LECTURE 2

Counting the Infinite, and the Geometry of Distance: Countability and Metric Spaces

MATH S4061 | Introduction to Modern Analysis I

Thursday, 28 May 2026

No one shall expel us from the paradise that Cantor has created.

David Hilbert

Contents of this lecture

Overview. Lecture 1 built $\mathbb{R}$ as the complete ordered field and drew from completeness the Archimedean property (1.7.1) and the density of $\mathbb{Q}$ (1.7.2); the same property gives infima as readily as suprema (1.5.3). Two new questions open here. First, *how large* is $\mathbb{R}$? Cantor's notion of countability lets us compare infinite sets, and his diagonal argument shows $\mathbb{R}$ is strictly larger than $\mathbb{Q}$—uncountable, with $\mathbb{Q}$ a vanishingly "thin" subset. Second, what is the right language for *distance*? Metric spaces distil the Euclidean norm into three axioms, and the open sets they generate are the gateway to topology, where the rest of the course lives. The companion text is Rudin, Chapter 2, especially R 2.1–R 2.19.

2.1 Counting the Infinite: Countable and Uncountable Sets

Definition 2.1.1

Injective, surjective, bijective

Let $f : X \to Y$ be a function (cf. R 2.1–R 2.3).

- f is *injective* if $f(x) = f(x') \Rightarrow x = x'$ (distinct inputs give distinct outputs);
- f is *surjective* if every $y \in Y$ equals $f(x)$ for some $x \in X$;
- f is *bijective* if it is both. A bijection has a two-sided inverse $f^{-1} : Y \to X$.

Definition 2.1.2

Equinumerous sets; finite, countable, uncountable

Sets X and Y are *equinumerous*, written $X \sim Y$, if there is a bijection $f : X \to Y$ (R 2.3). With $J_0 = \emptyset$ and $J_n = \{1, \dots, n\}$, a set X is

- *finite* if $X \sim J_n$ for some n, and *infinite* otherwise;
- *countable* if $X \sim \mathbb{N}$;
- *uncountable* if it is infinite and $X \not\sim \mathbb{N}$.[1]

This packages Rudin's finite/countable terminology R 2.4–R 2.7.

Proposition 2.1.3

Equinumerosity is an equivalence relation

The relation $\sim$ is reflexive, symmetric, and transitive. This is the basic relation behind R 2.3.

Proof 2.1.3

Direct Proof that Equinumerosity is an Equivalence Relation.

(1) For reflexivity, the identity $\mathrm{id}_X : X \to X$, $x \mapsto x$, is a bijection, so $X \sim X$.

(2) For symmetry, if $X \sim Y$ via a bijection f, then $f^{-1} : Y \to X$ is a bijection, so $Y \sim X$.

(3) For transitivity, if $X \sim Y$ via f and $Y \sim Z$ via g, then $g \circ f : X \to Z$ is a bijection (its inverse is $f^{-1} \circ g^{-1}$), so $X \sim Z$.

∎

[Direct] bijections invert and compose, with $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Example 2.1.4

$\mathbb{Z}$ is countable

[1] The definition can feel strange—how can $\mathbb{N}$ be the "same size" as a set that looks bigger or smaller? The shift is to measure size by *pairing* rather than counting: two sets match if their elements can be lined up in perfect one-to-one correspondence with none left over. For finite sets this is just counting; for infinite sets it is the only thing that still works, and it makes you give up the finite habit that a part must be smaller than the whole—$n \mapsto 2n$ already pairs all of $\mathbb{N}$ with only the even numbers.

The identity shows $\mathbb{N}$ is countable. For $\mathbb{Z}$, interleave the signs:

$$f : \mathbb{N} \to \mathbb{Z}, \qquad f(n) = \begin{cases} n/2 & n \text{ even}, \\ -(n+1)/2 & n \text{ odd}, \end{cases}$$

a bijection that lists $\mathbb{Z}$ as $0, -1, 1, -2, 2, -3, 3, \ldots$ (cf. R 2.12–R 2.13).

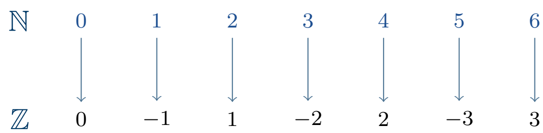

Figure 2.1.5. $\mathbb{N}$ COUNTS $\mathbb{Z}$. The bijection of 2.1.4 walks outward from 0 in alternating steps, reaching every integer exactly once.

Definition 2.1.6

Sequences and the listing criterion

A *sequence* in a set X is a function $f : \mathbb{N} \to X$, written $\{x_n\}_{n\in\mathbb{N}}$ with $x_n = f(n)$. Thus X is countable precisely when some sequence lists every element of X exactly once. Rudin's sequence definition is R 2.7.

Example 2.1.7

$\mathbb{Q}$ is countable

Identifying a fraction with its pair of integers exhibits $\mathbb{Q}$ inside $\mathbb{Z} \times \mathbb{Z}$, which a diagonal sweep lists as a single sequence.

Proof 2.1.7

Proof by Diagonal Enumeration that $\mathbb{Q}$ is Countable.

(1) Arrange the pairs (p, q) with $q \neq 0$ on a grid and traverse them along successive finite diagonals (each diagonal $|p| + |q| = \text{const}$ holds only finitely many pairs).

(2) This runs through every pair; deleting pairs with a common factor and any repeats leaves a sequence that lists each rational exactly once.

(3) Hence $\mathbb{Q} \sim \mathbb{N}$ (2.1.6).

■

[Constr] cf. R 2.12–R 2.13.

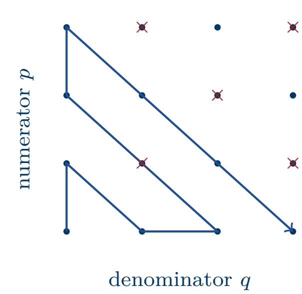

Figure 2.1.8. LISTING THE RATIONALS. Sweeping the lattice of pairs (p, q) along finite diagonals threads them into one sequence; pairs with a common factor (e.g. $\frac{2}{2}, \frac{2}{4}$) are skipped (2.1.7).

Proposition 2.1.9

An infinite subset of a countable set is countable

If X is countable and $A \subseteq X$ is infinite, then A is countable (R 2.8).

Proof 2.1.9

Proof by Recursive Selection of Indices that the Subset is Countable.

(1) Write $X = \{x_n\}_{n\in\mathbb{N}}$, every element listed once (2.1.6).

(2) Choose indices recursively: let n_0 be the least index with $x_{n_0} \in A$; having chosen $n_0 < \cdots < n_k$, let n_{k+1} be the least index $> n_k$ with $x_{n_{k+1}} \in A$. Such an index exists because A, being infinite, cannot lie inside the finite set $\{x_0, \ldots, x_{n_k}\}$, so some $x_m \in A$ has $m > n_k$.

(3) The map $k \mapsto x_{n_k}$ is strictly increasing, hence injective, and reaches every element of A (each $a = x_m \in A$ is selected no later than stage m); so it is a bijection $\mathbb{N} \to A$, and $A \sim \mathbb{N}$.

■

[A10] cf. R 2.8; the recursion is an induction, justified at each step by A being infinite.

Theorem 2.1.10

Cantor: the binary sequences are uncountable

The set $2^{\mathbb{N}} = \{\, f : \mathbb{N} \to \{0, 1\} \,\}$ of infinite binary sequences—equivalently, the subsets of $\mathbb{N}$—is uncountable.[2]

Proof 2.1.10

Proof by Cantor's Diagonal Argument that the Binary Sequences are Uncountable.

(1) Suppose $2^{\mathbb{N}}$ were countable, listed as $a_1, a_2, a_3, \ldots$, each $a_k = (a_k^1, a_k^2, a_k^3, \ldots)$ a binary sequence.

(2) Define $\bar{b} \in 2^{\mathbb{N}}$ by flipping the diagonal: $\bar{b}_n = 1 - a_n^n$.

(3) For each n, $\bar{b}$ differs from a_n in its nth entry ($\bar{b}_n \neq a_n^n$), so $\bar{b} \neq a_n$ for every n.

(4) Thus $\bar{b}$ appears nowhere in the list—contradicting that the list was all of $2^{\mathbb{N}}$. Hence $2^{\mathbb{N}}$ is uncountable.

■

[A8+A9] diagonalisation, by contradiction; cf. R 2.14.

[2] The diagonal proof below can look like a sleight of hand. Against *any* proposed list of binary sequences it produces one the list left out, so no list can hold them all—and you read that missing sequence straight off the list: make it disagree with the kth listed sequence in its kth entry, and it differs from the first (in entry 1), the second (entry 2), and every one after. The move is worth keeping on its own: to escape an enumeration, differ from item k in coordinate k.

$$\begin{array}{c|cccccc}
a_1 & 0 & 1 & 1 & 0 & 1 & 0 \\
a_2 & 1 & 1 & 1 & 1 & 0 & 1 \\
a_3 & 0 & 0 & 0 & 1 & 0 & 1 \\
a_4 & 0 & 1 & 1 & 0 & 1 & 0 \\
a_5 & 1 & 0 & 0 & 0 & 0 & 1 \\
\\
\bar{b} & 1 & 0 & 1 & 1 & 1 & \cdots
\end{array}$$

Figure 2.1.11. Cantor's diagonal. Flipping the boxed diagonal entries yields a sequence $\bar{b}$ differing from each listed a_k in place k, so $\bar{b}$ is absent from the list (2.1.10).

Theorem 2.1.12

The interval $[0,1]$ is uncountable

The interval $[0,1] \subseteq \mathbb{R}$ is uncountable.

Proof 2.1.12

Proof by Reduction to Binary Sequences that $[0,1]$ *is Uncountable.*

(1) Since $[0,1] \supseteq [0,1)$, it suffices to show $[0,1)$ is uncountable. Send each $x \in [0,1)$ to the binary expansion $(b_n) \in 2^{\mathbb{N}}$ that does *not* end in all 1s; on $[0,1)$ this representative exists and is unique, so the map is injective, with image exactly the sequences not ending in all 1s.

(2) If $[0,1)$ were countable, that image would be countable, so $2^{\mathbb{N}}$—the image together with the countably many all-1-tail sequences—would be countable too (R 2.12), contradicting 2.1.10.

(3) Hence $[0,1)$, and therefore $[0,1]$, is uncountable.

■

[Direct] uses 2.1.10; barring an all-1 tail makes binary expansions unique ($0.0111\ldots_2 = 0.1000\ldots_2$), and the barred eventually-all-1 sequences are countable.

Corollary 2.1.13

$\mathbb{R}$ is uncountable

Were $\mathbb{R}$ countable, its infinite subset $[0,1]$ would be countable too (2.1.9); but it is not (2.1.12). So $\mathbb{R}$ is uncountable.[3] Compare Rudin's diagonal proof for binary sequences R 2.14 and later perfect-set result R 2.43.

Remark 2.1.14

$\mathbb{Q}$ *is thin in* $\mathbb{R}$

[3] "Uncountable" says something stronger than "a bigger infinity": *no list can ever catch all the reals*—hand over any enumeration and the diagonal argument hands back a real it missed. To gauge the size of the gap: the numbers you can pin down as a root of a polynomial with integer coefficients (the *algebraic* numbers) form only a countable set, so *almost every* real lies beyond them—*transcendental*—even though such numbers are hard to point to.

Thus $\mathbb{Q}$ is countable yet dense (1.7.2), while $\mathbb{R}$ is uncountable: the rationals are a vanishingly small part of the line. Made precise by measure, $\mathbb{Q}$—indeed any countable set—has Lebesgue length zero, so a real drawn from any distribution with a density is irrational with probability 1. Compare the countability results in R 2.12–R 2.13 and the diagonal uncountability result R 2.14; the measure language is later perspective.

Theorem 2.1.15

A countable union of countable sets is countable

If $A_1, A_2, A_3, \ldots$ are countable sets, then $\bigcup_{n=1}^{\infty} A_n$ is countable (R 2.12).[4]

Proof 2.1.15

Proof by Diagonal Enumeration of the Array that the Union is Countable.

(1) List each set as a sequence (2.1.6), say $A_n = \{x_{n,1}, x_{n,2}, x_{n,3}, \ldots\}$, and arrange all the elements in the doubly-indexed array

$$\begin{matrix} x_{1,1} & x_{1,2} & x_{1,3} & \cdots \\ x_{2,1} & x_{2,2} & x_{2,3} & \cdots \\ x_{3,1} & x_{3,2} & x_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \end{matrix}$$

whose nth row exhausts A_n.

(2) Enumerate the array along its successive anti-diagonals—the entries $x_{n,m}$ with $n + m = k$, taken for $k = 2, 3, 4, \ldots$:

$$x_{1,1}; \quad x_{1,2}, x_{2,1}; \quad x_{1,3}, x_{2,2}, x_{3,1}; \quad \ldots$$

Each diagonal is finite, so this is a single sequence, and it visits every entry of the array, hence every element of $\bigcup_n A_n$.

(3) This sequence may repeat an element (the sets need not be disjoint). If the union is finite there is nothing to prove; otherwise discard repetitions, leaving an infinite subsequence that lists $\bigcup_n A_n$ without repetition (2.1.9). Hence $\bigcup_{n=1}^{\infty} A_n$ is countable.

■

[A8] the array threaded along finite diagonals, repetitions pruned by 2.1.9; cf. R 2.12.

[4] The trick is to defeat *two* indices at once. The elements come labelled by a row n (which set) and a column m (which element of that set), so they fill a square array—and a square array is "two-dimensional," seemingly too big to list. The diagonal sweep is the resolution: each anti-diagonal $n + m = k$ is *finite*, so walking the diagonals one after another threads the whole grid onto a single line. The same diagonal that listed the pairs for $\mathbb{Q}$ (2.1.7) is doing the work here.

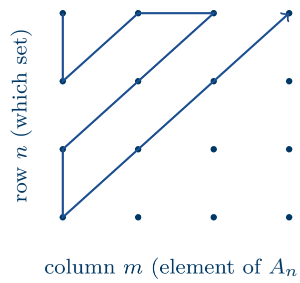

Figure 2.1.16. Threading the array. Sweeping the array $(x_{n,m})$ along finite anti-diagonals $n+m=k$ threads the whole double-indexed family onto one sequence (2.1.15).

Corollary 2.1.17

A finite product of countable sets is countable

If $A_1, \ldots, A_k$ are countable sets, then the product $A_1 \times \cdots \times A_k$ is countable (R 2.13). In particular $\mathbb{Z} \times \mathbb{Z}$ is countable, giving a second route to $\mathbb{Q}$ countable (2.1.7).

Proof 2.1.17

Proof by Induction on the Number of Factors that the Product is Countable.

(1) The case $k=1$ is immediate. We induct on k, taking $k=2$ as the base recorded by the array.

(2) *Base* $k=2$. Write $A_1 \times A_2 = \bigcup_{a \in A_1} (\{a\} \times A_2)$. Each slice $\{a\} \times A_2$ is in bijection with the countable set A_2, hence countable, and the union runs over the countable index set A_1. So $A_1 \times A_2$ is a countable union of countable sets, countable by 2.1.15.

(3) *Step.* Suppose $A_1 \times \cdots \times A_{k-1}$ is countable. Decompose

$$A_1 \times \cdots \times A_k = \bigcup_{a \in A_k} ((A_1 \times \cdots \times A_{k-1}) \times \{a\}),$$

again a countable union (over A_k) of countable sets, so it is countable by 2.1.15.

(4) By induction $A_1 \times \cdots \times A_k$ is countable for every k. Taking $A_1 = A_2 = \mathbb{Z}$ shows $\mathbb{Z} \times \mathbb{Z}$ is countable; since $\mathbb{Q}$ embeds in it as an infinite subset (a fraction in lowest terms is a pair (p,q), $q>0$), $\mathbb{Q}$ is countable by 2.1.9.

■

[A10] base case $k=2$ via 2.1.15; cf. R 2.13.

2.2 Metric Spaces

The Euclidean distance on $\mathbb{R}^n$, $d(x,y) = \|x-y\| = \left(\sum_{i=1}^n |x_i - y_i|^2\right)^{1/2}$, built from the inner product $\langle x, y \rangle = \sum_i x_i y_i$, is the model. A metric space keeps only the features of d that analysis actually uses.

Definition 2.2.1

Metric space

A *metric space* (R 2.15)[5] is a set X with a *metric* (or distance) $d : X \times X \to \mathbb{R}^{\geq 0}$ satisfying, for all $x, y, z \in X$:

1. $d(x, y) \geq 0$, with $d(x, y) = 0 \iff x = y$;
2. $d(x, y) = d(y, x)$ (symmetry);
3. $d(x, z) \leq d(x, y) + d(y, z)$ (the triangle inequality).

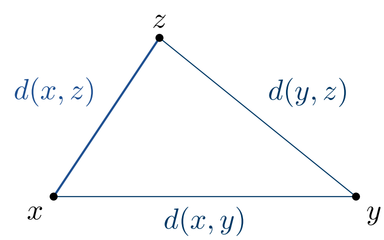

Figure 2.2.2. THE TRIANGLE INEQUALITY. The direct distance $d(x, z)$ never exceeds the detour $d(x, y) + d(y, z)$ through y—axiom (3) of 2.2.1.

Example 2.2.3

The standard metrics

$(\mathbb{R}, |\cdot|)$ with $d(x, y) = |x - y|$ is a metric space, as are $(\mathbb{Z}, |\cdot|)$, $(\mathbb{N}, |\cdot|)$, and $(\mathbb{Q}, |\cdot|)$. More generally any subset Y of a metric space (X, d) is itself a metric space under the restricted distance $d|_{Y \times Y}$. The Euclidean space $(\mathbb{R}^n, d_2)$, $d_2(x, y) = \big(\sum_{i=1}^n |x_i - y_i|^2\big)^{1/2}$, is the prototype. These are Rudin's basic metric examples (cf. R 2.16).

Example 2.2.4

The p-metrics and the discrete metric

On $\mathbb{R}^n$, for $p \geq 1$,[6]

$$d_p(x, y) = \Big(\sum_{i=1}^n |x_i - y_i|^p\Big)^{1/p}, \qquad d_\infty(x, y) = \max_i |x_i - y_i|.$$

On any set X, the *discrete metric*—$d(x, y) = 0$ if $x = y$ and 1 otherwise—is a metric. The triangle inequality for d_p is Minkowski's inequality, taken here as a standard fact (cf. R 2.16).

Example 2.2.5

Metrics on a function space

[5] Read the three axioms not as arbitrary rules but as the *least* a notion of "distance" must obey for analysis to run on it. (1) distances are honest numbers that vanish only when two points genuinely coincide, so d can tell points apart; (2) distance doesn't care which direction you travel; (3) the triangle inequality says a detour is never shorter than going straight—the one axiom with real force, and the workhorse of almost every proof to come (already 2.3.5). That is the whole point of abstracting: keep only these, and every notion built from them works on *any* such space at once—Euclidean space, the function spaces of 2.2.5, and more.

[6] The board writes $p > 1$; in fact d_p is a metric for every $p \geq 1$, with d_1 and d_∞ the boundary cases.

Let $C = C[0,1]$ be the continuous functions $f : [0,1] \to \mathbb{R}$.[7] Both

$$d_\infty(f,g) = \sup_{[0,1]} |f-g| \qquad \text{and} \qquad d_1(f,g) = \int_0^1 |f(x)-g(x)|\,dx$$

are metrics on C. Here d_∞ is finite because a continuous function on $[0,1]$ is bounded, and $d_1(f,g)=0$ forces $f=g$ since $|f-g|$ is continuous. As metric examples, compare R 2.16.

2.3 Open Sets and Open Balls

Basic *topology* studies a space through its *open subsets*[8]—concepts that can be phrased using open sets alone. Everything starts with the open ball.

Definition 2.3.1

Open ball

In a metric space (X,d), the *open ball* of radius $r > 0$ about $p \in X$ is

$$B_r(p) = \{\, x \in X : d(x,p) < r \,\}.$$

Rudin calls this a neighbourhood of p in R 2.18.

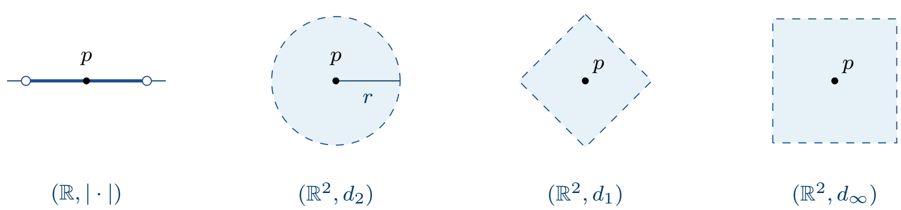

Figure 2.3.2. An open ball takes the shape of its metric. Same centre, same radius: an interval in $(\mathbb{R}, |\cdot|)$, a round disk under d_2, a diamond under d_1, a square under d_∞—each the set $\{x : d(x,p) < r\}$ of 2.3.1.

Definition 2.3.3

Interior point; open set

Let (X,d) be a metric space and $E \subseteq X$. A point $p \in E$ is an *interior point* of E if $B_r(p) \subseteq E$ for some $r > 0$, and E is *open in* X if every point of E is interior (R 2.18). Conversely, p fails to be interior exactly when $B_r(p) \not\subseteq E$ for *every* $r > 0$; so E is *not* open as soon as one of its points is not interior.

[7] The leap to make here is to stop picturing a function as a *graph* and start treating it as a single *point* in a space whose points happen to be functions. Then "two functions are close" just means their distance is small—and which metric you pick decides what "close" means: under d_∞ two functions are close only if they stay near each other *everywhere at once* (uniformly), while under d_1 they may disagree sharply on a thin set so long as the *average* gap is small. Same functions, two genuinely different geometries.

[8] The picture to carry: a set is *open* when none of its points sits on the "edge"—every point has a whole little ball of room around it that is still inside the set, so from any point you can step a little in every direction without leaving. That "room to move" is what makes open sets the right language for analysis: continuity, convergence, and compactness can all be expressed through open sets alone, never mentioning the particular distance that produced them (indeed the d_1, d_2, d_∞ of 2.2.4 produce the very same open sets on $\mathbb{R}^n$).

Remark 2.3.4

Openness is relative

Openness is meaningful only *relative to the ambient space*: one says "E is open in X," never merely "E is open." The same set can be open in one metric space and not in another. This anticipates Rudin's relative-openness theorem R 2.30.

Proposition 2.3.5

Every open ball is open

In any metric space (X, d), the ball $B_r(p)$ is open in X, for every $p \in X$ and $r > 0$.

Proof 2.3.5

Proof by the Triangle Inequality that Every Open Ball is Open.

(1) Fix $q \in B_r(p)$; then $d(p, q) < r$, so $r_q := r - d(p, q) > 0$.

(2) Take any $x \in B_{r_q}(q)$. By the triangle inequality,

$$d(x, p) \leq d(x, q) + d(q, p) < r_q + d(q, p) = r.$$

(3) Hence $x \in B_r(p)$, so $B_{r_q}(q) \subseteq B_r(p)$. Every $q \in B_r(p)$ is therefore interior, and $B_r(p)$ is open.

■

[Direct] the template for openness proofs; uses 2.2.1; cf. R 2.19.

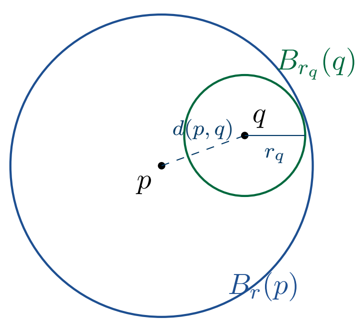

Figure 2.3.6. WHY A BALL IS OPEN. For $q \in B_r(p)$, the radius $r_q = r - d(p, q)$ keeps $B_{r_q}(q)$ inside $B_r(p)$—the triangle inequality at work (proof of 2.3.5).

Remark 2.3.7

A recipe for proving openness

To show E is open in X: (1) take an arbitrary $p \in E$; (2) produce a radius $r_p > 0$ (guided by the geometry); (3) verify $B_{r_p}(p) \subseteq E$. The proof of 2.3.5 is the model, and this is the working form of R 2.18–R 2.19.

Index of Objects