Econ 4111
Professor: John Nachbar
September 30, 2011

Metric Spaces

1 Metric Spaces Basics.

1.1 Metric spaces.

A *metric space* (X, d) consists of a set of points, X together with a distance function, or metric, $d : X \times X \to \mathbb{R}$. The interpretation is that $d(a, b)$ is the distance between a and b. To qualify as a distance function, d must satisfy three properties.

1. For any $a, b \in X$, $d(a, b) \geq 0$ with $d(a, b) = 0$ iff $a = b$.
2. For any $a, b \in X$, $d(a, b) = d(b, a)$.
3. For any $a, b, c \in X$, $d(a, c) \leq d(a, b) + d(b, c)$. (The *triangle inequality.*)

The triangle inequality gets its name because, in standard geometry, if the three points a, b, and c form a triangle then the length of the side from a to c is less then the sum of the lengths of the other two sides.

Think of X as fundamental while the metric d is a kind of overlay, like the grid on a map, that we add to help with our analysis. Any X has an infinity of possible metrics. At a minimum, metrics can differ because of units (inches rather than centimeters). But it is possible for different metrics to give different answers to questions like, "Is a further from c than b is from c?" (I give an example of this below.) Which metric we choose depends entirely on which is most helpful to us. For some spaces, notably $X = \mathbb{R}^N$, there is a default metric that is convenient for almost any application. For other spaces, such as variants of $X = \mathbb{R}^\infty$, there is no default metric.

1.2 The space $\mathbb{R}^N$.

The most familiar example of a metric space is $\mathbb{R}^N$, which is the space of points of the form $(x_1, \dots, x_N)$, where each $x_n \in \mathbb{R}$. Recall that if $x \in \mathbb{R}^N$ then the *Euclidean norm* of x is

$$\|x\| = (x \cdot x)^{1/2}.$$

If $N = 1$ then $\|x\| = |x|$. $\|x\|$ measures the distance, in the everyday use of the term, from x to the origin. The *Euclidean metric* is then defined by

$$d_E(a, b) = \|a - b\| = [(a - b) \cdot (a - b)]^{1/2} = \sqrt{\sum_n (a_n - b_n)^2}.$$

To be careful, I have to verify that d_E *is* a metric.

Theorem 1. *The Euclidean metric on $\mathbb{R}^N$ satisfies the three metric properties.*

Proof. Only the triangle inequality is non-trivial. For any $a, b, c \in \mathbb{R}^N$,

$$\begin{aligned}\|a-c\|^2 &= \|(a-b)+(b-c)\|^2 \\ &= \|a-b\|^2 + 2(a-b)\cdot(b-c) + \|b-c\|^2 \\ &\leq \|a-b\|^2 + 2\|a-b\|\,\|b-c\| + \|b-c\|^2 \\ &= \left(\|a-b\| + \|b-c\|\right)^2,\end{aligned}$$

where the inequality comes from the Cauchy-Schwartz inequality. Taking the square root of both sides yields the result. ■

Whenever I write $\mathbb{R}^N$ you can assume that the metric is d_E unless I explicitly state otherwise. I also use d_E as the metric whenever I work with $\mathbb{Q}$, the set of rational numbers.

Although d_E is the default metric for $\mathbb{R}^N$, there are many (infinitely many) other possible metrics. For later reference, here is another useful metric. For any $a, b, \in \mathbb{R}^N$,

$$d_{\max}(a,b) = \max_n\{|a_n - b_n|\}.$$

Geometrically, (a, b) forms an N-dimensional rectangle with sides of length $|a_n - b_n|$. $d_{\max}$ measures the distance between a and b as the length of the longest side.

The metrics $d_{\max}$ and d_E are different, not just in the trivial sense of giving different numbers (in that sense, $2d_E$ is a different metric than d_E) but in the more interesting sense that they can rank the proximity of points differently. In $\mathbb{R}^2$, for example, consider the points $a = (100, 0)$ and $b = (90, 90)$. Then $d_{\max}(a, 0) = 100 > d_{\max}(b, 0) = 90$. On the other hand, $\|a\| = 100 < \|b\| = 90\sqrt{2} \approx 127$. Thus, a is further from the origin than b under $d_{\max}$, but closer to the origin under d_E.

Again, to be careful, I have to verify that $d_{\max}$ really is a metric.

Theorem 2. *The metric $d_{\max}$ on $\mathbb{R}^N$ satisfies the three metric properties.*

Proof. Again, of the metric properties, only the triangle inequality requires any work. Consider any $a, b, c \in \mathbb{R}^N$. For each n, $|a_n - c_n| \leq |a_n - b_n| + |b_n - c_n|$. Since $|a_n - b_n| \leq d_{\max}(a, b)$ and $|b_n - c_n| \leq d_{\max}(b, c)$, it follows that $|a_n - c_n| \leq d_{\max}(a, b) + d_{\max}(b, c)$ for all n. Hence $d_{\max}(a, c) \leq d_{\max}(a, b) + d_{\max}(b, c)$. ■

1.3 The space $\mathbb{R}^\infty$.

A point in $\mathbb{R}^\infty$ is of the form $(x_1, x_2, x_3, \dots)$ where each $x_n \in \mathbb{R}$. In contrast to $\mathbb{R}^N$, $\mathbb{R}^\infty$ has no default metric. Later in the course, I discuss the two most commonly used metrics for $\mathbb{R}^\infty$, one of which is an extension of $d_{\max}$ and one of which is new. Certain facts about $\mathbb{R}^N$ with the default Euclidean metric fail in $\mathbb{R}^\infty$ for at least one of these two metric and sometimes for both.

1.4 Balls.

Let (X, d) be a metric space. For $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, the ε *ball around* x is

$$N_\varepsilon(x) = \{a \in X : d(x, a) < \varepsilon\}.$$

As a matter of notation, I'm using $N_\varepsilon(x)$ rather than the more obvious $B_\varepsilon(x)$ (B for ball) because I will be using B later in the course for "budget set." Think of N as mnemonic for "nearby."

Example 1. In $\mathbb{R}$, $N_\varepsilon(x)$ is simply the interval $(x - \varepsilon, x + \varepsilon)$. In $\mathbb{R}^2$, $N_\varepsilon(x)$ is a disk of radius ε, centered at x, excluding the boundary circle. In $\mathbb{R}^3$, $N_\varepsilon(x)$ is a solid ball of radius ε, centered at x, excluding the boundary sphere. □

1.5 Bounded and totally bounded sets.

Let (X, d) be a metric space. $A \subseteq X$ is *bounded* iff there is an $x \in X$ and $r > 0$, such that $A \subseteq N_r(x)$. In words, A is bounded iff it is contained in some ball of large enough radius. In $\mathbb{R}^N$, one can always take x to be the origin.[1]

A set A is *totally bounded* iff for any $\varepsilon > 0$ there is a *finite* set of points $E \subseteq X$ such that

$$A \subseteq \bigcup_{x \in E} N_\varepsilon(x).$$

If $A \subseteq \bigcup_{x \in E} N_\varepsilon(x)$ then the set of balls $\{N_\varepsilon(x)\}_{x \in E}$ is said to *cover* A. In words, then, A is totally bounded iff, for any $\varepsilon > 0$ no matter how small, A can be covered by a finite number of open balls of radius ε (the number of balls required will typically go up as ε shrinks).

The next result establishes that any totally bounded set is bounded.

Theorem 3. *Let (X, d) be a metric space. If $A \subseteq X$ is totally bounded then it is bounded.*

Proof. Since A is totally bounded there is a finite set $E \subseteq X$ such that A is covered by the balls $N_1(x)$ for $x \in E$. Fix any $x^* \in X$ and let $r = \max_{x \in E} d(x^*, x) + 1$. I claim that for any $a \in A$, $a \in N_r(x^*)$, which shows that A is bounded. For any $a \in A$, there is an $x_a \in E$ such that $a \in N_1(x_a)$. By the triangle inequality, $d(x^*, a) \leq d(x^*, x_a) + d(x_a, a) < d(x^*, x_a) + 1 \leq r$. ■

The converse, that every bounded set is totally bounded, is true in $\mathbb{R}^N$ but not in general. I return to this issue later in the course.

---

[1] Suppose that $A \subseteq N_r(x)$. Let $\hat{r} = r + d(x, 0)$. Then, by the triangle inequality, for any $a \in A$, $d(a, 0) \leq d(a, x) + d(x, 0) < r + d(x, 0) = \hat{r}$. Therefore, $A \subseteq N_{\hat{r}}(0)$.

2 Sequences, Cauchy Sequences, and Completeness.

2.1 Sequences.

Let (X, d) be a metric space. An infinite *sequence* $\{x_t\}$ is a function from the positive integers, $\{1, 2, \dots\}$, to X. The value that this function takes at 1 is denoted x_1, the value at 2 is denoted x_2, and so on. The domain is called the set of *indices.* One can also use other sets of integers as indices. For example, one can use $\{0, 1, \dots\}$ (i.e., start the indexing with 0 instead of 1). And one can work with finite sequences. All sequences here will be infinite, however.

The notation for a sequence $\{x_1, x_2, \dots\}$ looks identical to the notation for a set $\{x_1, x_2, \dots\}$. Order matters, however, for a sequence but not for a set. For example, the sequence $\{1, 0, 1, 0, \dots\}$ is very different from the set $\{1, 0, 1, 0, \dots\}$, which is equal to the set $\{1, 0\}$. You'll have to use context to avoid confusion.

I use subscripts to denote both elements of a sequence and, in the case of $\mathbb{R}^N$ or $\mathbb{R}^\infty$, coordinates of points in the sequence. Thus, if $\{x_t\}$ is a sequence in $\mathbb{R}^N$ then $x_{t,n}$ is the n coordinate of the t term of the sequence. For example, if the sequence is $\{(0, 1), (1, 5), (-1, 2), \dots\}$ then $x_{3,1}$, is the first coordinate of the third term in the sequence, namely -1. This notation is awkward, but the alternatives are also awkward, confusing, or both.

The *range* of a sequence is the set of values that the sequences takes. A sequence is said to be in a set $C \subseteq X$ iff the range of the sequence is a subset of C. A sequence is *bounded* iff its range is bounded.

A sequence $\{x_t\}$ in X is said to *converge* to a point $x \in X$, also written

$$x_t \to x,$$

iff for all $\varepsilon > 0$ there is a $T \in \mathbb{N}$ such that for all $t > T$, $x_t \in N_\varepsilon(x)$. In words, $x_t \to x$ iff all terms late in the sequence are close to x.

In a metric space, a sequence cannot converge to two different points.

Theorem 4. *Let $\{x_t\}$ be a sequence in a metric space (X, d). If $x_t \to x^*$ and $x_t \to \hat{x}$ then $x^* = \hat{x}$.*

Proof. Fix any $\varepsilon > 0$. If $x_t \to x^*$ then there is a T^* such that for all $t > T^*$, $d(x_t, x^*) < \varepsilon/2$. If $x_t \to \hat{x}$ then there is a $\hat{T}$ such that for all $t > \hat{T}$, $d(x_t, \hat{x}) < \varepsilon/2$. Therefore, by the triangle inequality, for any $t > \max\{T^*, \hat{T}\}$,

$$d(x^*, \hat{x}) \leq d(x^*, x_t) + d(x_t, \hat{x}) < \varepsilon.$$

Since ε was arbitrary, this implies $d(x^*, \hat{x}) = 0$, hence $x^* = \hat{x}$. ■

Given a sequence $\{x_t\}$, one can construct a new sequence, called a subsequence, out of the original sequence. The subsequence is of the form $\{x_{t_k}\} = \{x_{t_1}, x_{t_2}, \dots\}$ where $\{t_1, t_2, \dots\} \subseteq \{1, 2, \dots\}$, with $t_1 < t_2 < \dots$. Even if $\{x_t\}$ fails to converge, some of its subsequences may converge.

Example 2. Consider the sequence $\{x_t\} = \{1, 0, 1, 0, \dots\}$. This sequence does not converge. I can form the subsequence $\{x_{t_k}\} = \{x_1, x_3, x_5, \dots\} = \{1, 1, 1, \dots\}$. This subsequence converges to 1. There are, of course, other subsequences. For example, consider $\{x_{10}, x_{11}, x_{20}, x_{21}, \dots\} = \{0, 1, 0, 1, \dots\}$. This does not converge. □

2.2 Cauchy sequences and completeness.

Given a metric space (X, d), a sequence $\{x_t\}$ in X is *Cauchy* iff for any $\varepsilon > 0$ there is a T such that for any $s, t > T$, $d(x_s, x_t) < \varepsilon$. In words, a sequence is Cauchy iff all terms late in the sequence are close to each other.

Theorem 5. *Let (X, d) be a metric space. Let $\{x_t\}$ be a sequence in X. If $\{x_t\}$ is convergent then it is Cauchy.*

Proof. Suppose $x_t \to x$. Consider any $\varepsilon > 0$. Choose T such that for all $t > T$, $d(x_t, x) < \varepsilon/2$. Then, by the triangle inequality, for any $s, t > T$, $d(x_s, x_t) \leq d(x_s, x) + d(x, x_t) < \varepsilon$. ■

If the converse is always true, if Cauchy sequences always converge, then the set is called complete.

Definition 1. *Let (X, d) be a metric space. A set $A \subseteq X$ is* complete *iff every Cauchy sequence in A converges to a point in A.*

If X is complete then (X, d) is called a *complete metric space.* Later in the course, I establish that $\mathbb{R}$ is complete. In contrast, the space of rational numbers, $\mathbb{Q}$, is not complete.

Example 3. Consider the Cauchy sequence $\{3, 3.1, 3.14, 3.141, \dots\}$. Considered as a sequence in $\mathbb{R}$, this sequence converges to π. But $\pi \notin \mathbb{Q}$. As a sequence in the metric space $\mathbb{Q}$, therefore, this Cauchy sequence does not converge. □

To sum up, informally, a Cauchy sequence is a sequence that looks like it converges. If the sequence lies in a complete set then a Cauchy sequence actually does converge. But a Cauchy sequence in a set that is not complete need not converge.

For later use, I record the following additional facts about Cauchy sequences.

Theorem 6. *Let (X, d) be a metric space. Let $\{x_t\}$ be a Cauchy sequence in X. If $\{x_t\}$ has a convergent subsequence then it is convergent.*

Proof. Suppose $x_{t_k} \to x$ and choose any $\varepsilon > 0$. Since $x_{t_k} \to x$, there is a K such that if $k > K$ then $x_{t_k} \in N_{\varepsilon/2}(x)$. Since $\{x_t\}$ is Cauchy there is a T such that if $s, t > T$, $d(x_s, x_t) < \varepsilon/2$. Then, by the triangle inequality, for any $t, t_k > \max\{T, t_K\}$, $d(x_t, x) \leq d(x_t, x_{t_k}) + d(x_{t_k}, x) < \varepsilon$. ■

Theorem 7. *Let (X, d) be a metric space. Let $\{x_t\}$ be a Cauchy sequence in X. Then $\{x_t\}$ is bounded.*

Proof. Since $\{x_t\}$ is Cauchy, there is a T such that for all $t, s > T$, $d(x_s, x_t) < 1$. In particular, for all $t > T$, $d(x_t, x_{T+1}) < 1$. Let

$$r = \max\{d(x_1, x_{T+1}), \ldots, d(x_T, x_{T+1})\} + 1$$

Then the range of $\{x_t\}$ lies in $N_r(x_{T+1})$.■

3 Open and Closed Sets.

3.1 Open sets.

Definition 2. *Let (X, d) be a metric space. A point $x \in A \subseteq X$ is* interior to A *iff it is contained in a ball inside A: there is an $\varepsilon > 0$ such that $N_\varepsilon(x) \subseteq A$.*

Theorem 8. *Let (X, d) be a metric space and consider any $O \subseteq X$. The following are equivalent.*

1. *O is either empty or the union of balls.*
2. *Every point $x \in O$ is interior to O.*

Proof. $\Rightarrow$. Suppose that the first property holds. If O is empty then the condition holds vacuously. Otherwise, consider any $x \in O$. By property 1, there is an $a \in O$ and an $\varepsilon_a > 0$ such that $x \in N_{\varepsilon_a}(a) \subseteq O$. I need to show that there is an $\varepsilon > 0$ such that $N_\varepsilon(x) \subseteq O$. Let $\varepsilon = \varepsilon_a - d(a, x)$. Consider any $b \in N_\varepsilon(x)$. By the triangle inequality, $d(b, a) \leq d(b, x) + d(x, a) < \varepsilon_a$. Hence $b \in N_{\varepsilon_a}(a)$, which implies that $b \in O$, which implies that $N_\varepsilon(x) \subseteq O$, as was to be shown.

$\Leftarrow$. If O is empty then it is open. Otherwise, for each $x \in O$, choose $\varepsilon_x > 0$ such that $N_{\varepsilon_x}(x) \subseteq O$. Then

$$O = \bigcup_{x \in O} N_{\varepsilon_x}(x),$$

as was to be shown. ■

Any set that satisfies either of the properties in Theorem 8, and hence both of these properties, is called *open*. Formally, I define openness using property 1.

Definition 3. *Let (X, d) be a metric space. A set $O \subseteq X$ is* open *iff it is either empty or the union of balls.*

Example 4. In $\mathbb{R}$, any bounded interval (a, b) is a ball and therefore is open (set $\varepsilon = (b - a)/2$ and $x = (a + b)/2$; then $(a, b) = N_\varepsilon(x)$). Any interval $(-\infty, b)$ or (a, ∞) is open. For example, (a, ∞) is the union of balls of the form $N_{\varepsilon_x}(x)$ where $x \in (a, \infty)$ and $\varepsilon_x = (x - a)/2$. □

The next result establishes the fundamental properties of open sets.

Theorem 9. *Let (X, d) be a metric space.*

1. *Both X and $\emptyset$ are open.*
2. *For any two open sets O_1 and O_2, $O_1 \cap O_2$ is open. By induction,* finite *intersections of open sets are open.*
3. *For any set $\mathcal{O}$ of open sets,*

$$\bigcup_{O \in \mathcal{O}} O$$

is open. Thus, arbitrary *unions of open sets are open.*

Proof.

1. X is open since it is the union of all balls. $\emptyset$ is open by definition.
2. Let O_1 and O_2 be open. If $O_1 \cap O_2 = \emptyset$ then $O_1 \cap O_2$ is open by definition. Otherwise, let $x \in O_1 \cap O_2$. Since $x \in O_1$ and O_1 is open there is an $\varepsilon_1 > 0$ such that $N_{\varepsilon_1}(x) \subseteq O_1$ (Theorem 8). Similarly, there is an $\varepsilon_2 > 0$ such that $N_{\varepsilon_2}(x) \subseteq O_2$. Set $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. Then $N_\varepsilon(x) \subseteq O_1 \cap O_2$.
3. If the O are open then $\bigcup_{O \in \mathcal{O}} O$ is a union of unions of balls, and is therefore open by definition.

■

The next example shows that arbitrary *intersections* of open sets need not be open.

Example 5. Suppose $\mathcal{O}$ consists of intervals of the form $(0, 1+1/t)$ for $t \in \{1, 2, \dots\}$. Then

$$\bigcap_{O \in \mathcal{O}} O = (0, 1],$$

which is not open. □

A metric space is a special case of a more general type of space called a *topological space*. A topological space specifies a set of points X and a set of open sets τ (τ for 'topology'), with the requirement that the sets in τ must satisfy the conditions in Theorem 9. The twist is that the sets in τ don't have to be derived from a metric and in fact there are topological spaces that do not correspond to any metric space. For example, let $X = \mathbb{R}$ and let $\tau = \{X, \emptyset\}$: the only open sets are the space itself and the empty set. There is no metric on $\mathbb{R}$ that that can generate this τ. The main result on existence of optima generalizes to arbitrary topological spaces. Of the other results, some generalize and some do not.

3.2 Open sets and sequences

Theorem 10. *Let (X, d) be a metric space, let $\{x_t\}$ be a sequence in X and let $x^* \in X$. $x_t \to x^*$ iff for any open set O containing x^*, there is a $T \in \mathbb{N}$ such that for all $t > T$, $x_t \in O$.*

Proof. $\Rightarrow$. Let O be an open set containing x^*. Since O is open, x^* is interior, and hence there is an $\varepsilon > 0$ such that $N_\varepsilon(x^*) \subseteq O$. Since $x_t \to x^*$, there is a T such that for all $t > T$, $x_t \in N_\varepsilon(x^*) \subseteq O$, which is what I needed to show.

$\Leftarrow$. The claim is immediate since $N_\varepsilon(x^*)$ is open for any $\varepsilon > 0$. ■

In practice, it is usually easier to use the $N_\varepsilon(x^*)$ definition of convergence to verify whether a particular sequence convergence, while the open set characterization, which is the definition of convergence in general topological spaces, sometimes provides cleaner proofs in applications.

3.3 Limit Points.

Definition 4. *Let (X, d) be a metric space. Given a set $A \subseteq X$, x is a* limit point *of A iff for any $\varepsilon > 0$, $N_\varepsilon(x)$ contains some point $a \in A$, $a \neq x$.*

For example, 0 is a limit point of both $(0, 1]$ and $[0, 1]$ but 0 is not a limit point of the set $\{0\}$. A limit point is also known as a *cluster* point or an *accumulation* point. A point x^* is a limit point of A iff there are infinitely many points of A arbitrarily close to x^*.

Theorem 11. *Let (X, d) be a metric space, let $A \subseteq X$. A point x^* is a limit point of A iff for any $\varepsilon > 0$, the set $N_\varepsilon(x^*) \cap A$ is infinite.*

Proof. The $\Leftarrow$ direction is immediate from the definition. As for the $\Rightarrow$ direction, let $\varepsilon_1 = 1$ and let x_1 be any point of $N_{\varepsilon_1}(x) \cap A$, $x_1 \neq x^*$. Such an x_1 must exist since x is a limit point of A. Let $\varepsilon_2 = \min\{1/2, d(x_1, x^*)\} \in (0, \varepsilon_1)$. Let x_2 be any point of A, $x_2 \neq x^*$, in $N_{\varepsilon_2}(x)$. Again, such an x_2 must exist since x is a limit point of A. And so on. Continuing in this way, I construct a sequence $\{x_t\}$ converging to x^*, with all elements of the sequence distinct. For any $\varepsilon > 0$, let $T > 1/\varepsilon$. Then for all $t > T$, $x_t \in N_\varepsilon(x^*)$. ■

As in Theorem 10, on can replace arbitrary open balls with arbitrary open sets in the definition of limit point and in Theorem 11. To avoid clutter, I won't do so explicitly.

As the proof of Theorem 11 suggests, limit points of sets and limits of sequences are closely related concepts. I formalize this in the following.

Theorem 12. *Let (X, d) be a metric space.*

1. *If x is a limit point of $A \subseteq X$ then there is a sequence $\{x_t\}$ in A with $x_t \to x$.*

2. *Let $\{x_t\}$ be a sequence in X and let A be the range of $\{x_t\}$. If $x_t \to x$ then either x is a limit point of A or A is finite and $x \in A$.*

Proof. $\Rightarrow$. Let x_1 be any point of A in $N_1(x)$. Let x_2 be any point of A in $N_{1/2}(x)$. And so on. These points must exist since x is a limit point of A. By construction, $x_t \to x$.

$\Leftarrow$. Suppose that $x_t \to x$ and suppose that x is not a limit point of A. Then there is an $\varepsilon > 0$ such that no element of A, other than x, is in $N_\varepsilon(x)$. Since $x_t \to x$, this implies that there is a T such that for all $t > T$, $x_t = x$, in which case A is finite and $x \in A$. ■

Although limit points and limits are related, they are distinct concepts.

Example 6. Let $\{x_t\}$ be the sequence $\{1/2, 1/4, 1/8, \dots\}$ and let A be the set $\{1/2, 1/4, 1/8, \dots\}$. Then $x_t \to 0$ and the unique limit point of A is 0. □

Example 7. Let $\{x_t\}$ be the sequence $\{1, 1, 1, \dots\}$ and let A be the set $\{1, 1, 1, \dots\} = \{1\}$. Then $x_t \to 1$ but A has no limit points. □

3.4 Closed sets.

Theorem 13. *Let (X, d) be a metric space and consider any $C \subseteq X$. The following are equivalent.*

1. *For any sequence $\{x_t\}$ in C, if $x_t \to x^*$ then $x^* \in C$.*
2. *C contains all its limit points.*
3. *C^c is open.*

Proof. I prove this by showing that (1) implies (2), (2) implies (3), and (3) implies (1).

The fact that (1) implies (2) follows from the first part of Theorem 12.

To prove (2) implies (3), I argue by contraposition. Suppose that C^c is not open. Then there exists an $x^* \in C^c$ such that x^* is not interior to C^c. This means that for any $\varepsilon > 0$, $N_\varepsilon(x^*) \not\subseteq C^c$, hence $N_\varepsilon(x^*) \cap C \neq \emptyset$. Since $x^* \notin C$, this implies that x^* is a limit point. Again, since $x^* \notin C$, C has a limit point that is not in C.

Finally, to prove that (3) implies (1), I again argue by contraposition. Consider a sequence $\{x_t\}$ in C such that $x_t \to x^*$ but $x^* \notin C$. Since $x^* \notin C$, $x^* \in C^c$. Since $x_t \to x^*$, for any ε there is an $x_t \in N_\varepsilon(x^*)$. Since $x_t \in C$, hence $x_t \notin C^c$, this shows that x^* is not interior, hence C^c is not open. ■

Any set satisfying any of the properties in Theorem 13, hence satisfying all of them, is called *closed.* Formally, I define closed using property 3.

Definition 5. *Let (X, d) be a metric space. A set $C \subseteq X$ is* closed *iff its complement, C^c, is open.*

It is important to understand that this definition does *not* say that a set is closed iff it is not open. For example, the set $(0, 1]$ is not open, but it is not closed either. Moreover, there are some sets that are *both* open and closed, namely $\emptyset$ and X. (In the metric spaces we will be working with, these are the only sets that are both open and closed.)

The next result is the analog of Theorem 9.

Theorem 14. *Let (X, d) be a metric space.*

1. *Both X and $\emptyset$ are closed.*
2. *For any two closed sets C_1 and C_2, $C_1 \cup C^2$ is closed. By induction,* finite *unions of closed sets are closed.*
3. *For any set $\mathcal{C}$ of closed sets*
$$\bigcap_{C \in \mathcal{C}} C$$
is closed. Thus, arbitrary *intersections of closed sets are closed.*

Proof. Immediate from Theorem 9 and DeMorgan's laws (see the Set Theory notes). ■

The next example shows that arbitrary *unions* of closed sets need not be closed.

Example 8. Suppose $\mathcal{O}$ consists of intervals of the form $[0, 1-1/t]$ for $t \in \{1, 2, \dots\}$. Then
$$\bigcup_{O \in \mathcal{O}} O = [0, 1),$$
which is not closed. □

Being closed is similar to, but slightly weaker than, being complete.

Theorem 15. *Let (X, d) be a metric space and let $C \subseteq X$ be complete.*

1. *C is closed.*
2. *If $A \subseteq C$ is closed then A is complete.*

Proof.

1. Let $\{x_t\}$ be a sequence in C. If $\{x_t\}$ converges to a point $x \in X$ then $\{x_t\}$ is Cauchy (Theorem 5). Since C is complete, $x \in C$, hence C is closed.
2. Let $\{x_t\}$ be a Cauchy sequence in A. Since C is complete, $\{x_t\}$ converges to a point $x \in C$. Since A is closed, $x \in A$. Hence A is complete.

■

The converse of Theorem 15 is not true in general: closed subsets of arbitrary metric spaces need not be complete.

Example 9. Closed subsets of $\mathbb{Q}$, considered as its own metric space rather than as a subset of $\mathbb{R}$, need not be complete. In particular, $\mathbb{Q}$ itself is closed (Theorem 14) but not complete (Example 3). $\square$

Given a set A, let $\overline{A}$, called the *closure* of A, denote the intersection of all closed sets containing A. As an intersection of closed sets, $\overline{A}$ is closed.

Given a set A, let A' denote the set of limit points of A.

Theorem 16. *Let (X, d) be a metric space. For any set $A \subseteq X$,*

$$\overline{A} = A \cup A'.$$

Proof. Since $\overline{A}$ is closed and $A \subseteq \overline{A}$, it follows from Theorem 13 that $A \cup A' \subseteq \overline{A}$.

Since $A \subseteq A \cup A'$, it suffices to prove that $A \cup A'$ is closed, since then $\overline{A} \subseteq \overline{A \cup A'} = A \cup A'$ and hence $\overline{A} = A \cup A'$.

To show that $A \cup A'$ is closed, I argue by contraposition. Consider any $x \notin A \cup A'$. Therefore, $x \notin A$ and x is not a limit point of A. Hence there is an $\varepsilon > 0$ such that $N_\varepsilon(x) \subseteq A^c$. Consider any point $b \in N_\varepsilon(x)$. Then $b \notin A$. Also, since $N_\varepsilon(x)$ is open, b is interior to $N_\varepsilon(x)$, hence there is an ε_b such that $N_{\varepsilon_b}(b) \subseteq N_\varepsilon(x) \subseteq A^c$. This implies that $b \notin A'$. Therefore, $b \notin A \cup A'$. Since this holds for any $b \in N_\varepsilon(x)$, x is not a limit point of $A \cup A'$. In summary, if x is not in $A \cup A'$ then it is not a limit point of $A \cup A'$. By contraposition, of x is a limit point of $A \cup A'$ then $x \in A \cup A'$, as was to be shown. $\blacksquare$

In $\mathbb{R}$, finite unions of closed intervals are closed, but closed sets in $\mathbb{R}$ can be *much* more complicated than finite unions of closed intervals.

Example 10. Let $E_0 = [0,1]\backslash(1/3, 2/3) = [0, 1/3] \cup [2/3, 1]$, let $E_1 = E_0 \backslash ((1/9, 2/9) \cup (7/9, 8/9)) = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$ and so on. In words, at each stage, remove the (open) middle third from each remaining subinterval. Let $A = \bigcap E_t$. A is called the *Cantor set*.

A is closed, since it is an intersection of closed sets. In fact, one can show that A is *perfect*, meaning that $A = A'$. And A is uncountable.[2] But A^c is open (since A is closed) and one can show that A^c is *dense* in $[0, 1]$, meaning that $\overline{A^c} = [0, 1]$. So A is large in one sense (it is uncountable) but small in another (it is the complement of a set that is open and dense). $\square$

Intuitively, $\overline{N_\varepsilon(x)}$ should equal $\{a \in X : d(a, x) \leq \varepsilon\}$. This turns out to be true in $\mathbb{R}^N$ but not in general.

---

[2] This is true in general of perfect sets in $\mathbb{R}$. But it can be seen as follows. Write the elements of [0,1] as ternary expansions (i.e., in the form $x = x_1/3 + x_2/9 + x_3/27 + \dots$) and then note that $x \in A$ iff it has a ternary expansion with no 1s in it. Thus $x \in A$ iff it has a ternary expansion that is a string of 0s and 2s. Writing a 1 in place of a 2 puts the set of such strings into 1-1 correspondence with the set of strings of 0s and 1s, which can be identified with the set of binary expansions of elements of $[0, 1]$.

Theorem 17. *Let (X,d) be a metric space. For any $x \in X$ and any $\varepsilon > 0$, $\overline{N_\varepsilon(x)} \subseteq \{a \in X : d(a,x) \leq \varepsilon\}$.*

Proof. Since $N_\varepsilon(x) \subseteq \{a \in X : d(a,x) \leq \varepsilon\}$, it remains to show if a is a limit point of $N_\varepsilon(x)$, then $d(a,x) \leq \varepsilon$. I argue by contraposition. Suppose $d(a,x) > \varepsilon$. Let $\delta = d(a,x) - \varepsilon > 0$. Then for any $b \in N_\delta(a)$, $b \notin N_\varepsilon(x)$.[3] Therefore, a is not a limit point of $N_\varepsilon(x)$. ■

Theorem 18. *For any $x \in \mathbb{R}^N$ and any $\varepsilon > 0$, $\overline{N_\varepsilon(x)} = \{a \in \mathbb{R}^N : d_E(a,x) \leq \varepsilon\}$.*

Proof. For ease of notation, take $x = 0$. Take any $\varepsilon > 0$. By Theorem 17, it suffices to show that if $\|a\| = \varepsilon$, then a is a limit point of $N_\varepsilon(0)$. Since $a \notin N_\varepsilon(0)$, this means showing that for any $\delta > 0$, $N_\varepsilon(0) \cap N_\delta(a) \neq \emptyset$. It suffices to restrict attention to δ that is small, specifically $\delta < \varepsilon$. Take any $t \in (1 - \delta/\varepsilon, 1)$. Since $t < 1$, $ta \in N_\varepsilon(0)$. And since $t > 1 - \delta/\varepsilon$, $d_E(a, ta) = \|a - ta\| = (1-t)\|a\| = (1-t)\varepsilon < \delta$, hence $ta \in N_\delta(a)$. ■

Theorem 18 generalizes easily to any normed vector space over the reals, including $\mathbb{R}^\infty$ with standard metrics. But in other metric spaces, the set inclusion $\overline{N_\varepsilon(x)} \subseteq \{a \in X : d(a,x) \leq \varepsilon\}$ can be strict.

Example 11. Consider any non-empty set X with at least two elements and let d be defined by $d(a,b) = 1$ if $a \neq b$ and $d(a,b) = 0$ if $a = b$ (this is called the *discrete metric*). Then (X,d) is a metric space. For any $x \in X$, $N_1(x) = \{x\}$, $\overline{N_1(x)} = \{x\}$, but $\{a \in X : d(a,x) \leq 1\} = X$. □

Finally, I note that in $\mathbb{R}$ the (finite) least upper bound of a closed set must be in the set.

Theorem 19. *Let $C \subseteq \mathbb{R}$ be closed. If $\sup C < \infty$ then $\sup C \in C$.*

Proof. Let $b = \sup C < \infty$. Since b is the least upper bound, for any $t \in \{1, 2, \dots\}$, $b - 1/t$ is not an upper bound, and hence there is an $x_t \in C$ such that $x_t > b - 1/t$. Since b is an upper bound, $x_t \leq b$. Therefore, $x_t \in N_{1/t}(b)$. Thus $x_t \to b$. Since C is closed, $b \in C$. ■

---

[3] Since, $d(a,x) \leq d(a,b) + d(b,x) < \delta + d(b,x)$, it follows that $d(b,x) > d(a,x) - \delta = \varepsilon$, hence $b \notin N_\varepsilon(x)$.