Chapter 3

The Euclidean Closure Operator

This chapter presents a standard way of defining a closure operator on spaces where it is possible to measure the distance between two points. In particular, we will see how to construct the Euclidean closure operator out of the Euclidean metric on familiar spaces. Imagining what this closure operator should do was a major motivation and test case when considering the axioms for such objects. The Euclidean closure is a good mathematical way of capturing the notion of closeness on a piece of paper since it is built out of the idea of measuring distances as you might with a ruler.

As a first step, we need a scale for our ruler. A scale should be an ordered set so that we can distinguish small distances from large ones. The power set of X is ordered by inclusion, for example, but this is called a partial order since it is not always possible to compare two elements and say which is bigger. (Think of disjoint sets.) The only totally ordered set we have met so far is the integers. We could, of course, mark off our ruler with positive integers. The problem is that each one seems separated from the next, and so it would not be easy to measure small distances. We definitely need to compare very small distances if we are going to talk about closeness, so the natural thing to do is to mark off our ruler using fractions, too. A set on which it is possible to perform the necessary division operations along with addition, subtraction, and multiplication is called a field, and the smallest field containing a copy of $\mathbb{Z}^+$ is the set of rational numbers denoted $\mathbb{Q}$.

3.1 Ordered fields

Characterizing the rationals $\mathbb{Q}$ as the smallest field containing a copy of $\mathbb{Z}^+$ can be useful, but how do we think of the elements of $\mathbb{Q}$? Are we sure such a set exists at all? The usual answer is to say that the rationals are fractions with nonzero denominators. In other words, consider the set $\mathbb{Q}^* = \mathbb{Z} \times (\mathbb{Z} - \{0\})$ consisting of all ordered pairs (m, n) of integers whose second entry is not zero.

In this context, we usually write $\frac{m}{n}$ as an abbreviation for such a pair. But is this $\mathbb{Q}^*$ consisting of all these fractions the same as $\mathbb{Q}$? Almost, but not quite. The problem is that we have too many duplicates. In other words, $\frac{3}{4}$ and $\frac{15}{20}$ are different elements of $\mathbb{Q}^*$ as we have defined it so far.

Here is a construction for eliminating the unwanted duplication that will also turn out to be useful in many other settings. First we have to say when two elements of $\mathbb{Q}^*$ should be considered equivalent. More generally, we define relations and equivalence relations as follows.

Definition 3.1 *Given two sets A and B, a subset R of $A \times B$ is called a relation between A and B. If $A = B$, it is called a relation on A. We often write aRb to mean $(a, b) \in R$. An equivalence relation on A is a relation on A that satisfies the following three axioms for all $a, b, c \in A$.*

E1. (Reflexivity on A) aRa.

E2. (Symmetry) *If aRb, then bRa.*

E3. (Transitivity) *If aRb and bRc, then aRc.*

Given an equivalence relation R on A and an element a in A, the set

$$[a] = \{b \in A \mid aRb\}$$

is called the equivalence class of a.

The most familar example of an equivalence relation is equality. For another example, consider congruence on the set of all triangles in the plane.

On $\mathbb{Z}$, for each fixed an integer p we can define an equivalence relation called congruence modulo p by declaring $n \equiv m \pmod{p}$ to mean that p divides $n - m$. For $p = 12$, there are exactly 12 equivalence classes, one of which is $\{\dots, -12, 0, 12, 24, \dots\}$ and another is $\{\dots, -11, 1, 13, 25, \dots\}$. In general, the equivalence classes of A form a partition of A, meaning that the equivalence classes provide a mutually exclusive and exhaustive decomposition of A. The set of all equivalence classes of A under the relation R is often called the quotient of A by R. The quotient of $\mathbb{Z}$ by the equivalence relation congruence modulo p is denoted $\mathbb{Z}_p$ and called the integers mod p. Notice that $\mathbb{Z}_p$ has exactly p elements if p is not zero, and that addition, subtraction, and multiplication "descend" coherently from $\mathbb{Z}$ to well-defined operations on the quotient $\mathbb{Z}_p$ it "covers." For example, this means that to add two equivalence classes, you just pick an element of each, add them, and then form the equivalence class of the result and your answer will not depend on the representatives you chose. For $p = 12$, what we have is just the familiar clock arithmetic where we identify thirteen o'clock with one o'clock. Division works, too, sometimes and you can figure out when.

On the set $\mathbb{Q}^*$ of fractions, the equivalence relation $\sim$ we want to study is determined by declaring that $\frac{m}{n} \sim \frac{j}{k}$ if and only if $mk = nj$. The quotient of $\mathbb{Q}^*$ by $\sim$ is what we call the rationals $\mathbb{Q}$. Its elements officially consist of equivalences class like $[\frac{3}{4}] = \{\ldots, \frac{-6}{-8}, \frac{-3}{-4}, \frac{3}{4}, \frac{6}{8}, \frac{12}{16}, \frac{15}{20}, \ldots\}$. You can verify that addition, subtraction, multiplication, and division descend from $\mathbb{Q}^*$ to well-defined operations on $\mathbb{Q}$. This just amounts to the fact that you get the same answer regardless of whether you reduce fractions to lowest terms before or after you make a calculation. Hence $\mathbb{Q}$ really is a field. Having done all this, it is convenient not to distinguish between $[\frac{3}{4}]$ and $\frac{3}{4}$ in practice, and so we will usually not hesitate to write $\frac{3}{4} \in \mathbb{Q}$. Similarly, it is in this sense we can say that $\mathbb{Q}$ contains a copy of $\mathbb{Z}$, whereas literally $\mathbb{Q}$ contains an equivalence class of the form $[\frac{n}{1}]$ for each integer n.

The upshot is that $\mathbb{Q}$ really is the smallest field containing a copy of $\mathbb{Z}^+$. To serve as a scale, we also wanted it to be ordered. Declaring $[\frac{m}{n}] \leq [\frac{j}{k}]$ if and only if $mk \leq nj$ makes sense on $\mathbb{Q}$ since the way this orders two equivalence classes does not depend on our choice of representatives from each. In general, here is what we mean by an ordering on a set

Definition 3.2 *A relation R on a set A is called a partial ordering of A if it satisfies the following three axioms for all $a, b, c \in A$.*

E1. (Reflexivity on A) aRa.

E2. (Antisymmetry) *If* aRb *and* bRa, *then* $a = b$.

E3. (Transitivity) *If* aRb *and* bRc, *then* aRc.

A partial ordering is called a total ordering of A if, for every a and b in A, either aRb or bRa.

So, for example, inclusion $\subset$ determines a partial ordering of the power set of a set, and the relation $\leq$ provides a total ordering of $\mathbb{Q}$ that derives from the total ordering of the integers. How about $\mathbb{Z}_p$ for nonzero p?

The real numbers $\mathbb{R}$ as we will eventually define them are another field containing a copy of $\mathbb{Z}^+$ and hence also a copy of $\mathbb{Q}$ because the rationals are the smallest such field. Once we have defined a suitable notion of completeness, we will see that the reals are the smallest complete field containing a copy of $\mathbb{Z}^+$. In fact, the real numbers are the only field that is totally ordered, complete, and Archimedean. We will introduce what we mean by complete and Archimedean when we need these ideas. For the moment, the only properties of the reals we will use is that they, like the rationals, are a totally ordered field containing a copy of $\mathbb{Z}^+$. We picture $\mathbb{R}$ intuitively as filling out the real line, and $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ as filling out the plane, etc. The set $\mathbb{R}^n$ consisting of all ordered n-tuples of real numbers is called Euclidean n-space.

3.2 Distance

A ruler associates to any two points on a piece of paper a real number representing the distance between them. On any set, what plays the role of a generalized ruler is called a metric. To qualify as a metric on X, a candidate must assign real numbers to pairs of members of X in a way that satisfies our intuition about distance as captured by the following axioms.

Definition 3.3 *We call d a metric on X if it associates a real number $d(x, y)$ to each ordered pair (x, y) in $X \times X$ so that for all x, y, z in X we have:*

M1. (Strict Nonnegativity) $d(x, y) \geq 0$ *with* $d(x, y) = 0$ *if and only if* $x = y$

M2. (Symmetry) $d(x, y) = d(y, x)$

M3. (Triangle Inequality) $d(x, z) \leq d(x, y) + d(y, z)$.

A pair (X, d) consisting of a set with such a metric is called a metric space.

A simple and familiar example of a metric on the real line $\mathbb{R}$ is obtained by setting $d(x, y) = |x - y|$. If we label three points x, y, and z in the plane, we can envision the triangle inequality as capturing the idea that the shortest distance between two points is a straight line. Thinking about right triangles suggests that, if we want to calculate the straight line distance between two points, we can do so in terms of their coordinates using the Pythagorean Theorem. Here is the notation that expresses this procedure in general.

For $n \in \mathbb{Z}^+$, the n-dimensional **Euclidean space**, denoted by $\mathbb{R}^n$, is defined to be the set of all ordered n-tuples of real numbers $(x_1, x_2, \dots, x_n)$. For example, $\mathbb{R}$ is the set of real numbers, and the set $\mathbb{R}^2$ is the "Euclidean Plane" or the "Cartesian plane." We usually would not refer to it as the "x-y plane" since, in our notation, a point in the plane is represented by a pair (x_1, x_2) instead of a pair (x, y). We reserve variable names without subscripts for the points themselves instead of coordinates. For points in $\mathbb{R}^1 = \mathbb{R}$, there is only one coordinate, so we write it without a subscript or superscript.

Let us now define a "reasonable" expression for distance between points in $\mathbb{R}^n$. For $x, y \in \mathbb{R}^n$, we set

$$\begin{aligned} d_e(x, y) &\stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \\ &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}. \end{aligned}$$

To confirm our earlier intuition, note that the distance between two points in $\mathbb{R}^1 = \mathbb{R}$ is the absolute value of their difference, i.e. $d_e(x, y) = |x-y|$ for $x, y \in \mathbb{R}$.

It is very tempting to think about this distance as being the length of the hypotenuse of some sort of "right triangle" in $\mathbb{R}^n$. After all, for $n = 2$, the formula just expresses the Pythagorean Theorem $a^2+b^2 = c^2$ where $a = |x_1-y_1|$, $b = |x_2 - y_2|$ and c is the length of the hypotenuse of the right triangle formed by x, y and an auxiliary point, say $z = (x_1, y_2)$. At the moment, however, we will not supply definitions of terms like angle, right angle, or hypotenuse, either in the plane or in $\mathbb{R}^n$. Independent of any such interpretations, what we call d_e counts as a metric as long as we can verify that it satisfies the three axioms in the definition above. The first two are straightforward, and a proof of the triangle inequality appears in the appendix to this chapter. We therefore call d_e the Euclidean metric on $\mathbb{R}^n$. There are other metrics on $\mathbb{R}^n$, but this one is so standard that it is the one we always have in mind, with or without the subscript e, unless stated otherwise.

3.3 The Euclidean closure operator

Suppose $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$. What should it mean for x to be really close to the set A? Clearly, x should be "very close" to some chunk of A, some $a \in A$. One way of saying this is

> There exists a point $a \in A$ such that $d(x, a) < 1/100$.

That makes it pretty close, although perhaps we want it even closer:

> There exists $a \in A$ such that $d(x, a) < 1/1000000$.

So, in general, we would like to say

> There exists $a \in A$ such that $d(x, a) < \varepsilon$

where ε is some small positive number. How small should ε be? Why not say that ε can be *arbitrarily* small? In other words, we require the above statement to be true for *every* positive real number ε. We express this idea in the following definition.

Definition 3.4 *For each $A \subset \mathbb{R}^n$, define*

$$\mathbf{K}_e A \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid (\forall \varepsilon > 0)\ (\exists a \in A)\ d(x, a) < \varepsilon\}\,.$$

This $\mathbf{K}_e$ is called the Eucliedean or standard closure operator *on $\mathbb{R}^n$.*

The notation '$(\forall\varepsilon > 0)$' means 'for every ε in the set of positive real numbers'. See §A for more about the quantifiers '$\forall$' and '$\exists$'. Make sure that you can say this definition in English before continuing, and think carefully about what it means (try mimicking our discussion after the definition of a closure operator, back in Chapter 1). This operator will be ubiquitous in our future work, so our convention henceforth will be that, whenever we are working in $\mathbb{R}^n$, $\mathbf{K}$ with the subscript e suppressed will denote this standard closure operator unless stated otherwise.

We will now show that $\mathbf{K}$ is a closure operator. The proof is a bit long and introduces a couple of simple but important tricks. We will first explain how to find the proof; then we will write it up concisely.

Scrap paper. To show that $\mathbf{K}$ is a closure operator, we must check to see that it satisfies the four axioms C1-C4.

Axiom C1: $A \subset \mathbf{K}A$.

Let's assume that $A \subset \mathbb{R}^n$ and $x \in A$ (i.e., x could be any element of A). We need to show that $x \in \mathbf{K}A$ also. Thus, inserting the definition of the standard closure operator, we need to show that

$$(\forall\varepsilon > 0)\ (\exists a \in A)\ d(x, a) < \varepsilon.$$

From the definition, $\varepsilon > 0$ must be assumed; we need to find $a \in A$ such that $d(x, a) < \varepsilon$. The clear choice is $a = x$, for, since $x \in A$, $d(x, a) = 0 < \varepsilon$. So C1 holds.

Axiom C2: $\mathbf{K}(A \cup B) = \mathbf{K}A \cup \mathbf{K}B$.

Recall that two sets are equal if each is a subset of the other. Let $A, B \subset \mathbb{R}^n$. Let us first show that $\mathbf{K}(A \cup B) \supset \mathbf{K}A \cup \mathbf{K}B$. Suppose $x \in \mathbf{K}A \cup \mathbf{K}B$. Then $x \in \mathbf{K}A$ or $x \in \mathbf{K}B$ (or both!). Without loss of generality, $x \in \mathbf{K}A$. We wish to show $x \in \mathbf{K}(A \cup B)$, i.e.

$$(\forall\varepsilon > 0)\ (\exists a \in A \cup B)\ d(x, a) < \varepsilon.$$

Let $\varepsilon > 0$ be given. Since $x \in \mathbf{K}A$, by definition there exists $a \in A$ such that $d(x, a) < \varepsilon$. But this a is also in $A \cup B$, so we have the found the a we wanted.

Now we wish to show that $\mathbf{K}(A \cup B) \subset \mathbf{K}A \cup \mathbf{K}B$. Suppose $x \in \mathbf{K}(A \cup B)$. We wish to show that $x \in \mathbf{K}A \cup \mathbf{K}B$, i.e. $x \in \mathbf{K}A$ or $x \in \mathbf{K}B$. Since we do not know beforehand whether x will be in $\mathbf{K}A$ or $\mathbf{K}B$, it is not clear how to proceed. But, consulting Table A.1, we see that there is another way we can prove that $\mathbf{K}(A \cup B) \subset \mathbf{K}A \cup \mathbf{K}B$: we will assume that x is in *neither* $\mathbf{K}A$ nor $\mathbf{K}B$, and we will try to prove that $x \notin \mathbf{K}(A \cup B)$. This is, after all, a logically equivalent way of stating our original conjecture. The assumption that $x \notin \mathbf{K}A$ means

$$(\exists \varepsilon > 0)\ (\forall a \in A)\ d(x,a) \geq \varepsilon.$$

So we can choose an $\varepsilon_1 > 0$ such that $d(x,a) \geq \varepsilon_1$ for every $a \in A$. (We introduce the subscript to avoid confusing ε_1 with other ε's.) Likewise, since $x \notin \mathbf{K}B$, we can choose $\varepsilon_2 > 0$ such that $d(x,a) \geq \varepsilon_2$ for all $x \in B$.

Now we wish to show that $x \notin \mathbf{K}(A \cup B)$; in other words, we wish to find $\varepsilon > 0$ such that $d(x,a) \geq \varepsilon$ for all $x \in A \cup B$. The key is to let ε be the *smaller* of ε_1 and ε_2 (or either one if $\varepsilon_1 = \varepsilon_2$). The notation for this is $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. To complete the proof, we need to show that $d(x,a) \geq \varepsilon$ for every $a \in A \cup B$. If $a \in A \cup B$, then $a \in A$ or $a \in B$. If $a \in A$, then $d(x,a) \geq \varepsilon_1 \geq \varepsilon$; if $a \in B$ then $d(x,a) \geq \varepsilon_2 \geq \varepsilon$. So we have proved Axiom C2.

Axiom C3: $\mathbf{KK}A = \mathbf{K}A$.

Let $A \subset \mathbb{R}^n$ be given. Since we have already proved Axiom C1, we know that $\mathbf{K}A \subset \mathbf{KK}A$; hence it is enough to show that $\mathbf{KK}A \subset \mathbf{K}A$. Suppose $x \in \mathbf{KK}A$; we wish to show that $x \in \mathbf{K}A$. So let $\varepsilon > 0$ be given; we need to find $a \in A$ such that $d(x,a) < \varepsilon$, so let's start following the "trail" backwards through the closures. Since $x \in \mathbf{KK}A$, there exists $a_1 \in \mathbf{K}A$ such that $d(x,a_1) < \varepsilon$. Since $a_1 \in \mathbf{K}A$, there exists $a \in A$ such that $d(a_1,a) < \varepsilon$. By the triangle inequality,

$$d(x,a) \leq d(x,a_1) + d(a_1,a) < \varepsilon + \varepsilon = 2\varepsilon.$$

This is almost what we want, except that we have $d(x,a) < 2\varepsilon$ instead of $d(x,a) < \varepsilon$. But we can fix this; in the above argument, after the phrase 'let $\varepsilon > 0$ be given', replace every occurrence of ε with $\varepsilon/2$, and then the final result will be $d(x,a) < \varepsilon$, which is what we want. The argument is still legitimate, because $\varepsilon/2$ is a positive number.

Axiom C4: $\mathbf{K}\emptyset = \emptyset$.

We wish to show that

$$\{x \in \mathbb{R}^n \mid (\forall \varepsilon > 0)\ (\exists a \in \emptyset)\ d(x,a) < \varepsilon\} = \emptyset.$$

In other words, we need to prove that there is no $x \in \mathbb{R}^n$ such that

$$(\forall \varepsilon > 0)\ (\exists a \in \emptyset)\ d(x,a) < \varepsilon.$$

What if there were such an x? Then we could let ε be any positive number (for example $\varepsilon = 1$), and there would be $a \in \emptyset$ such that $d(x,a) < \varepsilon$. But this is absurd, because the empty set has no elements. So no such x can exist. The proof is now done, so we get to write '$\Box$'.

It was perhaps a bit misleading to call the above text "scrap paper". Our personal scrap paper arguments are nonlinear and full of mistakes, dead ends, unanswered questions, and fundamental confusions, all of which we have suppressed here for readability and ease of typesetting. Notice several things about our proof:

1. From the beginning, the problem is broken down into smaller problems (eg., verifying the axioms in turn).

2. These smaller statements are constantly being translated into logically equivalent statements. If any of these statements is proven true, the proposition is true; if any is proven false, the proposition is false.

Summary. Now that we have worked out this proof, we can summarize it in a short space. There were two main "tricks": the first was to use $\min\{\varepsilon_1, \varepsilon_2\}$ in the proof of Axiom C2, and the second was the "$\varepsilon/2$ trick" in the proof of C3. It is important to understand these techniques; the $\varepsilon/2$ trick is especially ubiquitous in mathematics. Our proof summary will focus on these less straightforward parts. The only facts we have used about $\mathbb{R}^n$ is that it is a metric space, and so we can state our result in this more general context.

Theorem 3.5 *Let (X, d) denote a metric space. Setting*

$$\mathbf{K}A \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid (\forall \varepsilon > 0)\ (\exists a \in A)\ d(x, a) < \varepsilon\}.$$

defines a closure operator on X called the standard closure induced by the metric d. Thus, every metric space can be considered a topological space in this standard way.

Proof.

To prove Axiom C1, suppose $x \in A$; we need to show that $x \in \mathbf{K}A$, i.e., $(\forall \varepsilon > 0)\ (\exists a \in A)\ d(x, a) < \varepsilon$. Given any $\varepsilon > 0$, we can just set $a = x$, and $d(x, a) = 0 < \varepsilon$.

For Axiom C2, it is not hard to see that $\mathbf{K}A \cup \mathbf{K}B \subset \mathbf{K}(A \cup B)$. To show $\mathbf{K}(A \cup B) \subset \mathbf{K}A \cup \mathbf{K}B$, assume $x \notin \mathbf{K}A \cup \mathbf{K}B$; we will show that $x \notin \mathbf{K}(A \cup B)$. Since $x \notin \mathbf{K}A$, we can choose $\varepsilon_1 > 0$ such that $(\forall a \in A)\ d(x, a) \geq \varepsilon_1$. Since $x \notin \mathbf{K}B$, we can choose $\varepsilon_2 > 0$ such that $(\forall a \in B)\ d(x, a) \geq \varepsilon_2$. Let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. Then $\varepsilon > 0$, but $d(x, a) \geq \varepsilon$ for every $a \in A \cup B$, so $x \notin \mathbf{K}(A \cup B)$.

Since we have already proved Axiom C1, we know that $\mathbf{K}A \subset \mathbf{K}\mathbf{K}A$. So to check Axiom C3, it is enough to show that $\mathbf{K}\mathbf{K}A \subset \mathbf{K}A$. Suppose $x \in \mathbf{K}\mathbf{K}A$ and $\varepsilon > 0$; we need to find $a \in A$ such that $d(x, a) < \varepsilon$. Since $x \in \mathbf{K}(\mathbf{K}A)$ and $\varepsilon/2 > 0$, we can choose $a_1 \in \mathbf{K}A$ such that $d(x, a_1) < \varepsilon/2$. Since $a_1 \in \mathbf{K}A$, we can choose $a \in A$ such that $d(a_1, a) < \varepsilon/2$. By the triangle inequality, $d(x, a) \leq d(x, a_1) + d(a_1, a) < \varepsilon/2 + \varepsilon/2 = \varepsilon$.

The verification of axiom C4 is left to the reader.

□

Why have we gone to all this trouble to prove that $\mathbf{K}$ *is a closure operator?* The answer is that everything we proved about closure operators in Chapter 1 (and everything we will prove about closure operators later) is automatically true for the standard closure operator. This includes everything we proved in §1.4. For example, we now know without any further work that

$$\mathbf{K}(A \cap B) \subset \mathbf{K}A \cap \mathbf{K}B.$$

While it might be amusing to prove this directly from the definition of **K**; the axiomatic approach makes this unnecessary.

Moreover, careful inspection of the proof above shows that we have used nothing about $\mathbb{R}^n$ or the Euclidean metric beyond the fact that $(\mathbb{R}^n, d_e)$ is a metric space. Our argument therefore gives us a standard way of defining a closure operator on any metric space so as to make it into a topological space. When there is a metric around, this will always be the closure operator assumed when discussing topological matters unless stated otherwise. As we will see in the exercises, different metrics on the same set may generate the same or different closure operators through this process.

Balls. You are probably familiar with the equation for a disk of radius r about the point (a, b) in the plane:

$$\sqrt{(y-b)^2 + (x-a)^2} < r$$

We see that this is equivalent to the equation

$$d((x, y), (a, b)) < r$$

Let us now generalize on this notion. If $x \in \mathbb{R}^n$ and $r > 0$, the **open ball of radius r centered at x**, denoted by $B(x; r)$, is defined as follows:

$$B(x; r) \stackrel{\text{def}}{=} \{y \in \mathbb{R}^n \mid d(x, y) < r\}.$$

The semicolon in the notation is a reminder that x and r are different kinds of objects; the first is an ordered n-tuple of real numbers, and the second is a single real number.

We can now redefine our closure operator in terms of balls. Definition 3.4 is equivalent to the statement that for any subset $A \subset \mathbb{R}^n$ and any point $x \in \mathbb{R}^n$,

$$x \in \mathbf{K}A \iff (\forall \varepsilon > 0)\ B(x; \varepsilon) \cap A \neq \emptyset;$$
$$x \notin \mathbf{K}A \iff (\exists \varepsilon > 0)\ B(x; \varepsilon) \cap A = \emptyset.$$

This notation is common in the fields of topology and analysis. Again, the same definitions work not just in the Euclidean case as we have stated them, but in any metric space.

(See figure 3.1.)

Figure 3.1: $x_1 \in \mathbf{K}A$ because every ball around x_1 intersects A, but $x_2 \notin \mathbf{K}A$ because there is a ball around x_2 that does not intersect A.

Exercises

1. Try to prove directly from the definition of the standard closure operator $\mathbf{K}$ that $A \subset B \Longrightarrow \mathbf{K}A \subset \mathbf{K}B$ and $\mathbf{K}(A \cap B) \subset \mathbf{K}A \cap \mathbf{K}B$, even though we already know that these are true because $\mathbf{K}$ is a closure operator.

2. Suppose $x_1, x_2 \in \mathbb{R}^n$ and $r_1, r_2 > 0$. Use the triangle inequality to prove the following:

   (a) Show that $d(x_1, x_2) \leq r_2 - r_1 \Longrightarrow B(x_1; r_1) \subset B(x_2; r_2)$.

   (b) Show that $d(x_1, x_2) \geq r_1 + r_2 \Longrightarrow B(x_1; r_1) \cap B(x_2; r_2) = \emptyset$.

   (c) Do you think the converses of these statements are true? For example, if $B(x_1; r_1) \subset B(x_2; r_2)$, does it follow that $d(x_1, x_2) \leq r_2 - r_1$? Try drawing pictures in $\mathbb{R}^2$.

3. Given a nonempty X, show that you can define a metric by setting $d(x, y) = 1$ whenever $x \neq y$ in X. What closure operator does this induce?

4. Consider the set $\mathbb{L} = \mathbb{Z}^+ \cup \{\infty\}$. Show that you can define a metric on $\mathbb{L}$ by setting

$$d(n, m) = \begin{cases} |1/n - 1/m| & \text{if } m, n \in \mathbb{Z}^+ \\ 1/n & \text{if } m = \infty \ . \end{cases}$$

   Prove that this metric induces the closure we have studied before on this space, namely

$$\mathbf{K}A = \begin{cases} A & \text{if } A \text{ is finite,} \\ A \cup \{\infty\} & \text{if } A \text{ is infinite.} \end{cases}$$

3.4 Examples

We can now apply our closure operator to some common topological objects - single points, intervals on the real line, balls in $\mathbb{R}^n$.

Example 3.6 If $a \in \mathbb{R}^n$, then $\mathbf{K}\{a\} = \{a\}$.

Proof.
Because we automatically know $\{a\} \subset \mathbf{K}\{a\}$ by Axiom C1, we need only show that $\mathbf{K}\{a\} \subset \{a\}$. Suppose $x \notin \{a\}$; we will show $x \notin \mathbf{K}\{a\}$. Since $x \notin \{a\}$, $x \neq a$, so $d(x, a) > 0$. Let $\varepsilon = d(x, a)$; then $B(x; \varepsilon) \cap \{a\} = \emptyset$, since $a \notin B(x; \varepsilon)$, so $x \notin \mathbf{K}\{a\}$.

□

The following definitions for sets on the real line are probably familiar to you already:

Let a and b be any real numbers with $a < b$. The **open interval** (a, b) is defined by

$$(a, b) \stackrel{\text{def}}{=} \{x \in \mathbb{R} \mid a < x < b\}.$$

As early as Chapter 1, we hinted that the closure operator would allow us to add the endpoints, a and b, to this sort of interval. Here, we will demonstrate that this fundamental intuition was correct.

As a notation check, observe that

$$(a, b) = B\left(\frac{a+b}{2}; \frac{b-a}{2}\right).$$

For $a \leq b$, the **closed interval** $[a, b]$ is defined to be

$$[a, b] \stackrel{\text{def}}{=} \{x \in \mathbb{R} \mid a \leq x \leq b\}.$$

Let's now see how they are related.

Example 3.7 If $a < b$, then $\mathbf{K}(a, b) = [a, b]$.

Proof.
We know that $\mathbf{K}(a, b) \subset \mathbb{R}$. So we need to show that $(\forall x \in \mathbb{R})\ x \in \mathbf{K}(a, b) \iff x \in [a, b]$. Let x be any real number. We will need to consider five cases, each for a different value of x.

Case $x < a$. Let $\varepsilon = a - x$; then $\varepsilon > 0$ and $B(x; \varepsilon) \cap (a, b) = \emptyset$ (why?), so $x \notin \mathbf{K}(a, b)$.

Case $x = a$. Given $\varepsilon > 0$, let $y = x + \min\{\varepsilon/2, (b-a)/2\}$; then $y \in B(x;\varepsilon) \cap (a,b)$, so $B(x;\varepsilon) \cap (a,b) \neq \emptyset$. Since this is true for any $\varepsilon > 0$, $x \in \mathbf{K}(a,b)$.

Case $a < x < b$. By Axiom C1, $x \in \mathbf{K}(a,b)$.

Cases $x = b, x < b$. These are analogous to previous cases.

□

Example 3.8 If $a \leq b$, then $\mathbf{K}[a,b] = [a,b]$.

Proof.
This follows from the previous example and Axiom C3.

□

Our next important example concerns calculating $\mathbf{K}A$ where $A = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4} \ldots\} = \{\frac{1}{n} | n \in \mathbb{Z}^+\}$. In many ways, this example serves as a model for many other closures we will take on the real line. We expect to find, of course, that $\mathbf{K}A$ will turn out to be $A \cup \{0\}$ since zero is all but in the set A already. For a careful proof, however, we must invoke one of the defining properties of $\mathbb{R}$ we have not used yet, which we state generally as follows:

Definition 3.9 *We say that an ordered field* $\mathbb{F}$ *containing* $\mathbb{Z}$ *has the* **Archimedean property** *if, for all* $\epsilon \geq 0$ *in* $\mathbb{F}$, $\epsilon < \frac{1}{n}$ *for all* $n \in \mathbb{Z}^+$ *implies that* $\epsilon = 0$.

Thus, for every positive real number, there exists a positive integer n such that $1/n < \varepsilon$. What does this property really say about the real line? If we think about flipping the inequality above, it says that you cannot invert a real number, no matter how small, to create another number that is bigger than all the integers. This rules out the existence of any "infinitesmals" - infinitely small numbers - or infinite numbers which are larger than any integer. These sorts of numbers can be allowed to exist, and they form an important part of "non-standard analysis" which is a bona-fide field of mathematical study. However, we will deal only with standard analysis in this text.

Perhaps more strikingly, this property is equivalent to the statement that every interval of length one in $\mathbb{F}$ contains an integer. If this property holds, for any $\epsilon < 0$ we can write

$$\frac{1}{\epsilon} < n \leq \frac{1}{\epsilon} + 1,$$

so that $0 < \frac{1}{n} < \epsilon$. This means that as long as the numbers in $\mathbb{F}$ run on, out to infinity, the integers will keep pace with them, marking off intervals of length one in their midst.

The ordered field $\mathbb{Q}$ is Archimidean since, a rational $\epsilon > 0$ must be of the form $\frac{p}{q}$ for some positive integers p and q and so $1/n < \varepsilon$ holds for $n = 2q$.

In what follows, we will posit that $\mathbb{R}$ also satisfies this Archimedean condition and refer to this as the **Archimedean Principle for $\mathbb{R}$.** Not until Chapter **??** will we show that there actually exists a set $\mathbb{R}$ with all the properties we desire, including this one. Let us proceed to our example.

Example 3.10 Let $A = \{1, 1/2, 1/3, 1/4, \ldots\}$. Then $\mathbf{K}A = A \cup \{0\}$.

Proof. We must show that $(\forall x \in \mathbb{R})\ x \in \mathbf{K}A \iff x \in A \cup \{0\}$. Let $x \in \mathbb{R}$ be given. Again, we will consider five cases.

Case $x < 0$. Let $\varepsilon = -x$; then $\epsilon > 0$ and $B(x;\varepsilon) \cap A = \emptyset$, so $x \notin \mathbf{K}A$.

Case $x = 0$. Let $\varepsilon > 0$ be given. Let n be a positive integer such that $1/n < \varepsilon$. Then $1/n \in B(x;\varepsilon) \cap A$, so $B(x;\varepsilon) \cap A \neq \emptyset$. Since this is true for every $\varepsilon > 0$, $x \in \mathbf{K}A$.

Case $x \in A$. By Axiom C1, $x \in \mathbf{K}A$.

Case $0 < x < 1$ and $x \notin A$. There exists an integer n such that $1/n < x$, so let m be the smallest positive integer with this property. (We can do this by the Well-Ordering Principle, which says that every nonempty set of positive integers has a smallest element. This can be thought of as one of the axioms *defining* the integers. For elucidation, see Appendix C.) Then

$$\frac{1}{m} < x < \frac{1}{m-1}.$$

(There is no division by zero here: $m - 1 > 0$ because $x < 1$ implies $m > 1$.) Let

$$\varepsilon = \min\left\{x - \frac{1}{m}, \frac{1}{m-1} - x\right\};$$

then $\varepsilon > 0$, and it is easy to check that $B(x;\varepsilon) \cap A = \emptyset$. Thus $x \notin \mathbf{K}A$.

Case $x > 1$. Let $\varepsilon = x - 1$; then $\varepsilon > 0$ and $B(x;\varepsilon) \cap A = \emptyset$, so $x \notin \mathbf{K}A$.

□

Notice in this example that we could eliminate any finite number of points from A and still find 0 in the closure of the set that is left. In other words,

$$0 \in \cap_{n=1}^{\infty} \mathbf{K}\{1/m \mid m \geq n\}$$

so that the property of belonging to the closure in this sense depends on the "tails" of the sequence rather than its beginnings.

Figure 3.2: Proof of Example 3.11.

Here is some more useful terminology. If you remember back to when we discussed "balls" in the last section, you will recall that we only defined an "open ball" in $\mathbb{R}^n$. Generalizing on Example 3.7, it is possible to see how "closed balls" are made from "open balls."

Example 3.11 Given $x \in \mathbb{R}^n$ and $r > 0$,

$$\mathbf{K}B(x;r) = \{y \in \mathbb{R}^n \mid d(x,y) \le r\}.$$

Proof. Suppose $d(x,y) > r$. Let $\varepsilon = d(x,y) - r$; then $\varepsilon > 0$ and $B(y;\varepsilon) \cap B(x;r) = \emptyset$. (See Exercise 3.3.2b.) Thus $y \notin \mathbf{K}B(x;r)$.

If $d(x,y) < r$, then $y \in B(x;r)$, so $y \in \mathbf{K}B(x;r)$.

Finally, suppose $d(x,y) = r$. Given $\varepsilon > 0$, let $l = \min\{\varepsilon/2, r\}$. Let a be the point at distance l from y, in the direction of x. Then $a \in B(y;\varepsilon) \cap B(x;r)$, so $B(y;\varepsilon) \cap B(x;r) \neq \emptyset$. (See Figure 3.2.) Since this is true for every $\varepsilon > 0$, $y \in \mathbf{K}B(x;r)$.

□

Thus, the closure of an open ball can be thought of as being a closed ball, by analogy to open and closed intervals on $\mathbb{R}$.

Exercises

1. For $a < b$, define

$$[a,b) \overset{\text{def}}{=} \{x \in \mathbb{R} \mid a \le x < b\}.$$

$$(a,b] \overset{\text{def}}{=} \{x \in \mathbb{R} \mid a < x \le b\}.$$

These are called **half-open intervals**. What is the closure of each of these sets, and why?

2. Find two sets $A, B \subset \mathbb{R}$ such that $\mathbf{K}(A \cap B) \neq \mathbf{K}A \cap \mathbf{K}B$. *Hint:* use two suitably chosen open intervals.

3. What is $\mathbf{K}\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 < 0\}$? Why?

4. Suppose $x \in \mathbb{R}^n$ and $r > 0$. Determine the interior, exterior, and boundary of $B(x; r)$. (See Exercise 1.4.3 for the definitions of these terms.)

5. Define a function d' on pairs of points in $\mathbb{R}^n$ as follows:

$$d'(x, y) = \sum_{i=1}^{n} |x_i - y_i| .$$

Check that d' is a metric. Now if $A \subset \mathbb{R}^n$, define

$$\mathbf{K}'A = \left\{x \in \mathbb{R}^n \mid (\forall \varepsilon > 0)\ (\exists a \in A)\ d'(x, a) < \varepsilon\right\} .$$

Show that $\mathbf{K}' = \mathbf{K}$; that is, $\mathbf{K}'A = \mathbf{K}A$ for all $A \subset \mathbb{R}^n$. (The metrics d and d' are said to be "equivalent." There are many metrics that are equivalent to the standard metric, so the definition of closure in $\mathbb{R}^n$ does not depend too strongly on the definition of distance.)

6. (challenge problem, hard) Let $(X, \mathbf{K})$ be a topological space and let $A \subset X$. Show that there are at most 14 sets that can be obtained by starting from A and using the operations of complementation and closure. (If $Y \subset X$, the *complement* of Y is defined to be $X - Y$.) Given an example of a subset of $\mathbb{R}$ from which 14 different sets can be so obtained.

3.5 A proof of the triangle inequality*

In this section we will show how to prove the triangle inequality. The algebraically disinclined reader may skip this and take the triangle inequality on faith. However, our proof is an instructive example of the proof-technique of "working backwards", regardless of the particular algebraic details. In addition, many of the inequalities generated in the course of the proof perennially crop up in the fields of topology and analysis.

Let $x, y, z \in \mathbb{R}^n$. We wish to show that

$$\sqrt{\sum (x_i - y_i)^2} + \sqrt{\sum (y_i - z_i)^2} \geq \sqrt{\sum (x_i - z_i)^2}.$$

Our strategy will be to work backwards and reduce this to succesively simpler statements, until we obtain something we know to be true.

Let $a_i = x_i - y_i$ and $b_i = y_i - z_i$. Notice that $a_i + b_i = x_i - z_i$, so we just need to show that

$$\sqrt{\sum a_i^2} + \sqrt{\sum b_i^2} \geq \sqrt{\sum (a_i + b_i)^2}.$$

(The geometric idea here is that we are translating x, y and z in order to make $y = 0$, which simplifies the algebra.) Since both sides are nonnegative, it is enough to show that

$$\left(\sqrt{\sum a_i^2} + \sqrt{\sum b_i^2}\right)^2 \geq \left(\sqrt{\sum (a_i + b_i)^2}\right)^2.$$

Multiplying out and cancelling, we find that the above inequality is equivalent to

$$\sqrt{\sum a_i^2 \sum b_i^2} \geq \sum a_i b_i.$$

Since the left side is nonnegative, it is enough to show that

$$\sum a_i^2 \sum b_i^2 \geq \left(\sum a_i b_i\right)^2.$$

(This is called the Cauchy-Schwartz inequality.) This multiplies out to

$$\sum_{i,j=1}^{n} a_i^2 b_j^2 \geq \sum_{i,j=1}^{n} a_i b_i a_j b_j.$$

(Here both sums range over all n^2 possible pairs (i, j).) If we can show that for every i and j, the sum of the (i, j) and (j, i) terms on the left is greater than or equal to the sum of the (i, j) and (j, i) terms on the right, then we are done. In other words, we wish to show that

$$a_i^2 b_j^2 + a_j^2 b_i^2 \geq a_i b_i a_j b_j + a_j b_j a_i b_i.$$

Moving everything over to the left side, we find that this is equivalent to

$$(a_i b_j - a_j b_i)^2 \geq 0.$$

But we know this is true, since the square of any real number is nonnegative. This completes the proof.

A logical caveat: this sort of proof works only when all of the steps are reversible! The logical progression of our proof goes something like this: $a \iff b$; b is true, so therefore, a is true (be sure you can identify a and b in the proof above!). Without logical reversibility, the first arrow in our argument would only point from a to b (instead of being bi-directional), and our (erroneous) conclusion would be based on the logical fallacy of *affirming the consequent*. It is not difficult to think of ridiculous and amusing conclusions based on this fallacy.

For example, in our zeal to prove that $x = y$, we might be tempted to multiply both sides by zero to obtain $0 = 0$. $x = y \Longrightarrow 0 = 0$ is certainly true, but difficulty arises in proving $x = y \Longleftarrow 0 = 0$ (showing reversibility), since division by zero is undefined. No conclusion can be made based on this operation, and the "proof" is invalid.

Some mathematicians prefer a writing style in which one always proceeds forwards. For example, the above proof would begin, "Since the square of any real number is nonnegative, $(a_i b_j - a_j b_i)^2 \geq 0$. Moving some terms over to the right, we obtain ..." Such a style can make it harder to see how one thought of the proof, but easier to see that the proof is correct. It is often helpful to include an informal discussion of strategy preceding the formal proof.

Exercises

1. Let a and b be any positive numbers. By working backwards, find proofs of the following three inequalities:

$$\sqrt{\frac{a^2+b^2}{2}} \geq \frac{a+b}{2} \geq \sqrt{ab} \geq \frac{2}{\frac{1}{a}+\frac{1}{b}}.$$

Also show that in each inequality, equality holds if and only if $a = b$.

(The four quantities here are called the *root-mean-square*, *arithmetic mean*, *geometric mean*, and *harmonic mean* of a and b, respectively. You have probably seen examples of the arithmetic and geometric means before. The root-mean-square arises as follows: if you have two square plots of land with side lengths a and b, then the side length of a plot of land whose area is the average (arithmetic mean) of the areas of the original two plots is equal to the root-mean-square of a and b. Also, if you drive a kilometer at velocity a and drive another kilometer at velocity b, then your average speed over time is equal to the harmonic mean of a and b.)