under the red rectangles.



Riemann integral

The more tradition form of the integral is the Riemann integral. The intuition is that of limit of an infinite sum of infinitesimally small rectangles,

$$\int_A f(x)dx = \sum_n f(x_n)\Delta x.$$

Integrals in the Riemann sense require continuous or piecewise continuous functions, the Lebesgue from shown previously relaxes this. Thus, the integral

$$\int_0^1 f(x)dx$$

with $f : [0,1] \to \mathbb{R}$ defined as

$$f = \begin{cases} 1 & \text{if } t \text{ is rational} \\ \\ 0 & \text{otherwise} \end{cases}$$

does not exist in the Riemann sense.

Lebesgue-Stieltjes integral

Let F be a nondecreasing function defined on a closed interval $[a, b]$ and suppose F is continuous from the left at every point $[a, b)$. F is called the generating function of the Lebesgue-Stieltjes measure ν_F.

The Lebesgue-Stieltjes integral of a function f is denoted by

$$\int_a^b f(x)\, dF(x)$$

which is the Lebesgue integral

$$\int_{[a,b]} f(x)\, d\nu_F.$$

An example of $d\nu_F$ is a probability density $p(x)dx$. Then ν_F would correspond to the cumulative distribution function.

Dense

Let A and B be subspaces of a metric space $\mathbb{R}$. A is said to be **dense** in B if $\bar{A} \subset B$. $\bar{A}$ is the closure of the subset A. In particular A is said to be *everywhere dense* in $\mathbb{R}$ if $\bar{A} = R$.

A point $x \in \mathbb{R}$ is called a *contact point* of a set $A \in \mathbb{R}$ if every neighborhood of x contains at least on point of A. The set of all contact points of a set A denoted by $\bar{A}$ is called the *closure* of A.

Examples

1. The set of all rational points is dense in the real line.

2. The set of all polynomials with rational coefficients is dense in $C[a, b]$.

3. Let K be a positive definite Radial Basis Function then the functions
$$f(x) = \sum_{i=1}^{n} c_i K(x - x_i)$$
is dense in L_2.

Note: A hypothesis space that is dense in L_2 is a desired property of any approximation scheme.

Separable

A metric space is said to be **separable** if it has a countable everywhere dense subset.

Examples:

1. The spaces $\mathbb{R}^1$, $\mathbb{R}^n$, $L_2[a,b]$, and $C[a,b]$ are all separable.

2. The set of real numbers is separable since the set of rational numbers is a countable subset of the reals and the set of rationals is is everywhere dense.

Completeness

A sequence of functions f_n is *fundamental* if $\forall \epsilon > 0 \; \exists N_\epsilon$ such that

$$\forall n \text{ and } m > N_\epsilon, \quad \rho(f_n, f_m) < \epsilon.$$

A metric space is **complete** if all fundamental sequences converge to a point in the space.

C, L^1, and L^2 are complete. That C_2 is not complete, instead, can be seen through a counterexample.

Incompleteness of C_2

Consider the sequence of functions $(n = 1, 2, ...)$

$$\phi_n(t) = \begin{cases} -1 & \text{if } -1 \le t < -1/n \\ nt & \text{if } -1/n \le t < 1/n \\ 1 & \text{if } 1/n \le t \le 1 \end{cases}$$

and assume that ϕ_n converges to a continuous function ϕ in the metric of C_2. Let

$$f(t) = \begin{cases} -1 & \text{if } -1 \le t < 0 \\ 1 & \text{if } 0 \le t \le 1 \end{cases}$$

Incompleteness of C_2 (cont.)

Clearly,

$$\left(\int(f(t)-\phi(t))^2dt\right)^{1/2} \le \left(\int(f(t)-\phi_n(t))^2dt\right)^{1/2} + \left(\int(\phi_n(t)-\phi(t))^2dt\right)^{1/2}.$$

Now the l.h.s. term is strictly positive, because $f(t)$ is not continuous, while for $n \to \infty$ we have

$$\int(f(t)-\phi_n(t))^2dt \to 0.$$

Therefore, contrary to what assumed, ϕ_n cannot converge to ϕ in the metric of C_2.

Completion of a metric space

Given a metric space $\mathbb{R}$ with closure $\bar{\mathbb{R}}$, a complete metric space $\mathbb{R}^*$ is called a **completion** of $\mathbb{R}$ if $\mathbb{R} \subset \mathbb{R}^*$ and $\bar{\mathbb{R}} = \mathbb{R}^*$.

Examples

1. The space of real numbers is the completion of the space of rational numbers.

2. Let K be a positive definite Radial Basis Function then L_2 is the completion the space of functions

$$f(x) = \sum_{i=1}^{n} c_i K(x - x_i).$$

Compact spaces

A metric space is **compact** *iff* it is *totally bounded* and *complete*.

Let $\mathbb{R}$ be a metric space and ϵ any positive number. Then a set $A \subset \mathbb{R}$ is said to be an ϵ-*net* for a set $M \subset \mathbb{R}$ if for every $x \in M$, there is at least one point $a \in A$ such that $\rho(x, a) < \epsilon$.

Given a metric space $\mathbb{R}$ and a subset $M \subset \mathbb{R}$ suppose M has a finite ϵ-net for every $\epsilon > 0$. Then M is said to be *totally bounded*.

A compact space has a finite ϵ-net for all $\epsilon > 0$.

Examples

1. In Euclidean n-space, $\mathbb{R}^n$, total boundedness is equivalent to boundedness. If $M \subset \mathbb{R}$ is bounded then M is contained in some hypercube Q. We can partition this hypercube into smaller hypercubes with sides of length ϵ. The vertices of the little cubes from a finite $\sqrt{n}\epsilon/2$-net of Q.

2. This is not true for infinite-dimensional spaces. The unit sphere Σ in l_2 with constraint

$$\sum_{n=1}^{\infty} x_n^2 = 1,$$

is bounded but not totally bounded. Consider the points

$$e_1 = (1, 0, 0, ...),\ e_2 = (0, 1, 0, 0, ...), ...,$$

where the n-th coordinate of e_n is one and all others are zero. These points lie on Σ but the distance between any two is $\sqrt{2}$. So Σ cannot have a finite ϵ-net with $\epsilon < \sqrt{2}/2$.

3. Infinite-dimensional spaces maybe totally bounded. Let Π be the set of points $x = (x_1, ..., x_n, ..)$ in l_2 satisfying the inequalities

$$|x_1| < 1, \ |x_2| < \frac{1}{2}, ..., \ |x_n| < \frac{1}{2^{n-1}}, ...$$

The set Π called the *Hilbert cube* is an example of an infinite-dimensional totally bounded set. Given any $\epsilon > 0$, choose n such that

$$\frac{1}{2^{n+1}} < \frac{\epsilon}{2},$$

and with each point

$$x = (x_1, ..., x_n, ..)$$

is Π associate the point

$$x^* = (x_1, ..., x_n, 0, 0, ...). \qquad (1)$$

Then

$$\rho(x, x^*) = \sqrt{\sum_{k=n+1}^{\infty} x_k^2} < \sqrt{\sum_{k=n}^{\infty} \frac{1}{4^k}} < \frac{1}{2^{n-1}} < \frac{\epsilon}{2}.$$

The set Π^* of all points in Π that satisfy (1) is totally bounded since it is a bounded set in n-space.

4. The RKHS induced by a kernel K with an infinite number of positive eigenvalues that decay exponentially is compact. In this case, our vector $x = (x_1, ..., x_n, ..)$ can

be written in terms of its basis functions, the eigenvectors of K. Now for the RKHS norm to be bounded

$$|x_1| < \mu_1,\ |x_2| < \mu_2, ...,\ |x_n| < \mu_n, ...$$

and we know that $\mu_n = O(n^{-\alpha})$. So we have the case analogous to the Hilbert cube and we can introduce a point

$$x^* = (x_1, ..., x_n, 0, 0, ...) \tag{2}$$

in a bounded n-space which can be made arbitrarily close to x.

Compactness and continuity

A family Φ of functions ϕ defined on a closed interval $[a, b]$ is said to be *uniformly bounded* if for $K > 0$

$$|\phi(x)| < K$$

for all $x \in [a, b]$ and all $\phi \in \Phi$.

A family Φ of functions ϕ is *equicontinuous* of for any given $\epsilon > 0$ there exists $\delta > 0$ such that $|x - y| < \delta$ implies

$$|\phi(x) - \phi(y)| < \epsilon$$

for all $x, y \in [a, b]$ and all $\phi \in \Phi$.

Arzela's theorem: A necessary and sufficient condition for a family Φ of continuous functions defined on a closed interval $[a, b]$ to be (relatively) compact in $C[a, b]$ is that Φ is uniformly bounded and equicontinuous.

Linear space

A set L of elements $x, y, z, ...$ is a **linear space** if the following three axioms are satisfied:

1. Any two elements $x, y \in L$ uniquely determine a third element in $x + y \in L$ called the sum of x and y such that
   (a) $x + y = y + x$ (commutativity)
   (b) $(x + y) + z = x + (y + z)$ (associativity)
   (c) An element $0 \in L$ exists for which $x + 0 = x$ for all $x \in L$
   (d) For every $x \in L$ there exists an element $-x \in L$ with the property $x + (-x) = 0$

2. Any number α and any element $x \in L$ uniquely determine an element $\alpha x \in L$ called the product such that
   (a) $\alpha(\beta x) = \beta(\alpha x)$
   (b) $1x = x$

3. Addition and multiplication follow two distributive laws
   (a)$(\alpha + \beta)x = \alpha x + \beta x$
   (b)$\alpha(x + y) = \alpha x + \alpha y$

Linear functional

A functional, $\mathcal{F}$, is a function that maps another function to a real-value

$$\mathcal{F} : f \to \mathbb{R}.$$

A linear functional defined on a linear space L, satisfies the following two properties

1. Additive: $\mathcal{F}(f+g) = \mathcal{F}(f) + \mathcal{F}(g)$ for all $f, g \in L$

2. Homogeneous: $\mathcal{F}(\alpha f) = \alpha \mathcal{F}(f)$

Examples

1. Let $\mathbb{R}^n$ be a real n-space with elements $x = (x_1, ..., x_n)$, and $a = (a_1, ..., a_n)$ be a fixed element in $\mathbb{R}^n$. Then

$$\mathcal{F}(x) = \sum_{i=1}^{n} a_i x_i$$

is a linear functional

2. The integral

$$\mathcal{F}[f(x)] = \int_a^b f(x)p(x)dx$$

is a linear functional

3. Evaluation functional: another linear functional is the

Dirac delta function

$$\delta_t[f(\cdot)] = f(t).$$

Which can be written

$$\delta_t[f(\cdot)] = \int_a^b f(x)\delta(x-t)dx.$$

4. Evaluation functional: a positive definite kernel in a RKHS

$$\mathcal{F}_t[f(\cdot)] = (K_t, f) = f(t).$$

This is simply the reproducing property of the RKHS.

Normed space

A **normed** space is a linear (vector) space N in which a norm is defined. A nonnegative function $\|\cdot\|$ is a norm *iff* $\forall f, g \in N$ and $\alpha \in \mathbb{R}$

1. $\|f\| \geq 0$ and $\|f\| = 0$ *iff* $f = 0$;

2. $\|f + g\| \leq \|f\| + \|g\|$;

3. $\|\alpha f\| = |\alpha| \; \|f\|$.

Note, if all conditions are satisfied except $\|f\| = 0$ *iff* $f = 0$ then the space has a seminorm instead of a norm.

Measuring distances in a normed space

In a normed space N, the distance ρ between f and g, or a *metric*, can be defined as

$$\rho(f, g) = \|g - f\|.$$

Note that $\forall f, g, h \in N$

1. $\rho(f, g) = 0$ *iff* $f = g$.

2. $\rho(f, g) = \rho(g, f)$.

3. $\rho(f, h) \leq \rho(f, g) + \rho(g, h)$.

Example: continuous functions

A norm in $C[a,b]$ can be established by defining

$$\|f\| = \max_{a \le t \le b} |f(t)|.$$

The distance between two functions is then measured as

$$\rho(f,g) = \max_{a \le t \le b} |g(t) - f(t)|.$$

With this metric, $C[a,b]$ is denoted as C.

Examples (cont.)

A norm in $L_1[a, b]$ can be established by defining

$$\|f\| = \int_a^b |f(t)| dt.$$

The distance between two functions is then measured as

$$\rho(f, g) = \int_a^b |g(t) - f(t)| dt.$$

With this metric, $L_1[a, b]$ is denoted as L_1.

Examples (cont.)

A norm in $C_2[a,b]$ and $L_2[a,b]$ can be established by defining

$$\|f\| = \left(\int_a^b f^2(t)dt\right)^{1/2}.$$

The distance between two functions now becomes

$$\rho(f,g) = \left(\int_a^b (g(t) - f(t))^2 dt\right)^{1/2}.$$

With this metric, $C_2[a,b]$ and $L_2[a,b]$ are denoted as C_2 and L_2 respectively.

Convergence revisited

A sequence of functions f_n converge to a function f **almost everywhere** iff

$$\lim_{n\to+\infty} f_n(x) = f(x)$$

A sequence of functions f_n converge to a function f **in measure** iff $\forall\epsilon > 0$

$$\lim_{n\to+\infty} \mu\{x : \ |f_n(x) - f(x)| \geq \epsilon\} = 0.$$

A sequence of functions f_n converge to a function f **uniformly** iff

$$\lim_{n\to+\infty} \sup_x (f_n(x) - f(x)) = 0$$

Relationship between different types of convergence

In the case of bounded intervals: *uniform* convergence (C) implies

- convergence *in the quadratic mean* (L_2) which implies convergence *in the mean* (L_1) which implies convergence *in measure*;

- *almost everywhere* convergence which implies convergence *in measure*.

Relationship between different types of convergence

That uniform convergence implies all other type of convergence is clear.

Consider L_2 over a bounded interval of width A. Keeping in mind that the function $g = 1$ belongs to L_2 and that $\|g\|_{L_2} = A$, convergence in the quadratic mean implies convergence in the mean because for every function $f \in L_2$ we have

$$\|f\|_{L_1} = \int_A |f| dx = \int_A |f| \cdot 1 dx \leq \|f\|_{L_2} \ \|1\|_{L_2} = A\|f\|_{L_2}$$

and hence that $f \in L_1$.

Any convergence implies convergence in measure

Convergence in measure is obtained by convergence in the mean through Chebyshev's inequality:
For any real random variable X and $t > 0$,
$P(|X| \geq t) \leq E[X^2/t^2]$.

The proof that almost everywhere convergence implies convergence in measure is somewhat more complicated.

Almost everywhere convergence does not imply convergence in the (quadratic) mean

Over the interval $[0, 1]$ let f_n be

$$f_n = \begin{cases} n & x \in (0, 1/n] \\ 0 & \text{otherwise} \end{cases}$$

Clearly $f_n \to 0$ for all $x \in [0, 1]$. Note that each f_n is not a continuous function and that the convergence is not uniform (the closer the x to 0, the larger n must be for $f_n(x) = 0$). However,

$$\int_0^1 |f_n(x)| dx = 1 \text{ for all } n,$$

in both the Riemann or the Lebesgue sense.

Convergence in the quadratic mean does not imply convergence at all!

Over the interval $(0, 1]$, for every $n = 1, 2, ...,$ and $i = 1, ..., n$ let

$$f_i^n = \begin{cases} 1 & \frac{i-1}{n} < x \leq \frac{i}{n} \\ 0 & \text{otherwise} \end{cases}$$

Clearly the sequence

$$f_1^1, f_1^2, f_2^2, ..., f_1^n, f_2^n, ... f_{n-1}^n, f_n^n, f_1^{n+1}, ...,$$

converges to 0 both in measure and in the quadratic mean. However, the same sequence does not converge for any x!

Convergence in probability and almost surely

Any event with probability 1 is said to happen **almost surely**. A sequence of real random variables Y_n converges almost surely to a random variable Y iff $P(Y_n \to Y) = 1$.

A sequence Y_n converges in **probability** to Y iff for every $\epsilon > 0$, $\lim_{n\to\infty} P(|Y_n - Y| > \epsilon) = 0$.

Convergence almost surely implies convergence in probability.

A sequence $X_1, ...X_n$ satisfies the *strong law of large numbers* if for some constant c, $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges to c almost surely. The sequence satisfies the *weak law of large numbers* iff for some constant c, $\frac{1}{n}\sum_{i=1}^{n} X_i$ converges to c in probability.

Euclidean space

A **Euclidean** space is a linear (vector) space E in which a dot product is defined. A real valued function $(\cdot, \cdot)$ is a dot product *iff* $\forall f, g, h \in E$ and $\alpha \in \mathbb{R}$

1. $(f, g) = (g, f)$;

2. $(f + g, h) = (f, h^*) + (g, h)$ and $(\alpha f, g) = \alpha(f, g)$;

3. $(f, f) \geq 0$ and $(f, f) = 0$ *iff* $f = 0$.

A Euclidean space becomes a *normed linear space* when equipped with the norm

$$\|f\| = \sqrt{(f, f)}.$$

Orthogonal systems and bases

A set of nonzero vectors $\{x_\alpha\}$ in a Euclidean space E is said to be an *orthogonal system* if

$$(x_\alpha, x_\beta) = 0 \quad \text{f}or \ \ \alpha \neq \beta$$

and an *orthonormal system* if

$$\begin{aligned} (x_\alpha, x_\beta) &= 0 \quad \text{f}or \ \ \alpha \neq \beta \\ (x_\alpha, x_\beta) &= 1 \quad \text{f}or \ \ \alpha = \beta. \end{aligned}$$

An orthogonal system $\{x_\alpha\}$ is called an **orthogonal basis** if it is complete (the smallest closed subspace containing $\{x_\alpha\}$ is the whole space E). A complete orthonormal system is called an **orthonormal basis**.

Examples

1. $\mathbb{R}^n$ is a real n-space, the set of n-tuples $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n)$. If we define the dot product as
$$(x, y) = \sum_{i=1}^{n} x_i y_i$$
we get Euclidean n-space. The corresponding norms and distances in $\mathbb{R}^n$ are
$$\begin{aligned} \|x\| &= \sqrt{\sum_{i=1}^{n} x_i^2} \\ \rho(x, y) = \|x - y\| &= \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \end{aligned}$$

The vectors

$$\begin{array}{lcr} e_1 & = & (1,0,0,....,0) \\ e_2 & = & (0,1,0,....,0) \\ \multicolumn{3}{c}{\dotfill} \\ e_n & = & (0,0,0,....,1) \end{array}$$

form an orthonormal basis in $\mathbb{R}^n$.

2. The space l_2 with elements $x = (x_1, x_2, ..., x_n, ...)$, $y = (y_1, y_2, ..., y_n, ...)$, ..., where

$$\sum_{i=1}^{\infty} x_i^2 < \infty, \ \ \sum_{i=1}^{\infty} y_i^2 < \infty, \ \ ..., \ \ ...,$$

becomes an infinite-dimensional Euclidean space when equipped with the dot product

$$(x, y) = \sum_{i=1}^{\infty} x_i y_i.$$

The simplest orthonormal basis in l_2 consists of vectors

$$\begin{array}{ccc} e_1 & = & (1, 0, 0, 0, ...) \\ e_2 & = & (0, 1, 0, 0, ...) \\ e_3 & = & (0, 0, 1, 0, ...) \\ e_4 & = & (0, 0, 0, 1, ...) \\ \end{array}$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

there are an infinite number of these bases.

3. The space $C_2[a, b]$ consisting of all continuous functions on $[a, b]$ equipped with the dot product

$$(f, g) = \int_a^b f(t)g(t)dt$$

is another example of Euclidean space.

An important example of orthogonal bases in this space is the following set of functions

$$1, \ \cos\frac{2\pi nt}{b-a}, \ \sin\frac{2\pi nt}{b-a} \quad (n = 1, 2, \ldots).$$

Hilbert space

A **Hilbert space** is a Euclidean space that is *complete*, *separable*, and generally *infinite-dimensional*.

A Hilbert space is a set H of elements $f, g, ...$ for which

1. H is a Euclidean space equipped with a scalar product

2. H is complete with respect to metric $\rho(f, g) = \|f - g\|$

3. H is separable (contains a countable everywhere dense subset)

4. (generally) H is infinite-dimensional.

l_2 and L_2 are examples of Hilbert spaces.

The δ function

We now consider the functional which returns the value of $f \in C$ at the location t (an evaluation functional),

$$\Phi[f] = f(t).$$

Note that this functional is degenerate because it does not depend on the entire function f, but only on the value of f at the specific location t.

The $\delta(t)$ is not a functional but a distribution.

The δ function (cont.)

The same functional can be written as

$$\Phi[f] = f(t) = \int_{-\infty}^{\infty} f(s)\delta(s-t)ds.$$

No ordinary function exists (in L_2) that behaves like $\delta(t)$, one can think of $\delta(t)$ as a function that vanishes for $t \neq 0$ and takes infinite value at $t = 0$ in such a way that

$$\int_{-\infty}^{\infty} \delta(t)dt = 1.$$

The δ function (cont.)

The δ function can be seen as the limit of a sequence of ordinary functions. For example, if

$$r_\epsilon(t) = \frac{1}{\epsilon}\left(U(t) - U(t-\epsilon)\right)$$

is a rectangular pulse of unit area, consider the limit

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} f(s) r_\epsilon(s-t) ds.$$

By definition of r_ϵ this gives

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_t^{t+\epsilon} f(s) ds = f(t)$$

because f is continuous.

Fourier Transform

The *Fourier Transform* of a real valued function $f \in L_1$ is the complex valued function $\tilde{f}(\omega)$ defined as

$$\mathcal{F}[f(x)] = \tilde{f}(\omega) = \int_{-\infty}^{+\infty} f(x)\ e^{-j\omega x} dx.$$

The FT $\tilde{f}$ can be thought of as a representation of the information content of $f(x)$. The original function f can be obtained through the *inverse Fourier Transform* as

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(\omega)\ e^{j\omega x} d\omega.$$

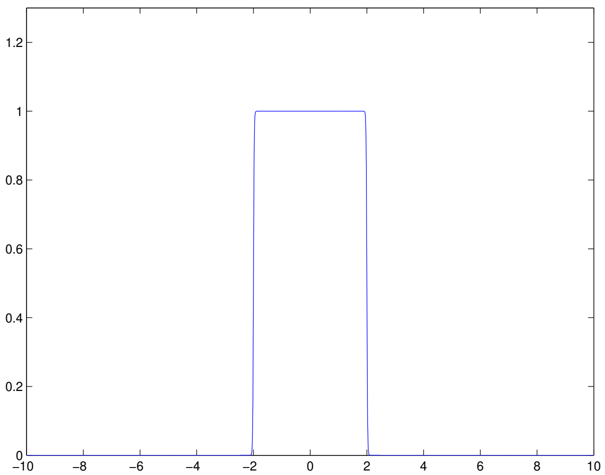
Properties

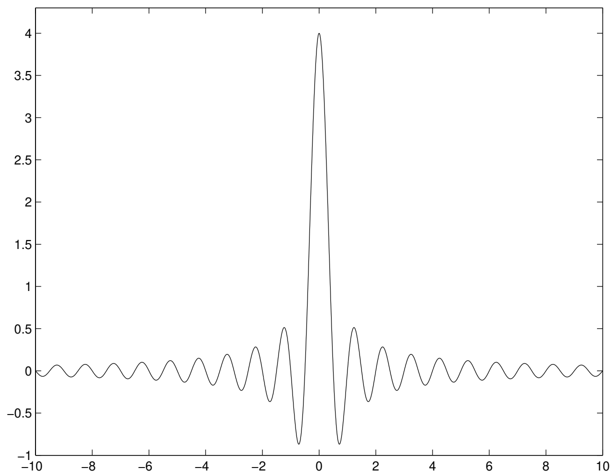
$$
\begin{aligned}
f(at) &\Leftrightarrow \frac{1}{|a|}F\left(\frac{\omega}{a}\right) \\
f^*(t) &\Leftrightarrow F^*(\omega) \\
F(t) &\Leftrightarrow 2\pi f(-\omega) \\
f(t-t_0) &\Leftrightarrow F(\omega)e^{-jt_0\omega} \\
f(t)e^{j\omega_0 t} &\Leftrightarrow F(\omega-\omega_0) \\
\frac{d^n f(t)}{dt^n} &\Leftrightarrow (j\omega)^n F(\omega) \\
(-jt)^n f(t) &\Leftrightarrow \frac{d^n F(\omega)}{d\omega^n} \\
\int_{-\infty}^{\infty} f_1(\tau)f_2(t-\tau)d\tau &\Leftrightarrow F_1(\omega)F_2(\omega) \\
\int_{-\infty}^{\infty} f^*(\tau)f(t+\tau)d\tau &\Leftrightarrow |F(\omega)|^2
\end{aligned}
$$

Properties

The box and the sinc

$$
\begin{aligned}
f(t) &= 1 \text{ if } -a \le t \le a \text{ and } 0 \text{ otherwise} \\
F(\omega) &= \frac{2\sin(a\omega)}{\omega}.
\end{aligned}
$$

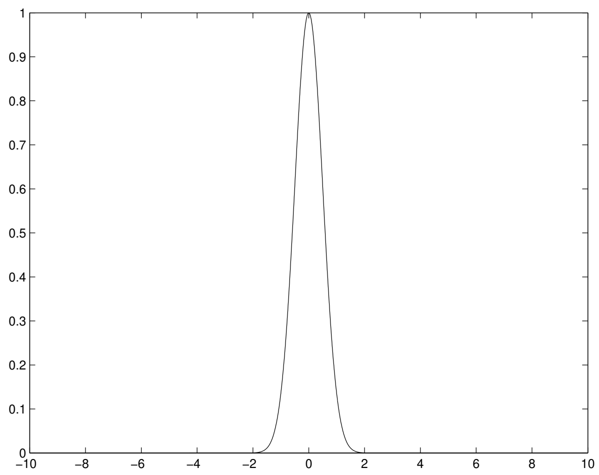


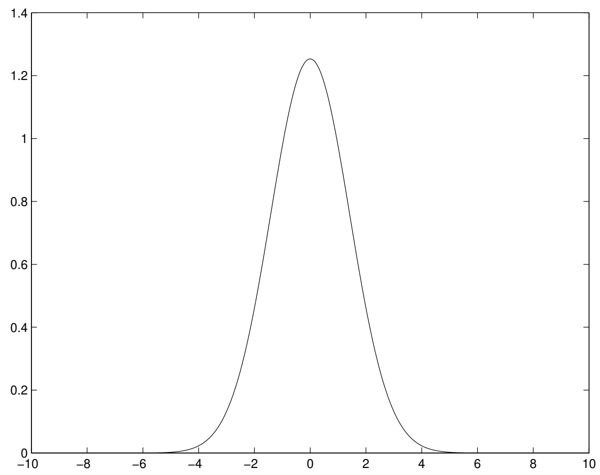


Properties

The Gaussian

$$\begin{aligned} f(t) &= e^{-at^2} \\ F(\omega) &= \sqrt{\frac{\pi}{a}} e^{-\omega^2/4a}. \end{aligned}$$

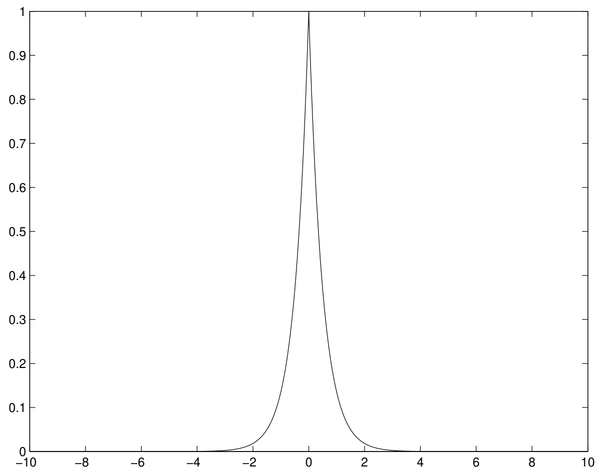


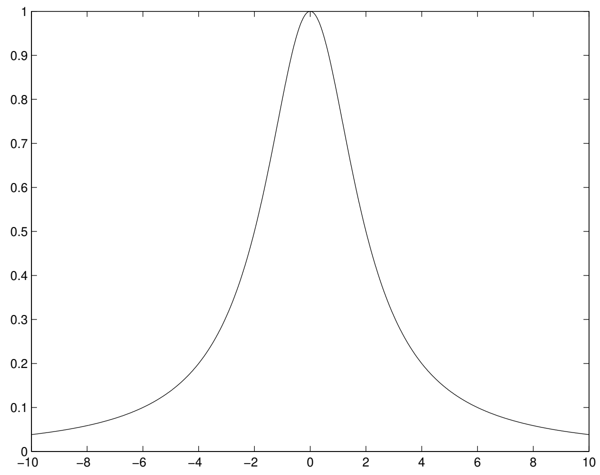


Properties

The Laplacian and Cauchy distributions

$$f(t) = e^{-a|t|}$$
$$F(\omega) = \frac{2a}{a^2 + \omega^2}.$$





Fourier Transform in the distribution sense

With due care, the Fourier Transform can be defined in the distribution sense. For example, we have

- $\delta(x) \iff 1$
- $\cos(\omega_0 x) \iff \pi(\delta(\omega - \omega_0) + \delta(\omega + \omega_0))$
- $\sin(\omega_0 x) \iff j\pi(\delta(\omega + \omega_0) - \delta(\omega - \omega_0))$
- $U(x) \iff \pi\delta(\omega) - j/\omega$
- $|x| \iff -2/\omega^2$

Parseval's formula

If f is also square integrable, the *Fourier Transform* leaves the norm of f unchanged. Parseval's formula states that

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\tilde{f}(\omega)|^2 d\omega.$$

Fourier Transforms of functions and distributions

The following are Fourier transforms of some functions and distributions

- $f(x) = \delta(x) \iff \tilde{f}(\omega) = 1$
- $f(x) = \cos(\omega_0 x) \iff \tilde{f}(\omega) = \pi(\delta(\omega - \omega_0) + \delta(\omega + \omega_0))$
- $f(x) = \sin(\omega_0 x) \iff \tilde{f}(\omega) = i\pi(\delta(\omega + \omega_0) - \delta(\omega - \omega_0))$
- $f(x) = U(x) \iff \tilde{f}(\omega) = \pi\delta(\omega) - i/\omega$
- $f(x) = |x| \iff \tilde{f}(\omega) = -2/\omega^2.$

Functional differentiation

In analogy with standard calculus, the minimum of a functional can be obtained by setting equal to zero the *derivative* of the functional. If the functional depends on the derivatives of the unknown function, a further step is required (as the unknown function has to be found as the solution of a differential equation).

Functional differentiation

The derivative of a functional $\Phi[f]$ is defined

$$\frac{D\Phi[f]}{Df(s)} = \lim_{h \to 0} \frac{\Phi[f(t) + h\delta(t-s)] - \Phi[f(t)]}{h}.$$

Note that the derivative depends on the location s. For example, if $\Phi[f] = \int_{-\infty}^{+\infty} f(t)g(t)dt$

$$\frac{D\Phi[f]}{Df(s)} = \int_{-\infty}^{+\infty} g(t)\delta(t-s)dt = g(s).$$

Intuition

Let $f : [a, b] \rightarrow \mathbb{R}$, $a = x_1$ and $b = x_N$. The intuition behind this definition is that the functional $\Phi[f]$ can be thought of as the limit for $N \rightarrow \infty$ of the function of N variables

$$\Phi_N = \Phi_N(f_1, f_2, ..., f_N)$$

with $f_1 = f(x_1)$, $f_2 = f(x_2)$, ... $f_N = f(x_N)$.

For $N \rightarrow \infty$, Φ depends on the entire function f. The dependence on the location brought in by the δ function corresponds to the partial derivative with respect to the variable f_k.

Functional differentiation (cont.)

If $\Phi[f] = f(t)$, the derivative is simply

$$\frac{D\Phi[f]}{Df(s)} = \frac{Df(t)}{Df(s)} = \delta(t - s).$$

Similarly to ordinary calculus, the minimum of a functional $\Phi[f]$ is obtained as the function solution to the equation

$$\frac{D\Phi[f]}{Df(s)} = 0.$$

Random variables

We are given a random variable $\xi \sim F$. To define a random variable you need three things:

1) a set to draw the values from, we'll call this Ω
2) a σ-algebra of subsets of Ω, we'll call this $\mathcal{B}$
3) a probability measure F on $\mathcal{B}$ with $F(\Omega) = 1$

So $(\Omega, \mathcal{B}, F)$ is a probability space and a random variable is a masurable function $X : \Omega \to \mathbb{R}$.

Expectations

Given a random variable $\xi \sim F$ the expectation is

$$\mathbb{E}\xi \equiv \int \xi dF.$$

Similarly the variance of the random variable $\sigma^2(\xi)$ is

$$\mathsf{var}(\xi) \equiv \mathbb{E}(\xi - \mathbb{E}\xi)^2.$$

Law of large numbers

The law of large numbers tells us:

$$\lim_{\ell\to\infty} \frac{1}{\ell} \sum_{i=1}^{\ell} I_{[f(x_i)\neq y_i]} \to \mathbb{E}_{x,y} I_{[f(x)\neq y]}.$$

If $\ell\sigma \to \infty$ the Central Limit Theorem states:

$$\frac{\sqrt{\ell}(\frac{1}{\ell}\sum I - \mathbb{E}I)}{\sqrt{\text{var}I}} \to N(0,1),$$

which implies

$$\left|\frac{1}{\ell}\sum I - \mathbb{E}I\right| \sim \frac{k}{\sqrt{\ell}}.$$

If $\ell\sigma \to c$ the Central Limit Theorem implies

$$\left|\frac{1}{\ell}\sum I - \mathbb{E}I\right| \sim \frac{k}{\ell}.$$