Introduction to Fractal Geometry

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 500

Subsection 1

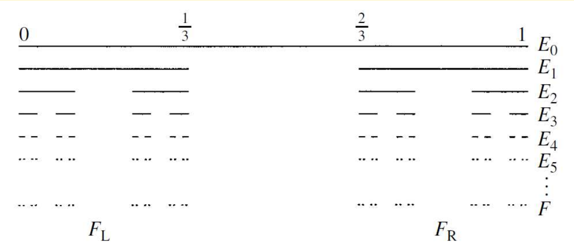
Introduction

From Smooth Sets to Fractal Geometry

- In the past, mathematics has been concerned largely with sets and functions to which the methods of classical calculus can be applied.
- Sets or functions that are not sufficiently smooth or regular have tended to be ignored as "pathological" and not worthy of study.
- They were regarded mostly as individual curiosities and rarely as a class to which a general theory might be applicable.
- More recently, it has been realized that a great deal can be said, and is worth saying, about the mathematics of non-smooth objects.
- Often, irregular sets provide a much better representation of natural phenomena than do the figures of classical geometry.
- Fractal geometry is a general framework for the study of such sets.

Construction of the Middle Third Cantor Set

- Let E_0 be the interval $[0, 1]$.
- Let E_1 be the set obtained by deleting the middle third of E_0, so that E_1 consists of the two intervals $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$.



- Deleting the middle thirds of these intervals gives E_2.
  Thus E_2 comprises the four intervals $[0, \frac{1}{9}]$, $[\frac{2}{9}, \frac{1}{3}]$, $[\frac{2}{3}, \frac{7}{9}]$, $[\frac{8}{9}, 1]$.
- We continue in this way, with E_k obtained by deleting the middle third of each interval in E_{k-1}.
  Thus, E_k consists of 2^k intervals each of length 3^{-k}.

The Middle Third Cantor Set

- The **middle third Cantor set** F consists of the numbers that are in E_k for all k.
- Mathematically,
$$F = \bigcap_{k=0}^{\infty} E_k.$$
- The Cantor set F may be thought of as the limit of the sequence of sets E_k as k tends to infinity.
- At first glance it might appear that we have removed so much of the interval $[0, 1]$ during the construction of F, that nothing remains.
- In fact, F is an infinite (and indeed uncountable) set, which contains infinitely many numbers in every neighborhood of each of its points.

Structure of the Middle Third Cantor Set

- The set F consists precisely of those numbers in $[0,1]$ whose base 3 expansion does not contain the digit 1.
- That is F consists of all numbers

$$a_1 3^{-1} + a_2 3^{-2} + a_3 3^{-3} + \cdots ,$$

  with $a_i = 0$ or 2 for each i.
- To see this, note that:
  - To get E_1 from E_0 we remove those numbers with $a_1 = 1$;
  - To get E_2 from E_1 we remove those numbers with $a_2 = 1$;

    ⋮

Features of the Middle Third Cantor Set

(i) F is self-similar.

- The part of F in the interval $[0, \frac{1}{3}]$ and the part of F in $[\frac{2}{3}, 1]$ are both geometrically similar to F, scaled by a factor $\frac{1}{3}$.
- The parts of F in each of the four intervals of E_2 are similar to F but scaled by a factor $\frac{1}{9}$.

$\vdots$

The Cantor set contains copies of itself at many different scales.

Features of the Middle Third Cantor Set (Cont'd)

(ii) The set F has a "fine structure".

That is, it contains detail at arbitrarily small scales.

The more we enlarge the picture of the Cantor set, the more gaps become apparent to the eye.

(iii) Although F has an intricate detailed structure, the actual definition of F is very straightforward.

(iv) F is obtained by a recursive procedure.

The construction consisted of repeatedly removing the middle thirds of intervals.

Successive steps give increasingly good approximations E_k to F.

Features of the Middle Third Cantor Set (Cont'd)

(v) The geometry of F is not easily described in classical terms.

It is not:

- The locus of the points that satisfy some simple geometric condition;
- The set of solutions of any simple equation.

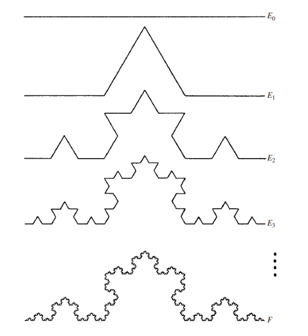
(vi) It is awkward to describe the local geometry of F.

Near each of its points, there are a large number of other points, separated by gaps of varying lengths.

(vii) Although F is in some ways quite a large set (it is uncountably infinite), its size is not quantified by the usual measures such as length.

By any reasonable definition F has length zero.

Construction of the Koch Curve



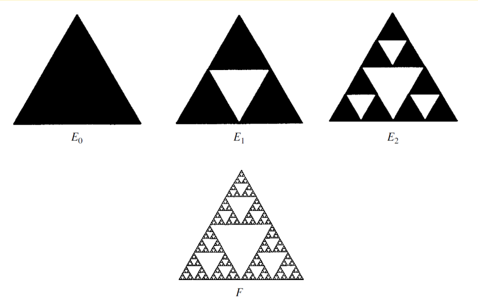
- Let E_0 be a line segment of unit length.
- The set E_1 consists of the four segments obtained by removing the middle third of E_0 and replacing it by the other two sides of the equilateral triangle based on the removed segment.
- We construct E_2 by applying the same procedure to each of the segments in E_1, and so on.
- Thus E_k comes from replacing the middle third of each straight line segment of E_{k-1} by the other two sides of an equilateral triangle.
- When k is large, the curves E_{k-1} and E_k differ only in fine detail.
- As k tends to infinity, the sequence of polygonal curves E_k approaches a limiting curve F, called the **von Koch curve**.

Features of the von Kock Curve

- The von Koch curve has similar features to those of the middle third Cantor set:
  - It is made up of four "quarters" each similar to the whole, but scaled by a factor $\frac{1}{3}$.
  - The fine structure is reflected in the irregularities at all scales.
  - This intricate structure stems from a basically simple construction.
  - Whilst it is reasonable to call F a curve, it is much too irregular to have tangents in the classical sense.
  - A simple calculation shows that E_k is of length $(\frac{4}{3})^k$.
    Letting k tend to infinity implies that F has infinite length.
    On the other hand, F occupies zero area in the plane.
    Therefore, neither length nor area provides a very useful description of the size of F.

The Sierpiński Triangle

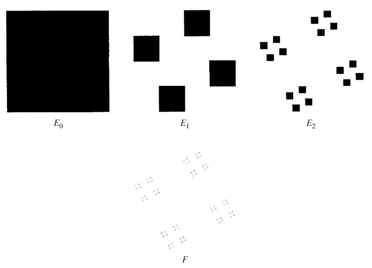
- The **Sierpiński triangle** or **gasket** is obtained by repeatedly removing (inverted) equilateral triangles from an initial equilateral triangle of unit side length.



- For many purposes, it is better to think of this procedure as repeatedly replacing an equilateral triangle by three triangles of half the height.

The Cantor Dust

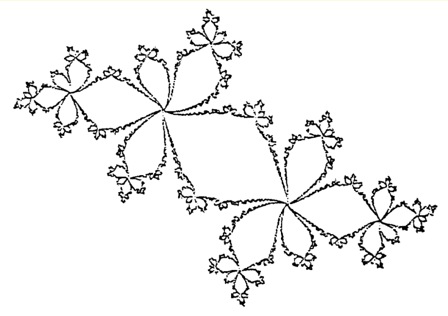
- A plane analogue of the Cantor set, a "**Cantor dust**" is obtained by dividing at each stage each remaining square into 16 smaller squares of which four are kept and the rest discarded.



- Of course, other arrangements or numbers of squares could be used to get different sets.
- Such examples have properties similar to those of the Cantor set and the von Koch curve.

Julia Sets

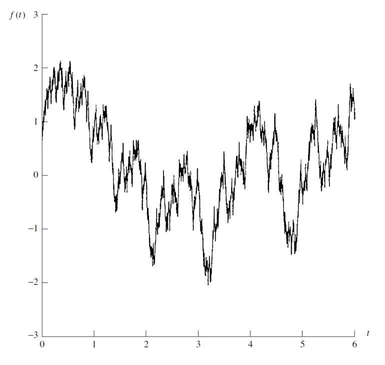
- The highly intricate structure of the **Julia set** stems from the single quadratic function $f(z) = z^2 + c$, for a suitable constant c.



- This set is not strictly self-similar in the sense of the Cantor set.
- It is "quasi-self-similar" in that arbitrarily small portions of the set can be magnified and then distorted smoothly to coincide with a large part of the set.

Graph of Function Defined by an Infinite Sum

- Consider the graph of the function $f(t) = \sum_{k=0}^{\infty} (\frac{3}{2})^{-k/2} \sin ((\frac{3}{2})^k t)$.



- It has a fine structure due to the infinite summation.
- It is not a smooth curve to which classical calculus is applicable.

Fractals

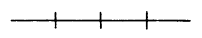
- These types of sets are commonly referred to as **fractals**.
- The word "fractal" was coined by Mandelbrot.
- It comes from the Latin **fractus**, meaning broken, to describe objects too irregular to fit into a traditional geometrical setting.
- Properties such as those listed before are characteristic of fractals.
  - Any fractal will have a fine structure, i.e., detail at all scales.
  - Many fractals have some degree of self-similarity.
    - They are made up of parts that resemble the whole in some way.
    - The resemblance may be weaker than strict geometrical similarity.

Dimension

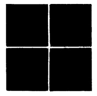
- Methods of classical geometry and calculus are unsuited to studying fractals and we need alternative techniques.
- The main tool of fractal geometry is **dimension** in its many forms.
- We are familiar with the idea that:
  - A (smooth) curve is a 1-dimensional object;
  - A surface is a 2-dimensional object.
- It is less clear that, for many purposes, we should regard:
  - The Cantor set as having dimension $\frac{\log 2}{\log 3} = 0.631\ldots$;
  - The von Koch curve as having dimension $\frac{\log 4}{\log 3} = 1.262\ldots$.
- The number $\frac{\log 4}{\log 3} = 1.262\ldots$ is, at least, consistent with the von Koch curve being:
  - "larger than 1-dimensional" (having infinite length);
  - "smaller than 2-dimensional" (having zero area).

The Similarity Dimension

- This notion of "dimension" reflects scaling and selfsimilarity.
  - The figure shows a line segment made up of four copies of itself, scaled by a factor $\frac{1}{4}$. The segment has dimension $\frac{-\log 4}{\log(1/4)} = 1$.



  - A square is made up of four copies of itself scaled by a factor $\frac{1}{2}$, i.e., with half the side length. It has dimension $\frac{-\log 4}{\log(1/2)} = 2$.



  - In the same way, the von Koch curve is made up of four copies of itself scaled by a factor $\frac{1}{3}$. It has dimension $\frac{-\log 4}{\log(1/3)} = \frac{\log 4}{\log 3}$.

  - The Cantor set may be regarded as comprising four copies of itself scaled by a factor $\frac{1}{9}$. It has dimension $\frac{-\log 4}{\log(1/9)} = \frac{\log 2}{\log 3}$.

- In general, a set made up of m copies of itself scaled by a factor r might be thought of as having dimension $\frac{-\log m}{\log r}$.
- This number is referred to as the **similarity dimension**.

Other Notions of Dimension

- Unfortunately, similarity dimension is meaningful only for a relatively small class of strictly self-similar sets.
- There are other definitions of dimension that are much more widely applicable.
- For example, *Hausdorff dimension* and the *box-counting dimensions* may be defined for any sets.
- Moreover, in these four examples, they may be shown to equal the similarity dimension.
- Roughly speaking, a dimension provides a description of how much space a set fills.
  - It measures the prominence of the irregularities of a set when viewed at very small scales;
  - It contains information about the geometrical properties of a set.

Properties Characteristic of Fractals

- When we refer to a set F as a fractal, we will typically have the following in mind:
  (i) F has a fine structure, i.e., detail on arbitrarily small scales.
  (ii) F is too irregular to be described in traditional geometrical language, both locally and globally.
  (iii) Often F has some form of self-similarity, perhaps approximate or statistical.
  (iv) Usually, the "fractal dimension" of F (defined in some way) is greater than its topological dimension.
  (v) In most cases of interest F is defined in a very simple way, perhaps recursively.

Subsection 2

Basic Set Theory

The n-Dimensional Euclidean Space

- We work in n-**dimensional Euclidean space**, $\mathbb{R}^n$, where $\mathbb{R}^1 = \mathbb{R}$ is the "real line" and $\mathbb{R}^2$ is the (Euclidean) plane.
- Points in $\mathbb{R}^n$ will be denoted by lower case letters x, y, etc., and in the coordinate form $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$.
- Addition and scalar multiplication are defined in the usual manner:

$$\begin{array}{rcl} x + y & = & (x_1 + y_1, \ldots, x_n + y_n), \\ \lambda x & = & (\lambda x_1, \ldots, \lambda x_n), \end{array}$$

  where λ is a real scalar.

The n-Dimensional Euclidean Space (Cont'd)

- We use the usual Euclidean **distance** or **metric** on $\mathbb{R}^n$.
- If x, y are points of $\mathbb{R}^n$, the distance between them is

$$|x - y| = \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{1/2}.$$

- In particular, we have, for all $x, y, z \in \mathbb{R}^n$:
  - The triangle inequality

$$|x + y| \leq |x| + |y|;$$

  - The reverse triangle inequality

$$|x - y| \geq ||x| - |y||;$$

  - The metric triangle inequality

$$|x - y| \leq |x - z| + |z - y|.$$

Set Notation

- Sets, which will generally be subsets of $\mathbb{R}^n$, are denoted by capital letters E, F, U, etc.
- $x \in E$ means that the point x belongs to the set E.
- $E \subseteq F$ means that E is a subset of the set F.
- $\{x : \text{condition}\}$ is the set of x for which "condition" is true.
- The empty set, which contains no elements, is written as $\emptyset$.
- The integers are denoted by $\mathbb{Z}$ and the rational numbers by $\mathbb{Q}$.
- We use a superscript $^+$ to denote the positive elements of a set.
  - $\mathbb{R}^+$ is the set of positive real numbers;
  - $\mathbb{Z}^+$ is the set of positive integers.
- Occasionally we refer to the complex numbers $\mathbb{C}$.
- $\mathbb{C}$ may be identified with the plane $\mathbb{R}^2$, with $x_1 + ix_2$ corresponding to the point (x_1, x_2).

Balls

- The **closed ball** of center x and radius r is defined by
$$B(x,r) = \{y : |y-x| \leq r\}.$$
- The **open ball** is
$$B^o(x,r) = \{y : |y-x| < r\}.$$
- The closed ball contains its bounding sphere, but the open ball does not.
- In $\mathbb{R}^2$ a ball is a disc.
- In $\mathbb{R}^1$ a ball is just an interval.

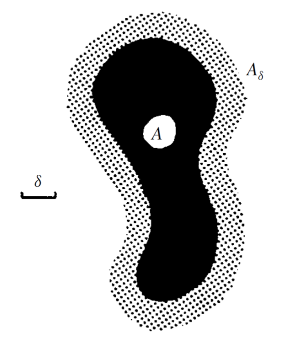
Cubes

- If $a < b$, we write:
  - $[a, b]$ for the **closed interval**
    $$\{x : a \leq x \leq b\};$$
  - (a, b) for the **open interval**
    $$\{x : a < x < b\}.$$
- Similarly $[a, b)$ denotes the **half-open interval** $\{x : a \leq x < b\}$, etc.
- The **coordinate cube** of side $2r$ and center $x = (x_1, \ldots, x_n)$ is the set
  $$\{y = (y_1, \ldots, y_n) : |y_i - x_i| \leq r, \text{ for all } i = 1, \ldots, n\}.$$
- A cube in $\mathbb{R}^2$ is just a square and in $\mathbb{R}^1$ is an interval.

δ-Neighborhoods

- The δ-**neighborhood** or δ-**parallel body**, A_δ, of a set A is the set of points within distance δ of A,

$$A_\delta = \{x : |x - y| \leq \delta, \text{ for some } y \text{ in } A\}.$$



Union Intersection, Difference, Cartesian Product

- We write $A \cup B$ for the **union** of the sets A and B, i.e. the set of points belonging to either A or B, or both.
- We write $A \cap B$ for their **intersection**, the set of points in both A and B.
- The union of an arbitrary collection of sets $\{A_\alpha\}$ is denoted

  $$\bigcup_\alpha A_\alpha.$$

  It consists of those points in at least one of the sets A_α.
- The intersection of an arbitrary collection of sets $\{A_\alpha\}$ is denoted

  $$\bigcap_\alpha A_\alpha.$$

  It consists of the set of points common to all of the A_α.

Union Intersection, Difference, Cartesian Product

- A collection of sets is **disjoint** if the intersection of any pair is the empty set.
- The **difference** $A \backslash B$ of A and B consists of the points in A but not B.
- The set $\mathbb{R}^n \backslash A$ is termed the **complement** of A.
- The (**Cartesian**) **product** of A and B is the set of all ordered pairs

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

- If $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$, then $A \times B \subseteq \mathbb{R}^{n+m}$.

Vector Sum and Scalar Multiple of Sets

- Let A and B be subsets of $\mathbb{R}^n$ and λ a real number.
- We define the **vector sum** of the sets as

$$A + B = \{x + y : x \in A \text{ and } y \in B\}.$$

- We define the **scalar multiple** by

$$\lambda A = \{\lambda x : x \in A\}.$$

Cardinalities

- An infinite set A is **countable** if its elements can be listed in the form $x_1, x_2, \ldots$ with every element of A appearing at a specific place in the list.
- Otherwise the set is **uncountable**.

  Example: The sets $\mathbb{Z}$ and $\mathbb{Q}$ are countable but $\mathbb{R}$ is uncountable.
- A countable union of countable sets is countable.

Suprema and Infima

- Let A be any non-empty set of real numbers.
- The **supremum** $\sup A$ is the least number m, such that $x \leq m$, for every x in A, or is $+\infty$ if no such number exists.
- The **infimum** $\inf A$ is the greatest number m, such that $m \leq x$, for all x in A, or is $-\infty$ if no such number exists.
- Intuitively the supremum and infimum are thought of as the maximum and minimum of the set, though it is important to realize that $\sup A$ and $\inf A$ need not be members of the set A itself.

  Example: $\sup (0,1) = 1$, but $1 \notin (0,1)$.
- We write

$$\sup_{x \in B} (\quad)$$

  for the supremum of the quantity in brackets, which may depend on x, as x ranges over the set B.

Diameter and Boundedness

- We define the **diameter** $|A|$ of a (non-empty) subset of $\mathbb{R}^n$ as the greatest distance apart of pairs of points in A,

$$|A| = \sup\{|x-y| : x, y \in A\}.$$

  Example: In $\mathbb{R}^n$:
  - A ball of radius r has diameter $2r$;
  - A cube of side length δ has diameter $\delta\sqrt{n}$.
- A set A is **bounded** if it has finite diameter.
- Equivalently, A is bounded if it is contained in some ball.

Convergence of Sequences

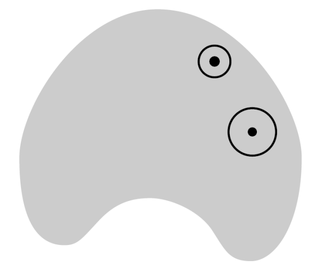
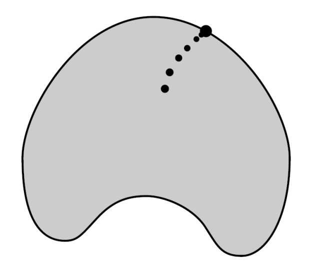
- A sequence $\{x_k\}$ in $\mathbb{R}^n$ **converges** to a point x of $\mathbb{R}^n$ as $k \to \infty$ if, given $\varepsilon > 0$, there exists a number K, such that

$$|x_k - x| < \varepsilon, \quad \text{for all } k > K.$$

- Equivalently, $\{x_k\}$ in $\mathbb{R}^n$ **converges** to a point x of $\mathbb{R}^n$ if and only if $|x_k - x|$ converges to 0.
- The number x is called the **limit** of the sequence.
- We write $x_k \to x$ or $\lim_{k\to\infty} x_k = x$.

Open and Closed Sets

- A subset A of $\mathbb{R}^n$ is **open** if, for all points x in A, there is some ball $B(x, r)$, centered at x and of positive radius, that is contained in A.

- A set is **closed** if, whenever $\{x_k\}$ is a sequence of points of A converging to a point x of $\mathbb{R}^n$, then x is in A.
- The empty set $\emptyset$ and $\mathbb{R}^n$ are regarded as both open and closed.

Properties of Open and Closed Sets and Neighborhoods

- A set is open if and only if its complement is closed.
- The union of any collection of open sets is open.
- The intersection of any finite number of open sets is open.
- The intersection of any collection of closed sets is closed.
- The union of any finite number of closed sets is closed.
- A set A is called a **neighborhood** of a point x if there is some (small) ball $B(x, r)$ centered at x and contained in A.

Closure, Interior, Boundary

- The intersection of all the closed sets containing a set A is called the **closure** of A, written $\overline{A}$.
- The closure of A is thought of as the smallest closed set containing A.
- The union of all open sets contained in A is the **interior** $\text{int}(A)$ of A.
- The interior is thought of as the largest open set contained in A.
- The **boundary** ∂A of A is given by $\partial A = \overline{A} \setminus \text{int}(A)$.
- Thus, $x \in \partial A$ if and only if the ball $B(x, r)$ intersects both A and its complement, for all $r > 0$.
- A set B is a **dense subset** of A if $B \subseteq A \subseteq \overline{B}$,
- Equivalently, B is a **dense subset** of A if and only if there are points of B arbitrarily close to each point of A.

Compactness

- A set A is **compact** if any collection of open sets which covers A (i.e., with union containing A) has a finite subcollection which also covers A.
- Technically, compactness is an extremely useful property that enables infinite sets of conditions to be reduced to finitely many.
- The Heine-Borel Theorem states that, a subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.
- So, for most of our purposes, it is enough to take the definition of a compact subset of $\mathbb{R}^n$ as one that is both closed and bounded.

Properties of Compactness

- The intersection of any collection of compact sets is compact.
- If $A_1 \supseteq A_2 \supseteq \cdots$ is a decreasing sequence of compact sets, then the intersection $\bigcap_{i=1}^{\infty} A_i$ is non-empty.
- If for compact sets A_i, $\bigcap_{i=1}^{\infty} A_i$ is contained in V for some open set V, then the finite intersection $\bigcap_{i=1}^{k} A_i$ is contained in V, for some k.

Connectedness

- A subset A of $\mathbb{R}^n$ is **connected** if there do not exist open sets U and V such that:
  - $A \subseteq U \cup V$;
  - $A \cap U$ and $A \cap V$ disjoint and non-empty.
- Intuitively, we think of a set A as connected if it consists of just one "piece".
- The largest connected subset of A containing a point x is called the **connected component** of x.
- The set A is **totally disconnected** if the connected component of each point consists of just that point.
- A sufficient condition for A to be totally disconnected is that, for every pair of points x and y in A, there exist disjoint open sets U and V, such that $x \in U$, $y \in V$ and $A \subseteq U \cup V$.

Borel Sets

- The class of **Borel sets** is the smallest collection of subsets of $\mathbb{R}^n$ with the following properties:
  (a) Every open set and every closed set is a Borel set;
  (b) The union of every finite or countable collection of Borel sets is a Borel set;
  (c) The intersection of every finite or countable collection of Borel sets is a Borel set.
- Virtually all of the subsets of $\mathbb{R}^n$ that will be of any interest to us will be Borel sets.
- Any set that can be constructed using a sequence of countable unions or intersections starting with the open sets or closed sets will certainly be Borel.

Subsection 3

Functions and Limits

Functions

- Let X and Y be any sets.
- A **mapping**, **function** or **transformation** f from X to Y is a rule or formula that associates a point $f(x)$ of Y with each point x of X.
- We write $f : X \to Y$ to denote this situation.
- X is called the **domain** of f.
- Y is called the **codomain**.
- If A is any subset of X, we write $f(A)$ for the **image** of A,

$$f(A) = \{f(x) : x \in A\}.$$

- If B is a subset of Y, we write $f^{-1}(B)$ for the inverse **image** or **pre-image** of B,

$$f^{-1}(B) = \{x \in X : f(x) \in B\}.$$

- The inverse image of a single point can contain many points.

Injections, Surjections and Bijections

- A function $f : X \to Y$ is called an **injection** or a **one-to-one function** if
$$x \neq y \quad \text{implies} \quad f(x) \neq f(y).$$
- That is, f is an injection if different elements of X are mapped to different elements of Y.
- The function is called a **surjection** or an **onto function** if, for every y in Y, there is an element x in X with $f(x) = y$.
- I.e., f is a surjection if every element of Y is the image of some point in X.
- A function that is both an injection and a surjection is called a **bijection** or **one-to-one correspondence** between X and Y.

Inverse Functions

- Suppose $f : X \to Y$ is a bijection.
- Then we may define the **inverse function** $f^{-1} : Y \to X$ by taking $f^{-1}(y)$ as the unique element x of X such that $f(x) = y$.
- In this situation, we have:
  - $f^{-1}(f(x)) = x$, for all x in X;
  - $f(f^{-1}(y)) = y$, for all y in Y.

Composition of Functions

- The **composition** of the functions $f : X \to Y$ and $g : Y \to Z$ is the function $g \circ f : X \to Z$, given by

$$(g \circ f)(x) = g(f(x)).$$

- This definition extends to the composition of any finite number of functions in the obvious way,

$$(f_n \circ f_{n-1} \circ \cdots \circ f_2 \circ f_1)(x) = f_n(f_{n-1}(\cdots(f_2(f_1(x)))\cdots)).$$

Transformations and Congruences

- Functions from $\mathbb{R}^n$ to $\mathbb{R}^n$ with a geometric significance are often referred to as **transformations** and are denoted by capital letters.
- The transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ is called a **congruence** or **isometry** if it preserves distances, i.e., if
$$|S(x) - S(y)| = |x - y|, \quad \text{for all } x, y \text{ in } \mathbb{R}^n.$$
- Congruences also preserve angles.
- Moreover, they transform sets into geometrically congruent ones.

Translations, Rotations and Reflections

- **Translations** are of the form

$$S(x) = x + a.$$

- They have the effect of shifting points parallel to the vector a.
- **Rotations** centered at a are such that

$$|S(x) - a| = |x - a|, \quad \text{for all } x.$$

- For convenience we also regard the identity transformation given by $I(x) = x$ as a rotation.
- **Reflections** map points to their mirror images in some $(n-1)$-dimensional plane.
- A congruence that may be achieved by a combination of a rotation and a translation, i.e., does not involve reflection, is called a **rigid motion** or **direct congruence**.

Similarities

- A transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ is a **similarity** of **ratio** or **scale** $c > 0$ if

$$|S(x) - S(y)| = c|x - y|, \quad \text{for all } x, y \text{ in } \mathbb{R}^n.$$

- A similarity transforms sets into geometrically similar ones with all lengths multiplied by the factor c.

Linear Transformations

- A transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is **linear** if, for all $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$:
  - $T(x+y) = T(x) + T(y)$;
  - $T(\lambda x) = \lambda T(x)$.
- Linear transformations may be represented by matrices in the usual way.
- Such a linear transformation is **nonsingular** if

$$T(x) = 0 \quad \text{if and only if} \quad x = 0.$$

Affine Transformations

- An **affine transformation** or an **affinity** is a transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ of the form

$$S(x) = T(x) + a,$$

  where:
  - T is a non-singular linear transformation;
  - a is a point in $\mathbb{R}^n$.
- An affinity's contracting or expanding effect need not be the same in every direction.
- If T is orthonormal, then S is a congruence.
- If T is a scalar multiple or an orthonormal transformation, then T is a similarity.

Groups of Transformations

- It is worth pointing out that such classes of transformations form groups under composition of mappings.

  Example: The composition of two translations is a translation.

  The identity transformation is trivially a translation.

  The inverse of a translation is a translation.

  Finally, the associative law $S \circ (T \circ U) = (S \circ T) \circ U$ holds for all translations S, T, U.
- Similar group properties hold for:
  - The congruences;
  - The rigid motions;
  - The similarities;
  - The affinities.

Hölder, Lipschitz and bi-Lipschitz Functions

- A function $f : X \to Y$ is called a **Hölder function of exponent** α if
$$|f(x) - f(y)| \leq c|x - y|^{\alpha}, \quad x, y \in X,$$
for some constant $c \geq 0$.
- The function f is called a **Lipschitz function** if α may be taken to be equal to 1, that is if
$$|f(x) - f(y)| \leq c|x - y|, \quad x, y \in X.$$
- f is called a **bi-Lipschitz function** if
$$c_1|x - y| \leq |f(x) - f(y)| \leq c_2|x - y|, \quad x, y \in X,$$
for $0 < c_1 \leq c_2 < \infty$.
- If f is bi-Lipschitz, both f and $f^{-1} : f(X) \to X$ are Lipschitz.

Limit

- Let X and Y be subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.
- Let $f : X \to Y$ be a function, and let a be a point of X.

  We say that $f(x)$ has **limit** y (or **tends to** y, or **converges to** y) **as** x **tends to** a, if, given $\varepsilon > 0$, there exists $\delta > 0$, such that, for all $x \in X$,

$$|x - a| < \delta \quad \text{implies} \quad |f(x) - y| < \varepsilon.$$

- We write $f(x) \to y$ as $x \to a$ or by $\lim_{x \to a} f(x) = y$.
- For a function $f : X \to \mathbb{R}$, we say that $f(x)$ **tends to infinity** (written $f(x) \to \infty$) **as** $x \to a$ if, given M, there exists $\delta > 0$, such that, for all $x \in X$,

$$|x - a| < \delta \quad \text{implies} \quad f(x) > M.$$

- The definition of $f(x) \to -\infty$ is similar.

Introducing Lower and Upper Limits

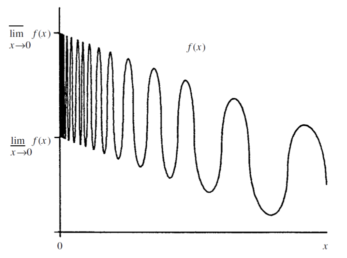
- Suppose that $f : \mathbb{R}^+ \to \mathbb{R}$.
- If $f(x)$ is increasing as x decreases, then $\lim_{x\to 0} f(x)$ exists either as a finite limit or as ∞.
- If $f(x)$ is decreasing as x decreases, then $\lim_{x\to 0} f(x)$ exists and is finite or $-\infty$.
- Of course, $f(x)$ can fluctuate wildly for small x and $\lim_{x\to 0} f(x)$ need not exist at all.
- The notions of lower and upper limits are used to describe such fluctuations.

Lower and Upper Limits

- We define the **lower limit** as
$$\underline{\lim}_{x\to 0} f(x) \equiv \lim_{r\to 0} (\inf \{f(x) : 0 < x < r\}).$$
- Since $\inf \{f(x) : 0 < x < r\}$ is either $-\infty$ for all positive r or else increases as r decreases, $\underline{\lim}_{x\to 0} f(x)$ always exists.
- The **upper limit** is defined as
$$\overline{\lim}_{x\to 0} f(x) \equiv \lim_{r\to 0} (\sup \{f(x) : 0 < x < r\}).$$

Properties of Lower and Upper Limits

- The lower and upper limits exist (as real numbers or $-\infty$ or ∞) for every function f.
- They are indicative of the variation in values of f for x close to 0.
  - $\underline{\lim}_{x\to 0} f(x) \leq \overline{\lim}_{x\to 0} f(x)$;
  - If the lower and upper limits are equal, then $\lim_{x\to 0} f(x)$ exists and equals this common value.



- If $f(x) \leq g(x)$ for $x > 0$, then $\underline{\lim}_{x\to 0} f(x) \leq \underline{\lim}_{x\to 0} g(x)$ and $\overline{\lim}_{x\to 0} f(x) \leq \overline{\lim}_{x\to 0} g(x)$.
- In the same way, it is possible to define lower and upper limits as $x \to a$ for functions $f : X \to \mathbb{R}$, where X is a subset of $\mathbb{R}^n$, a in X.

Comparing Functions

- We often need to compare two functions $f, g : \mathbb{R}^+ \to \mathbb{R}$ for small values.
- We write $f(x) \sim g(x)$ to mean that

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = 1.$$

- We will often have that $f(x) \sim x^s$.
- This means that f obeys an approximate power law of exponent s when x is small.
- On the other hand, the notation $f(x) \simeq g(x)$ is used more loosely.
- It means that $f(x)$ and $g(x)$ are approximately equal in some sense, to be specified in the particular circumstances.

Continuity and Homeomorphisms

- A function $f : X \to Y$ is **continuous** at a point a of X if
$$\lim_{x \to a} f(x) = f(a).$$
- f and is **continuous on** X if it is continuous at all points of X.
- Lipschitz and Hölder mappings are continuous.
- If $f : X \to Y$ is a continuous bijection with continuous inverse $f^{-1} : Y \to X$, then f is called a **homeomorphism**.
- Then, the sets X and Y are called **homeomorphic**.
- Congruences, similarities and affine transformations on $\mathbb{R}^n$ are examples of homeomorphisms.

Differentiability

- The function $f : \mathbb{R} \to \mathbb{R}$ is **differentiable** at x with the number $f'(x)$ as **derivative** if

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x).$$

- **Mean Value Theorem**: Given $a < b$ and f differentiable on $[a, b]$, there exists c with $a < c < b$, such that

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

  Intuitively, any chord of the graph of f is parallel to the slope of f at some intermediate point.

- f is **continuously differentiable** if $f'(x)$ is continuous in x.

Differentiability (Multiple Variables)

- Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$.
- We say that f is **differentiable** at x with **derivative** the linear mapping $f'(x) : \mathbb{R}^n \to \mathbb{R}^n$ if

$$\lim_{|h| \to 0} \frac{|f(x+h) - f(x) - f'(x)h|}{|h|} = 0.$$

Convergence and Uniform Convergence

- Consider a sequence of functions $f_k : X \to Y$, where X and Y are subsets of Euclidean spaces.
- We say that functions f_k **converge pointwise** to a function $f : X \to Y$ if, for every $x \in X$,

$$\lim_{k \to \infty} f_k(x) = f(x).$$

- We say that the convergence is **uniform** if

$$\sup_{x \in X} |f_k(x) - f(x)| \to 0 \text{ as } k \to \infty.$$

- Uniform convergence is stronger than pointwise convergence.

  The rate at which the limit is approached must be uniform across X.
- If the functions f_k are continuous and converge uniformly to f, then f is continuous.

Logarithms

- Logarithms will always be to base e.
- For $a, b > 0$, we have:

$$\begin{aligned} \log ab &= \log a + \log b; \\ \log a^c &= c \log a, \quad \text{for real numbers } c. \end{aligned}$$

- The identity $a^c = b^{c \log a / \log b}$ will often be used:

$$a^c = b^{\log_b a^c} = b^{\frac{\log a^c}{\log b}} = b^{\frac{c \log a}{\log b}}.$$

- The logarithm is the inverse of the exponential function.
  - $e^{\log x} = x$, for all $x > 0$;
  - $\log e^y = y$, for all $y \in \mathbb{R}$.

Subsection 4

Measures and Mass Distributions

Measures

- We call μ a **measure** on $\mathbb{R}^n$ if μ assigns a non-negative number, possibly ∞, to each subset of $\mathbb{R}^n$, such that:
  (a) $\mu(\emptyset) = 0$;
  (b) $\mu(A) \leq \mu(B)$ if $A \subseteq B$;
  (c) if $A_1, A_2, \ldots$ is a countable (or finite) sequence of sets then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i),$$

  with equality holding, i.e.,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i),$$

  if the A_i are disjoint Borel sets.
- We call $\mu(A)$ the **measure** of the set A, and think of $\mu(A)$ as the size of A measured in some way.

The Properties of Measures

- Condition (a) says that the empty set has zero measure;
- Condition (b) says "the larger the set, the larger the measure";
- Condition (c) says that if a set is a union of a countable number of pieces (which may overlap) then the sum of the measure of the pieces is at least equal to the measure of the whole.
- Moreover, if a set is decomposed into a countable number of disjoint Borel sets then the total measure of the pieces equals the measure of the whole.

Increasing Collections of Sets

- If $A \supseteq B$, then A may is the disjoint union $A = B \cup (A \backslash B)$.
  So, if A and B are Borel sets,
  $$\mu(A \backslash B) = \mu(A) - \mu(B).$$
- Similarly, if $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of Borel sets then $\lim\limits_{i \to \infty} \mu(A_i) = \mu(\bigcup_{i=1}^{\infty} A_i)$.
  To see this, note that $\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 \backslash A_1) \cup (A_3 \backslash A_2) \cup \cdots$, with this union disjoint. So we get
  $$\begin{array}{rcl} \mu(\bigcup_{i=1}^{\infty} A_i) & = & \mu(A_1) + \sum_{i=1}^{\infty}(\mu(A_{i+1}) - \mu(A_i)) \\ & = & \mu(A_1) + \lim_{k \to \infty} \sum_{i=1}^{k}(\mu(A_{i+1}) - \mu(A_i)) \\ & = & \lim_{k \to \infty} \mu(A_k). \end{array}$$
- If, for $\delta > 0$, A_δ are Borel sets that are increasing as δ decreases, i.e., $A_{\delta'} \subseteq A_\delta$, for $0 < \delta < \delta'$, then $\lim\limits_{\delta \to 0} \mu(A_\delta) = \mu(\bigcup_{\delta > 0} A_\delta)$.

Support and Mass Distribution

- The **support** of μ, written sptμ, is the smallest closed set X, such that
$$\mu(\mathbb{R}^n \backslash X) = 0$$
- The support of a measure is always closed.
- x is in the support if and only if $\mu(B(x, r)) > 0$, for all $r > 0$.
- We say that μ is a measure **on** a set A if A contains the support of μ.
- A measure on a bounded subset of $\mathbb{R}^n$ for which $0 < \mu(\mathbb{R}^n) < \infty$ will be called a **mass distribution**.
    - We think of $\mu(A)$ as the mass of the set A.
    - Intuitively, we take a finite mass and spread it in some way across a set X to get a mass distribution on X.

Examples

- **The counting measure**: For each subset A of $\mathbb{R}^n$ let $\mu(A)$ be:
  - The number of points in A if A is finite;
  - ∞, otherwise.

  Then μ is a measure on $\mathbb{R}^n$.
- **Point mass**: Let a be a point in $\mathbb{R}^n$.

  Define $\mu(A)$ to be:
  - 1, if A contains a;
  - 0, otherwise.

  Then μ is a mass distribution.

  It is thought of as a point mass concentrated at a

Example: Lebesgue Measure on $\mathbb{R}$

- **Lebesgue measure** $\mathcal{L}^1$ extends the idea of "length" to a large collection of subsets of $\mathbb{R}$ that includes the Borel sets.

  For open and closed intervals, we take

  $$\mathcal{L}^1(a, b) = \mathcal{L}^1[a, b] = b - a.$$

  If $A = \bigcup_i [a_i, b_i]$ is a finite or countable union of disjoint intervals, we let

  $$\mathcal{L}^1(A) = \sum (b_i - a_i)$$

  be the length of A thought of as the sum of the length of the intervals.

Example: Lebesgue Measure on $\mathbb{R}$ (Cont'd)

- This leads us to the definition of the **Lebesgue measure** $\mathcal{L}^1(A)$ of an arbitrary set A:

$$\mathcal{L}^1(A) = \inf\left\{\sum_{i=1}^{\infty}(b_i - a_i) : A \subseteq \bigcup_{i=1}^{\infty}[a_i, b_i]\right\}.$$

  I.e., we look at all coverings of A by countable collections of intervals, and take the smallest total interval length possible.
- Lebesgue measure on $\mathbb{R}$ is generally thought of as "length".
- We often write length(A) for $\mathcal{L}^1(A)$ to emphasize this meaning.

Example: Lebesgue measure on $\mathbb{R}^n$

- If $A = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : a_i \leq x_i \leq b_i\}$ is a "coordinate parallelepiped" in $\mathbb{R}^n$, the n-dimensional volume of A is given by

$$\mathrm{vol}^n(A) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n).$$

Of course, vol^1 is length, vol^2 is area and vol^3 is the usual 3-dimensional volume.

Then n-**dimensional Lebesgue measure** $\mathcal{L}^n$ is defined by extending n-dimensional volume to a large class of sets by

$$\mathcal{L}^n(A) = \inf \left\{ \sum_{i=1}^{\infty} \mathrm{vol}^n(A_i) : A \subseteq \bigcup_{i=1}^{\infty} A_i \right\},$$

where the infimum is taken over all coverings of A by coordinate parallelepipeds A_i.

Example: Lebesgue measure on $\mathbb{R}^n$ (Cont'd)

- We get that $\mathcal{L}^n(A) = \text{vol}^n(A)$ if A is a coordinate parallelepiped.
- The same holds, more generally, for any set for which the volume can be determined by the usual rules of mensuration.
- For intuition, we sometimes write:
  - area(A) in place of $\mathcal{L}^2(A)$;
  - vol(A) for $\mathcal{L}^3(A)$;
  - vol$^n(A)$ for $\mathcal{L}^n(A)$.
- Sometimes, we need to define "k-dimensional" volume on a k-dimensional plane X in $\mathbb{R}^n$.

  This may be done by identifying X with $\mathbb{R}^k$ and using $\mathcal{L}^k$ on subsets of X in the obvious way.

Example: Uniform Mass Distribution on a Line Segment

- Let L be a line segment of unit length in the plane.

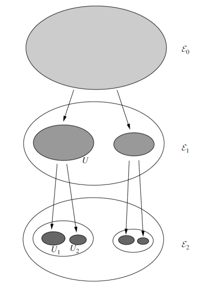
  Define
  $$\mu(A) = \mathcal{L}^1(L \cap A),$$
  i.e., the "length" of intersection of A with L.
- For all A, with $A \cap L = \emptyset$,
  $$\mu(A) = 0.$$
  So μ is a mass distribution with support L.
- We may think of μ as unit mass spread evenly along the line segment L.

Restriction of a Measure

- Let μ be a measure on $\mathbb{R}^n$.
- Let E be a Borel subset of $\mathbb{R}^n$.
- We may define a measure ν on $\mathbb{R}^n$, called the **restriction of μ to E**, by
$$\nu(A) = \mu(E \cap A), \text{ for every set } A.$$
- Then ν is a measure on $\mathbb{R}^n$ with support contained in $\overline{E}$.

Mass Distributions on Subsets of $\mathbb{R}^n$

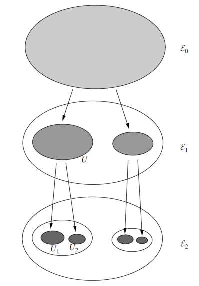
- Let $\mathcal{E}_0$ consist of the single Borel set E.
- For $k = 1, 2, \ldots$, let $\mathcal{E}_k$ be a collection of disjoint Borel subsets of E, such that each set U in $\mathcal{E}_k$:



  - Is contained in one of the sets of $\mathcal{E}_{k-1}$;
  - Contains a finite number of the sets in $\mathcal{E}_{k+1}$.
- We assume that the maximum diameter of the sets in $\mathcal{E}_k$ tends to 0 as $k \to \infty$.

Mass Distributions on Subsets of $\mathbb{R}^n$ (Cont'd)

- We define a mass distribution on E by repeated subdivision.
- We let $\mu(E)$ satisfy $0 < \mu(E) < \infty$.
- We split this mass between the sets $U_1, \ldots, U_m$ in $\mathcal{E}_1$ by defining $\mu(U_i)$ in such a way that
$$\sum_{i=1}^{m} \mu(U_i) = \mu(E).$$



- Similarly, we assign masses to the sets of $\mathcal{E}_2$ so that if $U_1, \ldots, U_m$ are the sets of $\mathcal{E}_2$ contained in a set U of $\mathcal{E}_1$, then
$$\sum_{i=1}^{m} \mu(U_i) = \mu(U).$$

Mass Distributions on Subsets of $\mathbb{R}^n$ (Cont'd)

- In general, we assign masses so that

$$\sum_i \mu(U_i) = \mu(U)$$

  for each set U of $\mathcal{E}_k$, where the $\{U_i\}$ are the disjoint sets in $\mathcal{E}_{k+1}$ contained in U.
- For each k, we let E_k be the union of the sets in $\mathcal{E}_k$.
- We define $\mu(A) = 0$, for all A, with $A \cap E_k = \emptyset$.
- Let $\mathcal{E}$ denote the collection of sets that belong to $\mathcal{E}_k$ for some k together with the subsets of $\mathbb{R}^n \backslash E_k$.
- The above procedure defines the mass $\mu(A)$ of every set A in $\mathcal{E}$.

Justification of the Construction

Proposition

Let μ be defined on a collection of sets $\mathcal{E}$ as above. Then the definition of μ may be extended to all subsets of $\mathbb{R}^n$ so that μ becomes a measure. The value of $\mu(A)$ is uniquely determined if A is a Borel set. The support of μ is contained in $\bigcap_{k=1}^{\infty} \overline{E}_k$.

- **Note on Proof**: If A is any subset of $\mathbb{R}^n$, let

$$\mu(A) = \inf\left\{\sum_i \mu(U_i) : A \subseteq \bigcup_i U_i \text{ and } U_i \in \mathcal{E}\right\}.$$

Thus we take the smallest value we can of $\sum_{i=1}^{\infty} \mu(U_i)$ where the sets U_i are in $\mathcal{E}$ and cover A.

Justification of the Construction (Cont'd)

- For A a subset of $\mathbb{R}^n$, we defined

$$\mu(A) = \inf\left\{\sum_i \mu(U_i) : A \subseteq \bigcup_i U_i \text{ and } U_i \in \mathcal{E}\right\}.$$

We have already defined $\mu(U_i)$ for $U_i \in \mathcal{E}$.

It is not difficult to see that if A is one of the sets in $\mathcal{E}$, then this reduces to the mass $\mu(A)$ specified in the construction.

Since $\mu(\mathbb{R}^n \backslash E_k) = 0$, we have $\mu(A) = 0$ if A is an open set that does not intersect E_k for some k.

This shows that the support of μ is in $\overline{E}_k$ for all k.

Example

- Let $\mathcal{E}_k$ denote the collection of "binary intervals" of length 2^{-k} of the form
$$[r2^{-k}, (r+1)2^{-k}), \quad 0 \leq r \leq 2^k - 1.$$
Take
$$\mu[r2^{-k}, (r+1)2^{-k}) = 2^{-k}.$$
Then the above construction gives the Lebesgue measure μ on $[0,1]$.
- **Note on calculation**: The requirements are satisfied.
If I is an interval in $\mathcal{E}_k$ of length 2^{-k} and I_1, I_2 are the two subintervals of I in $\mathcal{E}^{k+1}$ of length 2^{-k-1}, we have
$$\mu(I) = \mu(I_1) + \mu(I_2).$$
By the proposition, μ extends to a mass distribution on $[0,1]$.
We have $\mu(I) = \text{length}(\mu)$ for I in $\mathcal{E}$.
This implies that μ coincides with Lebesgue measure on any set.

Almost Everywhere

- We say that a property holds for **almost all** x, or **almost everywhere** (**with respect to a measure** μ) if the set for which the property fails has μ-measure zero.

  Example: We say almost all real numbers are irrational with respect to Lebesgue measure.

  The rational numbers $\mathbb{Q}$ are countable.

  They may be listed as $x_1, x_2, \ldots$, say.

  So $\mathcal{L}^1(\mathbb{Q}) = \sum_{i=1}^{\infty} \mathcal{L}^1\{x_i\} = 0$.

Hypothesis on Functions

- We would like to avoid technical difficulties involved in integrating functions with respect to measures.
- Let $f : D \to \mathbb{R}$ be a function defined on a Borel subset D of $\mathbb{R}^n$.
- We will assume that the set

$$f^{-1}(-\infty, a] = \{x \in D : f(x) \leq a\}$$

  is a Borel set for all real numbers a.
- A very large class of functions satisfies this condition.
- It includes all continuous functions, for which $f^{-1}(-\infty, a]$ is closed and therefore a Borel set.
- All functions to be integrated are taken to satisfy this condition.

Integration

- Suppose, first, that $f : D \to \mathbb{R}$ is a simple function, i.e., one that takes only finitely many values $a_1, \ldots, a_k$.
- We define the **integral with respect to the measure** μ of a non-negative simple function f as

$$\int f d\mu = \sum_{i=1}^{k} a_i \mu\{x : f(x) = a_i\}.$$

- The integral of more general functions is defined using approximation by simple functions.
- If $f : D \to \mathbb{R}$ is a non-negative function, we define its integral as

$$\int f d\mu = \sup\left\{\int g d\mu : g \text{ is simple, } 0 \le g \le f\right\}.$$

Integration (Cont'd)

- To complete the definition, if f takes both positive and negative values, we let

$$f^+(x) = \max\{f(x), 0\} \quad \text{and} \quad f^-(x) = \max\{-f(x), 0\}.$$

- Then, we have

$$f = f^+ - f^-.$$

- We define

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu,$$

provided that $\int f^+ d\mu$ and $\int f^- d\mu$ are both finite.

Properties of Integrals

- For functions $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$,
$$\int (f+g)d\mu = \int fd\mu + \int gd\mu.$$
- For a function $f : D \to \mathbb{R}$ and a scalar λ,
$$\int \lambda fd\mu = \lambda \int fd\mu.$$
- **Monotone Convergence Theorem**: If $f_k : D \to \mathbb{R}$ is an increasing sequence of non-negative functions converging (pointwise) to f, then
$$\lim_{k\to\infty} \int f_k d\mu = \int fd\mu.$$

Properties of Integrals (Cont'd)

- Given a Borel subset A of D, define its **indicator function** $\chi_A : \mathbb{R}^n \to \mathbb{R}$, by

$$\chi_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

- If A is a Borel subset of D, we define **integration over the set** A by

$$\int_A f d\mu = \int f \chi_A d\mu.$$

- If $f(x) \geq 0$ and $\int f d\mu = 0$, then

$$f(x) = 0 \text{ for } \mu\text{-almost all } x.$$

Integration Notation

- Integration is denoted in various ways, such as

$$\int f d\mu, \quad \int f \quad \text{or} \quad \int f(x) d\mu(x),$$

depending on the emphasis.

- When μ is n-dimensional Lebesgue measure $\mathcal{L}^n$, we usually write

$$\int f dx \quad \text{or} \quad \int f(x) dx$$

in place of $\int f d\mathcal{L}^n$.

Egoroff's Theorem

- Let D be a Borel subset of $\mathbb{R}^n$.
- Let μ a measure with $\mu(D) < \infty$.
- Let $f_1, f_2, \ldots$ and f be functions from D to $\mathbb{R}$, such that

$$f_k(x) \to f(x), \quad \text{for each } x \text{ in } D.$$

- Egoroff's Theorem asserts that, for any $\delta > 0$, there is a Borel subset E of D, such that $\mu(D \backslash E) < \delta$ and such that the sequence $\{f_k\}$ converges uniformly to f on E.
- I.e., $\{f_k\}$ satisfies

$$\sup_{x \in E} |f_k(x) - f(x)| \to 0, \text{ as } k \to \infty.$$

- For the measures that we shall be concerned with, it may be shown that we can always take the set E to be compact.