MATH 140A: FOUNDATIONS OF REAL ANALYSIS I

TODD KEMP

CONTENTS

1. Ordered Sets, Ordered Fields, and Completeness 1
1.1. Lecture 1: January 5, 2016 1
1.2. Lecture 2: January 7, 2016 4
1.3. Lecture 3: January 11, 2016 7
1.4. Lecture 4: January 14, 2014 9
2. Sequences and Limits 13
2.1. Lecture 5: January 19, 2016 13
2.2. Lecture 6: January 21, 2016 15
2.3. Lecture 7: January 26, 2016 18
2.4. Lecture 8: January 28, 2016 21
3. Extensions of $\mathbb{R}$: the Extended Real Numbers $\overline{\mathbb{R}}$ and the Complex Numbers $\mathbb{C}$ 25
3.1. Lecture 9: February 2, 2016 25
3.2. Lecture 10: February 4, 2016 28
4. Series 32
4.1. Lecture 11: February 9, 2016 32
4.2. Lecture 12: February 11, 2016 37
4.3. Lecture 13: February 16, 2016 42
5. Metric Spaces 44
5.1. Lecture 14: February 18, 2016 47
5.2. Lecture 15: February 23, 2016 50
5.3. Lecture 16: February 25, 2016 52
5.4. Lecture 17: February 29, 2016 55
6. Limits and Continuity 57
6.1. Lecture 18: March 3, 2016 57
6.2. Lecture 19: March 8, 2016 60
6.3. Lecture 20: March 10, 2016 63

1. ORDERED SETS, ORDERED FIELDS, AND COMPLETENESS

1.1. Lecture 1: January 5, 2016.

- $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$.
- $\mathbb{R}$ is the "Réal numbers". There is nothing *real* about them! That is the first, most important lesson to learn in this class. We will encounter many "obvious" statements that are, in fact, *false*. We will also see some counterintuitive statements that turn out to be true.

Date: March 9, 2016.

- Mathematicians roughly split into two groups: analysts and algebraists. (There's lots of overlap, though.) Roughly speaking, algebraists are largely concerned about *equalities*, while analysts are largely concerned about *inequalities*.

Definition 1.1. *A* **total order** *is a binary relation $<$ on a set S which satisfies:*

1. *transitive: if $x, y, z \in S$, $x < y$, and $y < z$, then $x < z$.*
2. *ordered: given any $x, y \in S$, exactly one of the following is true: $x < y$, $x = y$, or $y < x$.*

The usual order relation on $\mathbb{Q}$ (and its subsets $\mathbb{Z}$ and $\mathbb{N}$) is a total order. As usual, we write $x > y$ to mean $y < x$, and $x \leq y$ to mean "$x < y$ or $x = y$".

Definition 1.2. *Let $(S, <)$ be a totally ordered set. Let $E \subseteq S$. A* **lower bound** *for E is an element $\alpha \in S$ with the property that $\alpha \leq x$ for each $x \in E$. A* **upper bound** *for E is an element $\beta \in S$ with the property that $x \leq \beta$ for each $x \in E$. If E possesses an upper bound, we say E is* **bounded above***; if it possesses a lower bound, it is* **bounded below**.

For example, the set $\mathbb{N}$ is bounded below in $\mathbb{Z}$, but it is not bounded above. Any set that has a maximal element is bounded above by its maximum; similarly, any set with a minimal element is bounded below by its minimum.

Definition 1.3. *Let $(S, <)$ be a totally ordered set, and let $E \subseteq S$ be bounded above. The* **least upper bound** *or* **supremum** *of E, should it exist, is*

$$\sup E \equiv \min\{\beta \in S \colon \beta \text{ is an upper bound of } E\}.$$

Similarly, if F is bounded below, the **greatest lower bound** *or* **infimum** *of F, should it exsit, is*

$$\inf F \subseteq S \equiv \max\{\alpha \in S \colon \alpha \text{ is a lower bound of } F\}.$$

To work with the definition (of sup, say), we rewrite it slightly. A number $\sigma \in S$ is the supremum of E if the following two properties hold:

1. σ is an upper bound of E.
2. Given any $s \in S$ with $s < \sigma$, s is *not* an upper bound of E; i.e. there exists some $x \in E$ with $s < x \leq \sigma$.

Example 1.4. Consider the set $E = \{\frac{1}{n} \colon n \in \mathbb{N}\} \subset \mathbb{Q}$. This set has a maximal element: 1. So 1 is an upper bound. Moreover, if $s \in \mathbb{Q}$ is < 1, then s is not an upper bound of E (since $1 \in E$). Thus, $1 = \sup E$. (This argument shows in general that, *if* E has a maximal element, then $\max E = \sup E$.)

On the other hand, E has no minimal element. But note that all elements of E are positive, so 0 is a lower bound for E. If s is any rational number > 0, there is certainly some $n \in \mathbb{N}$ with $0 < \frac{1}{n} < s$ (this is the Archimedean property of the rational field). Hence, no such s is a lower bound for E. This shows that 0 is the *greatest* lower bound: $0 = \inf E$.

Example 1.5. It is well known that $\sqrt{2}$ is not rational: in other words, there is no rational number p satisfying $p^2 = 2$. You probably saw this proof in high school. Suppose, for a contradiction, that $p^2 = 2$. Since p is rational, we can write it *in lowest terms* as $p = m/n$ for $m, n \in \mathbb{Z}$. So we have $\frac{m^2}{n^2} = 2$, or $m^2 = 2n^2$. Thus m^2 is even, which means that m is even (since the square of an odd integer is odd). So $m = 2k$ for some $k \in \mathbb{Z}$, meaning $m^2 = 4k^2$, and so $4k^2 = 2n^2$, from which it follows that $n^2 = 2k^2$ is even. As before, this imples that n is even. But then both m and n are divisible by 2, which means they are not relatively prime. This contradicts the assumption that $p = m/n$ is in lowest terms.

A finer analysis of this situation shows that $\mathbb{Q}$ has "holes". Let

$$A = \{r \in \mathbb{Q}\colon r > 0, r^2 < 2\}, \quad \text{and} \quad B = \{r \in \mathbb{Q}\colon r > 0, r^2 > 2\}.$$

The set A is bounded above: if $q \geq \frac{3}{2}$ then $q^2 \geq \frac{9}{4} > 2$, meaning that $q \notin A$; the contrapositive is that if $q \in A$ then $q < \frac{3}{2}$, so $\frac{3}{2}$ is an upper bound for A. In fact, take any positive rational number r; then $r^2 > 0$ is also rational. By the total order relation, exactrly one of the following three statements is true: $r^2 < 2$, $r^2 = 2$, or $r^2 > 2$. In other words, $\mathbb{Q}_{>0} = A \sqcup \{r \in \mathbb{Q}\colon r > 0, r^2 = 2\} \sqcup B$. We just showed that the middle set is empty, so

$$\mathbb{Q}_{>0} = A \sqcup B.$$

- Every element $b \in B$ is an upper bound for A. Indeed, if $a \in A$ and $b \in B$, then $a^2 < 2 < b^2$ so $0 < b^2 - a^2 = (b-a)(b+a)$, and dividing through by the positive number $b + a$ shows $b - a > 0$ so $a < b$. (This also shows that every element $a \in A$ is a lower bound for B.)
- On the other hand, if $a \in A$, then a is *not* an upper bound for A; i.e. given $a \in A$, there exists $a' \in A$ with $a < a'$. To see this, we can just take
$$a' = a + \frac{2-a^2}{2+a} = \frac{2a+2}{a+2}.$$
Since $a \in A$, we know $a^2 < 2$ so $2 - a^2 > 0$, and the denominator $2 + a > 2 > 0$, so $a' > a$. But we also have
$$2 - (a')^2 = \frac{2(a+2)^2 - (2a+2)^2}{(a+2)^2} = \frac{2a^2 + 8a + 8 - 4a^2 - 8a - 4}{(a+2)^2} = \frac{2(2-a^2)}{(a+2)^2} > 0,$$
showing that $a' \in A$, as claimed.

Thus, B is equal to the set of upper bounds of A in $\mathbb{Q}_{>0}$, and similarly A is equal to the set of lower bounds of B in $\mathbb{Q}_{>0}$.

But then we have the following strange situation. The set A of lower bounds of B has no greatest element: we just showed that, given any $a \in A$, there is an $a' \in A$ with $a' > a$. Hence, B has no greatest lower bound: $\inf B$ does not exist in $\mathbb{Q}_{>0}$. Similarly, $\sup A$ does not exists in $\mathbb{Q}_{>0}$.

Example 1.5 viscerally demonstrates that there is a "hole" in $\mathbb{Q}$: the fact that $r^2 = 2$ has no solution in $\mathbb{Q}$ forces the ordered set to be disconnected into two pieces, each of which is very incomplete: not only does each fail to possess a $\max/\min$, they also fail to possess a $\sup/\inf$.

1.2. **Lecture 2: January 7, 2016.** We now set the stage for the formal study of the real numbers: it is the (unique) *complete ordered field*. To understand these words, we begin with *fields*.

Definition 1.6. *A* **field** *is a set* $\mathbb{F}$ *equipped with two binary operations* $+, \cdot \colon \mathbb{F} \times \mathbb{F} \to \mathbb{F}$*, called* **addition** *and* **multiplication***, satisfying the following properties.*

(1) **Commutativity***:* $\forall a, b \in \mathbb{F}$*,* $a + b = b + a$ *and* $a \cdot b = b \cdot a$*.*
(2) **Associativity***:* $\forall a, b, c \in \mathbb{F}$*,* $(a + b) + c = a + (b + c)$ *and* $(a \cdot b) \cdot c = a \cdot (b \cdot c)$*.*
(3) **Identity***: there exists elements* $0, 1 \in \mathbb{F}$ *s.t.* $\forall a \in \mathbb{F}$*,* $0 + a = a = 1 \cdot a$*.*
(4) **Inverse***: for any* $a \in \mathbb{F}$*, there is an element denoted* $-a \in \mathbb{F}$ *with the property that* $a + (-a) = 0$*. For any* $a \in \mathbb{F} \setminus \{0\}$*, there is an element denoted* a^{-1} *with the property that* $a \cdot a^{-1} = 1$*.*
(5) **Distributivity***:* $\forall a, b, c \in \mathbb{F}$*,* $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$*.*

Example 1.7. Here are some examples of fields.

1. The field $\mathbb{Z}_p = \{[0], [1], \ldots, [p-1]\}$ for any prime p, where the $+$ and $\cdot$ are the usual ones inherited from the $+$ and $\cdot$ on $\mathbb{Z}$ (namely $[a] + [b] = [a + b]$ and $[a] \cdot [b] = [a \cdot b]$ – you studied this field in Math 109). All finite fields have this form.
2. $\mathbb{Q}$ is a field.
3. $\mathbb{Z}$ is *not* a field: it fails item (4), lacking multiplicative inverses of all elements other than ± 1.
4. Let $\mathbb{Q}(t)$ denote the set of rational functions of a single variable t with coefficients in $\mathbb{Q}$:

$$\mathbb{Q}(t) = \left\{ \frac{p(t)}{q(t)} \colon p(t), q(t) \text{ are polynomials with coefficients in } \mathbb{Q} \text{ and } q(t) \text{ is not identically } 0 \right\}.$$

With the usual addition and multiplication of functions, $\mathbb{Q}(t)$ is a field. For example, $\left(\frac{p(t)}{q(t)}\right)^{-1} = \frac{q(t)}{p(t)}$, which exists so long as $p(t)$ is not identically 0 – i.e. as long as the original rational function $\frac{p(t)}{q(t)}$ is not the 0 function.

Fields are the kinds of number systems that behave the way you've grown up believing numbers behave, as summarized in the following lemma.

Lemma 1.8. *Let* $\mathbb{F}$ *be a field. The following properties hold.*

(1) **Cancellation***:* $\forall a, b, c \in \mathbb{F}$*, if* $a + b = a + c$ *then* $b = c$*. If* $a \neq 0$*, if* $a \cdot b = a \cdot c$ *then* $b = c$*.*
(2) **Hungry Zero***:* $\forall a \in \mathbb{F}$*,* $0 \cdot a = 0$*.*
(3) **No Zero Divisors***:* $\forall a, b \in \mathbb{F}$*, if* $a \cdot b = 0$*, then either* $a = 0$ *or* $b = 0$*.*
(4) **Negatives***:* $\forall a, b \in \mathbb{F}$*,* $(-a)b = -(ab)$*,* $-(-a) = a$*, and* $(-a)(-b) = ab$*.*

Proof. We'll just prove (2), leaving the others to the reader. For any $a \in \mathbb{F}$, note that

$$0 \cdot a + a = 0 \cdot a + 1 \cdot a = (0 + 1) \cdot a = 1 \cdot a = a = 0 + a.$$

Hence, by (1) (cancellation), it follows that $0 \cdot a = 0$. $\square$

Example 1.9. As in Example 1.7.1, we can consider $\mathbb{Z}_n$ for any positive integer n. This satisfies all of the properties of Definition 1.6 except (4): inverses don't always exist. For example, if n can be factored as $n = km$ for two positive integers $k, m > 1$, then we have two nonzero elements $[k], [m] \in \mathbb{Z}_n$ such that $[k] \cdot [m] = [km] = [n] = [0]$, which contradicts Lemma 1.8(3) – there are zero divisors. So $\mathbb{Z}_n$ is not a field when n is composite.

Now, we combine fields with ordered sets.

Definition 1.10. *An* **ordered field** *is a field $\mathbb{F}$ which is an ordered set $(\mathbb{F}, <)$, where the order relation also satisfies the following two properties:*

(1) $\forall a, b, c \in \mathbb{F}$, *if* $a < b$ *then* $a + c < b + c$.
(2) $\forall a, b \in \mathbb{F}$, *if* $a > 0$ *and* $b > 0$, *then* $a \cdot b > 0$.

From here, all the usual properties mixing the order relation and the field operations follow. For example:

Lemma 1.11. *Let $(\mathbb{F}, <)$ be an ordered field. Then*

(1) $\forall a \in \mathbb{F}$, $a > 0$ *iff* $-a < 0$.
(2) $\forall a \in \mathbb{F} \setminus \{0\}$, $a^2 > 0$. *In particular,* $1 = 1^2 > 0$.
(3) $\forall a, b \in \mathbb{F}$, *if* $a > 0$ *and* $b < 0$, *then* $a \cdot b < 0$.
(4) $\forall a \in \mathbb{F}$, *if* $a > 0$ *then* $a^{-1} > 0$.

Proof. For (1), simply add $-a$ to both sides of the inequality. Note, by the properties of $<$, this means $\mathbb{F}$ is the union of three disjoint subsets: the *positive* elements $a > 0$, the *negative* elements $a < 0$, and the zero element $a = 0$; and the operation of multiplication by -1 interchanges the positive and negative elements. So, for (2), we note that our given $a \neq 0$ must be either positive or negative; if $a > 0$ then $a^2 = a \cdot a > 0$ by Definition 1.10(2), while if $a < 0$ then $a^2 = (-a)^2 > 0$ by the same argument. For (3), we then have $a > 0$ and $-b > 0$, so $-(ab) = a \cdot (-b) > 0$, which means that $ab < 0$. Finally, for (4), suppose $a^{-1} < 0$. then by (3) we would have $1 = a \cdot a^{-1} < 0$; but by (2) we know $1 > 0$. This contradiction shows that $a^{-1} > 0$. □

Example 1.12.

1. $\mathbb{Q}$ is an ordered field, with its usual order: $\frac{m_1}{n_1} < \frac{m_2}{n_2}$ iff $m_1 n_2 < m_2 n_1$. In fact, this is the *unique* total order on the set $\mathbb{Q}$ which makes $\mathbb{Q}$ into an ordered field.
2. $\mathbb{Z}_p$ is not an ordered field for any prime p. For suppose it were; then by Lemma 1.11(2) we know that $[1] > [0]$. Then $[2] = [1] + [1] > [1] + [0] = [1]$, and so by transitivity $[2] > [0]$. Continuing this way by induction, we get to $[p-1] > [0]$. But we also have $[0] = [1] + [p-1] > [0] + [p-1] = [p-1]$. This is a contradiction.
3. Let $\mathbb{F}$ be an ordered field. Denote by $\mathbb{F}_c$ the following set of 2×2 matrices over $\mathbb{F}$:
$$\mathbb{F}_c = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} : a, b \in \mathbb{F} \right\}.$$
The determinant of such a matrix is $a^2 + b^2$. In an ordered field, we know that $a^2 > 0$ if $a \neq 0$, and thus we have the usual property that $a^2 + b^2 = 0$ iff $a = b = 0$. It follows that all nonzero matrices in $\mathbb{F}_c$ are invertible: we can easily verify that
$$(a^2 + b^2)^{-1} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$
If we define
$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$
then $\mathbb{F}_c = \{aI + bJ \colon a, b \in \mathbb{F}\}$. Note that $J^2 = -I$. It is now an easy exercise to show that $\mathbb{F}_c$ is a field, with $+, \cdot$ being given by matrix addition and multiplication, where I is the multiplicative identity and the additive identity is the 2×2 zero matrix. (Note: this is *not generally true* if $\mathbb{F}$ is not an ordered field. For example, in $\mathbb{Z}_2$ we have $1^2 + (-1)^2 = 0$, and as a result the matrix with $a = b = 1$ is not invertible in this case.) $\mathbb{F}_c$ is the *complexification* of $\mathbb{F}$. We will later construct the complex numbers $\mathbb{C}$ as $\mathbb{C} = \mathbb{R}_c$.

3.5 If $\mathbb{F}$ is any ordered field, then $\mathbb{F}_c$ cannot be ordered – there is no order relation that makes $\mathbb{F}_c$ into an ordered field. This is actually what Problem 4 on HW1 asks you to prove.

Item 2 above noted that the finite fields $\mathbb{Z}_p$ are not ordered fields. In fact, ordered fields must be infinite. The next results shows why this is true.

Lemma 1.13. *Let $(\mathbb{F}, <)$ be an ordered field. Then, for any $n \in \mathbb{Z} \setminus \{0\}$, $n \cdot 1_\mathbb{F} \neq 0_\mathbb{F}$.*

Here $n \cdot 1_\mathbb{F} = 1_\mathbb{F} + 1_\mathbb{F} + \cdots + 1_\mathbb{F}$. Note that this property is *not* automatic for fields: for example, in $\mathbb{Z}_p$, $p \cdot [1] = [0]$.

Proof. First, $1 \cdot 1_\mathbb{F} = 1_\mathbb{F} > 0_\mathbb{F}$ by Lemma 1.11(2). Proceeding by induction, suppose we've shown that $n \cdot 1_\mathbb{F} \neq 0_\mathbb{F}$. Then $(n+1) \cdot 1_\mathbb{F} = n \cdot 1_\mathbb{F} + 1_\mathbb{F} > 0 + 1_\mathbb{F} = 1_\mathbb{F} > 0_\mathbb{F}$. Thus, for every $n > 0$, $n \cdot 1_\mathbb{F} > 0_\mathbb{F}$, meaning it is $\neq 0$. If, on the other hand, $n < 0$ in $\mathbb{Z}$, then $n \cdot 1_\mathbb{F} = -(-n \cdot 1_\mathbb{F}) < 0_\mathbb{F}$, so also it is $\neq 0_\mathbb{F}$. □

Corollary 1.14. *Let $\mathbb{F}$ be an ordered field. The map $\varphi\colon \mathbb{Q} \to \mathbb{F}$ given by $\varphi(\frac{m}{n}) = (m \cdot 1_\mathbb{F}) \cdot (n \cdot 1_\mathbb{F})^{-1}$ is an injective ordered field homomorphism.*

An **ordered field homomorphism** is a function which preserves the field operations: $\varphi(a+b) = \varphi(a) + \varphi(b)$, $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$, and $\varphi(0) = 0$ and $\varphi(1) = 1$; and preserves the order relation: if $a < b$ then $\varphi(a) < \varphi(b)$. An injective ordered field homomorphism should be thought of as an **embedding**: we realize $\mathbb{Q}$ as a subset of $\mathbb{F}$, in a way that respects all the ordered field structure.

Proof. First we must check that φ is well defined: if $\frac{m_1}{n_1} = \frac{m_2}{n_2}$, then $m_1 n_2 = m_2 n_1$. It then follows (by an easy induction) that $(m_1 \cdot 1_\mathbb{F}) \cdot (n_2 \cdot 1_\mathbb{F}) = (m_2 \cdot 1_\mathbb{F}) \cdot (n_1 \cdot 1_\mathbb{F})$. Dividing out on both sides then shows that $(m_1 \cdot 1_\mathbb{F})(n_1 \cdot 1_\mathbb{F})^{-1} = (m_2 \cdot 1_\mathbb{F}) \cdot (n_2 \cdot 1_\mathbb{F})^{-1}$. Thus, φ is well-defined. It is similar and routine to verify that it is an ordered field homomorphism. Finally, to show it is one-to-one, suppose that $\varphi(q_1) = \varphi(q_2)$ for $q_1, q_2 \in \mathbb{Q}$. Using the homomorphism property, this means $\varphi(q_1 - q_2) = \varphi(q_1) - \varphi(q_2) = 0$. Let $q_1 - q_2 = \frac{m}{n}$; thus, we have $\varphi(\frac{m}{n}) = (m \cdot 1_\mathbb{F}) \cdot (n \cdot 1_\mathbb{F})^{-1} = 0_\mathbb{F}$. But then, multiplying through by the non-zero (by Lemma 1.13) element $n \cdot 1_\mathbb{F}$, we have $m \cdot 1_\mathbb{F} = 0_\mathbb{F}$, and again by Lemma 1.13, it follows that $m = 0$. but this means $q_1 - q_2 = \frac{m}{n} = 0$, so $q_1 = q_2$. Thus, φ is injective. □

Thus, we will from now on think if $\mathbb{Q}$ as a subset of any ordered field.

In Lecture 1, we saw that $\mathbb{Q}$ "has holes". In example 1.5, we found two subsets $A, B \subset \mathbb{Q}$ with the property that $B =$ the set of upper bounds of A, $A =$ the set of lower bounds of B, and A has no maximal element, while B has no minimal element. Thus, $\sup A$ and $\inf B$ do not exist. This turns out to be a serious obstacle to doing the kind of analysis we're used to in calculus, so we'd like to fill in these holes. This motivates our next definition.

1.3. **Lecture 3: January 11, 2016.**

Definition 1.15. *An ordered set* $(S,<)$ *is called* **complete** *if every nonempty subset* $\varnothing \neq E \subseteq S$ *that is bounded above possesses a supremum* $\sup E \in S$*. We also denote this by saying that* $(S,<)$ *has the* **least upper bound property**.

We could also formulate things in terms of $\inf$, with the greatest lower bound property. Example 1.5 demonstrates how these two are typically related. In fact, they are equivalent.

Proposition 1.16. *An ordered set* $(S,<)$ *has the least upper bound property if and only if, for every nonempty subset* $\varnothing \neq F \subseteq S$ *that is bounded below,* $\inf F \in S$ *exists.*

Proof. We will argue the forward implication: the least upper bound property implies the greatest lower bound property. The converse is very similar.

Let $F \neq \varnothing$ be bounded below; then $L \equiv \{\text{lower bounds for } F\}$ is a nonempty subset of S. If $x \in L$ and $y \in F$, then $x \leq y$, which shows that every $y \in F$ is an upper bound for L. Thus, L is bounded above and nonempty; by the least upper bound property of S, $\sigma = \sup L \in S$ exists. By definition of supremum, if $x < \sigma$ then x is not an upper bound for L; since every element of F *is* an upper bound for L, this means that such x is not in F. Taking contrapositives, this says that if $z \in F$ then $x \geq \sigma$. So σ is a lower bound for F – i.e. $\sigma \in L$. This shows that $\sigma = \max L$: i.e. σ is the greatest lower bound of F: $\sigma = \inf F$. So $\inf F$ exists, as claimed. $\square$

Let us now prove some important properties that complete ordered fields possess – properties that are critical for doing all of analysis.

Theorem 1.17. *Let* $\mathbb{F}$ *be a complete ordered field.*

(1) *(***Archimedean***) Let* $x, y \in \mathbb{F}$ *with* $x > 0$*. Then there exists* $n \in \mathbb{N}$ *so that* $nx > y$*.*
(2) *(***Density of** $\mathbb{Q}$*) Let* $x, y \in \mathbb{F}$*, with* $x < y$*. Then there exists* $r \in \mathbb{Q}$ *so that* $x < r < y$*.*

A field with property (1) is called **Archimedean**. It tells us (by setting $x = 1$) that the set $\mathbb{N}$ is *not bounded above* in the field: there is no $y \in \mathbb{F}$ that is $\geq$ every integer. It also tells us (by setting $y = 1$) that there are no "infinitesimals" – that is, no matter how small a positive number x is, there is always a positive integer n such that $0 < \frac{1}{n} < x$. This is an absolutely crucial property for a field to have if we want to talk about limits. And it does *not* hold in every ordered field.

Example 1.18. In the field $\mathbb{Q}(t)$ of rational functions with rational coefficients, it is always possible to uniquely express a function $f(t) \in \mathbb{Q}(t)$ in the form $f(t) = \lambda \cdot \frac{p(t)}{q(t)}$ where $\lambda \in \mathbb{Q}$ and $p(t), q(t)$ are *monic* polynomials: their highest order terms have coefficient 1. This allows us to define an order on $\mathbb{Q}(t)$: say $f(t) < g(t)$ iff $g(t) - f(t) = \lambda \frac{p(t)}{q(t)}$ where $p(t), q(t)$ are monic and $\lambda > 0$. (This is the same as insisting that the leading coefficients of the numerator and denominator of $f(t) - g(t)$ have the same sign.) For example $\frac{t^2-25t+7}{t^4-10^{23}} > 0$ while $\frac{-t^2-25t+7}{t^4-10^{23}} < 0$. Then it is easy but laborious to check that this makes $\mathbb{Q}(t)$ into an ordered field. Note: $t - n = 1 \cdot \frac{t-n}{1} > 0$ for any integer n; this means that, in the ordered field $\mathbb{Q}(t)$, the element t is *greater than every integer*. I.e. the set $\mathbb{Z} \subset \mathbb{Q}(t)$ actually has an upper bound (e.g. t) in $\mathbb{Q}(t)$. This means $\mathbb{Q}(t)$ is a non-Archimedean field. In particular, by Theorem 1.17, $\mathbb{Q}(t)$ is not a complete ordered field.

Proof of Theorem 1.17. (1) Suppose, for a contradiction, there there is no such n: that is, $nx \leq y$ for every $n \in \mathbb{N}$. Let $E = \{nx \colon n \in \mathbb{N}\}$. Then our assumption is that y is an upper bound for E, so E is bounded above. It is also non-empty (it contains x, for example). Thus, since $\mathbb{F}$ is complete, it follows that $\alpha = \sup E$ exists. In particular, since $\alpha - x < \alpha$, this means that $\alpha - x$ is not an upper bound for E, so there is some element $e \in E$ with $\alpha - x < e$. There is some integer $m \in \mathbb{N}$ so that $e = mx$, so we have $\alpha - x < mx$. But then $\alpha < (m+1)x$, and $(m+1)x \in E$. This contradicts $\alpha = \sup E$ being an upper bound. This contradiction proves the claim.

(2) Since $y - x > 0$, by (1) there is an $n \in \mathbb{N}$ so that $n(y-x) > 1$. Now, letting $y = \pm nx$ and applying (1) again, we can find two positive integers $m_1, m_2 \in \mathbb{N}$ so that $m_1 > nx$ and $m_2 > -nx$; in other words

$$-m_2 < nx < m_1.$$

This shows that the set $\{k \in \mathbb{Z} \colon nx < k \le m_1\}$ is finite: it is contained in the finite set $\{-m_2 + 1, -m_2 + 2, \ldots, m_1\}$. So, let $m = \min\{k \in \mathbb{Z} \colon nx < k\}$. Then since $m - 1 \in \mathbb{Z}$ and $m - 1 < m$, we must have $m - 1 \le nx$.

Thus, we have two inequalities:

$$n(y-x) > 1, \qquad m - 1 \le nx < m.$$

Combining these gives us

$$nx < m \le nx + 1 < ny.$$

Dividing through by (the positive) n shows that $x < \frac{m}{n} < y$, so setting $r = \frac{m}{n}$ completes the proof. □

Here is another extremely important property that holds in ordered fields; this is crucial for doing calculus.

Proposition 1.19. *Let $\mathbb{F}$ be a complete ordered field. For each $n \in \mathbb{N}$, let $a_n, b_n \in \mathbb{F}$ satisfy*

$$a_1 \le a_2 \le \cdots \le a_n \le \cdots \le b_n \le \cdots \le b_2 \le b_1.$$

Further, suppose that $b_n - a_n < \frac{1}{n}$. Then $\bigcap_{n \in \mathbb{N}} [a_n, b_n]$ is nonempty, and consists of exactly one point.

This is sometimes called the *nested intervals property*. It is actually equivalent to the least upper bound property. On HW2, you will prove the converse.

Proof. By construction, b_1 is an upper bound for $\{a_n \colon n \in \mathbb{N}\}$, which is a nonempty set. Thus, by completeness, $\alpha = \sup a_n$ exists in $\mathbb{F}$. Since α is an upper bound for $\{a_n\}$, we have $a_n \le \alpha$ for every n. On the other hand, since $b_m \ge a_n$ for every m, n, b_m is an upper bound for $\{a_n\}$, and since α is the *least* upper bound, it follows that $\alpha \le b_m$ as well. Thus $\alpha \in [a_n, b_n]$ for every n, and so it is in the intersection.

Now, suppose $\beta \in \bigcap_n [a_n, b_n]$. Then either $\alpha < \beta$, $\alpha > \beta$, or $\alpha = \beta$. Suppose, for the moment, that $\alpha < \beta$. Then we have $a_n \le \alpha < \beta \le b_n$ for every n, and since $b_n - a_n < \frac{1}{n}$, it follows that $0 < \beta - \alpha < \frac{1}{n}$ for every n. But this violates the Archimedean property of $\mathbb{F}$. A similar contradiction arises if we assume $\alpha > \beta$. Thus $\alpha = \beta$, and so α is the unique element of the intersection. □

Note: in the setup of the lemma, it is similar to see that the intersection consists of $\inf_n b_n$; so $\sup_n a_n = \inf_n b_n$.

1.4. **Lecture 4: January 14, 2014.** We have now seen several properties possessed by complete ordered fields. We would hope to find some examples as well. Here comes the big punchline.

Theorem 1.20. *There exists exactly one complete ordered field. We call this field $\mathbb{R}$, the Real numbers.*

We will talk about the proof of Theorem 1.20 as we proceed in the course. The textbook relegates an existence proof to the end of Chapter 1, through **Dedekind cuts**. This is an old-fashioned proof, and not very intuitive. We are not going to discuss it presently. Once we have developed a little more technology, we will prove the existence claim of the theorem using Cauchy's construction of $\mathbb{R}$ (through sequences).

We can, however, prove the uniqueness claim. To be precise, here is what uniqueness means in this case: suppose $\mathbb{F}$ and $\mathbb{G}$ are two complete ordered fields. Then there exists an ordered field *isomorphism* $\varphi\colon \mathbb{F} \to \mathbb{G}$. That means φ is an ordered field homomorphism that is also a bijection. So, from the point of view of ordered fields, $\mathbb{F}$ and $\mathbb{G}$ are indistinguishable.

The first question is: given two complete ordered fields $\mathbb{F}$ and $\mathbb{G}$, how do we define $\varphi\colon \mathbb{F} \to \mathbb{G}$? By Corollary 1.14, $\mathbb{Q}$ embeds in each of $\mathbb{F}$ and $\mathbb{G}$ via $\mathbb{Q}\cdot 1_{\mathbb{F}}$ and $\mathbb{Q}\cdot 1_{\mathbb{G}}$. So we can define φ as a *partial* function by its action on $\mathbb{Q}$:

$$\varphi(r1_{\mathbb{F}}) = r1_{\mathbb{G}}, \qquad r \in \mathbb{Q}.$$

The question is: how should we define φ on elements of $\mathbb{F}$ that are not necessarily in $\mathbb{Q}\cdot 1_{\mathbb{F}}$? Well, let $x \in \mathbb{F}\setminus\mathbb{Q}$. By Theorem 1.17(2), there are rationals $a_n, b_n \in \mathbb{Q}$ such that

$$x - \frac{1}{2n}1_{\mathbb{F}} < a_n 1_{\mathbb{F}} < x < b_n 1_{\mathbb{F}} < x + \frac{1}{2n}1_{\mathbb{F}}.$$

In particular, $b_n - a_n < \frac{1}{n}$. We should do this carefully and also make sure that $a_1 \le a_2 \le \cdots \le b_2 \le b_1$ – this can be achieved by choosing the a_n and b_n successively, increasing the a_n or decreasing the b_n each step as needed. It follows from Proposition 1.19 that $\bigcap_n [a_n 1_{\mathbb{G}}, b_n 1_{\mathbb{G}}]$ contains exactly one point, $\alpha = \sup_n(a_n 1_{\mathbb{G}}) = \inf_n(b_n 1_{\mathbb{G}})$. So we define

$$\varphi(x) = \alpha.$$

Note: if $x \in \mathbb{Q}$, then $x \cdot \mathbb{1}_{\mathbb{G}}$ is the unique element in the intersection, meaning that we can take the above nested intervals definition as the formula for φ on all of $\mathbb{F}$, not just the irrational elements. This will be our starting point.

Theorem 1.21. *If $\mathbb{F}$ and $\mathbb{G}$ are two complete ordered fields, then there exists an ordered field isomorphism $\varphi\colon \mathbb{F} \to \mathbb{G}$.*

Proof. Following our outline from above, we define φ as follows. To begin, using the denseness of $\mathbb{Q}$ in $\mathbb{F}$, select $a_1, b_1 \in \mathbb{Q}$ so that

$$x - \frac{1}{2}1_{\mathbb{F}} < a_1 1_{\mathbb{F}} < x < b_1 1_{\mathbb{F}} < x + \frac{1}{2}1_{\mathbb{F}}.$$

Now proceed inductively: once we've constructed $a_1, \ldots, a_{n-1}$ and $b_1, \ldots, b_{n-1}$, choose a_n and b_n so that

$$\max\left\{x - \frac{1}{2n}1_{\mathbb{F}}, a_{n-1}\right\} < a_n 1_{\mathbb{F}} < x < b_n 1_{\mathbb{F}} < \min\left\{x + \frac{1}{2n}1_{\mathbb{F}}, b_{n-1}\right\}. \tag{1.1}$$

Then we have $a_1 < a_2 < \cdots < a_n < \cdots < b_n < \cdots < b_2 < b_1$, and also

$$b_n - a_n < \left(x + \frac{1}{2n}1_{\mathbb{F}}\right) - \left(x - \frac{1}{2n}1_{\mathbb{F}}\right) = \frac{1}{n}.$$

So by the nested intervals property Proposition 1.19 applied in the field $\mathbb{G}$, we have

$$\bigcap_{n\in\mathbb{N}}[a_n 1_{\mathbb{G}}, b_n 1_{\mathbb{G}}] = \{\alpha\}$$

where $\alpha = \sup_n a_n 1_{\mathbb{G}} = \inf_n b_n 1_{\mathbb{G}}$. We thus define $\varphi(x) = \alpha$.

Now we must verify that:

- φ **is well-defined**: if a_n', b_n' are some other rational elements satisfying (1.1) then $\sup_n a_n 1_{\mathbb{G}} = \sup_n a_n' 1_{\mathbb{G}}$. In fact, this follows because we also then have the mixed inequalities
$$x - \frac{1}{2n}1_{\mathbb{F}} < a_n' 1_{\mathbb{F}} < x < b_n 1_{\mathbb{F}} < x + \frac{1}{2n}1_{\mathbb{F}}$$
and, as above, we have $\sup_n a_n' 1_{\mathbb{G}} = \inf_n b_n 1_{\mathbb{G}} = \sup_n a_n 1_{\mathbb{G}}$.
- φ **is an ordered field homomorphism**. This is laborious. Let's check one of the field homomorphism properties: preservation of addition. Let $x, y \in \mathbb{F}$, and let $a_n < x < b_n$ and $c_n < y < d_n$ where $b_n - a_n < \frac{1}{2n} < \frac{1}{n}$ and $d_n - c_n < \frac{1}{2n} < \frac{1}{n}$. Then $\varphi(x) = \sup_n a_n$ and $\varphi(y) = \sup_n c_n$. Now, on the other hand, we have
$$a_n + c_n < x + y < b_n + d_n, \quad \text{and} \quad (b_n + d_n) - (a_n + c_n) = (b_n - a_n) + (d_n - c_n) < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}.$$
It follows that $\varphi(x+y) = \sup(a_n + c_n)$. So, to see that $\varphi(x+y) = \varphi(x) + \varphi(y)$, it suffices to show that
$$\text{if } a_n \uparrow \ \& \ c_n \uparrow \quad \text{then } \sup_n (a_n + c_n) = \sup_n a_n + \sup_n c_n.$$
This is also on HW2. The other ordered field homomorphism properties are verified similarly.
- φ **is a bijection**. First, suppose that $x \neq y \in \mathbb{F}$. Then either $x < y$ or $x > y$; wlog $x < y$. Since φ is an ordered field homomorphism, it follows that $\varphi(x) < \varphi(y)$. In particular, $\varphi(x) \neq \varphi(y)$. A similar argument in the case $x > y$ shows that φ is one-to-one.

  Now, fix $y \in \mathbb{G}$. For each n, choose $a_n, b_n \in \mathbb{Q}$ nested so that $b_n - a_n < \frac{1}{n}$ and $a_n 1_{\mathbb{G}} < y < b_n 1_{\mathbb{G}}$. Mirroring the above arguments, we know that $a = \sup_n a_n 1_{\mathbb{F}} \in \bigcap_n [a_n 1_{\mathbb{F}}, b_n 1_{\mathbb{F}}]$. Since $a_n 1_{\mathbb{F}} < a < b_n 1_{\mathbb{F}}$, we have $a_n 1_{\mathbb{G}} = \varphi(a_n 1_{\mathbb{F}}) < \varphi(a) < \varphi(b_n 1_{\mathbb{F}}) = b_n 1_{\mathbb{G}}$. Thus $\varphi(a) \in \bigcap_n [a_n 1_{\mathbb{G}}, b_n 1_{\mathbb{G}}]$, and this intersection consists of the singleton element y, by Proposition 1.19. Hence, $\varphi(a) = y$, and so φ is onto.

□

So, we see that there can be *only one* complete ordered field. (They're like Highlanders.) A priori, that doesn't preclude the possibility that there aren't any at all. To prove that $\mathbb{R}$ exists, we need to first start talking about convergence properties of sequences. That will be our next task.

Before proceeding, let's return to our motivation for studying $\sup$ and $\inf$ and introducing completeness: we wanted to fill the "hole" in $\mathbb{Q}$ where $\sqrt{2}$ should be. To see that we've filled at least that hole, the next result shows that $\mathbb{R}$ (the complete ordered field) contains square roots, and in fact nth roots, of all positive numbers. First, let's state some standard results on "absolute value".

Lemma 1.22. *Let $\mathbb{F}$ be an ordered field. For $x \in \mathbb{F}$, define (as usual)*

$$|x| = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0 \end{cases}.$$

Then we have the following properties.

(1) *For all* $x \in \mathbb{F}$, $|x| \geq 0$, *and* $|x| = 0$ *iff* $x = 0$.
(2) *For all* $x, y \in \mathbb{F}$, $|x + y| \leq |x| + |y|$.
(3) *For all* $x, y \in \mathbb{F}$, $|xy| = |x||y|$.

All of these properties are straightforward but annoying to prove in cases. We will use the absolute value frequently in all that follows.

Theorem 1.23. *Let* $n \in \mathbb{N}$, $n \geq 1$. *For any* $x \in \mathbb{R}$, $x > 0$, *there is a unique* $y \in \mathbb{R}$, $y > 0$, *so that* $y^n = x$. *We denote it by* $y = x^{1/n}$.

Proof of Theorem 1.23. First, for uniqueness: let $y_1 \neq y_2$ be two positive real numbers, wlog $y_1 < y_2$. Then $y_1^2 = y_1 y_1 < y_1 y_2 < y_2 y_2 = y_2^2$; continuing by induction, we see that $y_1^n < y_2^n$. That is: the function $y \mapsto y^n$ is strictly increasing. In particular, it is one-to-one. It follows that there can be at most one y with $y^n = x$.

Now for existence. Let $E = \{y \in \mathbb{R} \colon y > 0, y^n < x\}$.

- $E \neq \varnothing$: note that $t = \frac{x}{x+1} \in (0, 1)$. This means that $0 < t^n < t$, and so since $\frac{x}{x+1} < x$, we have $0 < t^n < x$, meaning that $t \in E$.
- E is bounded above: let $s = 1 + x$. Then $s > 1$, and so $s^n > s > x$. Thus, if $y \in E$, then $y^n < x < s^n$, and so $0 < s^n - y^n = (s - y)(s^{n-1} + s^{n-2}y + \cdots + y^{n-1})$. The sum of terms is strictly positive, so we can divide out and find that $s - y > 0$. Thus s is an upper bound for E.

Hence, by completeness of $\mathbb{R}$, $\alpha = \sup E$ exists. Since α is the least upper bound, it follows that, for each k, there is an element $y_k \in E$ such that $y_k > \alpha - \frac{1}{k}$. Since $y_k^n < x$, we therefore have

$$\left(\alpha - \frac{1}{k}\right)^n < y_k^n < x, \qquad \text{for all } k \in \mathbb{N}.$$

But we can expand

$$\left(\alpha - \frac{1}{k}\right)^n = \sum_{j=0}^{n} \binom{n}{j} \alpha^{n-j} \left(-\frac{1}{k}\right)^{-j} = \alpha^n - \frac{1}{k} \sum_{j=1}^{n} \binom{n}{j} \alpha^{n-j} \left(-\frac{1}{k}\right)^{j-1}.$$

Thus, we have

$$\alpha^n < x + \frac{1}{k} \sum_{j=1}^{n} \binom{n}{j} \alpha^{n-j} \left(-\frac{1}{k}\right)^{j-1}$$

and so, applying the triangle inequality – Lemma 1.22(2) – repeatedly, we have

$$\alpha^n < x + \frac{1}{k} \left| \sum_{j=1}^{n} \binom{n}{j} \alpha^{n-j} \left(-\frac{1}{k}\right)^{j-1} \right| \leq x + \frac{1}{k} \cdot \sum_{j=1}^{n} \binom{n}{j} \alpha^{n-k} \left(\frac{1}{k}\right)^{j-1}.$$

Note that n is fixed, and $\frac{1}{k} \leq 1$, so for $k \geq 1$ we have $\left(\frac{1}{k}\right)^{j-1} \leq 1$. Let $M = \sum_{k=1}^{n} \binom{n}{j} \alpha^{n-k}$; then we have

$$\forall k \in \mathbb{N} \quad \alpha^n < x + \frac{M}{k}; \quad \text{i.e.} \quad \alpha^n - x < \frac{M}{k}.$$

By the Archimedean property, it follows that $\alpha^n - x \leq 0$; thus, we have shown that $\alpha^n \leq x$.

On the other hand, let $y \in E$. Then for any $k \in \mathbb{N}$ we have, by similar calculations,

$$\left(y + \frac{1}{k}\right)^n = y^n + \frac{1}{k} \sum_{j=1}^{n} \binom{n}{k} y^{n-j} \left(\frac{1}{k}\right)^{j-1} \leq y^n + \frac{1}{k} \cdot \sum_{j=1}^{n} \binom{n}{j} y^{n-j}.$$

Since $y \in E$, we know $y^n < x$, so $\epsilon = x - y^n > 0$. Let $L = \sum_{j=1}^n \binom{n}{j} y^{n-j}$, which is a positive constant; by the Archimedean property, there is some $k \in \mathbb{N}$ so that $\frac{1}{k} \cdot L < \epsilon$. Thus, for such k,

$$\left(y + \frac{1}{k}\right)^n \le y^n + \frac{L}{k} < y^n + \epsilon = x.$$

That is: $y + \frac{1}{k} \in E$. But $y + \frac{1}{k} > y$. That is, for any $y \in E$, there is $y' > y$ with $y \in E$. So E has no maximal element. This shows that $\alpha \notin E$, and hence $\alpha^n \ge x$.

In conclusion: we've shown that $\alpha^n \le x$ and $x \le \alpha^n$. It follows that $\alpha^n = x$. □

On Homework 2, you will flesh out extending this argument to defining x^r for $x > 0$ in $\mathbb{R}$ and $r \in \mathbb{Q}$, and then extending this further to define x^y for $x > 0$ and $y \in \mathbb{R}$. One can use similar arguments to define $\log_b(x)$ for $x, b > 0$. We will wait a little while until we have a firm grounding in sequences and limits before rigorously developing the calculus of these well-known functions.

2. Sequences and Limits

2.1. Lecture 5: January 19, 2016.

Definition 2.1. *Let X be a set. A* **sequence** *in X is a function $a\colon \mathbb{N} \to X$. Instead of the usual notation $a(n)$ for the value of the function at $n \in \mathbb{N}$, we usually use the notation $a_n = a(n)$; accordingly, we often refer to the function as $(a_n)_{n\in\mathbb{N}}$ or $\{a_n\}_{n\in\mathbb{N}}$, or (when being sloppy) simply (a_n) or $\{a_n\}$.*

In ordered fields, we can talk about limits of sequences. The following definition took half a century to finalize; its invention (by Weierstraß) is one of the greatest achievements of analysis.

Definition 2.2. *Let $\mathbb{F}$ be an ordered field, and let (a_n) be a sequence in $\mathbb{F}$. Let $a \in \mathbb{F}$. Say that a_n* **converges to** *a, written $a_n \to a$ or $\lim_{n\to\infty} a_n = a$, if the following holds true:*

$$\forall \epsilon > 0\ \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N\ |a_n - a| < \epsilon.$$

Let's decode the three-quantifier sentence here. What this say is, no matter how small a tolerance $\epsilon > 0$ you want, there is some time N after which all the terms a_n (for $n \geq N$) are within ϵ of a. Some convenient language for this is:

Given any $\epsilon > 0$, we have $|a_n - a| < \epsilon$ for **almost all** n.

Here we colloquially say that a set $S \subseteq \mathbb{N}$ contains **almost all** positive integers if the complement $\mathbb{N} \setminus S$ is finite. This is equivalent to saying that, after some N, all $n \geq N$ are in S. So, the limit definition is that, for any positive tolerance, no matter how small, almost all of the terms are within that tolerance of the limit.

If (a_n) is a sequence and there exists a so that $a_n \to a$, we say that (a_n) **converges**; if there is no such a, we say that (a_n) **diverges**. Here are some examples.

Example 2.3. Consider each of the following sequences in an Archimedean field.

(1) $a_n = 1$ converges to 1. More generally, if (a_n) is equal to a constant a for almost all n, then $a_n \to a$.
(2) $a_n = \frac{1}{n}$ converges to 0.
(3) $a_n = n + \frac{1}{n}$ diverges.
(4) $a_n = (-1)^n$ diverges.
(5) $a_n = 1 + \frac{1}{n}(-1)^n$ converges to 1.
(6) $a_n = \frac{4n+1}{7n-4}$ (defined for $n \geq 1$) converges to $\frac{4}{7}$.

In all these examples, we proved convergence (when the sequences converged) to a given value. However, a priori, it is not clear whether it might also have been possible to prove convergence to a *different* value as well. This is not the case: limits are unique.

Lemma 2.4. *Let $\mathbb{F}$ be an ordered field, and let (a_n) be a sequence in $\mathbb{F}$. Suppose $a, b \in \mathbb{F}$ and $a_n \to a$ and $a_n \to b$. Then $a = b$.*

Proof. Fix $\epsilon > 0$. We know that there is N_1 so that $|a_n - a| < \frac{\epsilon}{2}$ for all $n > N_1$, and there is N_2 so that $|a_n - b| < \frac{\epsilon}{2}$ for all $n > N_2$. Thus, for any $n > \max\{N_1, N_2\}$, we have

$$|a - b| = |a - a_n + a_n - b| \leq |a - a_n| + |a_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Now, suppose that $a \neq b$. Thus $a - b \neq 0$, which means that $|a-b| > 0$. So we can take $\epsilon = |a-b|$ above, and we find that $|a - b| < |a - b|$ – a contradiction. Hence, it must be true that $a = b$. $\square$

Remark 2.5. Note, in an *Archimedean* field, we are free to restrict $\epsilon = \frac{1}{k}$ for some $k \in \mathbb{N}$; that is, an equivalent statement of $a_n \to a$ is

$$\text{Given any } k \in \mathbb{N}, \text{ we have } |a_n - a| < \tfrac{1}{k} \text{ for almost all } n.$$

In non-Archimedean fields, this does not suffice. For example, in the field $\mathbb{Q}(t)$, to show $a_n(t) \to a(t)$ it does not suffice to show that, for any $k \in \mathbb{N}$, $|a_n(t) - a(t)| < \frac{1}{k}$ for all sufficiently large n. Indeed, what if $a_n(t) - a(t) = \frac{1}{t}$? This does not go to 0, but it is $< \frac{1}{k}$ for all $k \in \mathbb{N}$. Similarly, the sequence $a_n = \frac{1}{n}$ diverges in a non-Archimedean field.

2.2. **Lecture 6: January 21, 2016.**

Proposition 2.6. *Let $\mathbb{F}$ be a complete ordered field. Let (a_n) be a sequence in $\mathbb{F}$, and suppose $a_n \uparrow$ (i.e. $a_n \le a_{n+1}$ for all n) and bounded above. Let $\alpha = \sup\{a_n\}$. Then $a_n \to \alpha$. Similarly, if $b_n \downarrow$ and bounded below, then $\beta = \inf\{b_n\}$ exists and $b_n \to \beta$.*

Proof. Since $\mathbb{F}$ is a complete field, $\alpha = \sup\{a_n\}$ exists in $\mathbb{F}$. Let $\epsilon > 0$. Then $\alpha - \epsilon < \alpha$, and so by definition there exists some element $a_N \in \{a_n\}$ so that $\alpha - \epsilon < a_N \le \alpha$. Now, suppose $n \ge N$; then $a_n \le \alpha$ of course, but also since $a_n \uparrow$ we have $a_n \ge a_N > \alpha - \epsilon$. Thus, we have shown that $|a_n - \alpha| = \alpha - a_n < \epsilon$ for all $n \ge N$, which is to say that $a_n \to \alpha$.

The decreasing case is similar; alternatively, one can look at $a_n = -b_n$, which is increasing and bounded above; then we have by the first part that $-b_n = a_n \to \alpha$ where $\alpha = \sup\{-b_n\} = -\inf\{a_n\} = -\beta$. It follows that $b_n \to -\beta$, using the limit theorems below. □

In the proposition, we needed (a_n) to be bounded (above or below); indeed, the sequence $a_n = n$ is increasing, but not convergent. This is generally true: for any sequence to be convergent, it must be bounded (above and below). A sequence that is either increasing or decreasing is called **monotone**. So the proposition shows that monotone sequences either converge, or grow (in absolute value) without bound.

This gives us a new perspective on the motivating example that began our discussion of $\sup$ and $\inf$. Consider, again, the sets $A = \{r \in \mathbb{Q} \colon r > 0, r^2 < 2\}$ and $B = \{r \in \mathbb{Q} \colon r > 0, r^2 > 2\}$. We saw that the set of positive rationals is equal to $A \sqcup B$, and therefore $\sup A$ and $\inf B$ do not exist in $\mathbb{Q}$. Note that the sequence $1, 1.4, 1.41, 1.414, 1.4142, 1.42431, \ldots$ is in the set A. We recognize the terms as the decimal approximations to $\sqrt{2}$. This sequence looks like it's *going somewhere*; but in fact the only place it can go is stuck in between A and B, which is not in $\mathbb{Q}$. The question is: *why* does it look like it's going somewhere?

Definition 2.7. *A sequence (a_n) in an ordered set is called* **Cauchy**, *or is said to be a* **Cauchy sequence**, *if*

$$\forall \epsilon > 0\ \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \ge N\ |a_n - a_m| < \epsilon.$$

That is: a sequence is Cauchy if its terms get and stay close to *each other*. That is: for any given tolerance $\epsilon > 0$, there is some time N after which all the terms are within distance ϵ of a_N. This notion is very close to convergence. Indeed:

Lemma 2.8. *Any convergent sequence is Cauchy.*

Proof. Let (a_n) be a convergent sequence, with limit a. Fix $\epsilon > 0$, and choose N large enough so that $|a_n - a| < \frac{\epsilon}{2}$ for $n > N$. Then for any $n, m > N$,

$$|a_n - a_m| = |a_n - a + a - a_m| \le |a_n - a| + |a_m - a| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Hence, (a_n) is Cauchy. □

But the converse need not be true.

Example 2.9. In $\mathbb{Q}$, the sequence $1, 1.4, 1.41, 1.414, 1.4142, 1.42431, \ldots$ is Cauchy. Indeed, by the definition of decimal expansion, if a_n is the n-decimal expansion of a number, then a_{n+1} and a_n agree on the first n digits. This means exactly that $|a_m - a_n| < \frac{1}{10^n}$ for any $m > n$. So, fix $\epsilon > 0$. We can certainly find N so that $\frac{1}{10^N} < \epsilon$ (since, for example, $\frac{1}{10^N} < \frac{1}{N}$). Thus, for $n, m > N$, we have $|a_n - a_m| < \frac{1}{10^{\min\{m,n\}}} < \frac{1}{10^N} < \epsilon$.

Here are some more important facts about Cauchy sequences. Note that, by Lemma 2.8, any fact about Cauchy sequences is also a fact about convergent sequences.

Proposition 2.10. *Let (a_n) be a Cauchy sequences. Then (a_n) is bounded: there is a constant $M > 0$ so that $|a_n| \le M$ for all n.*

Proof. Taking $\epsilon = 1$, it follows from the definition of Cauchy that there is some $N \in \mathbb{N}$ so that $|a_n - a_m| < 1$ for all $n, m > N$. In particular, this shows that $|a_n - a_{N+1}| < 1$ for all $n > N$, which is to say that $a_{N+1} - 1 < a_n < a_{N+1} + 1$. Hence $|a_n| < \max\{|a_{N+1} - 1|, |a_{N+1} + 1|\}$ for $n > N$. So, define $M = \max\{|a_1|, \ldots, |a_N|, |a_{N+1} - 1|, |a_{N+1} + 1|\}$. If $n \le \mathbb{N}$, then $|a_n| \le M$ since $|a_n|$ appears in this list we maximize over; if $n > N$ then, as just shown, $|a_n| < \max\{|a_{N+1} - 1|, |a_{N+1} + 1|\} \le M$. The result follows. □

Another useful concept when working with sequences is **subsequences**.

Definition 2.11. *Let $\{n_k : k \in \mathbb{N}\}$ be a set of positive integers with the property that $n_k < n_{k+1}$ for all k; that is n_k is an increasing sequence in $\mathbb{N}$. Let (a_n) be a sequence. The function $k \mapsto a_{n_k}$ is called a* **subsequence** *of (a_n), usually denoted (a_{n_k}).*

Example 2.12. (a) Let $a_n = \frac{1}{n}$. Then $a_{2n} = \frac{1}{2n}$ and $a_{2^n} = \frac{1}{2^n}$ are subsequences. However

$$b_n = \begin{cases} a_n & \text{if } n \text{ is odd} \\ a_{n/2} & \text{if } n \text{ is even} \end{cases}$$

is **not** a subsequence of (a_n). Indeed, $b_k = a_{n_k}$ where $(n_k)_{k=1}^{\infty} = (1, 1, 3, 2, 5, 3, 7, 4, 9, 5, \ldots)$, and this is not an increasing sequence of integers.

(b) Let $a_n = (-1)^n$. Then $a_{2n} = 1$ and $a_{2n+1} = -1$ are subsequences.

Here is an extremely useful fact about the indices of subsequences: if (n_k) is an increasing sequence in $\mathbb{N}$, then $n_k \ge k$ for every k. (This follows by a simple induction.)

Proposition 2.13. *Let (a_n) be a sequence in an ordered set, and (a_{n_k}) a subsequence.*

(1) *If (a_n) is Cauchy, then (a_{n_k}) is Cauchy.*
(2) *If (a_n) is convergent with limit a, then (a_{n_k}) is convergent with limit a.*
(3) *If (a_n) is Cauchy, and (a_{n_k}) is convergent with limit a, then (a_n) is convergent with limit a.*

Proof. For (1): fix $\epsilon > 0$ and let $N \in \mathbb{N}$ be chosen so that $|a_n - a_m| < \epsilon$ for $n, m > N$. Then whenever $k, \ell > N$, we have $n_k \ge k > N$ and $n_\ell \ge \ell > N$, so by definition $|a_{n_k} - a_{n_\ell}| < \epsilon$. Thus (a_n) is Cauchy. The proof of (2) is very similar. Item (3) is on HW3. □

Before proceeding with the theory of Cauchy sequences, here are some useful facts about convergent sequences sequences.

Theorem 2.14. *Let (a_n) and (b_n) be convergent sequences in an ordered field $\mathbb{F}$.*

(1) *If $a_n \le b_n$ for all sufficiently large n, then $\lim_n a_n \le \lim_n b_n$.*
(2) *(Squeeze Theorem) Suppose also that $\lim_n a_n = \lim_n b_n$. If (c_n) is another sequence, and $a_n \le c_n \le b_n$ for all sufficiently large n, then (c_n) is convergent, and $\lim_n c_n = \lim_n a_n = \lim_n b_n$.*

Proof. Let $a = \lim_n a_n$ and $b = \lim_n b_n$. For (1), fix $\epsilon > 0$. There is $N_a \in \mathbb{N}$ so that $|a_n - a| < \frac{\epsilon}{2}$ for $n > N_a$, and there is $N_b \in \mathbb{N}$ so that $|b_n - b| < \frac{\epsilon}{2}$ for $n > N_b$. Thus, letting $N = \max\{N_a, N_b\}$, we have $a_n - a > -\frac{\epsilon}{2}$ and $b_n - b < \frac{\epsilon}{2}$ for $n > N$. But then

$$a_n - b_n > a - \frac{\epsilon}{2} - b - \frac{\epsilon}{2} = a - b - \epsilon.$$

Since $a_n \leq b_n$ for all large n, we therefore have $0 \geq a_n - b_n > a - b - \epsilon$ for such n, and therefore $a - b - \epsilon < 0$. This is true for any $\epsilon > 0$, and therefore $a - b \leq 0$, as claimed.

For (2), we have $a = b$. Choosing N_a, N_b, and N as above, we have $-\frac{\epsilon}{2} < a_n - a \leq c_n - a \leq b_n - a < \frac{\epsilon}{2}$ for all $n \geq N$. That is: $|c_n - a| < \frac{\epsilon}{2} < \epsilon$ for all $n \geq N$. This shows $c_n \to a$, as claimed. $\square$

Cauchy sequences give us a way of talking about completeness that is not so wrapped up in the order properties. As discussed in Example 2.9 last lecture, the "hole" in $\mathbb{Q}$ where $\sqrt{2}$ should be is the limit of a sequence in $\mathbb{Q}$ which is Cauchy, but does not converge in $\mathbb{Q}$. Instead of filling in the holes by demanding bounded nonempty sets have suprema, we could instead demand that Cauchy sequences have limits.

Definition 2.15. *Let S be an ordered set. Call S* **Cauchy complete** *if every Cauchy sequence in S actually converges in S.*

$\mathbb{Q}$ is not Cauchy complete. But, as we will see, $\mathbb{R}$ is. In fact, Cauchy completeness is *equivalent* to the least upper bound property *in any Archimedean field*. We can prove half of this assertion now.

2.3. **Lecture 7: January 26, 2016.**

Theorem 2.16. *Let $\mathbb{F}$ be an Archimedean field. If $\mathbb{F}$ is Cauchy complete, then $\mathbb{F}$ has the nested intervals property and hence is complete in the sense of Definition 1.15.*

Proof. That the nested intervals property implies the least upper bound property is the content HW2 Exercise 3; so it suffices to verify that $\mathbb{F}$ has the nested intervals property. Let (a_n) and (b_n) be sequences in $\mathbb{F}$ with $a_n \uparrow$, $b_n \downarrow$, $a_n \le b_n$, and $b_n - a_n < \frac{1}{n}$. Fix $\epsilon > 0$, and let $N \in \mathbb{N}$ be large enough that $\frac{1}{N} < \epsilon$ (here is where the Archimedean property is needed). Thus, for $n \ge N$, we have $b_n - a_n < \frac{1}{n} \le \frac{1}{N} < \epsilon$. Then for $m, n > N$, wlog $m \ge n$, we have

$$a_n \le a_m \le b_n$$

and so it follows that $|a_n - a_m| = a_m - a_n \le b_n - a_n < \epsilon$. Thus (a_n) is a Cauchy sequence. By the Cauchy completeness assumption on $\mathbb{F}$, we conclude that $a = \lim_n a_n$ exists in $\mathbb{F}$.

Now, fix n_0, and note that since $a_n \ge a_{n_0}$ for $n \ge n_0$, Theorem 2.14(1) shows that $a = \lim_n a_n \ge a_{n_0}$ (thinking of a_{n_0} as the limit of the constant sequences $(a_{n_0}, a_{n_0}, \ldots)$). Similarly, since $a_n \le b_{n_0}$ for all n, it follows that $a \le b_{n_0}$. Thus $a \in \bigcap_n [a_n, b_n]$, proving this intersection is nonempty. As usual, it follows that the intersection consists only of $\{a\}$. Indeed, if $x, y \in \bigcap_n [a_n, b_n]$, without loss of generality label them so that $x \le y$. Thus $a_n \le x \le y \le b_n$ for every n. For given $\epsilon > 0$, choose n so that $b_n - a_n < \epsilon$; then $y - x < \epsilon$. So $0 \le y - x < \epsilon$ for all $\epsilon > 0$; it follows that $x = y$. This concludes the proof of the nested intervals property for S. $\square$

Remark 2.17. The use of the Archimedean property is very subtle here. It is tempting to think that we can do without it. This is true if we replace the nested intervals property by a slightly weaker version: say an ordered S satisfies the *weak* nested intervals property if, given $a_n \uparrow$, $b_n \downarrow$, $a_n \le b_n$, and $b_n - a_n \to 0$, then $\bigcap_n [a_n, b_n]$ contains exactly one point. (This is weaker than the nested intervals property, because the assumption is stronger: we're assuming $b_n - a_n \to 0$ here, while in the usual nested intervals property we assume that $b_n - a_n < \frac{1}{n}$, which does not imply $b_n - a_n \to 0$ in the non-Archimedean setting.) The trouble is: this weak nested intervals property *does not imply* the least upper bound property in the absence of the Archimedean property. In fact, there *do exist* non-Archimedean fields (which therefore do not have the least upper bound property), but *are* Cauchy complete. (We may explore this a little later.) This is a prime example of how counterintuitive analysis can be without the Archimedean property. Soon enough, we will once-and-for-all demand that it holds true (in the Real numbers), and dispense with these weird pathologies.

We would like to show the converse is true: that the least upper bound property implies Cauchy completeness. (This turns out to be true in any ordered set: after all, the least upper bound property implies the Archimedean property in an ordered field.) Then we could characterize the real numbers as the unique Archimedean field that is Cauchy complete. To do this, we need to dig a little deeper into the connection between limits and suprema / infima.

Definition 2.18. *Let S be an ordered set with the least upper bound property. Let (a_n) be a bounded sequence in S. Define two new sequences from (a_n):*

$$\overline{a}_k = \sup\{a_n \colon n \ge k\}, \qquad \underline{a}_k = \inf\{a_k \colon n \ge k\}.$$

Since $\{a_n\}$ is bounded above (and nonempty), by the least upper bound property $\overline{a}_k$ exists for each k. Similarly, by Proposition 1.16, $\underline{a}_k$ exists for each k.

Note that $\{a_n \colon n \ge k+1\} \subseteq \{a_n \colon n \ge k\}$. Thus $\overline{a}_k$ is an upper bound for $\{a_n \colon n \ge k+1\}$. It follows that $\overline{a}_k$ is $\ge$ the least *upper bound of $\{a_n \colon n \ge k+1\}$, which is defined to be $\overline{a}_{k+1}$.*

This means that $\overline{a}_k \geq \overline{a}_{k+1}$: the sequence $\overline{a}_k$ is monotone decreasing. Similarly, the sequence $\underline{a}_k$ is monotone increasing.

By assumption, $\{a_n\}$ is bounded. Thus there is a lower bound $a_n \geq L$ for all n. Since $\overline{a}_1 \geq \overline{a}_k \geq a_k \geq L$ for all k, the sequence $\overline{a}_k$ is also bounded. Similarly, the sequence $\underline{a}_k$ is bounded.

Thus, $\overline{a}_k$ is a decreasing, bounded-below sequence. By Proposition 2.6, $\lim_{k\to\infty} \overline{a}_k$ exists, and is equal to $\inf\{\overline{a}_k\}$. *Similarly, $\lim_{k\to\infty} \underline{a}_k$ exists, and is equal to* $\sup\{\underline{a}_k\}$. *We define*

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \overline{a}_n = \lim_{k\to\infty} \sup\{a_n \colon n \geq k\} = \inf_{k\in\mathbb{N}} \sup_{n\geq k} a_n$$
$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \underline{a}_n = \lim_{k\to\infty} \inf\{a_n \colon n \geq k\} = \sup_{k\in\mathbb{N}} \inf_{n\geq k} a_n.$$

Example 2.19. Let $a_n = (-1)^n$. Note that $-1 \leq a_n \leq 1$ for all n. Now, for any k, there is some $k' \geq k$ so that $b_{k'} = 1$. Thus $\overline{b}_k = \sup_{n\geq k} a_k = 1$. Similarly $\underline{b}_k = -1$ for all k. Thus $\limsup_n b_n = 1$ and $\liminf b_n = -1$.

Here are a few more examples computing $\limsup$ and $\liminf$.

Example 2.20.

(1) Let $a_n = \frac{1}{n}$. Since $a_n \downarrow$, $\overline{a}_k = \sup_{n\geq k} a_n = a_k = \frac{1}{k}$. Thus $\limsup_n a_n = \lim_k a_k = 0$. On the other hand, for any k, $\inf_k \underline{a}_k = 0$ (by the Archimedean property), and so $\liminf_n a_n = \lim_k 0 = 0$. In this case, the $\limsup$ and $\liminf$ agree.

(2) Let $b_n = \frac{(-1)^n}{n}$. Note that $-1 \leq b_n \leq 1$ for all n, and more generally $|b_n| \leq \frac{1}{n}$. For any k, we therefore have $\overline{b}_k = \sup\{b_n \colon n \geq k\} \leq \sup\{|b_n| \colon n \geq k\} = \frac{1}{k}$ and similarly $\underline{b}_k \geq -\frac{1}{k}$. Now, $\underline{b}_k \leq \overline{b}_k$ (the $\sup$ of any set is $\geq$ its inf). Thus
$$-\frac{1}{k} \leq \underline{b}_k \leq \overline{b}_k \leq \frac{1}{k}.$$
Since $\pm\frac{1}{k} \to 0$, it follows from the Squeeze Theorem that $\lim_k \underline{b}_k = \lim_k \overline{b}_k = 0$. Thus $\limsup_n b_n = \liminf_n b_n = 0$.

(3) The sequence $(1, 2, 3, 1, 2, 3, 1, 2, 3, \ldots)$ has $\limsup = 3$ and $\liminf = 1$.

(4) Let $c_n = n$. This is not a bounded sequence, so it doesn't fit the mold for $\limsup$ and $\liminf$. Indeed, for any k, $\sup_{n\geq k} n$ does not exist for any k, and so $\limsup_n c_n$ does not exist. On the other hand, $\inf_{n\geq k} a_n = k$ does exists, but this sequence is unbounded and has no limit, so $\liminf c_n$ does not exist. This highlights the fact that we need *both* and upper *and* a lower bound in order for either $\limsup$ or $\liminf$ to exist.

In (1) and (2) in the example, $\limsup$ and $\liminf$ agree. This will always happen for a convergent sequence.

Proposition 2.21. *Let (a_n) be a bounded sequence. Then $\lim_n a_n$ exists iff $\limsup_n a_n = \liminf_n a_n$, in which case all three limits have the same value.*

Proof. Suppose that $\limsup_n a_n = \liminf_n a_n$. Thus $\underline{a}_k$ and $\overline{a}_k$ both converge to the same value. Since $\underline{a}_k \leq a_k \leq \overline{a}_k$ for each k, by the Squeeze Theorem, a_k also converges to this value, as claimed. Conversely, suppose that $\lim_n a_n = a$ exists. Let $\epsilon > 0$, and choose $N \in \mathbb{N}$ large enough that $|a_n - a| < \epsilon$ for all $n \geq N$. That is

$$a - \epsilon < a_n < a + \epsilon, \qquad n \geq N.$$

It follows that

$$a - \epsilon \leq \inf_{n\geq k} a_n \leq \sup_{n\geq k} a_n \leq a + \epsilon, \qquad k \geq N$$

which shows that both $\overline{a}_k$ and $\underline{a}_k$ are in $[a-\epsilon, a+\epsilon]$ for $k \geq N$. Thus they both converge to a, as claimed. $\square$

As with $\sup$ and $\inf$, there is a useful trick for transforming statements about $\limsup$ into statements about $\liminf$.

Proposition 2.22. *Let (a_n) be a bounded sequence. Then $\liminf_n(-a_n) = -\limsup_n a_n$.*

Proof. Recall that, for any bounded set A, if $-A = \{-a \colon a \in A\}$, then $\sup(-A) = -\inf A$ and $\inf(-A) = -\sup A$. Now, Let $b_n = -a_n$. Then $\underline{b}_k = \inf\{b_n \colon n \geq k\} = \inf\{-a_n \colon n \geq k\} = -\sup\{a_n \colon n \geq k\} = -\overline{a}_k$. Thus

$$\liminf_{n\to\infty} b_n = \sup\{\underline{b}_k \colon k \in \mathbb{N}\} = \sup\{-\overline{a}_k \colon k \in \mathbb{N}\} = -\inf\{\overline{a}_k \colon k \in \mathbb{N}\} = -\limsup_{n\to\infty} a_n.$$

$\square$

Here is a useful characterization of $\limsup$ and $\liminf$.

Proposition 2.23. *Let (a_n) be a bounded sequence in a complete ordered field. Denote $\overline{a} = \limsup_n a_n$ and $\underline{a} = \liminf_n a_n$. Then $\overline{a}$ and $\underline{a}$ are uniquely determined by the following properties: for all $\epsilon > 0$,*

$$a_n \leq \overline{a} + \epsilon \text{ for all sufficiently large } n \text{, and}$$
$$a_n \geq \overline{a} - \epsilon \text{ for infinitely many } n,$$

and

$$a_n \leq \underline{a} + \epsilon \text{ for infinitely many } n \text{, and}$$
$$a_n \geq \underline{a} - \epsilon \text{ for all sufficiently large } n.$$

Proof. This is an exercise on HW4. $\square$

To put this into words: there are many "approximate eventual upper bounds" for the sequence: numbers a large enough that the sequence eventually never gets much bigger than a. The $\limsup$, $\overline{a}$, is the *smallest* approximate eventual upper bound: it is the unique number that the sequence eventually never strays far above, but also regularly gets close to from below. Similarly, the $\liminf$, $\underline{a}$, is the *largest* approximate eventual lower bound.

2.4. **Lecture 8: January 28, 2016.** This brings us to an important understanding of $\limsup$ and $\liminf$: they are the *maximal and minimal subsequential limits*.

Theorem 2.24. *Let (a_n) be a bounded sequence in a complete ordered field. There exists a subsequence of (a_n) that converges to $\limsup_n a_n$, and there exists a subsequence of (a_n) that converges to $\liminf_n a_n$. Moreover, if (b_k) is any convergent subsequence of (a_n), then*

$$\liminf_{n\to\infty} a_n \le \lim_{k\to\infty} b_k \le \limsup_{n\to\infty} a_n.$$

Proof. Let $\overline{a} = \limsup_n a_n$. By Proposition 2.23, for any $k \in \mathbb{N}$ there are infinitely many n so that $a_n \ge \overline{a} - \frac{1}{k}$. So, we proceed inductively: choose some n_1 so that $a_{n_1} \ge \overline{a} - 1$. Then, since there are infiniely many of them, we can find some $n_2 > n_1$ so that $a_{n_2} \ge \overline{a} - \frac{1}{2}$. Proceeding, we find an increasing sequence $n_1 < n_2 < \cdots < n_k < \cdots$ so that $a_{n_k} \ge \overline{a} - \frac{1}{k}$ for each $k \in \mathbb{N}$. We therefore have

$$\overline{a} - \frac{1}{k} \le a_{n_k} \le \sup_{m \ge n_k} a_m = \overline{a}_{n_k}. \tag{2.1}$$

Note that $(\overline{a}_{n_k})$ is a subsequence of $(\overline{a}_n)$ which converges to $\overline{a}$; thus, by Proposition 2.13, $\lim_k \overline{a}_{n_k} = \overline{a}$. Hence, by (2.1) and the Squeeze Thoerem, it follows that $a_{n_k} \to \overline{a}$, and we have constructed the desired subsequence. The proof for $\liminf$ is very similar; alternatively, it can be reasoned using Proposition 2.22.

Now to prove the inequalities. Let (b_k) be a subsequence, so $b_k = a_{m_k}$ for some $m_1 < m_2 < m_3 < \cdots$. Then

$$\underline{a}_{m_k} = \inf_{n \ge m_k} a_n \le b_k \le \sup_{n \ge m_k} a_n = \overline{a}_{m_k}.$$

Thus, applying the Squeeze theorem, it follows that

$$\liminf_{n\to\infty} a_n = \lim_{k\to\infty} \underline{a}_{m_k} \le \lim_{k\to\infty} b_k \le \lim_{k\to\infty} \overline{a}_{m_k} = \limsup_{n\to\infty} a_n$$

as desired. □

This allows us to immediately prove our first "named theorem" in Real Analysis: the **Bolzano-Weierstrass Theorem**.

Theorem 2.25 (Bolzano-Weierstrass). *Let (a_n) be a bounded sequence in a complete ordered field, with $a_n \in [\alpha, \beta]$ for all n. Then (a_n) possesses a convergent subsequence, with limit in $[\alpha, \beta]$.*

Proof. Let $\overline{a} = \limsup_n a_n$. By Theorem 2.24, there is a subsequence (a_{n_k}) of (a_n) that converges to $\overline{a}$. Note, then, since $\alpha \le a_{n_k} \le \beta$ for all k, it follows from the Squeeze Theorem that $\alpha \le \lim_k a_{n_k} = \overline{a} \le \beta$, concluding the proof. □

This finally leads us to the converse of Theorem 2.16.

Theorem 2.26. *Let $\mathbb{F}$ be a complete ordered field (i.e. possessing the least upper bound property). Then $\mathbb{F}$ is Cauchy complete.*

Proof. Let (a_n) be a Cauchy sequence in $\mathbb{F}$. By Proposition 2.10, (a_n) is bounded. Thus, by the Bolzano-Weierstrass theorem, there is a subsequence a_{n_k} that converges. It then follows from Proposition 2.13 that (a_n) is convergent, concluding the proof. □

To summarize: we now have three equivalent characterizations of the notion of "completeness" in an Archimedean field:

$$\text{least upper bound property} \iff \text{nested intervals property} \iff \text{Cauchy completeness.}$$

We also know, by the half of Theorem 1.20 we've proved, that such a field is unique. So, to finally prove the existence of $\mathbb{R}$, it will suffice to give a construction of a Cauchy complete field that is Archimedean. The supplementary notes "Construction of $\mathbb{R}$" describe how this is done in gory detail.

Henceforth, we will deal with the field $\mathbb{R}$, which satisfies all of the three equivalent completeness properties.

Now comfortably working in $\mathbb{R}$, let us state a few more (standard) limit theorems.

Theorem 2.27 (Limit Theorems). *Let (a_n) and (b_n) be convergent sequences in $\mathbb{R}$, with $a_n \to a$ and $b_n \to b$.*

(1) *The sequence $c_n = a_n + b_n$ converges to $a + b$.*
(2) *The sequence $d_n = a_n b_n$ converges to ab.*
(3) *If $b \neq 0$, then $b_n \neq 0$ for almost all n, and $e_n = \frac{a_n}{b_n}$ converges to $\frac{a}{b}$.*

Proof. For (1), choose $N_a, N_b \in \mathbb{N}$ so that $|a_n - a| < \frac{\epsilon}{2}$ if $n \geq N_a$ and $|b_n - b| < \frac{\epsilon}{2}$ for $n \geq N_b$. For any $n \geq N = \max\{N_a, N_b\}$, we then have $|c_n - (a+b)| = |(a_n - a) + (b_n - b)| \leq |a_n - a| + |b_n - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$, proving that $\lim_n c_n = a + b$.

For (2), we need to be slightly more clever. Note that

$$|d_n - ab| = |a_n b_n - ab| = |a_n b_n - a_n b + a_n b - ab| \leq |a_n||b_n - b| + |a_n - a||b|.$$

By Proposition 2.10, there is some constant $M > 0$ so that $|a_n| \leq M$ for all n. So, for $\epsilon > 0$, fix N_1 large enough that $|b_n - b| < \frac{\epsilon}{2M}$ for all $n \geq N_1$, and fix N_2 large enough that $|a_n - a| < \frac{\epsilon}{2|b|}$ for all $n \geq N_2$. (If $b = 0$, we can take N_2 to be any number we like.) Then for $N = \max\{N_1, N_2\}$, if $n \geq N$ we have

$$|d_n - ab| \leq |a_n||b_n - b| + |a_n - a||b| < M \cdot \frac{\epsilon}{2M} + \frac{\epsilon}{2|b|} \cdot |b| = \epsilon,$$

proving that $\lim_n d_n = ab$.

For (3), first we need to show that (e_n) even makes sense. Note that $e_n = \frac{a_n}{b_n}$ is not well-defined for any n for which $b_n = 0$. But we're only concerned about tails of sequences for limit statements, so once we've proven that $b_n \neq 0$ for almost all n, we know that e_n is well-defined for all large n. For this, we use the assumption that $b \neq 0$, and so $|b| > 0$. Since $\lim_n b_n = b$, there is an $N_0 \in \mathbb{N}$ so that, for $n > N_0$, $|b_n - b| < \frac{|b|}{2}$; i.e. $-\frac{|b|}{2} < b_n - b < \frac{|b|}{2}$, and so $b_n < b + \frac{|b|}{2}$ and also $b_n > b - \frac{|b|}{2}$. Now, $b \neq 0$ so either $b < 0$ or $b > 0$. If $b < 0$, then $|b| = -b$ in which case $b_n < b + \frac{|b|}{2} = b - \frac{b}{2} = \frac{b}{2} < 0$; that is, for $n > N_0$, $b_n < 0$. If, on the other hand, $b > 0$, then $|b| = b$, and so $b_n > b - \frac{|b|}{2} = b - \frac{b}{2} = \frac{b}{2} > 0$; that is, for $n > N_0$, $b_n > 0$. Thus, in all cases, $b_n \neq 0$ for $n > N_0$, proving the first claim.

For the limit statement, note that $e_n = a_n \cdot \frac{1}{b_n}$. So, by (2), it suffices to show that $\frac{1}{b_n} \to \frac{1}{b}$. Compute that

$$\left|\frac{1}{b_n} - \frac{1}{b}\right| = \frac{|b_n - b|}{|b_n||b|}.$$

As shown above, there is N_0 so that, for $n > N_0$, then $b_n > \frac{b}{2} = \frac{|b|}{2}$ if $b_n > 0$ and $b_n < \frac{b}{2} = -\frac{|b|}{2}$ if $b_n < 0$; i.e. this means that $|b_n| > \frac{|b|}{2}$ for $n > N_0$. Hence, we have

$$\left|\frac{1}{b_n} - \frac{1}{b}\right| = \frac{|b_n - b|}{|b_n||b|} < 2\frac{|b_n - b|}{|b|^2}, \qquad n > N_0.$$

By assumption, $b_n \to b$, and so we can choose N' large enough that $|b_n - b| < \frac{|b|^2}{2}\epsilon$ for $n > N'$. Thus, letting $N = \max\{N_0, N'\}$, we have

$$\left|\frac{1}{b_n} - \frac{1}{b}\right| < 2\frac{|b_n - b|}{|b|^2} < \frac{2}{|b|^2} \cdot \frac{|b|^2}{2}\epsilon = \epsilon, \qquad n > N.$$

This proves that $\frac{1}{b_n} \to \frac{1}{b}$ as claimed. $\square$

One might hope that Theorem 2.27 carries over to $\limsup$ and $\liminf$; but this is not the case.

Example 2.28. Consider the sequences $a_n = (-1)^n$ and $b_n = -a_n = (-1)^{n+1}$. Then $\limsup_n a_n = \limsup b_n = 1$, $\liminf_n a_n = \liminf_n b_n = -1$, but $a_n + b_n = 0$ so $\limsup_n(a_n + b_n) = \liminf_n(a_n + b_n) = 0$. Hence, in this example we have

$$-2 = \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n < \liminf_{n\to\infty}(a_n + b_n) = 0 = \limsup_{n\to\infty}(a_n + b_n) < \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n = 2.$$

The *in*equalities in the example do turn out to be true in general.

Proposition 2.29. *Let (a_n) and (b_n) be bounded sequences in $\mathbb{R}$ The following always hold true.*

$$\liminf_{n\to\infty}(a_n + b_n) \ge \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n, \quad \textit{and}$$
$$\limsup_{n\to\infty}(a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n.$$

If $a_n \ge 0$ and $b_n \ge 0$ for all sufficiently large n, we also have the following.

$$\liminf_{n\to\infty}(a_n \cdot b_n) \ge \liminf_{n\to\infty} a_n \cdot \liminf_{n\to\infty} b_n, \quad \textit{and}$$
$$\limsup_{n\to\infty}(a_n \cdot b_n) \le \limsup_{n\to\infty} a_n \cdot \limsup_{n\to\infty} b_n.$$

Proof. The proofs of the $\limsup$ inequalities are exercises on HW4. Assuming these, the $\liminf$ statements follow from Proposition 2.22. For example, we have

$$\liminf_{n\to\infty}(a_n + b_n) = \liminf_{n\to\infty}[-(-a_n - b_n)] = -\limsup_{n\to\infty}[(-a_n) + (-b_n)].$$

Since $\limsup_n[(-a_n) + (-b_n)] \le \limsup_n(-a_n) + \limsup_n(-b_n)$ by HW4, taking negatives reverses the inequality, giving

$$-\limsup_{n\to\infty}[(-a_n) + (-b_n)] \ge -\limsup_{n\to\infty}(-a_n) - \limsup_{n\to\infty}(-b_n).$$

Now using Proposition 2.22 again on each term, we then have

$$\liminf_{n\to\infty}(a_n + b_n) \ge -\limsup_{n\to\infty}(-a_n) - \limsup_{n\to\infty}(-b_n) = \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n$$

as claimed. The proof of the inequality for products is very similar. $\square$

Let us close out our discussion (for now) of limits of real sequences with a rigorous treatment of the following special kinds of sequences.

Proposition 2.30. *Let $p > 0$ and $\alpha \in \mathbb{R}$.*

(1) $\displaystyle\lim_{n\to\infty} \frac{1}{n^p} = 0.$
(2) $\displaystyle\lim_{n\to\infty} p^{1/n} = 1.$
(3) $\displaystyle\lim_{n\to\infty} n^{1/n} = 1.$

(4) *If* $p>1$ *and* $\alpha\in\mathbb{R}$, *then* $\displaystyle\lim_{n\to\infty}\frac{n^\alpha}{p^n}=0$.

(5) *If* $|p|<1$, *then* $\displaystyle\lim_{n\to\infty}p^n=0$.

Proof. For (1): fix $\epsilon>0$, and choose $N\in\mathbb{N}$ large enough that $\frac{1}{N}<\epsilon^{1/p}$. Then for $n\geq N$, $\frac{1}{n}\leq\frac{1}{N}<\epsilon^{1/p}$, and so $0<\frac{1}{n^p}=\left(\frac{1}{n}\right)^p<\epsilon$. This shows that $\frac{1}{n^p}\to 0$ as claimed.

For (2): as above, in the case $p=1$ the sequence is constant $1^{1/n}=1$ with limit 1. If $p>1$, put $x_n=p^{1/n}-1$. Since $p>1$ we have $p^{1/n}>1$ and so $x_n>0$. From the binomial theorem, then,

$$(1+x_n)^n=\sum_{k=0}^n\binom{n}{k}x_n^k\geq 1+nx_n.$$

By definition $(1+x_n)^n=p$, and so

$$0<x_n<\frac{(1+x_n)^n-1}{n}=\frac{p-1}{n}.$$

Knowing that $\frac{p-1}{n}\to 0$, it now follows from the Squeeze Theorem that $x_n\to 0$. This proves the limit in the case $p>1$. If, on the other hand, $0<p<1$, then $r=\frac{1}{p}>1$, and $p^{1/n}=\left(\frac{1}{r}\right)^{1/n}=\frac{1}{r^{1/n}}$. We have just proved that $r^{1/n}\to 1$, and so it follows from Theorem 2.27(3) that $p^{1/n}\to\frac{1}{1}=1$.

For (3): we follow a similar outline. Let $x_n=n^{1/n}-1$, which is ≥ 0 (and >0 for $n>1$). We use the binomial theorem again, this time estimating with the quadratic term:

$$n=(1+x_n)^n=\sum_{k=1}^n\binom{n}{k}x_n^k\geq\binom{n}{2}x_n^2=\frac{n(n-1)}{2}x_n^2.$$

Thus, we have (for $n\geq 2$)

$$0\leq x_n\leq\sqrt{\frac{2}{n-1}}$$

and by the Squeeze Theorem $x_n\to 0$.

For (4): Choose a positive integer $\ell>\alpha$. Let $p=1+r$, so $r>0$. Applying the binomial theorem again, when $n>\ell$ we have

$$p^n=(1+r)^n=\sum_{k=0}^n\binom{n}{k}r^k>\binom{n}{\ell}r^\ell=\frac{n(n-1)\cdots(n-\ell+1)}{\ell!}r^\ell.$$

Now, if we choose $n\geq 2\ell$, each term $n-\ell+j\geq\frac{n}{2}$ for $1\leq j\leq\ell$, and so in this range

$$p^n>\frac{1}{\ell!}\left(\frac{n}{2}\right)^\ell r^\ell.$$

Hence, for $n\geq 2\ell$, we have

$$\frac{n^\alpha}{p^n}<n^\alpha\cdot\frac{\ell!2^\ell}{n^\ell r^\ell}=\frac{\ell!2^\ell}{r^\ell}\cdot n^{\alpha-\ell}.$$

This is a constant $\frac{\ell!2^\ell}{r^\ell}$ times $n^{\alpha-\ell}$, where $\alpha-\ell<0$; applying part (1) with $p=\alpha-\ell$ proves the result.

Finally, for (5): the special case of (4) with $\alpha=0$ yields $\frac{1}{r^n}\to 0$ when $r>1$. Thus, with $|p|<1$, setting $r=\frac{1}{|p|}$ gives us $|p^n|=|p|^n\to 0$. The reader should prove (if they haven't already) that $|a_n|\to 0$ iff $a_n\to 0$, so it follows that $p^n\to 0$ as claimed. □

3. Extensions of $\mathbb{R}$: the Extended Real Numbers $\overline{\mathbb{R}}$ and the Complex Numbers $\mathbb{C}$

3.1. **Lecture 9: February 2, 2016.** Now that we have a good understanding of real numbers, it is convenient to extend them a little bit to give us language about certain kinds of divergent sequences.

Definition 3.1. *Let (a_n) be a sequence in $\mathbb{R}$. Say that a_n* **diverges to** $+\infty$ *or $a_n \to +\infty$ if*

$$\forall M > 0\ \exists N \in \mathbb{N}\, s.t.\ \forall n \geq N\ a_n > M.$$

That is: no matter how large a bound M we choose, it is a lower bound for a_n for all sufficiently large n. Similarly, we say that a_n **diverges to** $-\infty$ *if $-a_n \to +\infty$; this is equivalent to*

$$\forall M > 0\ \exists N \in \mathbb{N}\, s.t.\ \forall n \geq N\ a_n < -M.$$

The expressions $a_n \to \pm\infty$ are also sometimes written as $\lim_{n\to\infty} a_n = \pm\infty$, and accordingly it is sometimes pronounced as a_n **converges to** $\pm\infty$.

Example 3.2. The sequence $a_n = n^p$ diverges to $+\infty$ for any $p > 0$. Indeed, fix a large $M > 0$. Then $M^{1/p} > 0$, and by the Archimedean property there is an $N \in \mathbb{N}$ with $N > M^{1/p}$. Thus, for $n \geq N$, $n > M^{1/p}$, and so $a_n = n^p > (M^{1/p})^p = M$, as desired.

On the other hand, the sequence $(a_n) = (1, 0, 2, 0, 3, 0, 4, 0, \ldots)$ does not diverge to $+\infty$: no matter how large N is, there is some integer $n \geq N$ with $a_n = 0$. (Indeed, we can either take $n = N$ or $n = N + 1$.) This sequence diverges, but it does not diverge to $+\infty$.

This suggests the we include the symbols $\pm\infty$ in the field $\mathbb{R}$. We must be careful how to do this, however. We have already proved that $\mathbb{R}$ is the *unique* complete ordered field, so no matter how we add $\pm\infty$, the resulting object cannot be a complete ordered field. In fact, it won't be a field at all, for we won't always be able to do algebraic operations.

Definition 3.3. *Let $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. We make $\overline{\mathbb{R}}$ into an ordered set as follows: given $x, y \in \overline{\mathbb{R}}$, if in fact $x, y \in \mathbb{R}$ then we use the order relation from $\mathbb{R}$ to compare x, y. If one of the two (say x) is in $\mathbb{R}$, then we declare $-\infty < x < +\infty$. Finally, we declare $-\infty < +\infty$.*

We make the following conventions. If $a \in \overline{\mathbb{R}}$ with $a > 0$, then $\pm\infty \cdot a = a \cdot \pm\infty = \pm\infty$; if $a \in \overline{\mathbb{R}}$ with $a < 0$ then $\pm\infty \cdot a = a \cdot \pm\infty = \mp\infty$. We also declare that $a + (\pm\infty) = \pm\infty$ for any $a \in \mathbb{R}$, and that $(+\infty) + (+\infty) = +\infty$ while $(-\infty) + (-\infty) = -\infty$. We leave all the following expressions **undefined***:*

$$(+\infty) + (-\infty),\ (-\infty) + (\infty),\ \frac{\infty}{\infty},\ 0 \cdot (\pm\infty),\ \textit{and}\ (\pm\infty) \cdot 0.$$

Example 3.4. Let $\alpha, \beta \in \mathbb{R}$ with $\alpha > 0$, and let $a_n = n$ while $b_n = -\alpha n + \beta$. Then

$$\lim_{n\to\infty} (a_n + b_n) = \begin{cases} +\infty, & \text{if } \alpha < 1 \\ \beta, & \text{if } \alpha = 1 \\ -\infty, & \text{if } \alpha > 1 \end{cases}.$$

Hence the value of the limit of the sum depends on the value of α. However, Example 3.2 shows that $a_n \to +\infty$ while a similar argument shows that $b_n \to -\infty$ for any α, β. So we ought to have

$$\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \text{``} = \text{''} (+\infty) + (-\infty).$$

This highlights why it is important to leave such expressions undefined: there is no way to consistently define them that respects the limit theorems.

We can also use these conventions to extend the notions of $\sup$ and $\inf$ to unbounded sets, and the notions of $\limsup$ and $\liminf$ to unbounded sequences.

Definition 3.5. *Let $E \subseteq \mathbb{R}$ be any nonempty subset. If E is not bounded above, declare* $\sup E = +\infty$*; if E is not bounded below, declare* $\inf E = -\infty$*. We also make the convention that* $\sup(\varnothing) = -\infty$ *while* $\inf(\varnothing) = +\infty$*. (Note: this means that, in the one special case $E = \varnothing$, it is not true that* $\inf E \le \sup E$*.)*

Similarly, let (a_n) be any sequence in $\mathbb{R}$. If (a_n) is not bounded above, declare $\limsup_n a_n = +\infty$*; if (a_n) is not bounded below, declare* $\liminf_n a_n = -\infty$*.*

With these conventions, essentially all of the theorems involving limits extend to unbounded sequences.

Proposition 3.6. *Using the preceding conventions, Lemma 2.4, Proposition 2.6, Proposition 2.13(2), Squeeze Theorem 2.14, Proposition 2.21, Proposition 2.22, and Theorem 2.24 all generalize to the cases where the limits in the statements are allowed to be in $\overline{\mathbb{R}}$ rather than just $\mathbb{R}$. Moreover, Theorem 2.27 and Proposition 2.29 also hold in this more general setting* whenever the statements make sense*: i.e. excluding the cases when the involved expressions are undefined (like $(+\infty) + (-\infty)$).*

Proof. It would take many pages to prove all of the special cases of all of these results remain valid in the extended reals. Let us choose just one to illustrate: Theorem 2.27(1): if $\lim_n a_n = a$ and $\lim_n b_n = b$, then $\lim_n (a_n + b_n) = a + b$. We already know this holds true when $a, b \in \mathbb{R}$. If $a = +\infty$ and $b = -\infty$, or $a = -\infty$ and $b = +\infty$, the sum $a + b$ is undefined, and so we exclude these cases from the statement of the theorem. So we only need to consider the cases that $a \in \mathbb{R}$ and $b = \pm\infty$, $a = \pm\infty$ and $b \in \mathbb{R}$, or $a = b = \pm\infty$.

- $a \in \mathbb{R}$ and $b = +\infty$: Since (a_n) is convergent in $\mathbb{R}$, it is bounded; thus say $|a_n| \le L$. Then fix $M > 0$ and choose N so that $b_n > M + L$ for $n \ge N$. Thus $a_n + b_n > -L + (M + L) = M$ for $n \ge N$, and so $a_n + b_n \to +\infty$. The argument is similar when $b = -\infty$.
- $a = \pm\infty$ and $b \in \mathbb{R}$: this is the same as the previous case, just reverse the roles of a_n and b_n and a and b.
- $a = b = +\infty$: let $M > 0$, and choose N_1 so that $a_n > M/2$ for $n \ge N_1$; choose N_2 so that $b_n > M/2$ for $n \ge N_2$. Thus, for $n \ge N = \max\{N_1, N_2\}$, it follows that $a_n + b_n \ge M/2 + M/2 = M$, proving that $a_n + b_n \to +\infty$ as required. The argument when $a = b = -\infty$ is very similar.

□

Now we turn to a very different extension of $\mathbb{R}$: the Complex Numbers. We've already discussed them a little bit, in Example 1.12(3-3.5) and HW1.4, so we'll start by reiterating that discussion. We will rely on our knowledge of linear algebra.

Definition 3.7. *Let $\mathbb{C}$ denote the following set of 2×2 matrices over $\mathbb{R}$:*

$$\mathbb{C} = \left\{ \begin{bmatrix} a & -b \\ b & a \end{bmatrix} : a, b \in \mathbb{R} \right\}.$$

Then $\mathbb{C} = \mathrm{span}_{\mathbb{R}}\{I, J\}$, where

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

As is customary, we denote $I = 1$ and $J = i$. We can compute that $J^2 = -I$, so $i^2 = -1$. Every complex number has the form $a1 + bi$ for unique $a, b \in \mathbb{R}$; we often suppress the 1 and write this as $a + ib$. We think of $\mathbb{R} \subset \mathbb{C}$ via the identification $a \leftrightarrow a + i0$ (so a is the matrix aI).

It is convenient to construct $\mathbb{C}$ this way, since, as a collection of matrices, we already have addition and multiplication built in; and we have all the tools of linear algebra to prove properties of $\mathbb{C}$.

Proposition 3.8. *Denote $1_{\mathbb{C}} = I$ and $0_{\mathbb{C}}$ the 2×2 zero matrix. Define $+$ and $\cdot$ on $\mathbb{C}$ by their usual matrix definitions. Then $\mathbb{C}$ is a field.*

Proof. Most of the work is done for us, since $+$ and $\cdot$ of matrices are associative and distributive, and $+$ is commutative, and $1_{\mathbb{C}}$ and $0_{\mathbb{C}}$ are multiplicative and additive identities. All that we are left to verify are the following three properties:

- $\mathbb{C}$ is closed under $+$ and $\cdot$, i.e. we need to check that if $z, w \in \mathbb{C}$ then $z + w \in \mathbb{C}$ and $z \cdot w \in \mathbb{C}$. Setting $z = a + ib$ and $w = c + id$, we simply compute

$$z + w = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} + \begin{bmatrix} c & -d \\ d & c \end{bmatrix} = \begin{bmatrix} a+c & -b-d \\ b+d & a+c \end{bmatrix} = (a+c) + (b+d)i \in \mathbb{C},$$

$$z \cdot w = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} c & -d \\ d & c \end{bmatrix} = \begin{bmatrix} ac - bd & -ad - bc \\ ad + bc & ac - bd \end{bmatrix} = (ac - bd) + (ad + bc)i \in \mathbb{C}.$$

- $\cdot$ is commutative: this follows from the computation above: if we exchange $z \leftrightarrow w$, meaning $a \leftrightarrow c$ and $b \leftrightarrow d$, the value of the product $z \cdot w$ is unaffected, so $z \cdot w = w \cdot z$.
- If $z \in \mathbb{C} \setminus \{0_{\mathbb{C}}\}$ then z^{-1} exists: here we use the criterion that a matrix z is invertible iff $\det(z) \neq 0$. We can readily compute that, for $z \in \mathbb{C}$,

$$\det(z) = \det \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = a^2 + b^2$$

  and this $= 0$ iff $a = b = 0$ meaning $a + ib = 0_{\mathbb{C}}$.

∎

Now more notation.

Definition 3.9. *Let $z = a + ib \in \mathbb{C}$. We denote $a = \mathrm{Re}(z)$ and $b = \mathrm{Im}(z)$, the* **Real** *and* **Imaginary** *parts of z. Define the* **modulus** *or* **absolute value** *of z to be*

$$|z| = \sqrt{\det(z)} = \sqrt{a^2 + b^2} = \sqrt{\mathrm{Re}(z)^2 + \mathrm{Im}(z)^2}.$$

For $z \in \mathbb{C}$, its **complex conjugate** *$\overline{z}$ is the complex number $\overline{z} = \mathrm{Re}(z) - i\mathrm{Im}(z)$; in terms of matrices, this is just the transpose $\overline{z} = z^\top$.*

Note that $z + \overline{z} = 2\mathrm{Re}(z)$ and $z - \overline{z} = 2i\mathrm{Im}(z)$. Since i is invertible (indeed $i^{-1} = -i$), it follows that

$$\mathrm{Re}(z) = \frac{z + \overline{z}}{2}, \qquad \mathrm{Im}(z) = \frac{z - \overline{z}}{2i}. \tag{3.1}$$

Note that, if $z \in \mathbb{C}$ happens to be in $\mathbb{R}$ (meaning that $\mathrm{Im} z = 0$ so $z = \mathrm{Re}(z)$), then $|z| = \sqrt{\mathrm{Re}(z)^2 + 0} = \sqrt{z^2} = |z|$ corresponds to the absolute value in $\mathbb{R}$; so the complex modulus generalizes the familiar absolute value.

3.2. **Lecture 10: February 4, 2016.** Here are some important properties of modulus and complex conjugate.

Lemma 3.10. *Let $z, w \in \mathbb{C}$. Then we have the following.*

(1) $\overline{\overline{z}} = z$.
(2) $\overline{z+w} = \overline{z} + \overline{w}$ *and* $\overline{zw} = \overline{z} \cdot \overline{w}$.
(3) $z\overline{z} = |z|^2$.
(4) $|\overline{z}| = |z|$.
(5) $|zw| = |z||w|$, *and so* $|z^n| = |z|^n$ *for all* $n \in \mathbb{N}$.
(6) $|\mathrm{Re}(z)| \le |z|$ *and* $|\mathrm{Im}(z)| \le |z|$.
(7) $|z+w| \le |z| + |w|$.
(8) $|z| = 0$ *iff* $z = 0$.
(9) *If* $z \ne 0$ *then* z^{-1} *(which we also write as* $\frac{1}{z}$*) is given by*
$$z^{-1} = \frac{\overline{z}}{|z|^2}.$$
(10) *If* $z \ne 0$ *then* $|z^{-1}| = |z|^{-1}$, *and so* $|z^n| = |z|^n$ *for all* $n \in \mathbb{Z}$.

Proof. (1) is the familiar linear algebra fact that $(z^\top)^\top = z$, and (2) follows similarly from linear algebra (and the commutativity of $\cdot$ in $\mathbb{C}$): $\overline{z+w} = (z+w)^\top = z^\top + w^\top = \overline{z} + \overline{w}$, and $\overline{zw} = (zw)^\top = w^\top z^\top = z^\top w^\top = \overline{zw}$. For (3), writing $z = a + ib$ we have
$$z\overline{z} = (a+ib)(a-ib) = a^2 + b^2 + (ab - ab)i = a^2 + b^2 = |z|^2.$$
(4) then follows that from this and (1), and commutativity of complex multiplication: $|\overline{z}|^2 = \overline{z}\overline{\overline{z}} = \overline{z}z = z\overline{z} = |z|^2$; taking square roots (using the fact that $|z| \ge 0$) shows that $|\overline{z}| = |z|$. (Alternatively, for (4), we simply note that $|\overline{z}| = \det(z^\top) = \det(z) = |z|$.)

(5) is a well-known property of determinants: $|zw| = \det(zw) = \det(z)\det(w) = |z||w|$; taking $z = w$ and doing induction shows that $|z^n| = |z|^n$. (6) follows easily from the fact that $|z| = \sqrt{|\mathrm{Re}(z)|^2 + |\mathrm{Im}(z)|^2}$. For (7), we have
$$|z+w|^2 = (z+w)(\overline{z+w}) = (z+w)(\overline{z}+\overline{w}) = z\overline{z} + z\overline{w} + w\overline{z} + w\overline{w} = |z|^2 + z\overline{w} + w\overline{z} + |w|^2.$$
The two middle terms can be written as $z\overline{w} + w\overline{z} = z\overline{w} + \overline{(z\overline{w})}$ and, by (3.1), this equals $2\mathrm{Re}(z\overline{w})$. Now, any real number x is $\le |x|$, and so
$$|z+w|^2 = |z|^2 + 2\mathrm{Re}(z\overline{w}) + |w|^2 \le |z|^2 + 2|\mathrm{Re}(z\overline{w})| + |w|^2 \le |z|^2 + 2|z\overline{w}|^2 + |w|^2$$
where we have used (6). From (4) and (5), $|z\overline{w}| = |z||\overline{w}| = |z||w|$, and so finally we have
$$|z+w|^2 \le |z|^2 + 2|z||w| + |w|^2 = (|z| + |w|)^2.$$
Taking square roots proves the result.

For (8), it is immediate that $|0| = 0$; the converse was shown in the proof of Proposition 3.8: $|z| = \det(z) = 0$ iff $(\mathrm{Re}(z))^2 + (\mathrm{Im}(z))^2 = 0$ which happens only when $\mathrm{Re}(z) = \mathrm{Im}(z) = 0$, so $z = 0$. Part (9) follows similarly from the matrix representation; alternatively we can simply check from (3) that
$$z \cdot \frac{\overline{z}}{|z|^2} = \frac{z\overline{z}}{|z|^2} = 1$$
showing that $z^{-1} = \frac{\overline{z}}{|z|^2}$. Finally, for (10), using (5) we have
$$|z^{-1}||z| = |z^{-1}z| = |1| = 1$$

so $|z|^{-1} = |z^{-1}|$. An induction argument combining this with (5) shows that $|z|^{-n} = |z^{-n}|$ for $n \in \mathbb{N}$, and coupling this with the second statement of (5) concludes the proof. □

Items (7) and (8) of Lemma 3.10 show that the complex modulus behaves just like the real absolute value: it satisfies the triangle inequality, and is only 0 at 0. These properties are all that were necessary to make most of the technology of limits of sequences in $\mathbb{R}$ work, and so we can now use the complex modulus to extend these notions to $\mathbb{C}$.

Definition 3.11. *Let (z_n) be a sequence in $\mathbb{C}$. Given $z \in \mathbb{C}$, say that $\lim_{n\to\infty} z_n = z$ iff*

$$\forall \epsilon > 0\ \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N\ |z_n - z| < \epsilon.$$

Say that (z_n) is a Cauchy sequence if

$$\forall \epsilon > 0\ \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \geq N\ |z_n - z_m| < \epsilon.$$

Note that these are, symbolically, *exactly the same* as the definitions (6.1 and 2.7) of limits and Cauchy sequences of real numbers; the only difference is, we now interpret $|z|$ to mean the modulus of the complex number z rather than the absolute value of a real number.

The properties of complex modulus mirroring those of real absolute value allow us to prove the results of Lemmas 2.4 and 2.8, Propositions 2.10 and 2.13, and the Limit Theorems (Theorem 2.27) with nearly identical proofs. To summarize:

Theorem 3.12.
(1) *Limits are unique: if $z_n \to z$ and $z_n \to w$, then $z = w$.*
(2) *Every convergent sequence in $\mathbb{C}$ is Cauchy.*
(3) *Every Cauchy sequence in $\mathbb{C}$ is bounded.*
(4) (a) *If $z_n \to z$ then any subsequence of (z_n) converges to z.*
  (b) *If (z_n) is Cauchy then any subsequence of (z_n) is Cauchy.*
  (c) *If (z_n) is Cauchy and has a convergent subsequence with limit z, then $z_n \to z$.*
(5) *If $z_n \to z$ and $w_n \to w$, then $z_n + w_n \to z + w$, $z_n w_n \to zw$, and if $z \neq 0$ then $z_n \neq 0$ for sufficiently large n and $\frac{1}{z_n} \to \frac{1}{z}$.*

To illustrate how to handle complex modulus in these proofs, let us look at the analog of Proposition 2.10: that Cauchy sequences are bounded. As before, we set $\epsilon = 1$ and let N be large enough that $|z_n - z_m| < 1$ whenever $n, m > N$. Thus, taking $m = N + 1$, for any $n > N$ we have $|z_n - z_{N+1}| < 1$. Now, $z_n = (z_n - z_{N+1}) + z_{N+1}$, and so by the triangle inequality

$$|z_n| = |(z_n - z_{N+1}) + z_{N+1}| \leq |z_n - z_{N+1}| + |z_{N+1}| < 1 + |z_{N+1}|, \qquad \forall\, n > N.$$

So, as in the previous proof, if we set $M = \max\{|z_1|, |z_2|, \ldots, |z_N|, 1 + |z_{N+1}|\}$ then $|z_n| \leq M$ for all n.

In fact, convergence and Cauchy-ness of complex sequences boils down to convergence and Cauchy-ness of the real and imaginary parts separately.

Proposition 3.13. *Let (z_n) be a sequence in $\mathbb{C}$. Then (z_n) is Cauchy iff the two real sequences $(\mathrm{Re}(a_n))$ and $(\mathrm{Im}(b_n))$ are Cauchy, and $z_n \to z$ iff $\mathrm{Re}(z_n) \to \mathrm{Re}(z)$ and $\mathrm{Im}(z_n) \to \mathrm{Im}(z)$.*

Proof. Let $z_n = a_n + ib_n$. Suppose (a_n) and (b_n) are Cauchy. Fix $\epsilon > 0$ and choose N_1 large enough that $|a_n - a_m| < \frac{\epsilon}{2}$ for $n, m > N_1$, and choose N_2 large enough that $|b_n - b_m| < \frac{\epsilon}{2}$ for

$n, m > N_2$. Then for $n, m > N = \max\{N_1, N_2\}$, we have

$$\begin{aligned}|z_n - z_m| = |(a_n + ib_n) - (a_m + ib_m)| = |(a_n - a_m) + i(b_n - b_m)| &\le |a_n - a_m| + |i(b_n - b_m)| \\ &= |a_n - a_m| + |b_n - b_m| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon\end{aligned}$$

where, in the second last step, we used the fact that $|i(b_n - b_m)| = |i||b_n - b_m|$ and $|i| = 1$. Thus, (z_n) is Cauchy. For the converse, suppose that (z_n) is Cauchy. Fix $\epsilon > 0$, and choose N large enough that $|z_n - z_m| < \epsilon$ for $n, m > N$. Then we also have

$$\begin{aligned}|\mathrm{Re}(z_n) - \mathrm{Re}(z_m)| &= |\mathrm{Re}(z_n - z_m)| \le |z_n - z_m| < \epsilon, \quad \text{and} \\ |\mathrm{Im}(z_n) - \mathrm{Im}(z_m)| &= |\mathrm{Im}(z_n - z_m)| \le |z_n - z_m| < \epsilon\end{aligned}$$

for $n, m > N$. Thus, both $(\mathrm{Re}(z_n))$ and $(\mathrm{Im}(z_n))$ are Cauchy, as claimed.

The proof of the limit statements is very similar, and is left as a homework exercise (on HW6). $\square$

Now, $\mathbb{C}$ is not an ordered field (you proved this on HW1), so it does not even make sense to ask if it has the least upper bound property (and likewise we cannot talk about a Squeeze Theorem, or $\limsup$ and $\liminf$). This is one of the main reasons we gave an equivalent characterization of the least upper bound property – Cauchy completeness – that does not explicitly require an order relation.

Theorem 3.14. *The field $\mathbb{C}$ is Cauchy complete: any Cauchy sequence is convergent.*

Proof. Let (z_n) be a Cauchy sequence in $\mathbb{C}$. By Proposition 3.13, the two real sequences $(\mathrm{Re}(z_n))$ and $(\mathrm{Im}(z_n))$ are both Cauchy. Since $\mathbb{R}$ is Cauchy complete, it follows that there are real numbers $a, b \in \mathbb{R}$ so that $\mathrm{Re}(z_n) \to a$ and $\mathrm{Im}(z_n) \to b$. It then follows, again by Proposition 3.13, that $z_n \to a + ib$. $\square$

In $\mathbb{R}$, we proved the Bolzano-Weierstrass theorem (that bounded sequences have convergent subsequences) using the technology of $\limsup$ and $\liminf$. As noted, since $\mathbb{C}$ is not ordered, there is no way to talk about $\limsup$ and $\liminf$ for a complex sequence. Nevertheless, the Bolzano-Weierstrass theorem holds true in $\mathbb{C}$. We conclude our discussion of $\mathbb{C}$ (for now) with its proof.

Theorem 3.15 (Bolzano-Weierstrass)**.** *Every bounded sequence in $\mathbb{C}$ contains a convergent subsequence.*

Proof. Let (z_n) be a bounded sequence. Letting $z_n = a_n + ib_n$, since $|a_n| \le |z_n|$ and $|b_n| \le |z_n|$, it follows that (a_n) and (b_n) are bounded sequences in $\mathbb{R}$. Now, by the Bolzano-Weierstrass theorem for $\mathbb{R}$, there is a subsequence a_{n_k} of (a_n) that converges to some real number a. Consider now the subsequence b_{n_k} of (b_n). Since (b_n) is bounded, so is (b_{n_k}), and so again applying the Bolzano-Weierstrass theorem for $\mathbb{R}$, there is a *further* subsequence $(b_{n_{k_\ell}})$ that converges to some $b \in \mathbb{R}$. The subsequence $(a_{n_{k_\ell}})$ is a subsequence of the convergent sequence a_{n_k} and hence also converges to a. Thus, by Propostion 3.13, the subsequence $z_{n_{k_\ell}}$ converges to $a + ib$ as $\ell \to \infty$. $\square$

Remark 3.16. This proof highlights an important technique with subsequences in higher dimensional spaces. We chose the second subsequence as a *sub*subsequence, not only a subsequence. Had we tried to select the subsequences of the real and imaginary parts independently, we could not have concluded anything about the two together. Indeed, the Bolzano-Weierstrass theorem gives us a convergent subsequence a_{n_k} and also gives us a convergent subsequence b_{m_k}. But we need to use the same index n for both a_n and b_n, which might not be possible with independent choices like this. A priori, the chosen convergent subsequence of a_n might have been $(a_1, a_3, a_5, \ldots)$, while from b_n we might have chosen $(b_2, b_4, b_6, \ldots)$, ne'er the 'tween shall meet.

4. SERIES

4.1. **Lecture 11: February 9, 2016.**

We now turn to a special class of sequences called **series**.

Definition 4.1. *Let (a_n) be a sequence in $\mathbb{R}$ or $\mathbb{C}$. The* **series** *associated to (a_n) is the new sequence (s_n) given by*

$$s_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n.$$

It is a bit of a misnomer to refer to series as special kinds of sequences; indeed, any sequence is the series associated to some other sequence. For let (a_n) be a sequence. Define a new sequence (b_n) by

$$b_1 = a_1, \qquad b_n = a_n - a_{n-1} \text{ for } n > 1.$$

Then $a_1 = b_1 = \sum_{k=1}^{1} b_k$, and for $n > 1$ we compute that

$$\sum_{k=1}^{n} b_k = b_1 + b_2 + \cdots + b_k = a_1 + (a_2 - a_1) + (a_3 - a_2) + \cdots + (a_n - a_{n-1}) = a_n.$$

Thus, (a_n) (the arbitrary sequence we started with) is the series associated to the sequence (b_n).

We will see, however, that the concept of convergence is quite different when applied to the series associated to a sequence rather than the sequence itself.

Definition 4.2. *Let (a_n) be a sequence in $\mathbb{R}$ or $\mathbb{C}$, and let $s_n = \sum_{k=1}^{n} a_k$ be its series. We say that the* **series converges** *if the sequence (s_n) converges. If the limit is $s = \lim_{n\to\infty} s_n$, we denote it by*

$$s = \sum_{n=1}^{\infty} a_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_k.$$

In this case, we will often use the cumbersome-but-standard notation "the series $\sum_{n=1}^{\infty} a_n$ converges".

Example 4.3 (Geometric Series). Let $r \in \mathbb{C}$, and consider the sequence $a_n = r^n$ (in this case it is customary to start at $n = 0$). We can compute the terms in the series exactly, following a trick purportedly invented by Gauss at age 10.

$$\begin{aligned} s_n &= \sum_{k=0}^{n} a_k = 1 + r + r^2 + \cdots + r^n. \\ \therefore \quad r s_n &= \qquad\quad r + r^2 + \cdots + r^n + r^{n+1}. \end{aligned}$$

So, subtracting the two lines, we have

$$(1-r)s_n = s_n - r s_n = 1 - r^{n+1}.$$

Now, if $r = 1$, this gives no information. In that degenerate case, we simply have $s_n = 1 + 1 + \cdots + 1 = n$, and this series does not converge. In all other cases, we have the explicit formula

$$s_n = \frac{1 - r^{n+1}}{1 - r}.$$

Using the limit theorems, we can decide whether this converges, and to what, just looking at the shifted sequence $a_{n+1} = r^{n+1}$. If $|r| < 1$, then this converges to 0. If $|r| \geq 1$, this sequence does not converge. (This is something you should work out.) Thus, we have

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \qquad \text{if } |r| < 1$$

while the series diverges if $|r| \geq 1$.

Example 4.4. Consider the sequence $a_n = \frac{1}{n(n+1)}$. We can employ a trick here: the partial fractions decomposition:

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}.$$

Thus, taking the series $s_n = \sum_{k=1}^{n} a_n$, we have

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1}\right) = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n} - \frac{1}{n+1}\right).$$

This is a *telescoping sum*: all terms except for the first and the last cancel in pairs. Thus, we have a closed formula

$$s_n = 1 - \frac{1}{n+1}.$$

Hence, the series converges, and we have explicitly

$$\sum_{n=1}^{\infty} = \lim_{n\to\infty} s_n = \lim_{n\to\infty} \left(1 - \frac{1}{n+1}\right) = 1.$$

Example 4.5 (Harmonic Series). Consider the series $s_n = \sum_{k=1}^{n} \frac{1}{k}$. That is

$$s_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

If you add up the first *billion* terms (i.e. s_{10^9}) you get about 21.3. This seems to suggest convergence; after all, the terms are getting arbitrarily small. However, this series *does not converge*. To see why, look at terms s_N with $N = 2^m + 2^{m-1} + \cdots + 2 + 1$ for some positive integer m. (By the way, from the previous example, this could be written explicitly as $N = 2^{m+1} - 1$.) Then we can group terms as

$$s_N = (1) + \left(\frac{1}{2} + \frac{1}{3}\right) + \left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right) + \cdots + \left(\frac{1}{2^m+1} + \frac{1}{2^m+2} + \cdots + \frac{1}{2^{m+1}-1}\right).$$

That is: we break up the sum into $m+1$ groups, the first group with 1 term, the second with 2, the third with 4, up to the last with 2^m terms. Now, $1 > \frac{1}{2}$. In the sec ond group of terms, both $\frac{1}{2}$ and $\frac{1}{3}$ are $> \frac{1}{4}$. In the next group, each of the four terms is $> \frac{1}{8}$. That is, we have

$$\begin{aligned} s_N &> \left(\frac{1}{2}\right) + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^{m+1}} + \frac{1}{2^{m+1}} + \cdots + \frac{1}{2^{m+1}}\right) \\ &= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2} = \frac{m+1}{2}. \end{aligned}$$

Now, $s_{n+1} = s_n + \frac{1}{n+1} \geq s_n$, so (s_n) is an increasing sequence. We've just shown that, for any integer m, we can find some time N so that $s_N \geq \frac{m+1}{2}$, and so it follows that for all *larger* $n \geq N$, $s_n \geq s_N \geq \frac{m+1}{2}$. Since $\frac{m+1}{2}$ is arbitrarily larger, we've just proved that $s_n \to +\infty$ as $n \to \infty$. So the series diverges.

In Example 4.3, we were able to compute the nth term in the series as a closed formula, and compute the limit directly. It is rare that we can do this explicitly; more often, we will need to make estimates like we did in Example 4.5. So we now begin to discuss some general tools for attacking such limits.

Proposition 4.6 (Cauchy Criterion). *Let a_n be a sequence in $\mathbb{R}$ or $\mathbb{C}$. Then the series $\sum_{n=1}^{\infty} a_n$ converges if and only if: for every $\epsilon > 0$, there is a natural number $N \in \mathbb{N}$ so that, for all $m \geq n \geq N$,*

$$\left| \sum_{k=n+1}^{m} a_k \right| < \epsilon.$$

Proof. This is just a restatement of the Cauchy completeness of $\mathbb{R}$ and $\mathbb{C}$. Let $s_n = \sum_{k=1}^{n} a_k$. Then

$$\sum_{k=n+1}^{m} = a_{n+1} + \cdots + a_m = s_m - s_n.$$

Thus, having decided to always use m to denote the larger of m, n, the statement is that, for every $\epsilon > 0$ there is $N \in \mathbb{N}$ such that, for $m \geq n \geq N$, $|s_m - s_n| < \epsilon$; this is precisely the statement that (s_n) is a Cauchy sequence. In $\mathbb{R}$ or $\mathbb{C}$, this is equivalent to (s_n) being convergent, as desired. $\square$

Corollary 4.7. *Let (a_n) be a sequence such that the series $\sum_{n=1}^{\infty} a_n$ converges. Then $a_n \to 0$.*

Proof. By the Cauchy Criterion (Proposition 4.6), given $\epsilon > 0$ we may find $N \in \mathbb{N}$ so that (letting $n = m - 1$) for $m > N$,

$$\epsilon > \left| \sum_{k=(m-1)+1}^{m} a_k \right| = |a_m|.$$

This is precisely the statement that $a_m \to 0$ as $m \to \infty$. $\square$

As Example 4.5 points out, the converse to Corollary 4.7 is false: there are sequences, such as $a_n = \frac{1}{n}$, that tend to 0, but for which the series $\sum_{n=1}^{\infty} a_n$ diverges.

It is often impossible to compute the exact value of the sum $\sum_{n=1}^{\infty} a_n$ of a convergent series. More often, we use estimates to approximate the value. More basically, we use estimates to determine whether the series converges or not, without any direct knowledge of the value if it does converge. The most basic test for convergence is the **comparison theorem**.

Theorem 4.8 (Comparison). *Let (a_n) and (b_n) be sequences in $\mathbb{C}$.*

(1) *If $b_n \geq 0$ and $\sum_n b_n$ converges, and if $|a_n| \leq b_n$ for all sufficiently large n, then $\sum_n a_n$ converges, and $|\sum_n a_n| \leq \sum_n b_n$.*
(2) *If $a_n \geq b_n \geq 0$ for all sufficiently large n and $\sum_n b_n$ diverges, then $\sum_n a_n$ diverges.*

Proof. For item 1: by assumption $\sum_n b_n$ converges, and so by the Cauchy criterion, for given $\epsilon > 0$ we can choose $N_0 \in \mathbb{N}$ so that, for $m \geq n \geq N_0$,

$$\sum_{k=n+1}^{m} b_k < \epsilon$$

(here we have used the fact that $b_n \geq 0$ to drop the modulus). Now, let N_1 be large enough that $|a_n| \leq b_n$ for $n \geq N_1$. Then for $m \geq n \geq \max\{N_0, N_1\}$, we have

$$\left|\sum_{k=n+1}^{m} a_k\right| \leq \sum_{k=n+1}^{m} |a_k| \leq \sum_{k=n+1}^{m} b_k < \epsilon.$$

So the series $\sum_n a_n$ satisfies the Cauchy criterion, and therefore is convergent.

Item 2 follows from item 1 by contrapositive: if $\sum_n a_n$ converges, then since $b_n = |b_n| \leq a_n$ for all large n, we have just proven that $\sum_n b_n$ converges. Thus, if $\sum_n b_n$ diverges, so must $\sum_n a_n$. □

Example 4.9. The series $\sum_n \frac{1}{\sqrt{n}}$ diverges, since $\frac{1}{\sqrt{n}} \geq \frac{1}{n}$ and, by Example 4.5, $\sum_n \frac{1}{n}$ diverges. On the other hand, note that

$$n^2 = \frac{1}{2}n^2 + \frac{1}{2}n^2 \geq \frac{1}{2}n^2 + \frac{1}{2}n = \frac{1}{2}n(n+1)$$

for $n \geq 1$. Thus $\frac{1}{n^2} \leq \frac{2}{n(n+1)}$. As we computed in Example 4.4, $\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1$, so by the limit theorems, $\sum_{n=1}^{\infty} \frac{2}{n(n+1)} = 2$. That is: this series converges. It follows from the comparison test that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges.

In showing that the harmonic series diverges, we broke the terms up into groups of exponentially increasing size. This is an important trick known as the **lacunary technique**, and it works well when the sequence of terms is positive and decreasing.

Proposition 4.10 (Lacunary Series). *Suppose (a_n) is a sequence of non-negative numbers that is decreasing: $a_n \geq a_{n+1} \geq 0$ for all n. Then $\sum_{n=1}^{\infty} a_n$ converges if and only if the series*

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots$$

converges.

Proof. Since $a_k \geq 0$ for all k, the series of partial sums $s_n = \sum_{k=1}^{n} a_k$ is monotone increasing. Hence, convergence of s_n is equivalent to the *boundedness* of (s_n). Let $t_k = a_1 + 2a_2 + 4a_4 + \cdots + 2^k a_k$. We will show that (s_n) is bounded iff (t_k) is bounded.

Note that $2^k \leq 2^{k+1} - 1$, and so if $n < 2^k$ then $n < 2^{k+1} - 1$. Then we have for such n

$$\begin{aligned} s_n = a_1 + a_2 + a_3 + \cdots + a_n &\leq a_1 + a_2 + a_3 + \cdots + a_{2^{k+1}-1} \\ &= (a_1) + (a_2 + a_3) + \cdots + (a_{2^k} + \cdots + a_{2^{k+1}-1}) \\ &\leq a_1 + 2a_2 + \cdots + 2^k a_{2^k} = t_k. \end{aligned}$$

(In the last inequality, we used the fact that a_n is decreasing.) This shows that if (t_k) is bounded, then so is (s_n). For the converse, we just group terms slightly differently (exactly as we did in the proof of the divergence of the harmonic series): if $n > 2^k$, then

$$\begin{aligned} s_n = a_1 + a_2 + a_3 + \cdots + a_n &\geq a_1 + a_2 + a_3 + \cdots + a_{2^k} \\ &= (a_1) + (a_2) + (a_3 + a_4) + \cdots + (a_{2^{k-1}+1} + \cdots + a_{2^k}) \\ &\geq \frac{1}{2}a_1 + a_2 + 2a_4 \cdots + 2^{k-1} a_{2^k} = \frac{1}{2}t_k. \end{aligned}$$

Thus $t_k \leq 2s_n$ whenever $n > 2^k$. This shows that if (s_n) is bounded then so is (t_k), concluding the proof. □

Example 4.11. Let $p \in \mathbb{R}$, and consider the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$. We've already seen that this series diverges when $p = 1$. If $p < 1$, then $\frac{1}{n^p} \geq \frac{1}{n}$; it follows by the comparison theorem that the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ diverges for $p \leq 1$.

On the other hand, consider $p > 1$. Here the sequence of terms $a_n = \frac{1}{n^p}$ is positive and decreasing, so we may use the lacunary series test to determine whether the series converges. Compute that

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = \sum_{k=0}^{\infty} 2^k \frac{1}{2^{kp}} = \sum_{k=0}^{\infty} 2^{(1-p)k}.$$

This is a geometric series, with base $r = 2^{1-p}$. So $0 < r < 1$ provided that $p > 1$, in which case the series converges. Hence, by Proposition 4.10, the series $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if and only if $p > 1$.

4.2. **Lecture 12: February 11, 2016.** Two generally effective tools for deciding convergence, that you already saw in your calculus class, are the **Root Test** and the **Ratio Test**. Both of them are predicated on rough comparison with a geometric series, cf. Example 4.3. If $a_n = r^n$, then $\sum_n a_n$ converges iff $|r| < 1$. Now, note for this series that this important constant $|r|$ can be computed either as $|a_n|^{1/n}$ or as $|\frac{a_{n+1}}{a_n}|$. Even when these quantities are not constant, they still can give a lot of information about the convergence of the series.

Theorem 4.12 (Root Test)**.** *Let (a_n) be a sequence in $\mathbb{C}$. Define*

$$\alpha = \limsup_{n\to\infty} |a_n|^{1/n}.$$

If $\alpha < 1$, then $\sum_n a_n$ converges. If $\alpha > 1$, then $\sum_n a_n$ diverges.

Remark 4.13. It is important to note that the theorem gives no information when $\alpha = 1$. Indeed, consider Examples 4.5 and 4.9, showing that $\sum_{n=1}^\infty \frac{1}{n}$ diverges, while $\sum_{n=1}^\infty \frac{1}{n^2}$ converges. But, in both cases, we have

$$\lim_{n\to\infty} \left(\frac{1}{n}\right)^{1/n} = \lim_{n\to\infty} \left(\frac{1}{n^2}\right)^{1/n} = 1$$

(see Theorem 3.20(c) in Rudin). Thus, $\limsup_n |a_n|^{1/n} = 1$ can happen whether $\sum_n a_n$ converges or diverges.

Proof. Suppose $\alpha < 1$. Then choose any $r \in \mathbb{R}$ with $\alpha < r < 1$. That is, we have $\limsup_n |a_n|^{1/n} < r$. Let $b_n = |a_n|^{1/n}$; then the statement is that $\limsup_n b_n = \lim_n \overline{b}_n < r$. That means that, for all sufficiently large n, $\overline{b}_n < r$, and so since $b_n \leq b_n$, we have $|a_n|^{1/n} < r$ for all sufficiently large n. That is: there is some N so that $|a_n| < r^n$ for $n \geq N$. Since the series $\sum_n r^n$ converges (as $0 < r < 1$), it now follows that $\sum_n a_n$ converges by the comparison theorem.

Now, suppose $\alpha > 1$. As above, let $b_n = |a_n|^{1/n}$. Since $\alpha = \limsup_n b_n$, from Theorem 2.24 there is a subsequence b_{n_k} that converges to α. (This is even true of $\alpha = +\infty$; in this case, it is quite easy to see that the series diverges.) In particular, this means that $b_{n_k} > 1$ for all k, and so $|a_{n_k}| = b_{n_k}^n > 1$ as well. It follows that a_n does not converge to 0, and so by Corollary 4.7, $\sum_n a_n$ diverges. □

The Ratio Test, which we state and prove below, is actually weaker than the Root Test. Its proof is based on comparison with the Root Test, using the following result.

Lemma 4.14. *Let c_n be a sequence of positive real numbers. Then*

$$\limsup_{n\to\infty} c_n^{1/n} \leq \limsup_{n\to\infty} \frac{c_{n+1}}{c_n}, \quad \textit{and}$$
$$\liminf_{n\to\infty} c_n^{1/n} \geq \liminf_{n\to\infty} \frac{c_{n+1}}{c_n}.$$

Proof. We prove the $\limsup$ inequality, and leave the similar $\liminf$ case as an exercise. Let $\gamma = \limsup_n \frac{c_{n+1}}{c_n}$. If $\gamma = +\infty$, there is nothing to prove, since every extended real number x satisfies $x \leq +\infty$. So, assume $\gamma \in \mathbb{R}$. Then we can choose some $\beta > \gamma$, and as in the proof of the Root Test above, it follows that $\frac{c_{n+1}}{c_n} < \beta$ for all sufficiently large n, say $n \geq N$. But then, by induction, we have

$$\frac{c_{N+k}}{c_N} = \frac{c_{N+k}}{c_{N+k-1}} \frac{c_{N+k-1}}{c_{N+k-2}} \cdots \frac{c_{N+1}}{c_N} < \beta^k.$$

Thus, for $n \geq N$, letting $k = n - N$, we have

$$c_n = c_{N+k} < C_N\beta^k = c_N\beta^{n-N} = (c_N\beta^{-N}) \cdot \beta^n$$

and so

$$c_n^{1/n} < (c_N\beta^{-N})^{1/n}\cdot\beta.$$

From the Squeeze Theorem, it follows that

$$\limsup_{n\to\infty} c_n^{1/n} \le \limsup_{n\to\infty}(c_N\beta^{-N})^{1/n}\cdot\beta = \beta\cdot\lim_{n\to\infty}(c_N\beta^{-N})^{1/n} = \beta.$$

(Here we have used the fact that $p = c_N\beta^{-N}$ is a positive constant, and $\lim_n p^{1/n} = 1$ for any $p > 0$; this last well-known limit can be found as Theorem 3.20(b) in Rudin.) Thus, for any $\beta > \gamma$, we have $\limsup_n c_n^{1/n} \le \beta$. It follows that $\limsup_n c_n^{1/n} \le \gamma$, as claimed. □

Theorem 4.15 (Ratio Test)**.** *Let (a_n) be a sequence in $\mathbb{C}$.*

(1) *If $\limsup_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| < 1$, then $\sum_n a_n$ converges.*
(2) *If $\liminf_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| > 1$, then $\sum_n a_n$ diverges.*

Proof. For (1): from Lemma 4.14, $\limsup_n |a_n|^{1/n} \le \limsup_n \frac{|a_{n+1}|}{|a_n|} < 1$, and so by the Root Test, $\sum_n a_n$ converges. For (2): from Lemma 4.14, $\limsup_n |a_n|^{1/n} \ge \liminf_n |a_n|^{1/n} \ge \liminf_n \frac{|a_{n+1}|}{|a_n|} > 1$, and so by the Root Test, $\sum_n a_n$ diverges. □

Remark 4.16. Once again, if the $\limsup$ or $\liminf$ of the ratio of successive terms $= 1$, the test cannot give any information: letting $a_n = \frac{1}{n}$ and $b_n = \frac{1}{n^2}$, in both cases we have $\lim_n |\frac{a_{n+1}}{a_n}| = \lim_n |\frac{b_{n+1}}{b_n}| = 1$, and yet $\sum_n a_n$ diverges while $\sum_n b_n$ converges.

Example 4.17. Consider the sequence $(a_n) = (\frac{1}{2}, \frac{1}{3}, \frac{1}{2^2}, \frac{1}{3^2}, \ldots)$. That is: $a_{2n-1} = \frac{1}{2^n}$ and $a_{2n} = \frac{1}{3^n}$ for $n \ge 1$. Thus $|a_{2n-1}|^{1/(2n-1)} = \left(\frac{1}{2}\right)^{n/(2n-1)} \to \frac{1}{\sqrt{2}}$ while $|a_{2n}|^{1/2n} = \left(\frac{1}{3}\right)^{n/2n} = \frac{1}{\sqrt{3}}$. Thus $\limsup_n |a_n|^{1/n} = \frac{1}{\sqrt{2}}$, and so by the Root Test, the series $\sum_n a_n$ converges. But the Ratio Test is no use here. Note that

$$\frac{a_{(2n-1)+1}}{a_{2n-1}} = \frac{1/3^n}{1/2^n} = \left(\frac{2}{3}\right)^n \to 0,$$
$$\frac{a_{2n+1}}{a_{2n}} = \frac{1/2^{n+1}}{1/3^n} = \frac{1}{2}\left(\frac{3}{2}\right)^n \to +\infty.$$

Thus $\limsup_n \frac{a_{n+1}}{a_n} = +\infty > 1$ while $\liminf_n \frac{a_{n+1}}{a_n} = 0 < 1$; so the Ratio Test gives no information.

Remark 4.18. You may remember the Ratio and Root Tests as being described as equivalent in your calculus class. This is only true if you restrict to the case when $\lim_n |\frac{a_{n+1}}{a_n}|$ exists. In this case, the limit is equal to both the $\liminf$ and the $\limsup$, and then Lemma 4.14 shows that $\lim_n |a_n|^{1/n}$ also exists. But this rules out series like the one above, that somehow "alternate" between different kinds of terms, all of which are shrinking fast enough for the series to converge.

Example 4.19 (The number e). Consider the series

$$\sum_{n=0}^{\infty}\frac{1}{n!}.$$

Note that the sequence of terms a_n satisfies $\frac{a_{n+1}}{a_n} = \frac{1/(n+1)!}{1/n!} = \frac{1}{n+1} \to 0$ as $n \to \infty$, and so by the ratio test the series converges. Its exact value is called e. It is sometimes called *Napier's* *constant*, since it was first alluded to in a table of logarithms in an appendix of a book written by the Scottish mathematician / natural philosopher John Napier, circa 1618. It was first directly studied by Jacob Bernoulli, who used the letter b to denote it. But, like everything else from that era, it was eventually Euler who proved much of what we know about it, and Euler called it e.

The approximate value is

$$e \approx 2.718281828459045235360287471352662497757247093699 95.$$

Fun fact: when Google went public in 2004, their IPO (initial public offering) was $2,718,281,828. *Nerrrrrrds.*

Lemma 4.20. *The number e is given by $e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$.*

Proof. Let $s_n = \sum_{k=0}^n \frac{1}{k!}$ be the nth partial sum of the series defining e. Let $t_n = (1+\frac{1}{n})^n$. Now, for fixed m, we can use the binomial theorem to expand

$$\left(1+\frac{1}{n}\right)^m = \sum_{k=0}^m \binom{m}{k}\frac{1}{n^k} = \sum_{k=0}^m \frac{m(m-1)\cdots(m-k+1)}{k!}\cdot\frac{1}{n^k}.$$

Write the kth term as

$$\frac{1}{k!}\cdot\frac{m}{n}\cdot\frac{m-1}{n}\cdots\frac{m-k+1}{n}.$$

Thus, specializing to the case $m=n$, we have

$$t_n = \left(1+\frac{1}{n}\right)^n = \sum_{k=0}^n \frac{1}{k!}\cdot 1\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n} < \sum_{k=0}^n \frac{1}{k!} = s_n.$$

We have yet to prove that $\lim_{n\to\infty} t_n$ exists, but since s_n converges to the finite number e, it follows from $t_n < s_n$ that t_n is bounded above, and so $\limsup_n t_n$ exists, and (by HW5)

$$\limsup_{n\to\infty} t_n \le \limsup_{n\to\infty} s_n = e.$$

Now, on the other hand, let m be fixed. Then for $n \ge m$,

$$t_n = \sum_{k=0}^n \frac{1}{k!}\cdot 1\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n} \ge \sum_{k=0}^m \frac{1}{k!}\cdot 1\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n} \equiv t_n^m.$$

So, for fixed m, the two sequences (t_n) and (t_n^m) are comparable: $t_n \ge t_n^m$. Again by HW5, and using the limit theorems (for the finite sum with m terms), we have

$$\liminf_{n\to\infty} t_n \ge \liminf_{n\to\infty} t_n^m = \sum_{k=0}^m \frac{1}{k!}\lim_{n\to\infty}\frac{n-1}{n}\cdots\frac{n-k+1}{n} = \sum_{k=0}^m \frac{1}{m!} = s_m.$$

As this holds true for every m, it follows from the squeeze theorem that

$$\liminf_{n\to\infty} t_n \ge \lim_{m\to\infty} s_m = e.$$

Thus

$$e \le \liminf_{n\to\infty} t_n \le \limsup_{n\to\infty} t_n \le e$$

which implies that the $\limsup$ and $\liminf$ are both equal to e, as claimed. $\square$

The second form of e, as a limit, is (one of the) reason(s) it is so important: this shows that e shows up in many problems related to compound interest or exponential decay. However, as a means of approximating e, this limit is *very* slow: for example

$$t_{10} \approx 2.5937 \text{ (4.6\% error)}, \quad t_{100} \approx 2.7048 \text{ (0.50\% error)}.$$

That is: you need 100 terms in order to get 2 digits of accuracy. On the other hand, s_{10} is within 10^{-8} of e, and s_{100} is so close to e my computer cannot compute the difference. But we can give a bound on this tiny error as follows. First note that s_n is increasing, so $|e - s_n| = e - s_n$. Now

$$e - s_n = \sum_{k=0}^{\infty} \frac{1}{k!} - \sum_{k=0}^{n} \frac{1}{k!} = \sum_{k=n+1}^{\infty} \frac{1}{k!}.$$

Now, we factor the terms (all of which have $k \ge n+1$) as

$$\frac{1}{k!} = \frac{1}{(n+1)!(n+2)(n+3)\cdots k} = \frac{1}{(n+1)!} \cdot \frac{1}{(n+2)(n+3)\cdots k} < \frac{1}{(n+1)!} \cdot \frac{1}{(n+1)^{k-n-1}}.$$

Thus

$$e - s_n < \frac{1}{(n+1)!} \sum_{k=n+1}^{\infty} \frac{1}{(n+1)^{k-n-1}} = \frac{1}{(n+1)!} \sum_{j=0}^{\infty} \frac{1}{(n+1)^j}.$$

This is a geometric series, and $0 < \frac{1}{n+1} < 1$, so we know the sum is

$$\sum_{j=0}^{\infty} \frac{1}{(n+1)^j} = \frac{1}{1 - \frac{1}{n+1}} = \frac{n+1}{n}.$$

Thus, we have our estimate:

$$e - s_n < \frac{1}{(n+1)!} \frac{n+1}{n} = \frac{1}{n! \cdot n}.$$

This is a *tiny* number. Since $10! = 3,628,800$, this shows that $e - s_{10} < \frac{1}{3\times 10^7}$ (and in fact it's 3 times smaller than this). For $n = 100$, we have $100! \cdot 100 \approx 10^{160}$, so s_{100} differs from e only after the 160th decimal digit!

This is one of the rare occasions where a perfectly practical question of error approximation actually allows us to prove something entirely theoretical.

Proposition 4.21. *The number e is irrational.*

Proof. For a contradiction, let us suppose $e \in \mathbb{Q}$. Since $e > 0$, this means there are positive integers m, n so that $e = \frac{m}{n}$. Now, from the above estimate, we have

$$0 < e - s_n < \frac{1}{n! \cdot n}, \qquad \therefore\ 0 < n!e - n!s_n < \frac{1}{n}.$$

Now,

$$n!s_n = n! \sum_{k=0}^{n} \frac{1}{k!} = \sum_{k=0}^{n} \frac{n!}{k!} = \sum_{k=0}^{n} n(n-1)\cdots(n-k+1) \in \mathbb{N}.$$

Also, by assumption $e = \frac{m}{n}$, and so $n!e = m \cdot (n-1)! \in \mathbb{N}$. Thus $\ell = n!e - n!s_n \in \mathbb{N}$. But this means $0 < \ell < \frac{1}{n}$ for some $n \in \mathbb{N}$, and that is a contradiction (there are no integers between 0 and $\frac{1}{n}$). $\square$

Moving to our final topic on the subject of series, let us consider *absolute convergence*.

Definition 4.22. *Let (a_n) be a sequence in $\mathbb{C}$. We say that the series $\sum_{n=1}^{\infty} a_n$* **converges absolutely** *if, in fact, $\sum_{n=1}^{\infty} |a_n|$ converges.*

Lemma 4.23. *If $\sum_{n=1}^{\infty} a_n$ converges absolutely, then it converges.*

Proof. This follows immediately from the Cauchy criterion. Fix $\epsilon > 0$ and choose $N \in \mathbb{N}$ large enough that, for $m > n \geq N$, $\sum_{k=n+1}^{m} |a_k| < \epsilon$. Then by the triangle inequality

$$\left| \sum_{k=n+1}^{m} a_k \right| \leq \sum_{k=n+1}^{m} |a_k| < \epsilon$$

and so $\sum_{n=1}^{\infty} a_n$ converges. □

4.3. **Lecture 13: February 16, 2016.** The converse of Lemma 4.23 is quite false. To see why, let us study one particular class of real series known as *alternating series*.

Proposition 4.24 (Alternating Series)**.** *Let $a_n \geq 0$ be a monotone decreasing sequence with limit $a_n \to 0$. Then $\sum_{n=1}^{\infty}(-1)^{n-1}a_n = a_1 - a_2 + a_3 - a_4 + \cdots$ converges.*

Proof. Fix $m > n \in \mathbb{N}$ and consider the tail sum

$$|(-1)^n a_{n+1} + (-1)^{n+1}a_{n+2} + (-1)^{m-1}a_m| = |a_{n+1} - a_{n+2} + \cdots \pm a_m|.$$

We consider two cases: either $m-n$ is even or odd. If it is even, then we can group the terms as

$$|(a_{n+1} - a_{n+2}) + \cdots + (a_{m-1} - a_m)| = (a_{n+1} - a_{n+2}) + \cdots + (a_{m-1} - a_m),$$

where we have used the fact that $a_n \downarrow$. On the other hand, we may group terms as

$$= a_{n+1} - (a_{n+2} - a_{n+3}) - (a_{n+4} - a_{n+5}) - \cdots - a_m \leq a_{n+1}.$$

On the other hand, if $n-m$ is odd, then by similar reasoning

$$\left|\sum_{k=n+1}^{m}(-1)^{k-1}a_k\right| = (a_{n+1} - a_{n+2}) + \cdots + (a_{m-2} - a_{m-1}) + a_m$$

and we may group this as

$$= a_{n+1} - (a_{n+2} - a_{n+3}) - \cdots - (a_{m-1} - a_m) \leq a_{n+1}.$$

Hence, in all cases, we have

$$\left|\sum_{k=n+1}^{m}(-1)^{k-1}a_k\right| \leq a_{n+1}.$$

Thus, fix $\epsilon > 0$. Since $a_n \to 0$, we may choose $N \in \mathbb{N}$ so that, for $n \geq N$, $a_n = |a_n| < \epsilon$. Since $a_{n+1} \leq a_n$, we therefore have $|\sum_{k=n+1}^{m}(-1)^{k-1}a_k| \leq a_{n+1} < \epsilon$ whenever $m > n \geq N$, which verifies the Cauchy criterion showing that $\sum_{n=1}^{\infty}(-1)^{n-1}a_n$ converges. $\square$

Example 4.25. The sequence $a_n = \frac{1}{n}$ is positive, decreasing, and satisfies $a_n \to 0$. Therefore, by Proposition 4.24,

$$\sum_{n=1}^{\infty} a_n = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

converges. (Remembering your calculus, it in fact converges to $\ln 2$.) This is known as the **alternating harmonic series**. Note that the absolute series $\sum_{n=1}^{\infty}\frac{1}{n}$ diverges. So this is an example of a series that is convergent but not absolutely convergent. These are sometimes called **conditionally convergent** series.

Conditionally convergent series have strange properties, particularly with regard to *rearrangements*. That is: suppose we reorder the terms. Continuing Example 4.25, rearrange the terms as follows.

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \frac{1}{7} - \cdots + \frac{1}{2n-1} - \frac{1}{4n-2} - \frac{1}{4n} + \cdots$$

Each of the terms in the alternating harmonic series appears exactly once in this sum. It is no longer alternating, so we cannot apply a theorem to tell whether it converges; but we can in fact sum it as follows: in each three-term group, simplify

$$\frac{1}{2n-1} - \frac{1}{4n-2} - \frac{1}{4n} = \left(\frac{1}{2n-1} - \frac{1}{4n-2}\right) - \frac{1}{4n} = \frac{1}{4n-2} - \frac{1}{4n} = \frac{1}{2}\left(\frac{1}{2n-1} - \frac{1}{2n}\right).$$

So, the sum of the whole rearranged series is

$$\frac{1}{2}\left(1-\frac{1}{2}\right)+\frac{1}{2}\left(\frac{1}{3}-\frac{1}{4}\right)+\cdots+\frac{1}{2}\left(\frac{1}{2n-1}-\frac{1}{2n}\right)+\cdots=\frac{1}{2}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}=\frac{1}{2}\ln 2.$$

That is: this rearrangement produces *half* the value of the original series!

This is always possible for a conditionally convergent series of real numbers. Riemann proved this: if $\sum_{n=1}^{\infty} a_n$ is conditionally convergent and $a_n \in \mathbb{R}$, then there is a rearrangement a'_n of the terms so that the sequence $s'_n = \sum_{k=1}^{n} a'_n$ has *any tail behavior possible*: given any α, β with $-\infty \le \alpha \le \beta \le +\infty$, one can find a rearrangement so that $\limsup_n s'_n = \beta$ and $\liminf_n s'_n = \alpha$. (This is proved as Theorem 3.54 in Rudin.) Fortunately, this kind of craziness is not possible for absolutely convergent series, as our final theorem in this section attests to.

Theorem 4.26. *Let (a_n) be a complex sequence such that $\sum_{n=1}^{\infty} |a_n|$ converges. Then for any rearrangement a'_n of a_n, $\sum_{n=1}^{\infty} a'_n = \sum_{n=1}^{\infty} a_n$.*

Proof. Fix ϵ, and choose $N \in \mathbb{N}$ so that $\sum_{k=n=1}^{m} |a_k| < \epsilon$ for $m > n \ge N$. Let $s_n = \sum_{k=1}^{n} a_k$ and $s'_n = \sum_{k=1}^{n} a'_k$. The numbers $1, 2, \ldots, N$ appear as indices in the rearranged sequence (a'_n) each exactly once, so there must be some finite p so that they all appear by time p in (a'_n). Thus, for $m > n \ge p$, in the difference $s_m - s'_m$ the N terms $a_1, \ldots, a_N$ cancel leaving only with (original) indices $> N$. Thus, by the choice of N, this difference is $\le |\sum_{k=N+1}^{m} a_k| \le \sum_{k=N+1}^{m} |a_k| < \epsilon$. This shows that the sequence $s_n - s'_n$ converges to 0, and it follows, since we know s_n converges to $\sum_{n=1}^{\infty} a_n$, that s'_n also converges to the sum. $\square$

5. METRIC SPACES

For the remainder of this course, we are going to generalize the concepts we've worked with (notably *convergence*) beyond the case of $\mathbb{R}$ or $\mathbb{C}$. The key to this generalization was already discussed in the generalization from $\mathbb{R}$ to $\mathbb{C}$: we replace the absolute value in $\mathbb{R}$ (defined in terms of the order relation) with the complex modulus in $\mathbb{C}$. For all of the same technology to work, only a few basic properties of the absolute value / modulus were needed: that $|x| \geq 0$, that $|x| = 0$ only when $x = 0$, and finally the triangle inequality $|x + y| \leq |x| + |y|$.

This last property requires a notion of addition, and we'd like to move beyond vector spaces. The trick is that, in the notion of convergence, the absolute value / modulus only ever comes up as a means of measuring *distance* between two elements: $|x - y|$. Thinking of it this way, what do the three key properties say?

- For any two elements x, y, the distance $|x - y|$ is ≥ 0.
- If the distance $|x - y|$ is 0, then actually $x = y$.
- The triangle inequality: for any *three* elements x, y, z, the distance $|x - z|$ is bounded above by $|x - y| + |y - z|$. (This is really why it's called the triangle inequality: draw the associated picture.) Indeed, we have

$$|x - z| = |(x - y) + (y - z)| \leq |x - y| + |y - z|.$$

Interpreted in this light, we don't need a notion of addition: everything can be stated purely in terms of the notion of distance (in this case given by $(x, y) \mapsto |x - y|$). We generalize thus.

Definition 5.1. *Let X be a nonempty set. A function $d\colon X \times X \to \mathbb{R}$ is called a* **metric** *if it satisfies the following three properties.*

(1) *For any $x, y \in X$, $d(x, y) = d(y, x) \geq 0$.*
(2) *For any $x, y \in X$, if $d(x, y) = 0$, then $x = y$.*
(3) *For any $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.*

The pair (X, d) is called a **metric space***.*

Example 5.2. (1) As above, if we let $d_{\mathbb{C}}(x, y) = |x - y|$, then $(\mathbb{C}, d_{\mathbb{C}})$ is a metric space. Same goes for $\mathbb{R}$ equipped with the restriction of $d_{\mathbb{C}}$ to $\mathbb{R}$.

(2) More generally, fix n, and consider the set $\mathbb{C}^n$ of n-tuples of real numbers. Define the **Euclidean norm** on $\mathbb{C}^n$ as follows:

$$\|(x_1, \ldots, x_n)\|_2 = \left(|x_1|^2 + \cdots + |x_n|^2\right)^{1/2}.$$

It is a simple but laborious exercise to verify that the **Euclidean metric** $d_2(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$ is a metric on $\mathbb{C}^n$. As above, the restriction to $\mathbb{R}^n$ is also a metric.

(3) There are many other, *different* metrics on $\mathbb{R}^n$. The best known are the p-metrics: for $1 \leq p < \infty$,

$$\|(x_1, \ldots, x_n)\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

There is also the ∞-norm, aka the sup norm

$$\|(x_1, \ldots, x_n)\|_\infty = \max\{|x_1|, \ldots, |x_n|\}.$$

As above, all of these norms yield metrics in the usual way, $d_p(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p$. Note: the definition still makes sense when $p < 1$, but it no longer gives a metric: the triangle inequality is violated. For example, taking $p = \frac{1}{2}$, we have

$$\|(9,1) + (16,0)\|_{1/2} = \|(25,1)\|_{1/2} = (25^{1/2} + 1^{1/2})^2 = 36$$
$$\|(9,1)\|_{1/2} + \|(16,0)\|_{1/2} = (9^{1/2} + 1^{1/2})^2 + (16^{1/2} + 0^{1/2})^2 = 32 < 36.$$

(4) Let $B[0,1]$ consist of all bounded functions $[0,1] \to \mathbb{R}$. Then define a function $d_u\colon B[0,1] \times B[0,1] \to \mathbb{R}$ by
$$d_u(f,g) = \sup_{x\in[0,1]} |f(x) - g(x)|.$$
This is well-defined: since f and g are bounded, the set $\{f(x) - g(x)\colon x \in [0,1]\}$ is a bounded, nonempty set, so it has a sup. It is ≥ 0, and moreover if $d_u(f,g) = 0$, then for every $x_0 \in [0,1]$, $|f(x_0) - g(x_0)| \leq \sup_{x\in[0,1]} |f(x) - g(x)| = 0$, which implies that $f(x_0) - g(x_0) = 0$ – i.e. $f = g$. This verifies the first two properties of Definition 5.1. For the triangle inequality, we have
$$\begin{aligned} d_u(f,h) = \sup_{x\in[0,1]} |f(x) - h(x)| &= \sup_{x\in[0,1]} |f(x) - g(x) + g(x) - h(x)| \\ &\leq \sup_{x\in[0,1]} (|f(x) - g(x)| + |g(x) - h(x)|) \\ &\leq \sup_{x\in[0,1]} |f(x) - g(x)| + \sup_{x\in[0,1]} |g(x) - h(x)| \\ &= d_u(f,g) + d_u(g,h) \end{aligned}$$
using the properties of sup we now know well. Thus the triangle inequality holds for d_u as well, and so it is a metric. Note, like the above examples, it has the form $d_u(f,g) = \|f - g\|_u$ for a *norm* $\|\cdot\|_u$: a function on $B[0,1]$ which has the properties $\|f\|_u \geq 0$ and $= 0$ only if $f = 0$, and satisfies the triangle inequality $\|f + g\|_u \leq \|f\|_u + \|g\|_u$. Whenever we have a function like this defined on a vector space, it gives rise to a metric by subtraction.

(5) Not every metric is given in terms of a norm like this. For example, consider on $\mathbb{R}$ the function
$$d(x,y) = \min\{|x - y|, 1\}.$$
It is easy to verify that this satisfies properties (1) and (2) in Definition 5.1. The triangle inequality is also easy to see, by breaking into eight cases (depending whether $|x - y|$, $|x - z|$, and $|y - z|$ are ≤ 1 or > 1); this boring proof is left to the reader.

(6) Given any nonempty set X, one can define a metric on X by the silly rule
$$d(x,y) = \begin{cases} 0, & x = y \\ 1, & x \neq y \end{cases}.$$
This is known as the **discrete metric**. It says two points are close only if they are equal; otherwise they are far apart. It is again simple to verify this is a metric.

One important observation was made at several points in the examples: if (X,d) is a metric space, and $Y \subseteq X$, then $(Y, d|_Y)$ is a metric space – that is, the metric Y defined on all pairs $(x,y) \in X \times X$, also defines a metric when restricted only to pairs in $Y \times Y$, as is straightforward to verify. Thus, the Euclidean metric on $\mathbb{C}^n$ automatically gives us a metric (also called the Euclidean metric) on $\mathbb{R}^n$. Similarly, the usual metric on $\mathbb{R}$ restricts to a metric on $[0,1]$.

Usually thinking of metric spaces using our intuition from $\mathbb{R}^2$ and $\mathbb{R}^3$, we introduce the following notation.

Definition 5.3. *Let (X, d) be a metric space, and let $x_0 \in X$. For a fixed $r > 0$, the* **ball** *of radius r centered at x_0, denoted $B_r(x_0)$, is the set*

$$B_r(x_0) = \{x \in X : d(x_0, x) < r\}.$$

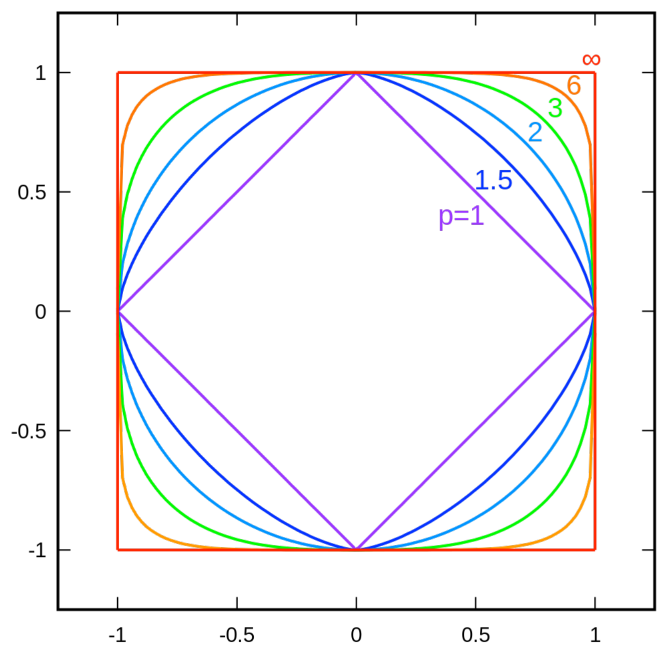
(Rudin calls this a **neighborhood** *$N_r(x_0)$.) With $r = 1$, we refer to this as the* **unit ball** *centered at x_0.*

Example 5.4. (1) In $\mathbb{R}^n$, using the definition of the Euclidean metric (and choosing the base point $\mathbf{0}$ to simplify things), we have

$$B_r(\mathbf{0}) = \{(x_1, \ldots, x_n) : x_1^2 + \cdots + x_n^2 < r^2\}$$

which is what we usually know as a ball (in n-dimensions).

(2) Consider $(\mathbb{R}^2, d_p)$, with the p-metric of Example 5.2(3). Here are some pictures of the unit ball:



(3) In a discrete metric space (X, d) as in Example 5.2(6), $B_r(x_0) = X$ if $r > 1$, and $B_r(x_0) = \{x_0\}$ if $r \leq 1$.

5.1. **Lecture 14: February 18, 2016.** Now, once more, we define **convergence** and **Cauchy** in the wider world of metric spaces.

Definition 5.5. *Let (X, d) be a metric space, and let (x_n) be a sequence in X, and let $x \in X$. Say that (x_n)* **converges to** *x, or $x_n \to x$, or $\lim_{n\to\infty} x_n = x$, if*

$$\forall \epsilon > 0\, \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N\ d(x_n, x) < \epsilon.$$

In other words, $x_n \to x$ means that the real sequence $d(x_n, x) \to 0$. Alternatively, we could state this as: for all sufficiently large n, $x_n \in B_\epsilon(x)$.

Similarly, say that (x_n) **is a Cauchy sequence** *in X if*

$$\forall \epsilon > 0\, \exists N \in \mathbb{N} \text{ s.t. } \forall n, m \geq N\ d(x_n, x_m) < \epsilon.$$

As discussed in the generalization from $\mathbb{R}$ to $\mathbb{C}$, limits are unique: if $x_n \to x$ and $x_n \to y$, then $x = y$ (this follows from the fact that $d(x, y) = 0$ implies that $x = y$; it is primarily for this reason that this *non-degeneracy* property is required in the definition of a metric).

Example 5.6. Consider a discrete metric space: (X, d) where d is given as in Example 5.2(6). Let $x_n \to x$. In particular, this means that there is some time N such that, for $n \geq N$, $d(x_n, x) < \frac{1}{2}$. But by the definition of d, either $d(x_n, x) = 0$ or $d(x_n, x) = 1$; so if $d(x_n, x) < \frac{1}{2}$ then $d(x_n, x) = 0$, and so $x_n = x$. Thus, if $x_n \to x$, then $x_n = x$ for all large n. In a discrete metric space, *convergence* is the same thing as *eventually constant*. The same holds true for Cauchy.

In general, there is a fundamental difference between convergent sequences that are eventually constant and convergent sequences that are *not* eventually constant. We use this difference to define one of the most important topological concepts.

Definition 5.7. *Let (X, d) be a metric space, and let $E \subseteq X$ be a subset. A point $x \in X$ (not necessarily in E) is called a* **limit point** *of E if there is a sequence $x_n \in E \setminus \{x\}$ that converges to x, $x_n \to x$. That is: a limit point of E is a limit of some* not eventually constant *sequence in E. A point $e \in E$ that is* not *a limit point of E is called an* **isolated point** *of E.*

Example 5.8. In $\mathbb{R}$ with the usual metric, take $E = (-1, 0] \cup \mathbb{N}$. Then -1 is a limit point of E: for example, $-1 = \lim(-1 + \frac{1}{n})$ and $-1 + \frac{1}{n} \in E$ for each n. Also, any point $x \in E$ is a limit point of E: take $x_n = x - \frac{1+x}{n}$ as the sequence. This is in E since $1 + x > 0$ and so $x - \frac{1+x}{n} < x \leq 0$, but also $x - \frac{1+x}{n} > x - (1 + x) = -1$.

On the other hand, the positive integers $\mathbb{N}$ are isolated points of E. For example, consider 1. If y_n is any sequence in $\mathbb{R}$ that converges to 1, then we must have $y_n \in (0.9, 1.1)$ for all large n; but then if $y_n \in E$ it follows that $y_n = 1$ for all large n, which isn't allowed. Thus, no sequence in $E \setminus \{1\}$ converges to 1, showing that the point $1 \in E$ is not a limit point of E – it is an isolated point.

The set of all limit points of a set E is denoted E'. So E is closed iff $E' \subseteq E$.

Definition 5.9. *A subset E of a metric space is called* **closed** *if it contains all of its limit points.*

Example 5.10.
(1) The set $E = (-1, 0] \cup \mathbb{N}$ from Example 5.8 is not closed: -1 is a limit point of E, but $-1 \notin E$.
(2) The set $F = [-1, 0] \cup \mathbb{N}$ *is* closed. The argument in Example 5.8 shows that each of the points in $[-1, 0]$ is a limit point of F, while each of the points $n \in \mathbb{N}$ is an isolated point of F. On the other hand, if x is a real number not in F, then either $x < -1$ or $x > 0$ and $x \notin \mathbb{N}$. In the former case, this means that no sequence in F can come within distance

$1+x>0$ of x, and so cannot converge to x; a similar argument with $x>0$ shows that x is not a limit point of F. Thus, the set of limit points of F consists exactly of the set $[-1,0]$, and this set is contained in F. So F is closed.

Definition 5.9 is stated in terms of limit points to make it clear that there are two kinds of points to consider in deciding whether a set is closed: isolated points and non-isolated points. For example, if one has a closed set, then adding to it a finite collection of isolated points will preserve closedness. But for the purposes of a concise definition, one need not be concerned about the distinction.

Proposition 5.11. *A subset E of a metric space (X,d) is closed if and only if, for any sequence (x_n) in E that converges in X, the limit $\lim_{n\to\infty} x_n$ is actually in E.*

That is: *closed* means *closed under limits of sequences*.

Proof. Suppose E is closed, so $E' \subseteq E$. Now, let (x_n) be any sequence in E that converges to some point x. If $x_n \neq x$ for any n, then by definition $x \in E'$, and therefore by assumption $x \in E$. If, on the other hand, there exists n with $x_n = x$, then since $x_n \in E$ for each n, we have $x \in E$. Thus, the E is closed under limits of sequences.

Conversely, suppose E is closed under limits of sequences. Let $x \in E'$; so by definition there is a sequence $x_n \in E \setminus \{x\}$ such that $x_n \to x$. Well, since $x_n \to x$ and $x_n \in E$, by assumption $x \in E$. Thus $E' \subseteq E$, and E is closed. $\square$

There is a complementary notion to *closed*, called *open*.

Definition 5.12. *A subset E of a metric space is called* **open** *if, for any point $x \in E$, there is a ball $B_r(x)$ (for some $r>0$) with $B_r(x) \subseteq E$.*

Example 5.13.
(1) The set $E = (-1,0] \cup \mathbb{N}$ from Example 5.8 is *not* open. Consider the point $0 \in E$. For any $0<r<1$, the ball $B_r(0) = (-r,r)$ contains some points (for example $\frac{r}{2}$) that are in $(0,1)$, and hence not in E. Similarly, any of the points in $\mathbb{N}$ are in E but none is contained in a ball contained in E. So E is not open.
(2) On the other hand, the set $U=(0,1)$ *is* open. Indeed, let $x \in U$. Let's consider two cases: either $0<x<\frac{1}{2}$ or $\frac{1}{2} \le x < 1$. In the former case, the ball $B_x(x) = (0,2x)$ is contained in $U=(0,1)$; in the latter case, the ball $B_{1-x}(x) = (2x-1,1)$ is contained in U. So every point of x is contained in a ball inside U, showing that U is open.
(3) Let (X,d) be a discrete metric space. If $x \in X$, then by Example 5.4(3), $B_1(x) = \{x\}$. Thus, every singleton point in a discrete metric space is an open set. On the other hand, by Example 5.6, there are no non-eventually-constant sequences converging to any point x, which means every point is isolated. That is: X has no limit points, which means that (vacuously) X contains all its limit points. So X is also closed.
(4) Consider the empty set $\varnothing$ in any metric space. It is both open and closed. Indeed, the definitions of "open" and "closed" each start with "for every point in the set. . . " and since there are no points in $\varnothing$ to check the condition, it follows that the condition holds vacuously.

Example 5.13(2) has a nice, important generalization to any metric space. Not that $(0,1)$ is itself a ball in $\mathbb{R}$: it is the ball $B_{1/2}(1/2)$. The fact is, any ball is open.

Proposition 5.14. *Let (X,d) be a metric space, let $x \in X$, and let $r>0$. Then the ball $B_r(x)$ is open in X.*

Proof. Let $y \in B_r(x)$. This means $d(x,y) < r$. Hence, there is some $\epsilon > 0$ so that $d(x,y) = r - \epsilon$. I claim that $B_\epsilon(y) \subset B_r(x)$. Indeed, suppose that $z \in B_\epsilon(y)$, meaning that $d(z,y) < \epsilon$. Then

$$d(x,z) \leq d(x,y) + d(y,z) = r - \epsilon + d(y,z) < r - \epsilon + \epsilon = r.$$

so $z \in B_r(x)$. We have thus shown that, for any $y \in B_r(x)$, there is a ball $B_\epsilon(y) \subset B_r(x)$. That is: $B_r(x)$ is open. $\square$

5.2. **Lecture 15: February 23, 2016.** We referred to *open* and *closed* as complementary properties. That doesn't mean that any set is either open or closed: for example, the set $(-1,0]$ considered above is neither open nor closed. But they concepts are complementary, in the following precise sense.

Proposition 5.15. *Let (X,d) be a metric space. A subset $E\subseteq X$ is open if and only if $E^c = X\setminus E$ is closed.*

Since $(E^c)^c = E$, it follows similarly that E is closed iff E^c is open. In the proof we will use the characterization of *closed* given in Proposition 5.11.

Proof. Suppose E is open. Let (x_n) be a sequence in E^c that converges to some point $x\in X$. We want to show that $x\in E^c$; to produce a contradition, we therefore assume that $x\notin E^c$, meaning $x\in E$. Since E is open, by definition there is some $r>0$ so that $B_r(x)\subseteq E$. On the other hand, since $x_n\to x$, there is certainly some N so that $d(x_N,x)<r$. Thus $x_N\in B_r(x)\subseteq E$, which means that $x_N\in E$. But we assumed that $x_N\in E^c$, so this is a contradition. Therefore $x\in E^c$. This shows that any convergent sequence in E^c has limit in E^c, which shows that E^c is closed.

Conversely, suppose E^c is closed. Let $x\in E$. We want to show that there is some $r>0$ so that $B_r(x)\subseteq E$; to produce a contradiction, we therefore assume that there is no such r. That means that, for say $r=\frac{1}{n}$, $B_{1/n}(x)\nsubseteq E$, which means precisely that there is som point $x_n\notin E$ such that $x_n\in B_{1/n}(x)$. So, we have produced a sequence $x_n\in E^c$ such that $d(x_n,x)<\frac{1}{n}$, meaning that $x_n\to x$. Thus x is the limit of a sequence in E^c, and so by assumption $x\in E^c$. This contradicts the assumption that $x\in E$. Therefore there must be some $r>0$ so that $B_r(x)\subseteq E$, and so E^c is closed. $\square$

Let us make a few more definitions that pertain the local properties of open and closed sets.

Definition 5.16. *Let (X,d) be a metric space, and let $E\subseteq X$.*

(1) *The* **closure** *of E is the set $\overline{E} = E\cup E'$.*
(2) *A point $x\in E$ is called an* **interior point** *if there is some $r>0$ with $B_r(x)\subseteq E$. The set of all interior points of E is called the* **interior** *of E, and is denoted $\mathring{E}$.*
(3) *The* **boundary** *of E is the set $\partial E = \overline{E}\setminus\mathring{E}$.*

Remark 5.17. Following the proof of Proposition 5.11, $\overline{E}$ can alternatively be described as the set of all limits of convergent sequences in E.

Example 5.18. Consider again the set $E=(-1,0]\cup\mathbb{N}$ in $\mathbb{R}$, considered in Examples 5.8 and 5.10. We've shown that the points in $(-\infty,-1)$ and $(0,\infty)$ are not limit points (the points $1,2,3,\ldots$ are in E but are isolated); on the other hand, we've shown that the points $[-1,0]$ are all limit points. Thus $E'=[-1,0]$, and so $\overline{E}=E\cup E'=[-1,0]\cup\mathbb{N}$. We've also shown in Example 5.13(1) that there is no ball centered at 0 contained in E; similarly, there are no balls centered at the points $1,2,3,\ldots$ contained in E, so these are not interior points. On the other hand, an argument very similar to Example 5.13(2) shows that the points in $(-1,0)$ are interior points. So $\mathring{E}=(-1,0)$. Finally, this shows that $\partial E=\overline{E}\setminus\mathring{E}=\{-1,0,1,2,\ldots\}$.

Example 5.19. Let $\mathbb{Q}$ denote the rational numbers as a subset of the metric space $\mathbb{R}$. By Theorem 1.17(2) (the density of $\mathbb{Q}$ in $\mathbb{R}$), given any real numbers $a<b$ there is a rational number $q\in\mathbb{Q}$ with $a<q<b$. In particular, fix $x\in\mathbb{R}$; then for $n\in\mathbb{N}$ there is a rational number q_n with

$x+\frac{1}{2n} < q_n < x+\frac{1}{n}$. In particular, we have $\frac{1}{2n} < |q_n - x| < \frac{1}{n}$. This shows that $q_n \to x$ but $q_n \neq x$ for any n; thus $x \in \mathbb{Q}'$. So every real number is a limit point of $\mathbb{Q}$, and so $\overline{\mathbb{Q}} = \mathbb{Q} \cup \mathbb{Q}' = \mathbb{Q} \cup \mathbb{R} = \mathbb{R}$. (This is another way of saying $\mathbb{Q}$ is dense in $\mathbb{R}$; in general, we say a subset $E \subseteq X$ is **dense** in a metric space X if $\overline{E} = X$.)

On the other hand, let $r > 0$, and let $q \in \mathbb{Q}$. The number $x = q + \frac{r}{\sqrt{2}}$ is $< x + r$, which shows that $x \in (q-r, q+r) = B_r(q)$. But $x \notin \mathbb{Q}$: indeed, we can solve $\sqrt{2} = \frac{r}{x-q}$, and so if x were rational $\sqrt{2}$ would also be rational, which we know it is not. Thus $B_r(q) \nsubseteq \mathbb{Q}$ for any $r > 0$. This shows q is not interior to $\mathbb{Q}$. This holds for any $q \in \mathbb{Q}$, and so, in fact, $\overset{\circ}{\mathbb{Q}} = \varnothing$.

Theorem 5.20. *Let (X, d) be a metric space, and let $E \subseteq X$.*

(1) *$\overline{E}$ is closed; E is closed iff $E = \overline{E}$.*
(2) *$\overset{\circ}{E}$ is open; E is open iff $E = \overset{\circ}{E}$.*

Proof. We begin with item 1. Let (x_n) be a sequence in $\overline{E}$ with limit x. We wish to show $x \in \overline{E}$. If $x \in E \subseteq \overline{E}$ we are done, so assume $x \notin E$. For each x_n, either $x_n \in E$ or $x_n \in E'$. In the latter case, by definition of E' there is some other sequence $y_k \in E$ such that $y_k \to x_n$; in particular, we can choose some k large enough that $d(y_k, x_n) < \frac{1}{n}$. So, we can define a new sequence (x_n') as follows: if $x_n \in E$ then $x_n' = x_n$; if $x_n \notin E'$, then $x_n' = y_k$ as above, so $x_n' \in E$ and $d(x_n, x_n') < \frac{1}{n}$. Then we have $d(x_n', x) \leq d(x_n', x_n) + d(x_n, x) < \frac{1}{n} + d(x_n, x) \to 0$, and so $x_n' \to x$. As $x_n' \in E$ and $x \notin E$, it follows that x is a limit point of E, and so $x \in E' \subseteq E \cup E' = \overline{E}$. Thus $\overline{E}$ is closed under limits; by Proposition 5.11, it follows that $\overline{E}$ is closed, as claimed. Now, by definition E is closed iff $E' \subseteq E$, and this happens iff $\overline{E} = E \cup E' = E$, proving the second point.

For item 2, let $x \in \overset{\circ}{E}$; thus, there is some ball $B_r(x)$ contained in E. But by Proposition 5.14, the ball $B_r(x)$ is open, which means all its points are interior points; thus $B_r(x) \subseteq \overset{\circ}{E}$. So, any point in $\overset{\circ}{E}$ is interior to $\overset{\circ}{E}$, which shows that $\overset{\circ}{E}$ is open. By definition $\overset{\circ}{E} \subseteq E$ for any set E; thus $\overset{\circ}{E} = E$ iff $E \subseteq \overset{\circ}{E}$, which is the statement that every point of E is an interior point, which is precisely the definition of E being open. $\square$

Example 5.21. Let $E \subset \mathbb{R}$ be nonempty and bounded above. Then $\alpha = \sup E$ exists. By definition, $\alpha - \frac{1}{n}$ is not an upper bound for E for any $n \in \mathbb{N}$, which shows that there is an element $x_n \in E$ with $\alpha - \frac{1}{n} < x_n \leq \alpha$. This shows that $x_n \to \alpha$. By Remark 5.17, it follows that $\alpha \in \overline{E}$: the supremum is always in the closure. On the other hand, if there were some $r > 0$ with $B_r(\alpha) \subseteq E$, then, for example, $\alpha + \frac{r}{2} \in E$. Since $\alpha + \frac{r}{2} > \alpha$, this contradicts α being an upper bound for E. Thus, α is not in $\overset{\circ}{E}$. That is: $\sup E \in \overline{E} \setminus \overset{\circ}{E} = \partial E$.

5.3. **Lecture 16: February 25, 2016.** Now we come to an important concept you may not have encountered before: *compactness*.

Definition 5.22. *Let (X,d) be a metric space. A subset $K \subseteq X$ is called* **compact** *if every sequence (x_n) in K has a convergent subsequence whose limit is in K.*

Example 5.23.
(1) Let $a < b$ be real numbers, and consider the set $K = [a,b]$. The Bolzano-Weierstrass Theorem for $\mathbb{R}$ (Theorem 2.25) is precisely the statement that $[a,b]$ is compact.
(2) On the other hand, $E = [a,b)$ is not compact: the sequence $x_n = b - \frac{b-a}{n}$ is in E, but converges to $b \notin E$, therefore all of its subsequences converge to b, and hence none of them converge in E. Similarly, an unbounded interval like $[0,\infty)$ is not compact: for example the sequence $x_n = n$ has no convergent subsequences at all.
(3) Let (X,d) be a discrete metric space. If K is a *finite* subset of X, say $K = \{y_1, y_2, \ldots, y_m\}$, then K is compact. Indeed, if (x_n) is any sequence in K, then there must be some (perhaps many) y_j so that $x_n = y_j$ for infinitely many n (by the pigeonhole principle). That means exactly that there is an increasing sequence n_k with $x_{n_k} = y_j$ for all k, which means $x_{n_k} \to y_j \in K$. Thus K is compact. On the other hand if $E \subseteq X$ is infinite, it is not compact: for then we can find an infinite sequence $x_1, x_2, x_3, \ldots \in E$ all distinct. Thus, any subsequence also has all distinct terms, which means it is not eventually constant. By Example 5.6, this means no subsequence converges.

Now, there is an alternate definition of compactness which is the only one used in Rudin; we refer to it as **topological compactness**, given in Definition 5.24 below. First, let us highlight the fact that Definition 5.22 was the *original* definition of compact, and predated the so-called "modern" definition by almost a century. Bolzano was already using our definition of compactness in 1817, although it would not be until 1906 that Definition 5.22 was written down formally (by Fréchet). It was around this time that Lebesgue proved (as a useful lemma) that Definition 5.24 also characterizes compactness; indeed, as we will see, it is a very useful tool. Much later, in 1929, the Russian school (led by Alexandrov and Urysohn) *redefined* compactness as what we are calling topological compactness. Our definition of the word *compact* is now often called **sequentially compact**.

Definition 5.24. *Let (X,d) be a metric space. Let $K \subseteq X$ be a subset. An* **open cover** *of K is a collection (finite or infinite) of open set $\mathscr{C}$ in X such that every point in K is in at least on $U \in \mathscr{C}$: that is $X \subseteq \bigcup \mathscr{U}$. We call K* **topologically compact** *if, given* any *open cover $\mathscr{C}$ of K, there is a* finite sub cover*: that is, there are finitely many $U_1, \ldots, U_m \in \mathscr{C}$ such that $K \subseteq U_1 \cup \cdots \cup U_m$.*

Example 5.25. Consider the interval $(0,1]$. We have already seen this is not compact. It is also not topologically compact. Indeed, consider the sets $U_n = (\frac{1}{n}, 2)$ for $n \in \mathbb{N}$. If $x \in (0,1)$ then $x > 0$ and so there is some $n \in \mathbb{N}$ with $\frac{1}{n} < x$. Therefore $x \in (\frac{1}{n}, 1) \subset (\frac{1}{n}, 2) = U_n$. This shows that the collection $\mathscr{C} = \{U_n : n \in \mathbb{N}\}$ is an open cover of $(0,1]$. Now, consider *any* finite collection of sets from $\mathscr{C}$: $U_{n_1}, U_{n_2}, \ldots, U_{n_k}$ for some $k \in \mathbb{N}$. Note that $\frac{1}{m} < \frac{1}{\ell}$ when $m > \ell$, and so $U_\ell \subset U_m$ in this case. What that means is that, if we let $m = \max\{n_1, \ldots, n_k\}$ then $U_{n_1} \cup \cdots \cup U_{n_k} = U_m = (\frac{1}{m}, 2)$. But then this does *not* cover $(0,1]$: there are points $x \in (0,1]$ with $x < \frac{1}{m}$. Thus, no finite subcover of $\mathscr{C}$ will cover all of $(0,1]$. The existence of such an open cover without any finite subcover shows that $(0,1]$ is not topologically compact.

Theorem 5.26. *Let K be a set in a subset of a metric space. Then K is sequentially compact iff K is topologically compact.*

We will not prove Theorem 5.26 here; this is the sort of thing that will be covered in an undergraduate topology course (such as Math 190). Rudin chooses to use the more abstract topological definition of compactness (for historical reasons that I find unsatisfactory), and this has the effect of both making everything more abstract, and also making all the proofs *harder than necessary*. We will stick exclusively with *sequential compactness*. This means all our proofs will be different from Rudin's – and generally shorter and easier to understand!

In Example 5.23(2), the absence of the point b from $[a, b)$ makes the set non-compact. Note that b is in the closure of $[a, b)$. This highlights the following proposition.

Proposition 5.27. *Compact sets are closed. Also, if K is compact and $F \subseteq K$ is closed, then F is compact.*

Proof. Suppose K is compact. Let (x_n) be a sequence in K which converges. By compactness, there is some subsequence (x_{n_k}) that converges to a point in K. But we know that every subsequence of (x_n) converges to $\lim x_n$, and hence $\lim x_n \in K$. Thus, K is closed under limits, and so K is closed.

Now, let F be a closed subset of a compact set K. Let (y_n) be any sequence in F. Then (y_n) is a sequence in K, and hence by compactness there is a subsequence (y_{n_k}) that converges in K. Note that $y_{n_k} \in F$ for each k, and hence since F is closed it follows that $\lim y_{n_k} \in F$. Thus, any sequence in F has a convergent subsequence with limit in F; i.e. F is compact. $\square$

Definition 5.28. *Let E be a subset of a metric space. The* **diameter** *of E, denoted $\mathrm{diam}(E)$ is defined to be*

$$\mathrm{diam}(E) = \sup\{d(x, y) \colon x, y \in E\}.$$

Note: this might well be $+\infty$. If $\mathrm{diam}(E) < +\infty$, we call E **bounded***; otherwise E is* **unbounded**.

Example 5.29.
1. $\mathrm{diam}(B_r(x)) \leq 2r$ for any ball in a metric space. But it could be less: for example in a discrete metric space with at least two elements, $\mathrm{diam}(B_r(x)) = 0$ if $r \leq 1$ and $= 1$ if $r > 1$.
2. In $\mathbb{R}$, $\mathrm{diam}(0, 1) = \mathrm{diam}(0, 1] = \mathrm{diam}[0, 1) = \mathrm{diam}[0, 1] = 1$.
3. In $\mathbb{R}$, $\mathrm{diam}(\mathbb{N}) = \infty$. Indeed, $d(0, n) = n$ so $\sup\{d(x, y) \colon x, y \in \mathbb{N}\} \geq n$ for every n.

Note: if E is a bounded set, with diameter $\delta > 0$, then for any point $x \in E$, $E \subseteq B_{2\delta}(x)$ (or $B_{1.0001\delta}(x)$, or $B_{\delta+0.0001}(x)$, etc.) Conversely, suppose there is some x in the metric space and some $r > 0$ with $E \subseteq B_r(x)$. Since $\mathrm{diam}(B_r(x)) \leq 2r$, it follows that $\mathrm{diam}(E) \leq 2r$. So, to say E is bounded is the same as saying it is contained in some ball.

Proposition 5.30. *Compact sets are bounded.*

Proof. We prove the contrapositive: *un*bounded sets are *not* compact. Let E be unbounded, and fix a point $x_0 \in E$. Consider the set of balls $B_n(x_0)$ for $n \in \mathbb{N}$. By assumption, $E \not\subseteq B_n(x_0)$ for any n, so we can choose a point $x_n \in E$ with $d(x_0, x_n) \geq n$.

In fact, the sequence (x_n) has no convergent subsequences. For let x be any point in the metric space. Let $n \in \mathbb{N}$ be large enough that $N > d(x_0, x)$. Then for $n \geq N + 1$, we have by the triangle inequality

$$d(x_n, x) \geq d(x_n, x_0) - d(x_0, x) \geq n - d(x_0, x) \geq 1 + N - d(x_0, x) > 1.$$

That is: for any point x in the metric space, eventually x_n never comes within distance 1 of x. It follows that no subsequence of (x_n) can converge to x. Since this holds for any x, it follows that (x_n) has no convergent subsequences. Since (x_n) is a sequence in E, this means E is not compact. $\square$

Thus, we have seen that compact sets are closed and bounded. One of the biggest theorems of this course, the Heine–Borel Theorem, states that the converse is true *in Euclidean space*.

Theorem 5.31 (Heine–Borel)**.** *Let $m \in \mathbb{N}$. A subset of $\mathbb{R}^m$ is compact iff it is closed and bounded.*

Proof. Let $K \subset \mathbb{R}^m$. If K is compact, then by Propositions 5.27 and 5.30 K is closed and bounded. We must prove the converse. Suppose K is a closed and bounded subset of $\mathbb{R}^m$. Let (x_n) be a sequence in K. We may write it in terms of its components

$$x_n = (x_n^1, x_n^2, \ldots, x_n^m).$$

Consider first the sequence $(x_n^1)_{n=1}^\infty$ in $\mathbb{R}$. Note that

$$|x_n^1| \le |x_n| = d(x_n, x_1) \le \mathrm{diam}(K).$$

So the sequence (x_n^1) is a bounded sequence in $\mathbb{R}$. By Theorem 2.25 (the Bolzano-Weierstrass Theorem for $\mathbb{R}$), there is a subsequence $x_{n_k}^1$ that converges. Now we proceed as in the proof of Theorem 3.15 (the Bolzano-Weierstrass Theorem for $\mathbb{C}$). Consider the subsequence $x_{n_k}^2$. Again we have $|x_{n_k}^2| \le \mathrm{diam}(K)$ is bounded, so by the Bolzano-Weierstrass Theorem for $\mathbb{R}$, it possesses a further subsequence $x_{n_{k_\ell}}^2$ that is convergent. Note that $x_{n_{k_\ell}}^1$ is a subsequence of the convergent subsequence $x_{n_k}^1$, so it is also convergent. Now we proceed to select a further convergent subsubsubsequence that makes $x_{n_{k_{\ell_s}}}^3$ converge, and so forth. The notation becomes ridiculous, but in the end (after m steps) we produce a single set of indices $1 \le \ell_1 < \ell_2 < \cdots$ such that all of the components $(x_{\ell_n}^1, x_{\ell_n}^2, \ldots, x_{\ell_n}^m)$ converge as $n \to \infty$. We now follow the proof of Proposition 3.13 exactly to see that convergence in $\mathbb{R}^m$ is equivalent to convergence of each component separately, and so we conclude that the subsequence (x_{ℓ_n}) converges to some element $x \in \mathbb{R}^m$. Finally, note that $x_{\ell_n} \in K$ by assumption, and K is closed; thus the point x is also in K. This shows that every sequence in K has a convergent subsequence with limit in K, concluding the proof that K is compact. $\square$

5.4. **Lecture 17: February 29, 2016.** So, closed intervals $[a, b]$ are compact (as we already knew), as are sets like $[0, 1] \cup [2, 3] \cup \{4, 5, 6, 7, 8\}$. In fact, there are much more complicated closed and bounded sets in $\mathbb{R}$ (e.g. the Cantor set of Example 5.34 below). Let's emphasize that the Heine–Borel Theorem is exclusively about the metric spaces $\mathbb{R}^d$; it does not apply in general.

Example 5.32.
(1) Let (X, d) be a discrete metric space. Then for any two points $x, y \in X$, either $d(x, y) = 0$ or $d(x, y) = 1$. Thus, for any subset $E \subseteq X$, $\mathrm{diam}(E) \le 1$, so E is bounded. We have also shown that any subset E is closed. However, if E is an infinite set, it is not compact, cf. Example 5.23(3). So any infinite discrete metric spaces contains closed and bounded sets that are not compact.
(2) For a less contrived example, consider again $B[0, 1]$, the set of all bounded, real-valued functions on $[0, 1]$, which is a metric space with respect to the metric

$$d(f, g) = \sup_{x\in[0,1]} |f(x) - g(x)|.$$

Consider the functions

$$f_n(x) = \begin{cases} 1, & x \ge \frac{1}{n} \\ 0, & x < \frac{1}{n} \end{cases}.$$

All of these functions are in $B[0, 1]$. We can also compute the function $f_n - f_m$; assuming $m > n$ we have

$$f_m(x) - f_n(x) = \begin{cases} 0, & x \ge \frac{1}{n} \\ 1, & \frac{1}{m} \le x < \frac{1}{n}, \\ 0, & x < \frac{1}{m} \end{cases}.$$

This shows that $d(f_n, f_m) = \sup_x |f_n(x) - f_m(x)| = 1$ for any $m \ne n$! Thus, the sequence (f_n) cannot have a convergent subsequence: no two terms in the sequence are ever closer to (or farther from) each other than 1.

Here is an important property of compact sets. This is the generalization of the nested intervals property that we used in the construction of $\mathbb{R}$.

Proposition 5.33. *Let $K_1, K_2, K_3, \ldots$ be nonempty compact sets in a metric space, and suppose they are nested: $K_{n+1} \subseteq K_n$ for all n. Then $\bigcap_n K_n$ is a nonempty compact set. If, in addition, $\mathrm{diam}(K_n) \to 0$ as $n \to \infty$, then $\bigcap_n K_n$ consists of exactly one point.*

Proof. Since $K_n \ne \varnothing$ for any n, we can choose a point $x_n \in K_n$ for each n. By the nested property, $x_n \in K_1$ for each n. Thus, (x_n) is a sequence in the compact set K_1, and therefore it has a convergent subsequence x_{n_k} with a limit $x \in K_1$. Now, for any $m \in \mathbb{N}$, the tail subsequence $(x_{n_k})_{k=m}^\infty$ also converges to x; but this is a sequence of terms in K_{n_m}, which is closed, and so $x \in K_{n_m}$. This holds for every m. Finally, for any n, there is $n_m > n$, and therefore $K_{n_m} \subseteq K_n$; thus, $x \in K_n$ for every n, which shows that $x \in \bigcap_n K_n$. This intersection is therefore nonempty. It is an intersection of compact sets, therefore it is compact (Exercise 1 on HW9).

For the second claim, let $x, y \in \bigcap_n K_n$. Fix $\epsilon > 0$; since $\mathrm{diam}(K_n) \to 0$, there is some n with $\mathrm{diam}(K_n) < \epsilon$. Thus, since $x, y \in K_n$, $d(x, y) \le \mathrm{diam}(K_n) < \epsilon$. So $0 \le d(x, y) < \epsilon$ for all $\epsilon > 0$; it follows that $d(x, y) = 0$ and so $x = y$. That is: there is at most one point in the intersection. As we've shown the intersection is nonempty, this proves that it consists of exactly one point, as claimed. ∎

Example 5.34 (Cantor set). The unit interval $K_0 = [0,1]$ is compact. Now remove the "middle third" and let $K_1 = [0,\frac{1}{3}]\cup[\frac{2}{3},1]$; this set is also compact, and $K_1 \subset K_0$. Now repeat this: remove the middle third from each of the two intervals in K_1, producing $K_2 = [0,\frac{1}{9}]\cup[\frac{2}{9},\frac{1}{3}]\cup[\frac{2}{3},\frac{7}{9}]\cup[\frac{8}{9},1]$. Again, this set is compact, and $K_2 \subset K_1$. We can repeat this "delete middle thirds" process indefinitely. Note a certain self-similarity: K_2 has two pieces, each of which looks like K_1 shrunk down by a factor of $\frac{1}{3}$. In fact, we can inductively define

$$K_n = \frac{1}{3}K_{n-1} \cup \left(\frac{2}{3} + \frac{1}{2}K_{n-1}\right).$$

All of these sets are finite collections of closed, bounded intervals, so they are all compact, and they are nested $K_{n+1} \subset K_n$. Hence, by Proposition 5.33, the set $C = \bigcap_n K_n$ is a nonempty compact set. This set is called the **Cantor set**.

What can we say about this set? Well, what is the length of the longest interval in it? Note that K_n consists of 2^n intervals, each of length $\frac{1}{3^n}$. Since the length of the intersection of two intervals is $\le$ the length of either interval, and since $C \subset K_n$ for every n, this means C contains no intervals of length $> \frac{1}{3^n}$ for any $n \in \mathbb{N}$; but $\frac{1}{3^n} \to 0$ as $n \to \infty$, and therefore C *contains no intervals of length* > 0. This proves that $\mathring{C} = \varnothing$. Indeed, if x were an interior point of C, that would mean $B_r(x) \subseteq C$ for some $r > 0$; but $B_r(x) = (x-r, x+r)$ is an interval of length $2r > 0$, which we know is not contained in C. Thus no point is interior to C. At the same time, C is compact, so it is closed. Thus $\overline{C} = C$, and so $\partial C = \overline{C} \setminus \mathring{C} = C$ – the Cantor set *is its own boundary*.

That also happens for discrete sets: if K consists entirely of isolated points, then K is closed and $\mathring{K} = \varnothing$, so $\partial K = K$. But the Cantor set is the *opposite* of a discrete set: it contains *no* isolated points, so C consists entirely of limit points, $C' = C$. To see this, fix $x \in C$; so $x \in K_n$ for every n. Now K_n is a collection of disjoint closed intervals, so there is some interval $I_n \subset K_n$ with $x \in I_n$. Either x is in the interior of this interval or it is one of the endpoints; either way, there is *one* endpoint x_n of I_n with $x_n \neq x$. Now, from the construction of C, the endpoints of all the intervals are in C, so $x_n \in C$. Also, as $\mathrm{diam}(I_n) = \frac{1}{3^n} \to 0$ and $x, x_n \in I_n$, we have $d(x, x_n) \to 0$. This $x_n \to x$, but $x_n \neq x$ for any n, and $x_n \in C$; this proves that $x \in C'$. Since x was an arbitrary element of C, this means $C \subseteq C'$, and since C is closed, we have $C' \subseteq C$, so $C = C'$.

6. Limits and Continuity

6.1. **Lecture 18: March 3, 2016.** We now begin to study functions. We have, of course, been studying functions (for example sequences, which are functions with domain $\mathbb{N}$); now we will concentrate on metric properties of functions. So we will set things up in terms of functions between metric spaces.

Definition 6.1. *Let X and Y be metric spaces. Let $E \subseteq X$, and let $x_0 \in E'$ be a limit point of E. Let $L \in Y$. Now, for any function $f\colon E \to Y$, we say $f(x)$* **tends to** *L* **as** *$x \to x_0$, or* **the limit as** *$x \to x_0$* **of** *$f(x)$* **is** *L, in symbols*

$$\lim_{x \to x_0} f(x) = L$$

if: given any sequence (x_n) in $E \setminus \{x_0\}$ that converges $x_n \to x_0$, it follows that the sequence $(f(x_n))$ in Y converges $f(x_n) \to L$.

This is a more general kind of limit than the limit of a sequence: we are letting the argument "tend to" a limit point through a set that may be quite different from $\mathbb{N}$. Our definition makes use of our knowledge of limits of sequences. This is useful, for example, in establishing some of the basic properties of limits. For example:

Lemma 6.2. *Limits are unique: if $\lim_{x\to x_0} f(x) = L_1$ and $\lim_{x\to x_0} f(x) = L_2$, then $L_1 = L_2$.*

Proof. Since x_0 is a limit point, there is a sequence (x_n) with $x_n \neq x_0$ for any n and $x_n \to x_0$. By definition of $\lim_{x\to x_0} f(x) = L_1$, this means that the sequence $f(x_n)$ converges to L_1; by definition of $\lim_{x\to x_0} f(x) = L_2$, this means that $f(x_n)$ converges to L_2. Thus, by uniqueness of limits of sequences, $L_1 = L_2$. $\square$

Remark 6.3.

(1) If we had not included in the definition the fact that x_0 is a limit point, this argument would fail. Indeed, if x_0 is an isolated point, vacuously it holds that $\lim_{x\to x_0} f(x) = L$ for *all* L.

(2) On the other hand, we might try to modify the definition of limit so that this wouldn't happen: we could, for example, insist that $f(x_n) \to L$ for *any* sequence x_n that converges to x_0, even if it does hit x_0 at some times. But this would rule out some of our intuition about limits, as the following example shows.

Example 6.4. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 0, & x \neq 0 \\ 1, & x = 0 \end{cases}.$$

We know from our calculus intuition that $\lim_{x\to 0} f(x) = 0$. Indeed, we can verify this from Definition 6.1: if x_n is any sequence in $\mathbb{R} \setminus \{0\}$, then $f(x_n) = 0$ for all n, and the constant sequence 0 does indeed converge to 0.

On the other hand, suppose we had left out the $x_n \neq x_0$ clause in Definition 6.1, and insisted that $f(x_n) \to L$ for every sequence $x_n \to x_0$. In this scenario, the function above would have no limit at 0. Indeed, we could take the sequence $x_n = \frac{1}{n}$ if n is even and $x_n = 0$ if n is odd. Then the sequence $f(x_n) = (0, 1, 0, 1, 0, 1, \ldots)$ has no limit.

This illustrates the fundamental idea of limits: a limit is where a function is *going* as you *approach* the limit point; it is unrelated to the actual *value* of the function at that point (if it is even defined).

We can use our theory of limits of sequences to calculate many limits. For example, if the range space for the function is the familiar $\mathbb{C}$, we have the following echo of the limit theorems for $\mathbb{C}$-sequences:

Theorem 6.5 (Limit Theorems)**.** *Let $f, g\colon X \to \mathbb{C}$, and let x_0 be a limit point in X. If $\lim_{x\to x_0} f(x) = L$ and $\lim_{x\to x_0} g(x) = M$, then*

$$\lim_{x\to x_0}[f(x)+g(x)] = L+M, \quad \textit{and} \quad \lim_{x\to x_0}[f(x)\cdot g(x)] = L\cdot M.$$

Proof. Let (x_n) be any sequence in $X\setminus\{x_0\}$ that converges to x_0. By assumption, $f(x_n)\to L$ and $g(x_n)\to M$. Thus, by the limit theorems for sequences in $\mathbb{C}$, cf. Theorem 2.27, $f(x_n)+g(x_n)\to L+M$ and $f(x_n)\cdot g(x_n)\to L\cdot M$. This is precisely what it means to say that $\lim_{x\to x_0}[f(x)+g(x)] = L+M$ and $\lim_{x\to x_0}[f(x)\cdot g(x)] = L\cdot M$. $\square$

There is an equivalent definition of limit which does not explicitly rely on sequences. This definition is one of the crowning achievements of 19th Century mathematics. The calculus was built on an intuitive understanding of limits in the minds of Newton and Liebnitz (and others), but it wasn't until Weierstrass came up with this modern definition that analysis of functions was finally put on rigorous footing.

Theorem 6.6. *Let (X, d_X) and (Y, d_Y) be metric spaces, let $E\subseteq X$, let $f\colon E\to Y$ be a function, let $x_0\in E'$, and let $L\in Y$. Then $\lim_{x\to x_0} f(x) = L$ if and only if the following holds true:*

$$\forall\epsilon>0\ \exists\delta>0 \text{ s.t. } \forall x\in E,\ x\in B_\delta(x_0)\setminus\{x_0\} \implies f(x)\in B_\epsilon(L).$$

I.e.

$$\forall\epsilon>0\ \exists\delta>0 \text{ s.t. } \forall x\in E,\ 0<d_X(x,x_0)<\delta \implies d_Y(f(x),L)<\epsilon. \tag{6.1}$$

In words: to say $f(x)$ tends to L as x tends to x_0 means that, for any tolerance $\epsilon>0$, no matter how small, there is some (potentially even smaller) tolerance $\delta>0$ so that, if x is δ-close to x_0 (but not *equal* to x_0), then $f(x)$ is ϵ-close to L.

Proof. First, suppose that (6.1) holds. Let $x_n\in E\setminus\{x_0\}$ be a sequence converging to x_0. Fix $\epsilon>0$, and let $\delta>0$ be the corresponding δ. Now, as $x_n\to x_0$, there is some $N\in\mathbb{N}$ so that, for $n\ge N$, $d_X(x_n,x_0)<\delta$. It follows from (6.1) that $d_Y(f(x_n),L)<\epsilon$ for all $n\ge N$. This proves that $f(x_n)\to L$. Thus, we have shown that $\lim_{x\to x_0} f(x) = L$ by definition.

Conversely, suppose (6.1) fails to hold. This means that there exists some $\epsilon>0$ so that, for all $\delta>0$, there is some point $x_\delta\in B_\delta(x_0)\setminus\{x_0\}$ such that $f(x_\delta)$ is *not* in $B_\epsilon(L)$. In particular, do this with $\delta=\frac{1}{n}$: for each $n\in\mathbb{N}$, choose some $x_n\in B_{\frac{1}{n}}(x_0)$ such that $d_Y(f(x_n),L)\ge\epsilon$. On the one hand, since $0<d_X(x_n,x_0)<\frac{1}{n}\to 0$, we have $x_n\to x_0$ but $x_n\ne x_0$. On the other hand, since $d_Y(f(x_n),L)\ge\epsilon$ for all n, this means that the sequence $f(x_n)$ does not converge to L. By definition, this means that the statement $\lim_{x\to x_0} f(x) = L$ is false. $\square$

Example 6.7. Let us work directly from the ϵ–δ definition of (6.1) to show that $\lim_{x\to 2} x^2 = 4$. Here the domain and range metric spaces are both $\mathbb{R}$. Fix $\epsilon>0$. We want to guarantee that $|x^2-4|<\epsilon$. Write this as $|x-2||x+2|<\epsilon$. We want to choose $\delta>0$ and force $0<|x-2|<\delta$, meaning $2-\delta<x<2+\delta$. So, as long as we assure that $\delta\le 2$, this means that $0\le x\le 4$, in which case $|x-2||x+2|\le 6|x-2|$. Thus, it suffices to make sure that $6|x-2|<\epsilon$, which is to say $|x-2|<\epsilon/6$. This tells us how to choose δ.

So, starting fresh: Let $\epsilon > 0$. Choose $\delta = \epsilon/6$ if this is < 2, or $\delta = 2$ otherwise. Then, so long as $0 < |x-2| < \delta$, we have $0 \le 2-\delta < x < 2+\delta \le 4$, and so

$$|x^2 - 4| = |x+2||x-2| \le 6|x-2| < 6 \cdot \frac{\epsilon}{6} = \epsilon.$$

Thus, by (6.1), we have proven that $\lim_{x\to 2} x^2 = 4$.

On the other hand, if we refer to Theorem 6.5, we see that this follows from the fact that $\lim_{x\to 2} x = 2$ (which is easy to verify by either definition of limit) and therefore $\lim_{x\to 2} x \cdot x = 2 \cdot 2 = 4$. Similar considerations show that $\lim_{x\to x_0} f(x) = f(x_0)$ holds for any point $x_0 \in \mathbb{R}$ (or $\mathbb{C}$) if f is a polynomial, for example.

6.2. **Lecture 19: March 8, 2016.** In Example 6.7, what we showed is that the function $f(x) = x^2$ satisfies $\lim_{x\to 2} f(x) = f(2)$. We should recognize this as saying that f *is continuous at* 2.

Definition 6.8. *Let X, Y be metric spaces, $E \subseteq X$, and $f\colon E \to Y$. Let $x_0 \in E'$ be a limit point. Say that f is* **continuous** *at x_0 if*

$$\lim_{x\to x_0} f(x) = f(x_0).$$

Note that this only defines continuity at limit points: we have left undefined what it would mean for f to be continuous at an isolated point of its domain of definition. Indeed, what should we mean by saying that a function is continuous on the set $\mathbb{N}$? This is, to some degree, up to debate. The standard answer is to say this is a vacuous condition: *every* function is continuous on a discrete set.

Now, consider again Example 6.7. To use the definition of limit, we assumed that $d(x, x_0) = |x - 2| > 0$ (as limits are about where you're going, not where you get to). However, observe that this requirement was never used in the proof. That is generically true in limits of continuous functions, as the next result demonstrates.

Proposition 6.9. *Let X, Y be metric spaces and $f\colon X \to Y$. Let $x_0 \in X'$. Then f is continuous at x_0 if and only if for* every *sequence (x_n) in X with $\lim_{n\to\infty} x_n = x_0$, it follows that $\lim_{n\to\infty} f(x_n) = f(x_0)$. Similarly, f is continuous at x_0 if and only if*

$$\forall \epsilon > 0\ \exists \delta > 0 \text{ s.t. } \forall x \in E,\ d_X(x, x_0) < \delta \implies d_Y(f(x), f(x_0)) < \epsilon.$$

That is: we need not assume that the sequence (x_n) never hits x_0; and we need not remove x_0 from the δ-ball in the ϵ–δ definition of the limit. In fact, with these assumption no longer required, there is no reason to assume $x_0 \in X'$; this definition makes perfect sense for isolated points as well, so we take it more generally as the definition of continuity. From this more general definition, it follows that any function is continuous at an isolated point of its domain (as you should work out).

Proof. Suppose that $x_n \to x_0$ implies $f(x_n) \to f(x_0)$ in general; then in particular this holds if we also assume that $x_n \neq x_0$ for any n, which means that $\lim_{x\to x_0} f(x) = f(x_0)$ by definition. Thus f is continuous at x_0. Conversely (in contrapositive form), suppose there is some sequence $x_n \to x_0$ such that $f(x_n) \not\to f(x_0)$. This means there is some $\epsilon > 0$ so that $d(f(x_n), f(x_0)) \geq \epsilon$ for infinitely many n. So let $n_1, n_2, n_3, \ldots$ be these infinitely many indices where, for each k, $d(f(x_{n_k}), f(x_0)) \geq \epsilon$; then $(x_{n_k})_{k=1}^\infty$ is a sequence in X that converges to x_0 (as a subsequence of a sequence x_n which converge to x_0), but $f(x_{n_k}) \not\to f(x_0)$ and, moreover, since $f(x_{n_k}) \neq f(x_0)$ for any k, it follows that $x_{n_k} \neq x_0$ for any k. This shows, from the definition, that it is false that $\lim_{x\to x_0} f(x) = f(x_0)$, completing the proof of the first statement.

The proof of the equivalence of the ε–δ statement is similar and left to the reader. □

The point is: when the putative limit is the *value* of the function at the limit point, there is no reason to exclude the limit point from consideration: where you are going and where you get to are the same in this case!

Example 6.10. Let (X, d) be a metric space, and let $y \in X$. Then the function $f(x) = d(x, y)$ is continuous at every point in X. Indeed, fix $x \in X$, and let (x_n) be a sequence in X with $x_n \to x$. Then

$$d(x_n, y) \leq d(x_n, x) + d(x, y)$$

and so $d(x_n, y) - d(x, y) \leq d(x_n, x)$. But also

$$d(x, y) \leq d(x, x_n) + d(x_n, y)$$

and so $d(x,y) - d(x_n,y) \le d(x,x_n) = d(x_n,x)$. Together, these give

$$0 \le |d(x_n,y) - d(x,y)| \le d(x_n,x).$$

Since $x_n \to x$, $d(x_n,x) \to 0$ by definition, and so by the squeeze theorem $|d(x_n,y)-d(x,y)| \to 0$, meaning that $f(x_n) = d(x_n,y) \to d(x,y) = f(x)$. This shows that f is continuous at x_0.

We would be remiss if we did not include some examples of *dis*continuous functions.

Example 6.11. Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function

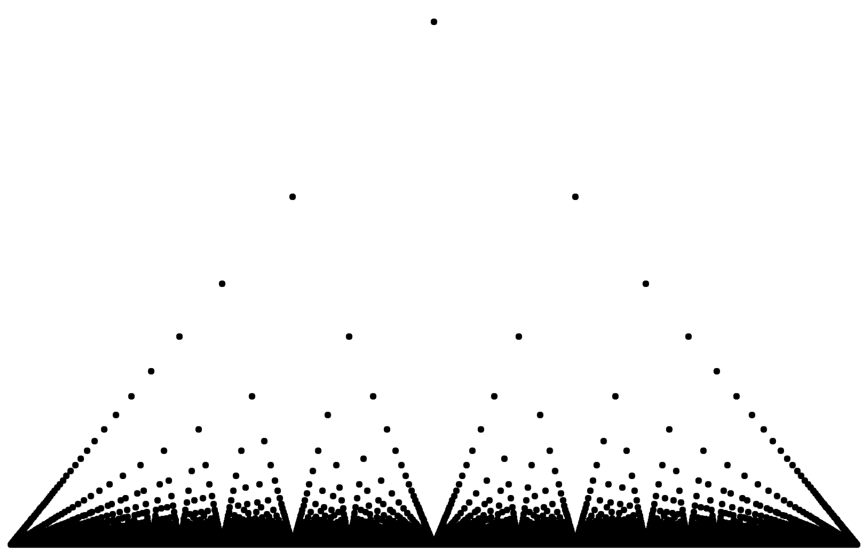
$$f(x) = \begin{cases} 0, & x \notin \mathbb{Q} \\ 1, & x \in \mathbb{Q} \end{cases}.$$

Then f is not continuous at *any* point. Indeed, fix $x \in \mathbb{R}$. For any $\delta > 0$, the ball $B_\delta(x) = (x-\delta, x+\delta)$ contains both rational and irrational numbers. So, if $x \in \mathbb{Q}$, choose some $y \notin \mathbb{Q}$ in the ball, and we have $|f(x) - f(y)| = |1-0| = 1$; if $x \notin \mathbb{Q}$, choose some $y \in \mathbb{Q}$ in the ball, and we have $|f(x) - f(y)| = |0-1| = 1$. In any case, we see that for any $\delta > 0$ there are points $y \in B_\delta(x)$ so that $|f(y) - f(x)| = 1$, so we can never force $f(y)$ to be in, for example, $B_{\frac{1}{2}}(f(x))$. This shows f is discontinuous at x, for any x.

Example 6.12. Consider the following function $f\colon [0,1] \to [0,1]$, sometimes called the *popcorn function*:

$$f(x) = \begin{cases} 0, & x \notin \mathbb{Q} \\ \frac{1}{q}, & x = \frac{p}{q} \text{ in lowest terms} \end{cases}.$$

The graph of this function looks like this:



In fact, f is discontinuous at all rational points, but it is actually continuous at all irrational points. Indeed, let $x = \frac{p}{q}$ be rational, and let $x_n = x + \frac{\sqrt{2}}{n}$ for all n large enough that this is in $[0,1]$; then $x_n \to x$. Then $x_n \notin \mathbb{Q}$ meaning that $f(x_n) = 0$; but $f(x) = \frac{1}{q} \neq 0$, so $f(x_n) \not\to f(x)$. On the other hand, let $x \notin \mathbb{Q}$; we want to show that f is continuous at x, meaning $\lim_{y\to x} f(y) = f(x) = 0$. Fix $\epsilon > 0$, and choose some integer $n \in \mathbb{N}$ with $\frac{1}{n} < \epsilon$. As $f(x) \ge 0$ for all x, it suffices to show that $f(y) < \frac{1}{n}$ for all y sufficiently close to x. Well, if y is a point in $[0,1]$

where $f(y) \ge \frac{1}{n}$, then $y \in \mathbb{Q}$ and, when written in lowest terms, $y = \frac{p}{q}$ with $q \le n$. There are only finitely many such rational numbers, and x (which is irrational) is not one of them. Thus, we can define $\delta = \min\{|x-y| \colon y = \frac{p}{q}$ in lowest terms, with $q \le n\}$; then for $|y-x| < \delta$, it follows that $f(y) < \frac{1}{n} < \epsilon$, proving that $\lim_{y\to x} f(y) = 0$, and so f is continuous at x.

In Examples 6.11 and 6.12, we looked at the set of points where a function is continuous. That is: if $f\colon X \to Y$ is a function and $E \subseteq X$, say that f **is continuous on** E if, for each $x \in E$, f is continuous at x.

Example 6.13. Let $X = (0,1)$, and let $f\colon (0,1) \to \mathbb{R}$ be the function $f(x) = \frac{1}{x}$. Then f is continuous on its whole domain: for every $x \in (0,1)$, f is continuous at x. We could see this by applying the limit theorems; but let's use this as an opportunity to practice our ϵ–δ proofs. Fix $\epsilon > 0$. We want to guarantee that, when y is close to x, we have $|\frac{1}{x} - \frac{1}{y}| < \epsilon$. That is

$$\frac{1}{x} - \epsilon < \frac{1}{y} < \frac{1}{x} + \epsilon.$$

We only need this to hold for all sufficiently small $\epsilon > 0$, so it's fine to assume ϵ is small enough that $\frac{1}{x} - \epsilon > 0$. Thus we can reciprocate to get

$$\frac{1}{1/x - \epsilon} > y > \frac{1}{1/x+\epsilon}.$$

Now, subtract x from both sides and we have

$$\frac{-\epsilon x}{1/x+\epsilon} = \frac{1}{1/x+\epsilon} - x < y - x < \frac{1}{1/x-\epsilon} - x = \frac{\epsilon x}{1/x-\epsilon}.$$

This shows us how to choose δ: we define

$$\delta = \min\left\{\frac{\epsilon x}{1/x+\epsilon}, \frac{\epsilon x}{1/x-\epsilon}\right\} = \frac{\epsilon x}{1/x+\epsilon}. \tag{6.2}$$

Then, reversing the above steps, we have that for any $y \in B_\delta(x)$, we have $|y-x| < \frac{\epsilon x}{1/x+\epsilon} < \frac{\epsilon x}{1/x-\epsilon}$, and this gives in particular the above two inequalities that can be reversed to say $|\frac{1}{x} - \frac{1}{y}| < \epsilon$. So we have proved that there is a $\delta > 0$ for any given $\epsilon > 0$ (as long as $\epsilon < \frac{1}{x}$; otherwise, if $\epsilon \ge \frac{1}{x} \ge 1$, we could take δ to be something silly and big), proving continuity at x.

6.3. **Lecture 20: March 10, 2016.** In Example 6.13, we showed explicitly that the function $f(x) = \frac{1}{x}$ is continuous at every point $x \in (0,1)$. But note: the δ we had to choose for each $\epsilon > 0$ in (6.2) *depends on* x as well as ϵ. This will generically be true. Look at the ϵ–δ definition of continuity: a function f is continuous on a set E if

$$\forall x \in E\ \forall \epsilon > 0\ \exists \delta > 0 \text{ s.t. } \forall y \in E,\ d_X(x,y) < \delta \implies d_Y(f(x), f(y)) < \epsilon. \tag{6.3}$$

Having chosen a x and $\epsilon > 0$, we must then find a suitable $\delta = \delta(x, \epsilon)$. In Example 6.13, not only does δ depend on x, but it does so in a bad way: as $x \to 0$, for given $\epsilon > 0$, the $\delta \to 0$ as well (quite fast, in fact: the numerator is shrinking and the denominator is growing). The closer x is to 0, the smaller δ must be to get the same control over the function. So, while the function is continuous, there is a lack of uniformity in how continuous it is. (Note: we have shown this δ works; one might ask whether a larger, possibly more uniform δ could work just as well. The answer is no: it is not hard to show in this example that the δ in (6.2) is the largest possible δ for the given x and ϵ; it is called the *modulus of continuity* of the function.)

Definition 6.14. *Let X, Y be metric spaces, $E \subseteq X$, and $f\colon E \to Y$. Call f* **uniformly continuous** *on E if*

$$\forall \epsilon > 0\ \exists \delta > 0 \textit{ s.t. } \forall x, y \in E,\ d_X(x,y) < \delta \implies d_Y(f(x), f(y)) < \epsilon. \tag{6.4}$$

Compare (6.4) with (6.3). The difference appears subtle: just the placement of the quantifier $\forall x$. This makes a world of difference: (6.4) says that, not only is f continuous at each point x, but one can choose a $\delta = \delta(\epsilon)$ that is **uniform**: it need not depend on x. This outlaws behavior like the function $f(x) = \frac{1}{x}$ near 0.

Example 6.15. Let $f(x) = 2x$ on $\mathbb{R}$. Then $|f(x) - f(y)| = 2|x - y|$, so for any $\epsilon > 0$, we may let $\delta = \epsilon/2$; then if $|x - y| < \delta = \epsilon/2$, it follows that $|f(x) - f(y)| = 2|x - y| < 2\delta = \epsilon$. This shows that f is continuous at all points; moreover, we may choose $\delta = \delta(\epsilon) = \epsilon/2$ uniformly over all $x, y \in \mathbb{R}$. Thus, f is uniformly continuous on $\mathbb{R}$.

Example 6.16. Let $f(x) = x^2$ on $[0, \infty)$. We want to make $|x^2 - y^2|$ small. We have $|x^2 - y^2| = (x+y)|x-y|$. Thus, in order for $|x^2 - y^2| < \epsilon$, we must have $|x - y| < \frac{\epsilon}{x+y}$ (these are equivalent). But this shows that f is not uniformly continuous. Indeed, in order for $|x - y| < \delta$ to imply that $|x - y| < \frac{\epsilon}{x+y}$, we must have $\delta \le \frac{\epsilon}{x+y}$; and there is no positive number $\delta = \delta(\epsilon)$ that is $\le \frac{\epsilon}{x+y}$ for all $x, y > 0$.

In Examples 6.13 and 6.16, we saw continuous functions on the intervals $(0,1)$ and $[0,\infty)$ that are not uniformly continuous. In both cases, the non-uniformity was manifest by uncontrolled growth near the "edge". As it turns out, if the domain of the continuous function is compact, this cannot happen. That will be our final big theorem of this class.

Theorem 6.17. *Let X, Y be metric spaces, $K \subseteq X$ compact, and $f\colon K \to Y$ a continuous function. Then f is uniformly continuous.*

Proof. Suppose, for a contradiction, that f is not uniformly continuous on K. Negating Definition 6.14, this means

$$\exists \epsilon > 0\ \forall \delta > 0\ \exists x, y \in K \text{ s.t. } d_X(x,y) < \delta, \text{ but } d_Y(f(x), f(y)) \ge \epsilon.$$

That is: there is a positive number $\epsilon > 0$ so that, for every positive number $\delta > 0$, we can find two points x and y that are within distance δ of each other, but such that $f(x)$ and $f(y)$ are at least ϵ apart. So, let's do this with $\delta = \frac{1}{n}$ for any given positive integer: we can find x_n, y_n with $d_X(x_n, y_n) < \frac{1}{n}$, and yet $d_Y(f(x_n), f(y_n)) \geq \epsilon$.

Now, we use the compactness of the domain K: the sequence (x_n) has a convergent subsequence (x_{n_k}) with a limit $x \in K$. Consider, now, the corresponding subsequence y_{n_k}; this has a convergent subsequence $y_{n_{k_\ell}}$ with a limit $y \in K$. Now, $x_{n_{k_\ell}}$ is a subsequence of x_{n_k} which converges to x, hence $x_{n_{k_\ell}} \to x$ as well. But we also have

$$d_X(x_{n_{k_\ell}}, y_{n_{k_\ell}}) < \frac{1}{n_{k_\ell}} < \frac{1}{\ell} \to 0.$$

Hence, it follows from the triangle inequality that $x = y$. On the other hand, by their very construction, the points $x_{n_{k_\ell}}$ and $y_{n_{k_\ell}}$ all satisfy

$$d_Y(f(x_{n_{k_\ell}}), f(y_{n_{k_\ell}})) \geq \epsilon. \tag{6.5}$$

But $x_{n_{k_\ell}} \to x$ and so, since f is continuous, $f(x_{n_{k_\ell}}) \to f(x)$; similarly, $y_{n_{k_\ell}} \to y = x$, and so by continuity $f(y_{n_{k_\ell}}) \to f(y) = f(x)$. Thus, by Problem 4 on Exam 2,

$$d_Y(f(x_{n_{k_\ell}}), f(y_{n_{k_\ell}})) \to d_Y(f(x), f(x)) = 0.$$

This contradicts (6.5). Thus, we have proven that f is, in fact, uniformly continuous. $\square$

Theorem 6.17 is typically the best way to prove uniform continuity of a function. For example: any polynomial is continuous on $\mathbb{R}$, but, as we saw in Example 6.16, they need not be uniformly continuous. By Theorem 6.17, polynomial functions on compact intervals $[a, b]$ are automatically uniformly continuous. What's more: once you know a function is uniformly continuous on a set K, it is then automatically uniformly continuous on any subset $E \subseteq K$ (the same $\delta = \delta(\epsilon)$ that works on all of K also works on all of $E \subseteq K$). So, for example, polynomials are uniformly continuous on all bounded intervals (a, b), $(a, b]$, etc. Similarly, the function $f(x) = \frac{1}{x}$ of Example 6.13 *is* uniformly continuous on $[\alpha, 1]$ for any $\alpha > 0$. We could see this directly from (6.2), since the modulus of continuity

$$\delta = \delta(x, \epsilon) = \frac{\epsilon x}{1/x + \epsilon}$$

decreases as x decreases; it follows that the uniform $\delta = \delta(\alpha, \epsilon)$ will work for all $x \geq \alpha$. However, this gets smaller as α shrinks, and if we include all of $(0, 1]$ in the domain, there is *no* uniform δ. For an alternate proof of the non-uniformity in this example, see HW10.4.

Here is another very useful property of continuous functions on compact sets.

Proposition 6.18. *Let X, Y be metric spaces, $K \subseteq X$ compact, and $f\colon K \to Y$ continuous (hence uniformly continuous). Then the image $f(K) \subseteq Y$ is compact.*

To be clear: $f(K)$ denotes the image of f on K:

$$f(K) = \{f(x) \in Y : x \in K\} = \{y \in Y : \exists x \in K \text{ s.t. } y = f(x)\}.$$

Proof. Let (y_n) be any sequence in $f(K)$. By definition of $f(K)$, for each y_n, there exists some (or potentially many) $x_n \in K$ such that $y_n = f(x_n)$. Since K is compact, it then follows that the sequence (x_n) has a convergent subsequence (x_{n_k}) with limit $x \in K$. Since f is continuous, it then follows that $f(x_{n_k}) \to f(x)$ as $k \to \infty$. Since $x \in K$, $f(x) \in f(K)$. Thus, the subsequence $y_{n_k} = f(x_{n_k})$ of $y_n = f(x_n)$ converges in K. We have thus shown that every sequence in $f(K)$ has a convergent subsequence with limit in $f(K)$; that is, $f(K)$ is compact. $\square$

Corollary 6.19 (Minimax Theorem)**.** *Let K be a nonempty compact metric space, and $f\colon K \to \mathbb{R}$. Then f attains its maximum and minimum values on K.*

Corollary 6.19 is a standard result stated in calculus classes, usually in the special case that $K = [a, b]$ is a compact interval in $\mathbb{R}$.

Proof. By Proposition 6.18, $f(K)$ is compact. In particular, it is closed and bounded. It is also nonempty since K is nonempty (so $f(K)$ contains $f(x)$ for any $x \in K$). Thus, by the least upper bound property of $\mathbb{R}$, the set $f(K) \subset \mathbb{R}$ has a supremum M and and infimum m. Now, for any $n \in \mathbb{N}$, $M - \frac{1}{n} < M$, which means that $M - \frac{1}{n}$ is not an upper bound for $f(K)$; thus, there is some $y_n \in f(K)$ with $M - \frac{1}{n} < y_n \leq M$ Hence, by the Squeeze Theorem, $y_n \to M$. By definition of $f(K)$, there exists some $x_n \in K$ with $y_n = f(x_n)$. Since K is compact, there is a convergent subsequence (x_{n_k}) of (x_n), with limit $x \in K$. Since f is continuous, $y_{n_k} = f(x_{n_k}) \to f(x)$. But y_{n_k} is a subsequence of y_n which converges to M; thus $f(x) = M$. We have therefore found a $x \in K$ for which $f(x) = M = \sup f(K)$. That is: $\sup f(K) = \max f(K)$, and the maximum is achieved at the point x. A very similar argument shows there is a point x' with $f(x') = m = \inf f(K)$, completing the proof. □