Introduction to Real Analysis
Fall 2014 Lecture Notes

Vern I. Paulsen

November 6, 2014

Contents

Chapter 1

Metric Spaces

These notes accompany the Fall 2011 *Introduction to Real Analysis* course

1.1 Definition and Examples

Definition 1.1. *Given a set X a* **metric** *on X is a function $d : X \times X \to \mathbb{R}$ satisfying:*

1. *for every $x, y \in X$, $d(x, y) \geq 0$,*
2. *$d(x, y) = 0$ if and only if $x = y$,*
3. *$d(x, y) = d(y, x)$,*
4. *(triangle inequality) for every $x, y, z \in X$, $d(x, y) \leq d(x, z) + d(z, y)$.*

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

Example 1.2. *Let $X = \mathbb{R}$ and set $d(x, y) = |x - y|$, then d is a metric on $\mathbb{R}$. We call this the* usual metric *on $\mathbb{R}$.*

To prove it is a metric we verify (1)–(4). For (1): $d(x, y) = |x - y| \geq 0$, by the definition of the absolute value functions so (1). Since $d(x, y) = 0$ if and only if $|x - y| = 0$ if and only if $x = y$, (2) follows. (3) follows since $d(x, y) = |x - y| = |y - x| = d(y, x)$. Finally, for (4), $d(x, y) = |x - y| = |x - z + z - y| \leq |x - z| + |z - y| = d(x, z) + d(z, y)$.

Example 1.3 (The taxi cab metric)**.** *Let $X = \mathbb{R}^2$. Given $x = (x_1, x_2), y = (y_1, y_2)$, set $d(x, y) = |x_1 - y_1| + |x_2 - y_2|$, then d is a metric on $\mathbb{R}^2$.*

We verify (1)–(4). (1) and (3) are obvious. For (2): $d(x,y) = 0$ iff $|x_1 - y_1| + |x_2 - y_2| = 0$. But since both terms in the sum are non-negative for the sum to be 0, each one must be 0. So $d(x,y) = 0$ iff $|x_1 - y_1| = 0$ AND $|x_2 - y_2| = 0$ iff $x_1 = y_1$ and $x_2 = y_2$ iff $x = (x_1, x_2) = (y_1, y_2) = y$. Finally to see (4):

$$\begin{aligned} d(x,y) &= |x_1 - y_1| + |x_2 - y_2| = |x_1 - z_1 + z_1 - y_1| + |x_2 - z_2 + z_2 - y_2| \\ &\le |x_1 - z_1| + |z_1 - y_1| + |x_2 - z_2| + |z_2 - y_2| = d(x,z) + d(z,y). \end{aligned}$$

We often denote the taxi cab metric by $d_1(x,y)$.

Example 1.4. *A different metric on $\mathbb{R}^2$. For $x = (x_1, x_2), y = (y_1, y_2)$ set $d_\infty(x,y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$. So the distance between two points is the larger of these two numbers.*

We only check the triangle inequality. Let $z = (z_1, z_2)$ be another point. We have two cases to check. Either $|x_1 - y_1| = d(x,y)$ OR $d(x,y) = |x_2 - y_2|$.

Case 1: $d(x,y) = |x_1 - y_1|$.

Now notice that $|x_1 - z_1| \le \max\{|x_1 - z_1|, |x_2 - z_2|\} = d(x,z)$. Similarly, $|z_1 - y_1| \le \max\{|z_1 - y_1|, |z_2 - y_2|\}$. Hence,

$$\begin{aligned} d(x,y) = |x_1 - y_1| = |x_1 - z_1 + z_1 - y_1| \le |x_1 - z_1| + |z_1 - y_1| \\ \le d(x,z) + d(z,y). \end{aligned}$$

Case 2: $d(x,y) = |x_2 - y_2|$. Now use $|x_2 - z_2| \le \max\{|x_1 - z_1|, |x_2 - z_2|\} = d(x,z)$. Similarly, $|z_2 - y_2| \le \max\{|z_1 - y_1|, |z_2 - y_2|\}$. Hence,

$$\begin{aligned} d(x,y) = |x_2 - y_2| = |x_2 - z_2 + z_2 - y_2| \le |x_2 - z_2| + |z_2 - y_2| \\ \le d(x,z) + d(z,y). \end{aligned}$$

So in each case the triangle inequality is true, so it is true.

Example 1.5. *In this case we let X be the set of all continuous real-valued functions on $[0,1]$. We use three facts from Math 3333:*

1. *if f and g are continuous on [0,1], then $f - g$ is continuous on [0,1],*

2. *if f is continuous on [0,1], then $|f|$ is continuous on [0,1],*

3. *if h is continuous on [0,1], then there is a point $0 \le t_0 \le 1$, so that $h(t) \le h(t_0)$ for every $0 \le t \le 1$. That is $h(t_0) = \max\{h(t) : 0 \le t \le 1\}$.*

Now given $f, g \in X$, we set $d(f,g) = \max\{|f(t) - g(t)| : 0 \leq t \leq 1\}$.

Note that by (1) and (2) $|f - g|$ is continuous and so by (3) there is a point where it acheives its maximum.

We now show that d is a metric on X. Clearly, (1) holds. Next, if $d(f,g) = 0$, then the maximum of $|f(t) - g(t)|$ is 0, so we must have that $|f(t) - g(t)| = 0$ for every t. But then this means that $f(t) = g(t)$ for every t, and so $f = g$. So $d(f,g) = 0$ implies $f = g$. Also $f = g$ implies $d(f,g) = 0$ so (2) holds. Clearly (3) holds. Finally to see the triangle inequality, we let f, g, h be three continuous functions on [0,1]. that is, $f, g, h \in X$. We must show that $d(f,g) \leq d(f,h) + d(h,g)$.

We know that there is a point $t_0, 0 \leq t_0 \leq 1$, so that

$$d(f,g) = \max\{|f(t) - g(t)| : 0 \leq t \leq 1\} = |f(t_0) - g(t_0)|.$$

Hence,

$$\begin{aligned} d(f,g) = |f(t_0) - g(t_0)| &= |f(t_0) - h(t_0) + h(t_0) - g(t_0)| \\ &\leq |f(t_0) - h(t_0)| + |h(t_0) - g(t_0)| \\ \leq \max\{|f(t) - h(t)| &: 0 \leq t \leq 1\} + \max\{|h(t) - g(t)| : 0 \leq t \leq 1\} \\ &= d(f,h) + d(h,g) \end{aligned}$$

Example 1.6 (Euclidean space, Euclidean metric)**.** *Let $X = \mathbb{R}^n$ the set of real n-tuples. For $x = (a_1, \ldots, a_n)$ and $y = (b_1, \ldots, b_n)$ we set*

$$d(x,y) = \sqrt{(a_1 - b_1)^2 + \cdots + (a_n - b_n)^2}.$$

This defines a metric on $\mathbb{R}^n$, which we will prove shortly. This metric is called the **Euclidean metric** *and $(\mathbb{R}^n, d)$ is called* **Euclidean space.** *Sometimes, we will write d_2 for the Euclidean metric.*

It is easy to see that the Euclidean metric satisfies (1)–(3) of a metric. It is harder to prove the triangle inequality for the Euclidean metric than some of the others that we have looked at. This requires some results first.

Lemma 1.7. *Let $p(t) = at^2 + bt + c$ with $a \geq 0$. If $p(t) \geq 0$ for every $t \in \underline{R}$, then $b^2 \leq 4ac$.*

Proposition 1.8 (Cauchy-Schwarz Inequality)**.** *Let $a_1, ..., a_n, b_1, ..., b_n$ be real numbers. Then*

$$|a_1 b_1 + \cdots a_n b_n| \leq \sqrt{a_1^2 + \cdots + a_n^2}\sqrt{b_1^2 + \cdots + b_n^2}.$$

Proof. Look at $p(t) = (ta_1 + b_1)^2 + \cdots + (ta_n + b_n)^2 \geq 0$ for all t. $\square$

Corollary 1.9 (Minkowski's Inequality). $\sqrt{(a_1 + b_1)^2 + \cdots + (a_n + b_n)^2} \leq \sqrt{a_1^2 + \cdots + a_n^2} + \sqrt{b_1^2 + \cdots + b_n^2}$.

Proof. Let LHS denote the left hand side, RHS the right hand side of the inequality. then

$$
\begin{aligned}
(LHS)^2 = (a_1+b_1)^2+\cdots+(a_n+b_n)^2 &= a_1^2+\cdots+a_n^2+2(a_1b_1+\cdots+a_nb_n)+b_1^2+\cdots+b_n^2 \\
&\leq a_1^2 + \cdots + a_n^2 + 2\sqrt{a_1^2 \cdots + a_n^2}\sqrt{b_1^2 + \cdots + b_n^2} + b_1^2 + \cdots + b_n^2 \\
&= (\sqrt{a_1^2 + \cdots + a_n^2} + \sqrt{b_1^2 + \cdots + b_n^2})^2 = (RHS)^2
\end{aligned}
$$

$\square$

Now prove the triangle inequality.

Example 1.10 (The discrete metric). *Let X be any non-empty set and define*

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

Then this is a metric on X called the **discrete metric** *and we call (X, d) a* **discrete metric space.**

Example 1.11. *When (X, d) is a metric space and $Y \subseteq X$ is a subset, then restricting the metric on X to Y gives a metric on Y, we call (Y, d) a* **subspace of (X,d).**

Problem 1.12. *Let $X = \mathbb{R}$, define $d(x, y) = |x - y| + 1$. Show that this is NOT a metric.*

Problem 1.13. *Let $X = \mathbb{R}$, define $d(x, y) = |x^2 - y^2|$. Show that this is NOT a metric.*

Problem 1.14. *Let $X = \mathbb{R}$, define $d(x, y) = |x - y| + |x^2 - y^2|$. Prove that this is a metric on $\mathbb{R}$.*

Problem 1.15. *Let $X = \mathbb{R}^n$ for $x = (a_1, ..., a_n), y = (b_1, ..., b_n)$ define*

$$d_1(x, y) = |a_1 - b_1| + \cdots + |a_n - b_n|.$$

Prove that this is a metric.

Problem 1.16. *Let $X = \mathbb{R}^n$ for x and y as before, define*

$$d_\infty(x, y) = \max\{|a_1 - b_1|, ..., |a_n - b_n|\}.$$

Prove that this is a metric.

1.1.1 Vector Spaces, Norms and Metrics

Definition 1.17. *Let X be a real vector space. A function $\|\cdot\| : X \to \mathbb{R}$ is called a* **norm** *provided that*

1. $\|x\| \geq 0$ *for all* x,
2. $\|x\| = 0$ *if and only if* $x = 0$,
3. $\|rx\| = |r|\|x\|$ *for every* $r \in \mathbb{R}$ *and* $x \in X$,
4. *(triangle inequality)* $\|x + y\| \leq \|x\| + \|y\|$.

The next result summarizes the relation between this concept and norms.

Proposition 1.18. *Let X be a real vector space and let $\|\cdot\|$ be a norm on X. Then setting $d(x,y) = \|x - y\|$ defines a metric on X.*

Proof. We have that $d(x,y) = \|x - y\| \geq 0$ so property (1) of a metric holds.

We have that $d(x,y) = 0$ iff $\|x - y\| = 0$ iff $x - y = 0$ iff $x = y$, so property (2) of a metric holds.

Also, $d(x,y) = \|x - y\| = \|(-1)(y - x)\| = |-1| \cdot \|y - x\| = \|y - x\| = d(y,x)$, so property (3) of a metric holds.

Finally, $d(x,y) = \|x - y\| = \|x - z + z - y\| \leq \|x - z\| + \|z - y\| = d(x,z) + d(z,y)$ so property (4) of a metric holds. □

The metric that one gets from a norm by using this result is called the **metric induced by the norm**.

Example 1.19. *For $x = (a_1, ..., a_n) \in \mathbb{R}^n$ if we set $\|x\|_2 = \sqrt{a_1^2 + \cdots + a_n^2}$ then this defines a norm called the* **Euclidean norm**.

To see that this is a norm, it is easy to check the first 3 properties needed for a norm. To see the fourth property, use Corollary 1.9. Thus, the Euclidean metric is the metric induced by the Euclidean norm via Proposition 1.18.

Example 1.20. *For $x = (a_1, a_2) \in \mathbb{R}^2$ if we set $\|x\|_1 = |a_1| + |a_2|$ and $\|x\|_\infty = \max\{|a_1|, |a_2|\}$ then these both define norms on $\mathbb{R}^2$ and the metrics that they induce are respectively, d_1–the taxi cab metric and d_∞.*

1.2 Open Sets

Definition 1.21. *Let (X,d) be a metric space, fix $x \in X$ and $r > 0$. The* **open ball of radius r centered at x** *is the set*

$$B(x;r) = \{y \in X : d(x,y) < r\}.$$

Example 1.22. *In $\mathbb{R}$ with the usual metric, $B(x;r) = \{y : |x-y| < r\} = \{y : x-r < y < x+r\} = (x-r, x+r)$.*

Example 1.23. *In $\mathbb{R}^2$ with the Euclidean metric, $x = (x_1, x_2)$, then $B(x;r) = \{(y_1,y_2) : (x_1-y_1)^2 + (x_2-y_2)^2 < r^2\}$, which is a disk of radius r centered at x.*

Example 1.24. *In $\mathbb{R}^3$ with the Euclidean metric, $B(x;r)$ really is an open ball of radius r. This example is where the name comes from.*

Example 1.25. *In $(\mathbb{R}^2, d_\infty)$ we have*

$$B(x;r) = \{(y_1,y_2) : |x_1-y_1| < r \text{ and } |x_2-y_2| < r\} = \\ \{(y_1,y_2) : x_1 - r < y_1 < x_1 + r \text{ and } x_2 - r < y_2 < y_2 + r\}.$$

So now an open "ball" is actually an open square, centered at x with sides of length 2r.

Example 1.26. *In $(\mathbb{R}^2, d_1)$ we have $B((0,0);1) = \{(x,y) : |x-0|+|y-0| < 1\}$ which can be seen to be the "diamond" with corners at (1,0),(0,1),(-1,0), (0,-1).*

Example 1.27. *When $X = \{f : [0,1] \to \mathbb{R} | f \text{ is continuous}\}$ and $d(f,g) = \max\{|f(t)-g(t)| : 0 \le t \le 1\}$, then $B(f;r) = \{g : g \text{ is continuous and } f(t) - r < g(t) < f(t) + r, \forall t\}$. This can be pictured as all continuous functions g whose graphs lie in a band of width r about the graph of f.*

Example 1.28. *If we let $\mathbb{R}$ have the usual metric and let $Y = [0,1] \subseteq \mathbb{R}$ be the subspace, then when we look at the metric space Y we have that $B(0;1/2) = [0,1/2) = (-1/2,1/2) \cap Y$.*

Example 1.29. *When we let X be a set with the discrete metric and $x \in X$, then $B(x;r) = \{x\}$ when $r \le 1$. When $r > 1$, then $B(x;r) = X$.*

Definition 1.30. *Given a metric space (X,d) a subset $\mathcal{O} \subseteq X$ is called* **open** *provided that whenever $x \in \mathcal{O}$, then there is an $r > 0$ such that $B(x;r) \subseteq \mathcal{O}$.*

Showing that sets are open really requires proof, so we do a few examples.

Example 1.31. *In* $\mathbb{R}$ *with the usual metric, an interval of the form* $(a, b) = \{x : a < x < b\}$ *is an open set.*

Proof. Given $x \in (a, b)$, let $r = \min\{x - a, b - x\}$. If $y \in B(x; r)$ then $x - r < y < x + r$. Since $y < x + r \leq x + (b - x) = b$ and $y > x - r \geq x - (x - a) = a$ we have $a < y < b$. So $B(x; r) \subseteq (a, b)$. Thus, (a, b) is open. □

So "open intervals" really are "open sets".

Example 1.32. *In* $\mathbb{R}^2$ *with the Euclidean metric, a rectangle of the form* $R = \{(y_1, y_2) : a < y_1 < b, c < y_2 < d\}$ *is an open set.*

Proof. Given $x = (x_1, x_2) \in R$, let $r = \min\{x_1 - a, b - x_1, x_2 - c, d - x_2\}$. We claim that $B(x; r) \subseteq R$. To see this let $y = (y_1, y_2) \in B(x; r)$, then $(y_1 - x_1)^2 + (y_2 - x_2)^2 < r^2$.

Looking at just one term in the sum, we see that $(y_1 - x_1)^2 < r^2$ and so $|y_1 - x_1| < r$. This implies that $x_1 - r < y_1 < x_1 + r$. As before we see that $x_1 + r \leq b$ and $x_1 - r \geq a$, so $a < y_1 < b$.

Similarly, we get that $c < y_2 < d$, so $y \in R$. □

Example 1.33. *In* $\mathbb{R}$ *with the usual metric an interval of the form* $[a, b)$ *is not open.*

Proof. Consider the point $a \in [a, b)$. Any ball about this point is of the form $B(a; r) = (a - r, a + r)$ and this contains points that are less than a and so not in the set $[a, b)$. □

CAREFUL! If we let $Y = [0, 1) \subseteq \mathbb{R}$ be the subspace. Then *in* Y the set $\mathcal{O} = [0, 1/2)$ is open!(Explain why.)

The next result justifies us calling $B(x; r)$ an *open* ball.

Proposition 1.34. *Let (X,d) be a metric space, fix* $x \in X$ *and* $r > 0$*. Then* $B(x; r)$ *is an open set.*

Theorem 1.35. *Let (X,d) be a metric space. Then*

1. *the empty set is open,*
2. *X is open,*
3. *the union of any collection of open sets is open,*
4. *the intersection of finitely many open sets is open.*

Proposition 1.36. *In a discrete metric space, every set is open.*

Problem 1.37. *Prove that the set* $\mathcal{O} = \{(y_1, y_2) : y_1 + y_2 > 0\}$ *is an open subset of* $\mathbb{R}^2$ *in the Euclidean metric.*

Problem 1.38. *Prove that the set* $\mathcal{O} = \{(y_1, y_2) : y_1 > 0\}$ *is an open subset of* $\mathbb{R}^2$ *in the Euclidean metric.*

Problem 1.39. *Prove that the disk* $D = \{(y_1, y_2) : y_1^2 + y_2^2 < 1\}$ *is an open subset of* $(\mathbb{R}^2, d_\infty)$.

Problem 1.40. *Let* X *be the set of continuous real-valued functions on [0,1] with the metric that we introduced. Prove that* $\mathcal{O} = \{g \in X : g(t) > 0 \forall t\}$ *is an open subset.*

1.2.1 Uniformly Equivalent Metrics

The definition of open set really depends on the metric. For example, on $\mathbb{R}$ if instead of the usual metric we used the discrete metric, then every set would be open. But we have seen that when $\mathbb{R}$ has the usual metric, then not every set is open. For example, $[a, b)$ is not an open subset of $\mathbb{R}$ in the usual metric. Thus, whether a set is open or not really can depend on the metric that we are using.

For this reason, if a given set X has two metrics, d and ρ, and we say that a set is open, we generally need to specify which metric we mean. Consequently, we will say that a set is **open with respect to** d or **open in (X, d)** when we want to specify that it is open when we use the metric d. In this case it may or may not be **open with respect to** ρ**.**

In the case of $\mathbb{R}^2$, we already have three metrics, the Euclidean metric d, the taxi cab metric d_1 and the metric d_∞. So when we say that a set is open in $\mathbb{R}^2$, we could potentially mean three different things. On the other hand it could be the case that all three of these metrics give rise to the same collection of open sets.

In fact, these three metrics do give rise to the same collections of open sets and the following definition and result explains why.

Definition 1.41. *Let* X *be a set and let* d *and* ρ *be two metrics on* X*. We say that these metrics are* **uniformly equivalent** *provided that there are constants* A *and* B *such that for every* $x, y \in X$,

$$\rho(x, y) \leq Ad(x, y) \text{ and } d(x, y) \leq B\rho(x, y).$$

Example 1.42. *On $\mathbb{R}^2$ the Euclidean d and the metric d_∞ are uniformly equivalent. In fact,*

$$d_\infty(x,y) \le d(x,y) \text{ and } d(x,y) \le \sqrt{2}d_\infty(x,y).$$

To see this, let $x = (a_1, a_2), y = (b_1, b_2)$. Since $|a_1 - b_1| \le \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2}$ and $|a_2 - b_2| \le \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2}$, we have that

$$d_\infty(x,y) = \max\{|a_1-b_1|, |a_2-b_2|\} \le \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2} = d(x,y).$$

On the other hand, since $|a_1 - b_1| \le d_\infty(x,y)$ and $|a_2 - b_2| \le d_\infty(x,y)$, we have that $(a_1-b_1)^2 + (a_2-b_2)^2 \le 2(d_\infty(x,y))^2$. Taking square roots of both sides, yields $d(x,y) \le \sqrt{2}d_\infty(x,y)$.

Example 1.43. *On $\mathbb{R}^2$ the Euclidean metric d and d_1 are uniformly equivalent. In fact,*

$$d(x,y) \le d_1(x,y) \text{ and } d_1(x,y) \le \sqrt{2}d(x,y).$$

We have that

$$\begin{aligned} &d_1(x,y)^2 = (|a_1-b_1| + |a_2-b_2|)^2 \\ &= |a_1-b_1|^2 + 2|a_1-b_1||a_2-b_2| + |a_2-b_2|^2 \ge (a_1-b_1)^2 + (a_2-b_2)^2 \\ &\qquad = (d(x,y))^2. \end{aligned}$$

Hence, $d(x,y) \le d_1(x,y)$.

To see the other inequality, we use the Schwarz inequality,

$$\begin{aligned} &d_1(x,y) = ||a_1-b_1| \cdot 1 + |a_2-b_2| \cdot 1| \le \\ &\qquad \sqrt{|a_1-b_1|^2 + |a_2-b_2|^2}\sqrt{1^2+1^2} = \sqrt{2}d(x,y) \end{aligned}$$

Given a set X with metrics d and ρ, a point $x \in X$ and $r > 0$, we shall write $B_d(x;r) = \{y \in X : d(x,y) < r\}$ and $B_\rho(x;r) = \{y \in X : \rho(x,y) < r\}$.

Lemma 1.44. *Let X be a set, let d and ρ be two metrics on X that are uniformly equivalent, and let A and B denote the constants that appear in Definition 1.41. Then $B_\rho(x;r) \subseteq B_d(x;Br)$ and $B_d(x;r) \subseteq B_\rho(x;Ar)$.*

Proof. If $y \in B_\rho(x;r)$, then $\rho(x,y) < r$ which implies that $d(x,y) \le B\rho(x,y) < Br$. Hence, $y \in B_d(x;Br)$ and so $B_\rho(x;r) \subseteq B_d(x;Br)$. The other case is proven similarly, using the other inequality. $\square$

Theorem 1.45. *Let X be a set and let d and ρ be metrics on X that are uniformly equivalent. Then a set is open with respect to d if and only if it is open with respect to ρ.*

Proof. Let $E \subseteq X$ be a set that is open with respect to d. We must prove that E is open with respect to ρ.

Since E is open with respect to d, given $x \in E$ there is an $r > 0$ so that $B_d(x; r) \subseteq E$. By the lemma, $B_\rho(x; r/B) \subseteq B_d(x; r)$. Thus, $B_\rho(x; r/B) \subseteq E$ and we have shown that given an arbitrary point in E that there is an open ball in the ρ metric about that point that is contained in E. Hence, E is open with respect to ρ.

The proof that a set that is open with respect to ρ is open with respect to d is similar and uses the other containment given in the lemma. $\square$

Problem 1.46. *Let X be a set with three metrics, d, ρ, and γ. Prove that if d and ρ are uniformly equivalent and ρ and γ are uniformly equivalent, then d and γ are uniformly equivalent.*

Problem 1.47. *Prove that on $\mathbb{R}^n$ the Euclidean metric, the metric d_1 and the metric d_∞ are all uniformly equivalent.*

1.3 Closed Sets

Definition 1.48. *Given a set X and $E \subseteq X$, the* **complement of E,** *denoted E^c is the set of all elements of X that are not in E, i.e.,*
$E^c = \{x \in X : x \notin E\}$.

Other notations that are used for the complement are $E^c = \mathcal{C}E = X \backslash E$.
Note that $(E^c)^c = E$.

Definition 1.49. *Let (X, d) be a metric space. Then a set $E \subseteq X$ is closed if and only if E^c is open.*

The following gives a useful way to re-state this definition.

Proposition 1.50. *Let (X, d) be a metric space. Then a set E is closed if and only if there is an open set $\mathcal{O}$ such that $E = \mathcal{O}^c$.*

Proof. If E is closed, then E^c is open. So let $\mathcal{O} = E^c$, then $E = \mathcal{O}^c$.

Conversely, if $E = \mathcal{O}^c$ for some open set $\mathcal{O}$, then $E^c = (\mathcal{O}^c)^c = \mathcal{O}$, which is open, so by the definition, E is closed. $\square$

Example 1.51. *In $\mathbb{R}$ with the usual metric, we have that (b, ∞) is open and $(-\infty, a)$ is open. So when $a < b$ we have that $\mathcal{O} = (-\infty, a) \cup (b, +\infty)$ is open. Hence, $\mathcal{O}^c = [a, b]$ is closed.*

This shows that our old calculus definition of a "closed interval", really is a closed set in this sense.

Definition 1.52. *Let (X, d) be a metric space, let $x \in X$ and let $r > 0$. The* **closed ball with center x and radius r** *is the set*

$$B^-(x; r) = \{y \in X : d(x, y) \leq r\}.$$

The following result explains this notation.

Proposition 1.53. *Let (X, d) be a metric space, let $x \in X$ and let $r > 0$. Then $B^-(x; r)$ is a closed set.*

Proof. Let $\mathcal{O} = (B^-(x; r))^c$ denote the complement. We must prove that $\mathcal{O}$ is open. Note that $\mathcal{O} = \{y \in X : d(x, y) > r\}$.

For $p \in \mathcal{O}$ set $r_1 = d(x, p) - r > 0$. We claim that $B(p; r_1) \subseteq \mathcal{O}$. If we can prove this claim, then we will have shown that $\mathcal{O}$ is open. So let $q \in B(p; r_1)$, then $d(x, p) \leq d(x, q) + d(q, p) < d(x, q) + r_1$. Subtracting r_1 from both sides of this last inequality yields, $d(x, p) - r_1 < d(x, q)$. But $d(x, p) - r_1 = r$. Hence, $r < d(x, q)$ and so $q \in \mathcal{O}$. Since q was an arbitrary element of $B(p; r_1)$, we have that $B(p; r_1) \subseteq \mathcal{O}$ and our proof is complete. $\square$

Because the definition of closed sets involves complements, it is useful to recall DeMorgan's Laws. Given subsets $E_i \subseteq X$ where i belongs to some set I, we have

$$\bigcup_{i \in I} E_i = \{x \in X : \text{ there exists } i \in I \text{ with } x \in E_i\}$$

and

$$\bigcap_{i \in I} E_i = \{x \in X : x \in E_i \text{ for every } i \in I\}.$$

Proposition 1.54 (DeMorgan)**.** *Let $E_i \subseteq X$ for $i \in I$. Then*

$$\Big(\bigcup_{i \in I} E_i\Big)^c = \bigcap_{i \in I} E_i^c \text{ and } \Big(\bigcap_{i \in I} E_i\Big)^c = \bigcup_{i \in I} E_i^c.$$

The following theorem about closed sets follows from DeMorgan's Laws and Theorem 1.31.

Theorem 1.55. *Let (X, d) be a metric space. Then:*

1. *the empty set is closed,*
2. *X is closed,*
3. *the intersection of any collection of closed sets is closed,*
4. *the union of finitely many closed sets is closed.*

Proof. By Theorem 1.31, the set X is open, so X^c is closed, but X^c is the empty set. Similarly, the empty set is open, so its complement, which is X, is closed.

Given a collection of closed set E_i $i \in I$, we have that each E_i^c is open. Since

$$(\bigcap_{i \in I} E_i)^c = \bigcup_{i \in I} E_i^c,$$

we see that the complement of their intersection is the union of a collection of open sets. By Theorem 1.31.3, this union of open sets is an open set. Hence, the complement of the intersection is open and so the intersection is closed.

Similarly, if we have only a finite collection of closed sets, $E_1, \dots, E_n$ then $(E_1 \cup \cdots \cup E_n)^c = E_1^c \cap \cdots E_n^c$, which is a finite intersection of open sets. By Theorem 1.31.4, this finite intersection of open sets is open. Hence, the complement of the union is open and so the union is closed. □

Proposition 1.56. *In a discrete metric space, every set is closed.*

As with open sets, when there is more than one metric on the set X, then we need to specify which metric we are referring to when saying that a set is closed. The following is the analogue of Theorem 1.41 for closed sets.

Proposition 1.57. *Let X be a set and let d and ρ be two metrics on X that are uniformly equivalent. Then a set is closed with respect to d if and only if it is closed with respect to ρ.*

Problem 1.58. *Prove that $\{(a_1, a_2) \in \mathbb{R}^2 : 0 \leq a_1 \leq 2, 0 \leq a_2 \leq 4\}$ is a closed set in the Euclidean metric.*

1.4 Convergent Sequences

Our general idea from calculus of a sequence $\{p_n\}$ converging to a point p is that as n grows larger, the points p_n grow closer and closer to p. When we say "grow closer" we really have in our minds that some distance is growing smaller. This leads naturally to the following definition.

Definition 1.59. *Let (X, d) be a metric space, $\{p_n\} \subseteq X$ a sequence in X and $p \in X$. We say that the sequence $\{p_n\}$* **converges to p** *and write*

$$\lim_{n \to +\infty} p_n = p,$$

provided that for every $\epsilon > 0$, there is a real number N so that when $n > N$, then $d(p, p_n) < \epsilon$.

Often out of sheer laziness, I will write $\lim_n p_n = p$ for $\lim_{n\to+\infty} p_n = p$.

Example 1.60. *When $X = \mathbb{R}$ and d is the usual metric, then $d(p, p_n) = |p - p_n|$ and this definition is identical to the definition used when we studied convergent sequences of real numbers in Math 3333.*

For a quick review, we will look at a couple of examples of sequences and recall how we would prove that they converge.

For a first example, consder the sequence given by $p_n = \frac{3n+1}{5n-2}$. Fo any natural number n, the denominator of this fraction is non-zero. So this formula defines a sequence of points. For large n, the $+1$ in the numerator and the -2 in the denominator are small in relation to the $3n$ and $5n$ so we expect that this sequence has limit $p = \frac{3}{5}$.

Now for some scrap work. To prove this we would need

$$d(p, p_n) = |p - p_n| = |\frac{3(5n-2) - (3n+1)5}{5(5n-2)}| < \epsilon.$$

Simplifying this fraction leads to the condition $\frac{11}{25n-10} < \epsilon$. Solving for n we see that this is true provided $\frac{11}{\epsilon} < 25n - 10$, i.e., $\frac{11}{25\epsilon} + \frac{2}{5} < n$. So it looks like we should choose, $N = \frac{11}{25\epsilon} + \frac{2}{5}$. This is not a proof, just our scrap work.

Now for the proof:

Given $\epsilon > 0$, define $N = \frac{11}{25\epsilon} + \frac{2}{5}$. For any $n > N$, we have that $25n - 10 > 25N - 10 = \frac{11}{\epsilon}$, and so $d(p, p_n) = |\frac{3}{5} - \frac{3n+1}{5n-2}| = \frac{11}{25n-10} < (\frac{11}{11/\epsilon})^{-1} = \epsilon$. Hence, $\lim_{n\to+\infty} p_n = p$.

For the next example, we look at $p_n = \sqrt{n^2 + 8n} - n$ and prove that this sequence has limit $p = 4$.

Now that we have recalled what this definition means in $\mathbb{R}$ and have seen how to prove a few examples. We want to look at what this concept means in some of our other favorite metric spaces.

Theorem 1.61. *Let $\mathbb{R}^k$ be endowed with the Euclidean metric d, let $\{p_n\} \subseteq \mathbb{R}^k$ be a sequence with $p_n = (a_{1,n}, a_{2,n}, \dots, a_{k,n})$ and let $p = (a_1, a_2, \dots, a_k) \in \mathbb{R}^k$. Then $\lim_{n \to +\infty} p_n = p$ in $(\mathbb{R}^k, d)$ if and only if for each $j, 1 \le j \le k$, we have $\lim_{n \to +\infty} a_{j,n} = a_j$ in $\mathbb{R}$ with the usual metric.*

Proof. First we assume that $\lim_{n \to +\infty} p_n = p$. Given $\epsilon > 0$, there is a N so that for $n > N$, we have that

$$d(p, p_n) = \sqrt{(a_1 - a_{1,n})^2 + \cdots + (a_k - a_{k,n})^2} < \epsilon.$$

Now fix j, we have that $|a_j - a_{j,n}| = \sqrt{(a_j - a_{j,n})^2}$ is smaller than the above square root, since it is just one of the terms. Thus, for the same N, we have that $n > N$ implies that $|a_j - a_{j,n}| \le d(p, p_n) < \epsilon$. Hence, for each ϵ, we have found an N and so $\lim_n a_{j,n} = a_j$. This proves the first implication.

Now conversely, suppose that for each j, we have that $\lim_n a_{j,n} = a_j$. Given an $\epsilon > 0$, we must produce an N so that when $n > N$, we have $d(p, p_n) < \epsilon$.

For each j, taking $\epsilon_1 = \frac{\epsilon}{\sqrt{k}}$ we can find an N_j so that for $n > N_j$ we have that $|a_j - a_{j,n}| < \epsilon_1$.(Why did I call it N_j ?) Now let $N = \max\{N_1, \dots, N_k\}$, then for $n > N$, we have that $(a_j - a_{j,n})^2 < \epsilon_1^2$ for every $j, 1 \le j \le k$.

Hence, $d(p - p_n) = \sqrt{(a_1 - a_{1,n})^2 + \cdots + (a_k - a_{k,n})^2} < \sqrt{\epsilon_1^2 + \cdots + \epsilon_1^2} = \sqrt{k\epsilon_1^2} = \epsilon$. This proves that $\lim_n p_n = p$ in the Euclidean metric. □

So the crux of the above theorem is that a sequence of points in $\mathbb{R}^k$ converges if and only if each of their components converge.

If we combine our first two examples with the above theorem, we see that if we define a sequence of points in $\mathbb{R}^2$ by setting $p_n = (\frac{3n+1}{5n-2}, \sqrt{n^2 + 8n} - n)$ then in the Euclidean metric these points converge to the point $p = (\frac{3}{5}, 4)$.

What if we had used the taxi cab metric or d_∞ metric on $\mathbb{R}^2$ instead of the Euclidean metric, would these points still converge to the same point? The answer is yes and the following result explains why and saves us having to prove separate theorems for each of these metrics.

Proposition 1.62. *Let X be a set with metrics d and ρ that are uniformly equivalent, let $\{p_n\} \subseteq X$ be a sequence and let $p \in X$ be a point. Then $\{p_n\}$ converges to p in the metric d if and only if $\{p_n\}$ converges to p in the metric ρ.*

Proof. Let $\rho(x, y) \leq Ad(x, y)$ and $d(x, y) \leq B\rho(x, y)$ for every $x, y \in X$.

First assume that $\{p_n\}$ converges to p in the metric d. Given $\epsilon > 0$, we can pick a N so that when $n > N$, we have that $d(p, p_n) < \frac{\epsilon}{A}$. Then when $n > N$, we have that $\rho(p, p_n) \leq Ad(p, p_n) < \epsilon$. Hence, $\{p_n\}$ converges to p in the metric ρ.

The proof of the converse statement is similar, using the constant b. □

We now look at the discrete metric.

Proposition 1.63. *Let (X, d) be the discrete metric space, let $\{p_n\} \subseteq X$ be a sequence, and let $p \in X$. Then the sequence $\{p_n\}$ converges to p if and only if there exists N so that for $n > N$, $p_n = p$.*

Proof. First assume that $\{p_n\}$ converges to p. Taking $\epsilon = 1$, there is N so that $n > N$ implies that $d(p, p_n) < \epsilon = 1$. But since this is the discrete metric, $d(x, y) < 1$ implies that $d(x, y) = 0$, i.e., that $x = y$. Hence, $p_n = p$ for every $n > N$.

Conversely, if there is an N_1 so that when $n > N_1$, we have $p_n = p$, then given any $\epsilon > 0$, pick $N = N_1$. Then $n > N$ implies that $d(p, p_n) = d(p, p) = 0 < \epsilon$. □

A sequence $\{p_n\}$ such that $p_n = p$ for every $n > N$, is often called **eventually constant.**

We now look at some general theorems about convergence in metric spaces.

Proposition 1.64. *Let (X, d) be a metric space. A sequence $\{p_n\} \subseteq X$ can have at most one limit.*

Proof. Suppose that $\lim_n p_n = p$ and $\lim_n p_n = q$, we need to prove that this implies that $p = q$. We will do a proof by contradiction. Suppose that $p \neq q$, then $d(p, q) > 0$.

Set $\epsilon = \frac{d(p,q)}{2}$. Since $\lim_n p_n = p$, there is N_1 so that $n > N_1$ implies that $d(p, p_n) < \epsilon$.

Since $\lim_n p_n = q$, there is N_2 so that $n > N_2$ implies that $d(q, p_n) < \epsilon$. So if we let $N = \max\{N_1, N_2\}$, then for $n > N$, both of these inequalities will be true.

Thus, for any $n > N$, we have that $d(p, q) \leq d(p, p_n) + d(p_n, q) < \epsilon + \epsilon = d(p, q)$. Thus, $d(p, q) < d(p, q)$, a contradiction. □

The above result is most often used to prove that two points are really the same point, since if $\lim_n p_n = p$ and $\lim_n p_n = q$, then $p = q$.

Convergence is an important way to characterize closed sets.

Theorem 1.65. *Let (X, d) be a metric space and let $S \subseteq X$ be a subset. Then S is a closed subset if and only if whenever $\{p_n\} \subseteq S$ is a convergent sequence, we have $\lim_n p_n \in S$.*

That is, a set is closed if and only if limits of convergent sequences stay in the set.

Proof. First assume that S is closed, let $\{p_n\} \subseteq S$ be a convergent sequence and let $p = \lim_n p_n$. We must show that $p \in S$. We do a proof by contradiction.

Suppose that $p \notin S$, then $p \in S^c$ which is open. Hence, there is $r > 0$ so that $B(p; r) \subseteq S^c$. Thus, for any $q \in S$, $d(p, q) \geq r$. Now set $\epsilon = r$, then for every n, $d(p, p_n) \geq \epsilon$, and hence we can find no N for this ϵ.

To prove the converse, we consider the logically equivalent contrapositive statement: if S is not closed, then there exists at least one convergent sequence $\{p_n\} \subseteq S$, with $\lim_n p_n \notin S$.

So we want to prove that this statement is true. To this end, look at S^c since S is not closed, S^c is not open. Hence, there exists $p \in S^c$, such that for every $r > 0$, the set $B(p; r)$ is not a subset of S^c. That is for every $r > 0$, $B(p; r) \cap S$ is not empty. Now for each n if we set $r = 1/n$, then we can pick a point $p_n \in B(p; 1/n) \cap S$. Thus, the sequence of points that we get this way satisfies, $\{p_n\} \subseteq S$. Also, since $d(p, p_n) < 1/n$, we have that $\lim_n p_n = p$, by Problem 1.65.

This completes the proof of the theorem. □

One of our other main results from Math 3333 about convergent seqeunces in $\mathbb{R}$ is that every convergent sequence of real numbers is bounded. This plays an important role in metric spaces too. But first we need to say what it means for a set to be bounded in a metric space.

Proposition 1.66. *Let (X, d) be a metric space and let $E \subseteq X$ be a subset. Then the following are equivalent:*

1. *there exists a point $p \in X$ and $r_1 > 0$ such that $E \subseteq B^-(p; r_1)$,*
2. *there exists a point $q \in X$ and $r_2 > 0$ such that $E \subseteq B(q; r_2)$,*
3. *there exists $M > 0$ so that every $x, y \in E$ satisfies $d(x, y) \leq M$.*

Proof. If 1) holds, then let $q = p$ and let r_2 be any number satisfying, $r_2 > r_1$. Then $E \subseteq B^-(p; r_1) \subseteq B(p; r_2)$, and so 2) holds.

If 2) holds, let $M = 2r_2$, then for any $x, y \in E$, we have that $d(x, y) \leq d(x, q) + d(q, y) < r_2 + r_2 = M$, and so 3) holds.

If 3) holds, fix any point $p \in E$ and let $r_1 = M$, then any $y \in E$ satisfies, $d(p, y) \leq M$ and so, $E \subseteq B^-(p, M)$. $\square$

Definition 1.67. *Let (X, d) be a metric space and let $E \subseteq X$. We say that E is a* **bounded set** *provided that it satisfies any of the three equivalent conditions of the above proposition.*

Proposition 1.68. *Let (X,d) be a metric space. If $\{p_n\}$ is a convergent sequence, then it is bounded.*

Proof. Let $\lim_n p_n = p$. For $\epsilon = 1$, there is N so that when $n > N$, then $d(p, p_n) < 1$.

Let $r_1 = \max\{1, d(p, p_n) : n \leq N\}$. Since there are only finitely many numbers in this set the maximum is well-defined and a finite number.

Now when, $n \leq N$, $d(p, p_n) \leq r_1$, since it is one of the numbers used in finding the maximum. While if $n > N$, then $d(p, p_n) < 1 \leq r_1$. Thus, for every $n, d(p, p_n) \leq r_1$ and so $\{p_n\} \subseteq B^-(p; r_1)$. $\square$

Problem 1.69. *Let (X, d) be a metric space, $\{p_n\} \subseteq X$ and $p \in X$. Prove that $\lim_n p_n = p$ if and only if the sequence of real numbers $\{d(p, p_n)\}$ satisfies $\lim_n d(p, p_n) = 0$.*

Problem 1.70. *Let $S \subseteq \mathbb{R}^2$ be the set defined by $S = \{(a_1, a_2) : 0 \leq a_2 \leq a_1\}$. Prove that S is a closed subset of $\mathbb{R}^2$ in the Euclidean metric.*

Problem 1.71. *Give an example of a bounded subset of $\mathbb{R}$ that is not closed, of a closed subset that is not bounded and of a bounded sequence that is not convergent.*

1.4.1 Further results on convergence in $\mathbb{R}^k$

Recall that $\mathbb{R}^k$ is a vector space, for $p = (a_1, \ldots, a_k), q = (b_1, \ldots, b_k) \in \mathbb{R}^k$ and $r \in \mathbb{R}$ we have

$$p + q = (a_1 + b_1, \ldots, a_k + b_k) \text{ and } rp = (ra_1, \ldots, ra_k).$$

There is also the "dot product",

$$p \cdot q = a_1 b_1 + \cdots + a_k b_k.$$

Note that in terms of the dot product and Euclidean metric, we can see that the Schwarz inequality says that

$$|p \cdot q| \leq \sqrt{a_1^2 + \cdots + a_k^2}\sqrt{b_1^2 + \cdots + b_k^2} = d(0, p)d(0, q).$$

Two other useful connection to notice between the Euclidean metric and the vector space operations is that

$$d(p,q) = \sqrt{(a_1-b_1)^2+\cdots+(a_k-b_k)^2} = d(0,p-q)$$

and

$$d(rp,rq) = \sqrt{(ra_1-rb_1)^2+\cdots+(ra_k-rb_k)^2} = |r|d(p,q).$$

Lemma 1.72. *Let $(\mathbb{R}^k, d)$ be Euclidean space. If $E \subseteq \mathbb{R}^k$ is a bounded set, then there is A so that $E \subseteq B^-(0;A)$.*

Proof. Since E is bounded, there is $p \in \mathbb{R}^k$ and r_1, so that $d(p,x) \le r_1$ for every $x \in E$. Let $A = r_1 + d(0,p)$, then for any $x \in E$, $d(0,x) \le d(0,p) + d(p,x) \le A$, and so $E \subseteq B^-(0;A)$. $\square$

The following result can be proved using Theorem 1.57 and results from Math 3333 about convergence of sums and product of real numbers. we give a proof that mimics the proofs for real numbers.

Theorem 1.73. *Let $(\mathbb{R}^k, d)$ be Euclidean space, let $p_n = (a_{1,n}, \ldots, a_{k,n})$ and $q_n = (b_{1,n}, \ldots, b_{k,n})$ be sequences in $\mathbb{R}^k$ with $\lim_n p_n = p = (a_1, \ldots, a_k)$ and $\lim_n q_n = q = (b_1, \ldots, b_k)$, and let $\{r_n\}$ be a sequence in $\mathbb{R}$ with $\lim_n r_n = r$. Then we have the following:*

1. $\lim_n p_n + q_n = p + q$,
2. $\lim_n r_n p_n = rp$,
3. $\lim_n p_n \cdot q_n = p \cdot q$.

Proof. To see 1), given $\epsilon > 0$, choose N_1, so that for $n > N_1$, $d(p,p_n) < \epsilon/2$ and choose N_2 so that for $n > N_2$, $d(q,q_n) < \epsilon/2$.

Let $N = \max\{N_1, N_2\}$, then for $n > N$, we have that

$$d(p+q, p_n+q_n) = \sqrt{(a_1+b_1-a_{1,n}-b_{1,n})^2+\cdots+(a_k+b_k-a_{k,n}-b_{k,n})^2} \le$$
$$\sqrt{(a_1-a_{1,n})^2+\cdots+(a_k-a_{k,n})^2} + \sqrt{(b_1-b_{1,n})^2+\cdots+(b_k-b_{k,n})^2} \le$$
$$d(p,p_n) + d(q,q_n) < \epsilon.$$

To prove 2), since $\{r_n\}$ is a bounded sequence of real numbers, there is $M > 0$, so that $|r_n| \le M$, for every n. Note that also $|r| \le M$. Since $\{p_n\}$ is a bounded set of vectors, by the lemma there is $A > 0$, so that $d(0,p_n) \le A$, for every n.

Now choose N_1 so that for $n > N_1$, $|r - r_n| < \frac{\epsilon}{2A}$, and choose N_2, so that for $n > N_2$, $d(p, p_n) < \frac{\epsilon}{2M}$.

Let $N = \max\{N_1, N_2\}$, then for $n > N$, we have

$$d(rp, r_n p_n) \le d(rp, rp_n) + d(rp_n, r_n p_n) = |r| d(p, p_n) + |r - r_n| d(0, p_n) < |r| \frac{\epsilon}{2M} + \frac{\epsilon}{2A} A \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Finally, to prove 3), let $d(0, p_n) \le A$, as before and let $d(0, q_n) \le B$, so that $d(0, p) \le A$ and $d(0, q) \le B$. We may choose N_1 so that for $n > N_1$ we have that $d(p, p_n) < \frac{\epsilon}{2B}$ and N_2 so that for $n > N_2$, we have $d(q, q_n) < \frac{\epsilon}{2A}$.

Thus, if we let $N = \max\{N_1, N_2\}$, then for $n > N$,

$$\begin{aligned} &|p \cdot q - p_n \cdot q_n| = |(p - p_n) \cdot q + p_n \cdot (q - q_n)| \\ &\le |(p - p_n) \cdot q| + |p_n \cdot (q - q_n)| \le d(0, p - p_n) d(0, q) + d(0, p_n) d(0, q - q_n) \\ &\le d(p, p_n) B + A d(q, q_n) < \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

□

1.4.2 Subsequences

We recall the definition of a subsequence.

Definition 1.74. *Given a set X, a sequence $\{p_n\}$ in X and numbers, $1 \le n_1 < n_2 < \cdots$, the new sequence that we get by setting $q_k = p_{n_k}$ is called a* **subsequence of** $\{p_n\}$.

For example, if $n_k = 2k$, $k = 1, 2, ...$, then the subsequence that we get is just the even numbered terms of the old sequence. If $n_k = 2k-1$, $k = 1, 2, ...$, then we get the subsequence of odd terms.

Often we simply denote the subsequence by $\{p_{n_k}\}$.

Proposition 1.75. *Let (X, d) be a metric space, $\{p_n\}$ a sequence in X and $p \in X$. If the sequence $\{p_n\}$ converges to p, then every subsequence of $\{p_n\}$ also converges to p.*

Proof. Let $1 \le n_1 < n_2 < \cdots$, be the numbers for our subsequence. Note that since these are integers and $n_j < n_{j+1}$, we have that they must be at least one apart. Since, $1 \le n_1$, we have that $2 \le n_2, 3 \le n_3$, and in general, $k \le n_k$.

Now let $q_k = p_{n_k}$ denote our subsequence and let $\epsilon > 0$ be given. We must show that there is a K so that $k > K$, implies that $d(p, q_k) < \epsilon$.

Since $\lim_n p_n = p$, there is N so that $n > N$ implies that $d(p, p_n) < \epsilon$. Let $K = N$, then for $k > K$, $n_k \geq k > N$. Hence, for $k > K$, $d(p, q_k) = d(p, p_{n_k}) < \epsilon$.

This proves that $\lim_k q_k = p$. □

1.5 Interiors, Closures, Boundaries of Sets

In this section we look at some other important concepts related to open and closed sets.

Definition 1.76. *Let (X,d) be a metric space and $S \subseteq X$. A point p is called an* **interior point of S** *provided that there exists $r > 0$ so that $B(p; r) \subseteq S$. The set of all interior points of S is called the* **interior of S** *and is denoted as* **int(S).**

Note that since $p \in B(p; r)$, every interior point of S is a point in S. Thus, $int(S) \subseteq S$.

Example 1.77. *Let $X = \mathbb{R}$ with the usual metric. Then $int([a, b]) = (a, b)$, while $int(\mathbb{Q})$ is the empty set.*

Example 1.78. *Let $X = \mathbb{R}^2$ with the Euclidean metric. Then $int(\{(x_1, x_2) : a \leq x_1 \leq b, c \leq x_2 \leq d\}) = \{(x_1, x_2) : a < x_1 < b, c < x_2 < d\}$ and $int(\{(x_1, 0) : a \leq x_1 \leq b\})$ is the empty set.*

Proposition 1.79. *Let (X,d) be a metric space and let $S \subseteq X$. Then:*

1. *$int(S)$ is an open set,*
2. *if $\mathcal{O} \subseteq S$ is an open set, then $\mathcal{O} \subseteq int(S)$,*
3. *$int(S)$ is the largest open set contained in S.*

Proof. If $int(S)$ is empty, then it is open. Otherwise, let $p \in int(S)$, so that there is $r > 0$ with $B(p : r) \subseteq S$. If we show that $B(p; r) \subseteq int(S)$, then that will prove that $int(S)$ is open. Let $q \in B(p; r)$ and set $r_1 = r - d(p, q)$. Using the triangle inequality, we have seen before that $B(q; r_1) \subseteq B(p; r) \subseteq S$. This shows that $q \in int(S)$. Hence, $B(p; r) \subseteq int(S)$.

To see the second statement, if $p \in \mathcal{O}$, then since $\mathcal{O}$ is open, there is $r > 0$, so that $B(p; r) \subseteq \mathcal{O} \subseteq S$. This shows that $p \in int(S)$ and so $\mathcal{O} \subseteq S$.

The third statement is a summary of the first two statements. □

Definition 1.80. *Let (X,d) be a metric space and let $S \subseteq X$. Then the* **closure of S,** *denoted $\overline{S}$ is the intersection of all closed sets containing S.*

Before giving any examples, it will be easiest to have some other descriptions of $\overline{S}$.

Proposition 1.81. *$\overline{S}$ is closed and if C is any closed set with $S \subseteq C$, then $\overline{S} \subseteq C$.*

Proof. $\overline{S}$ is closed because intersections of closed sets are closed. If C is closed and $S \subseteq C$, then C is one of the sets that is in the intersection defining $\overline{S}$ and so $\overline{S} \subseteq C$. □

Thus, $\overline{S}$ is the smallest closed set containing S.

Theorem 1.82. *Let $p \in X$, then $p \in \overline{S}$ if and only if for every $r > 0$, $B(p;r) \cap S$ is non-empty.*

Proof. Suppose that there was an $r > 0$ so that $B(p;r) \cap S$ was empty. This would imply that $S \subseteq B(p;r)^c$, which is a closed set. Thus, we have a closed set containing S and $p \notin B(p;r)^c$. This implies that $p \notin \overline{S}$. (This is the contrapositive.)

Conversely, if $B(p;r) \cap S$ is non-empty for every $r > 0$, then for each n, we may choose a point $p_n \in B(p;1/n) \cap S$. This gives a sequence $\{p_n\} \subseteq S$, and it is easy to see that $\lim_n p_n = p$. Now if C is any closed set with $S \subseteq C$, then $\{p_n\} \subseteq C$, and so by Theorem 1.61, $p \in C$. Hence, p is in every closed set containing S and so $p \in \overline{S}$. □

Corollary 1.83. *Let $p \in X$, then $p \in \overline{S}$ if and only if there is a sequence $\{p_n\} \subseteq S$ with $\lim_n p_n = p$.*

Example 1.84. *Let $X = \mathbb{R}$ with the usual metric. Then the closure of (a,b) is $[a,b]$ and $\overline{\mathbb{Q}} = \mathbb{R}$.*

Example 1.85. *If $X = \{x \in \mathbb{R} : x > 0\}$ with the usual metric and $S = \{x : 0 < x < 1\}$ then $\overline{S} = (0,1]$.*

Definition 1.86. *Let (X,d) be a metric space and let $S \subseteq X$. Then the* **boundary of S,** *denoted ∂S is the set $\partial S = \overline{S} \cap \overline{S^c}$.*

Proposition 1.87. *Let $p \in X$, then $p \in \partial S$ if and only if for every $r > 0$, $B(p;r) \cap S$ is non-empty and $B(p;r) \cap S^c$ is non-empty.*

Proof. Apply the characterization of the closure of sets to $\overline{S}$ and $\overline{S^c}$. □

Corollary 1.88. *Let $p \in X$. $p \in \partial S$ if and only if there is a sequence $\{p_n\} \subseteq S$ and a sequence $\{q_n\} \subseteq S^c$ with $\lim_n p_n = \lim_n q_n = p$.*

Example 1.89. *Let $X = \mathbb{R}$ with the usual metric and let $S = (0, 1]$. Then $\partial S = \{0, 1\}$ and $\partial\mathbb{Q} = \mathbb{R}$.*

Problem 1.90. *Prove Corollary 1.83.*

Problem 1.91. *Prove Corollary 1.88.*

Problem 1.92. *Let $X = \mathbb{R}^2$ with the Euclidean metric and let $S = \{(x_1, x_2) : x_1^2 + x_2^2 < 1\}$. Prove that $\overline{S} = \{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}$ and that $\partial S = \{(x_1, x_2) : x_1^2 + x_2^2 = 1\}$.*

Definition 1.93. *Let (X,d) be a metric space $S \subseteq X$. A point $p \in X$, is called a* **cluster point** *or* **accumulation point** *of S, provided that for every $r > 0$, $B(p; r) \cap S$ has infinitely many points.*

Problem 1.94. *Prove that p is a cluster point of S if and only if there is a sequence $\{p_n\} \subseteq S$ of distinct points, i.e., $p_i \neq p_j$ for $i \neq j$, such that $\lim_n p_n = p$.*

Problem 1.95. *Let (X,d) be a discrete metric space and let $S \subseteq X$ be any non-empty set. Find int$(S), \overline{S}, \partial S$ and give the cluster points of S.*

1.6 Completeness

One weakness of "convergence" is that when we want to prove that a sequence $\{p_n\}$ converges, then we need the point p that it converges to before we can prove that it converges. But often in math, one doesn't know yet that a problem has a "solution" and we can only produce a sequence $\{p_n\}$ that somehow is a better and better approximate solution and we want to claim that necessarily a point exists that *is* the limit of this sequence. It is for these reasons that mathematicians introduced the concepts of *Cauchy sequences* and *complete metric spaces.*

Definition 1.96. *Let (X,d) be a metric space. A sequence $\{p_n\} \subseteq X$ is called* **Cauchy** *provided that for each $\epsilon > 0$ there exists N so that whenever $m, n > N$, then $d(p_n, p_m) < \epsilon$.*

We look at a few properties of Cauchy sequences.

Proposition 1.97. *Let (X,d) be a metric space and let $\{p_n\} \subseteq X$ be a sequence. If $\{p_n\}$ is a convergent sequence, then $\{p_n\}$ is a Cauchy sequence.*

Proof. Let $p = \lim_n p_n$. Given $\epsilon > 0$ there is N so that $n > N$ implies that $d(p, p_n) < \frac{\epsilon}{2}$. Now let $m, n > N$, then $d(p_n, p_m) \le d(p_n, p) + d(p, p_m) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. □

Proposition 1.98. *Let (X,d) be a metric space, $\{p_n\} \subseteq X$ a sequence and $\{p_{n_k}\}$ a subsequence. If $\{p_n\}$ is Cauchy, then $\{p_{n_k}\}$ is Cauchy, i.e., every subsequence of a Cauchy sequence is also Cauchy.*

Proof. Given $\epsilon > 0$, there is N so that $n, m > N$ then $d(p_n, p_m) < \epsilon$. Let $K = N$ and recall that $n_k \ge k$. Thus, if $k, j > K$ then $n_k \ge k > N$ and $n_j \ge j > N$, and so $d(p_{n_k}, p_{n_j}) < \epsilon$. □

Proposition 1.99. *Let (X,d) be a metric space and let $\{p_n\} \subseteq X$ be a Cauchy sequence. Then $\{p_n\}$ is bounded.*

Proof. Set $\epsilon = 1$ and choose N so that for $n, m > N$, $d(p_n, p_m) < 1$. Fix any $n_1 > N$, we will show that there is a radius r so that $\{p_n\} \subseteq B^-(p_{n_1}; r)$. To this end, set $r = \max\{1, d(p_{n_1}, p_n) : 1 \le n \le N\}$. This is a finite set of numbers so it does have a maximum. Now given any n, either $1 \le n \le N$, in which case $d(P_{n_1}, p_n) \le r$, since it is one of the numbers occurring the max, or $n > N$, and then $d(p_{n_1}, p_n) < 1 \le r$. □

Definition 1.100. *Let (X,d) be a metric space. If for each Cauchy sequence in (X,d), there is a point in X that the sequence converges to, then (X,d) is called a* **complete metric space.**

Example 1.101. *Let $X = (0, 1] \subseteq \mathbb{R}$ be endowed with the usual metric $d(x, y) = |x - y|$. Then X is not complete, since $\{\frac{1}{n}\} \subseteq X$ is a Cauchy sequence with no point in X that it can converge to.*

Example 1.102. *Let $\mathbb{Q}$ denote the rational numbers with metric $d(x, y) = |x - y|$. We can take a sequence of rational numbers converging to $\sqrt{2}$, which we know is irrational. Then that sequence will be Cauchy, but not have a limit in $\mathbb{Q}$. Thus, $(\mathbb{Q}, d)$ is not complete.*

In Math 3333, we proved that $\mathbb{R}$ with the usual metric has the property that every Cauchy sequence converges, that is, $(\mathbb{R}, d)$ is a complete metric space. This fact is so important that we repeat the proof here. First, we need to recall a few important facts and definitions.

Recall that a set $S \subseteq \mathbb{R}$ is called **bounded above** if there is a number $b \in \mathbb{R}$ such that $s \in S$ implies that $s \le b$. Such a number b is called an **upper bound for S.** An upper bound for S that is smaller than every other upper bound of S is called a **least upper bound for S** and denoted **lub(S)** in

Rosenlicht's book. In the book that we used for Math 3333, a least upper bound for S was called a **supremum for S** and denoted **sup(S).**

Similarly, a set $S \subseteq \mathbb{R}$ is called **bounded below** if there is a number $a \in \mathbb{R}$ such that $s \in S$ implies that $a \leq s$. Such a number a is called a **lower bound for S.** A lower bound for S that is larger than every other lower bound is called a **greatest lower bound for S,** and denoted **glb(S)** in Rosenlicht's book. In the book that we used for Math 3333, a greatest lower bound was called an **infimum for S** and denoted **inf(S).**

A key property of $\mathbb{R}$ is that every set that is bounded above has a least upper bound and that every set that is bounded below has a greatest lower bound.

Proposition 1.103. *Let $\{a_n\}$ be a sequence of real numbers such that the set $S = \{a_n : n \geq 1\}$ is bounded above and such that $a_n \leq a_{n+1}$ for every n. Then $\{a_n\}$ converges and $\lim_n a_n = lub(S)$. If $\{b_n\}$ is a sequence of real numbers such that $S = \{b_n : n \geq 1\}$ is bounded below and $b_n \geq b_{n+1}$, then $\{b_n\}$ converges and $\lim_n b_n = glb(S)$.*

Proof. Let $a = lub(S)$, then $a_n \leq a$ for every n. Given $\epsilon > 0$, since $a - \epsilon < a$, we have that $a - \epsilon$ is not an upper bound for S and so there is a N so that $a - \epsilon < a_N$. If $n > N$, then $a - \epsilon < a_N \leq a_n \leq a < a + \epsilon$. Thus, $a - \epsilon < a_n < a + \epsilon$ and so $|a_n - a| < \epsilon$ for every $n > N$. This proves that $\lim_n a_n = a$.

The proof of the other case is similar. □

Theorem 1.104. *Let $(\mathbb{R}, d)$ denote the real numbers with the usual metric. Then $(\mathbb{R}, d)$ is complete, i.e., every Cauchy sequence of real numbers converges.*

Proof. Let $\{p_n\}$ be a Cauchy sequence of real numbers. By the above results $S_1 = \{p_j : j \geq 1\}$ is a bounded set of real numbers, so it is bounded above and bounded below. For each $n \geq 1$, let $S_n = \{p_k : k \geq n\}$. Note that $S_1 \supseteq S_2 \supseteq \ldots$, so that each set S_n is a bounded set of real numbers.

Let $a_n = glb(S_n)$, and $b_n = lub(S_n)$. Since $S_{n+1} \subseteq S_n$ we have that $a_n \leq a_{n+1}$ and $b_{n+1} \leq b_n$. Since $p_n \in S_n$, $a_n \leq p_n \leq b_n$. In particular, $a_n \leq b_1$ for every n, and so $\{a_n\}$ converges and $a = \lim_n a_n = lub(\{a_n : n \geq 1\})$. Similarly, $a_1 \leq b_n$ for every n and so $\{b_n\}$ converges and $b = \lim_n b_n = glb(\{b_n : n \geq 1\})$. Also, since $a_n \leq b_n$ we have that $a \leq b$.

Now given any $\epsilon > 0$, there is N so that when $j, n > N$, $|p_n - p_j| < \epsilon$. Hence, for any $j \geq n$, we have that $p_n - \epsilon < p_j < p_n + \epsilon$. This shows that $p_n - \epsilon$ is a lower bound for S_n and $p_n + \epsilon$ is an upper bound for S_n. Thus, $p_n - \epsilon \leq a_n \leq a \leq b \leq b_n \leq p_n + \epsilon$.

This shows that a and b are trapped in an interval of length 2ϵ. Since ϵ was arbitrary, by the "vanishing ϵ principle", $|a-b|=0$, so $a=b$. Thus, $\lim_n a_n = \lim_n b_n$, but $a_n \le p_n \le b_n$, so by the "Squeeze Theorem" for limits, $\lim_n a_n = \lim_n p_n = \lim_n b_n$. □

Another set of important examples of compelte metric spaces are the Euclidean spaces.

Theorem 1.105. *Let $(\mathbb{R}^k, d)$ denote k-dimensional Euclidean space. Then $(\mathbb{R}^k, d)$ is complete.*

Proof. To simplify notation, we will only do the case $k=2$. So let $p_n = (a_n, b_n)$ be a Cauchy sequence of points in $\mathbb{R}^2$ with the Euclidean metric d. Given $\epsilon > 0$, there is N so that $n, m > N$ implies that $d(p_n, p_m) < \epsilon$. But $|a_n - a_m| \le \sqrt{|a_n - a_m|^2 + |b_n - b_m|^2} = d(p_n, p_m)$, so for $n, m > N$, $|a_n - a_m| < \epsilon$, and so $\{a_n\}$ is a Cauchy sequence of real numbers. Hence $\{a_n\}$ converges, let $a = \lim_n a_n$.

Similarly, $|b_n - b_m| < \epsilon$ for $n, m > N$ and so $\{b_n\}$ converges and let $b = \lim_n b_n$.

Since $\lim_n a_n = a$ and $\lim_n b_n = b$, by Theorem 1.57, $\lim_n p_n = (a, b)$. Hence, this arbitrary Cauchy sequence has a limit and the proof is done. □

Problem 1.106. *Let X be a set and let d and ρ be two uniformly equivalent metrics on X. Prove that (X, d) is a complete metric space if and only if (X, ρ) is complete.*

By the above theorem and problem, $(\mathbb{R}^k, d_1)$ and $(\mathbb{R}^k, d_\infty)$ are also complete metric spaces.

Example 1.107. *Let $X = (0, 1]$. Recall that X is not complete in the usual metric $d(x, y) = |x - y|$. Given $x, y \in X$ we set $\gamma(x, y) = |\frac{1}{x} - \frac{1}{y}|$. It is easy to check that γ is a metric on X. We claim that (X, γ) is complete! Thus, d and γ are not uniformly equivalent.*

We sketch the proof. To prove this we must show that if $\{x_n\} \subseteq X$ is Cauchy in the γ metric, then it converges to a point in X. Given $\epsilon > 0$, suppose that for $n, m > N$, $\gamma(x_n, x_m) < \epsilon$. Since $\gamma(x_n, x_m) = |\frac{1}{x_n} - \frac{1}{x_m}| = |\frac{x_n - x_m}{x_n x_m}|$ we have that $|x_n - x_m| < \epsilon |x_n x_m| \le \epsilon$. Thus, $\{x_n\}$ is Cauchy in the usual metric. Let $x = \lim_n x_n$. Since, $0 < x_n \le 1$ we have that $0 \le x \le 1$. We claim that $x \ne 0$. Because if $x = 0$, then for any N, when $n, m > N$, if we fix m and we let $n \to +\infty$, $x_n \to 0$, then $\frac{1}{x_n} \to +\infty$. Thus, $\gamma(x_n, x_m) = |\frac{1}{x_n} - \frac{1}{x_m}| \to +\infty$. This prevents us making $\gamma(x_n, x_m) < \epsilon$ and so violates the Cauchy condition.

Thus, $x \neq 0$. But also $x_n \neq 0$ for every n. By one of our basic results from 3333, when $x \neq 0, x_n \neq 0$ and $\lim_n x_n = x$, then $\lim_n \frac{1}{x_n} = \frac{1}{x}$. But this last limit being true means that for any $\epsilon > 0$, we can pick N so that when $n > N$, $|\frac{1}{x} - \frac{1}{x_n}| < \epsilon$. But this implies that for $n > N$, $\gamma(x, x_n) < \epsilon$ and so $\{x_n\}$ converges to x in the γ metric! We are done.

Now that we have a few examples of complete metric spaces, the following result gives us many more examples.

Proposition 1.108. *Let (X,d) be a complete metric space. If $Y \subseteq X$ is a closed subset, then (Y,d) is a complete metric space.*

Proof. Let $\{y_n\} \subseteq Y$ be a Cauchy sequence. Since X is complete, there is a point $x \in X$, so that $\lim_n y_n = x$. But since Y is closed, $x \in Y$. Thus, each Cauchy sequence in Y has a limit in Y. □

1.7 Compact Sets

Definition 1.109. *Let (X,d) be a metric space, $S \subseteq X$. A collection $\{U_\alpha\}_{\alpha \in A}$ of subsets of X is called a* **cover of S** *provided that $S \subseteq \cup_{\alpha \in A} U_\alpha$ and an* **open cover of S** *provided that it is a cover of S and every set U_α is open. A subset $S \subseteq X$ is called* **compact** *provided that whenever $\{U_\alpha\}_{\alpha \in A}$ is an open cover of S, then there is a finite subset $F \subseteq A$ such that $S \subseteq \cup_{\alpha \in F} U_\alpha$. The collection $\{U_\alpha\}_{\alpha \in F}$ is called a* **finite subcover.**

Example 1.110. *Let $\mathbb{R}$ have the usual metric. Let $U_n = B(0; n) = (-n, +n)$, $n \in \mathbb{N}$. Then these sets are open and $\mathbb{R} = \cup_{n \in \mathbb{N}} U_n$. Suppose that there was a finite subset $F = \{n_1, ..., n_L\} \subseteq \mathbb{N}$ so that $\mathbb{R} \subseteq \cup_{n \in F} U_n = U_{n_1} \cup \cdots \cup U_{n_L}$. If we let $N = \max\{n_1, ..., n_L\}$, then since $n < m$ implie that $U_n \subseteq U_m$, we would have that $\mathbb{R} \subseteq \cup_{n \in F} U_n = U_N$. But this implies that every real number is in $B(0; N)$, a contradiction. Hence, no finite subcover of $\{U_n\}_{n \in \mathbb{N}}$ covers $\mathbb{R}$ and so $\mathbb{R}$* **is not compact.**

Proposition 1.111. *Let (X, d) be a discrete metric space and let $K \subseteq X$. Then K is compact if and only if K is a finite set.*

Proof. Assume that K is compact. For each $x \in K$, let $U_x = \{x\}$. Because X is discrete each U_x is an open set. Also $K = \cup_{x \in K} U_x$, so this is an open cover of K. But if you leave out any of the sets U_x then you no longer cover K. So this cover must be a finite cover, in which case K has only finitely many points.

Conversely, assume that $K = \{x_1, ..., x_n\}$ is a finite set and let $\{U_\alpha\}_{\alpha \in A}$ be any open cover of K. There must be some α_J so that $x_j \in U_{\alpha_j}$. But then $\{U_{\alpha_1}, ..., U_{\alpha_n}\}$ is a finite subcover for K. □

Before we can give many more examples of compact sets, we need some theorems.

Proposition 1.112. *Let (X,d) be a metric space. If $K \subseteq X$ is compact, then K is closed.*

Proof. We will prove that if K is not closed, then K is not compact. Since K is not closed $\overline{K} \neq K$. Hence, there is $p \in \overline{K}$, $p \notin K$. Since $p \in \overline{K}$, then for every $r > 0$, $B(p;r) \cap K$ is non-empty.

Let $U_n = B^-(p;1/n)^c = \{q \in X : d(p,q) > 1/n\}$. Then each of these sets is open and $\cup_{n \in \mathbb{N}} U_n = \{q \in X : d(p,q) > 0\} = \{p\}^c$. Since $p \notin K$, we have that thi is an open cover of K. Not that $n < m$ implies that $U_n \subseteq U_m$, so as in the proof of Example 1.106, $U_{n_1} \cup \cdots \cup U_{n_L} = U_N$ where $N = \max\{n_1, ..., n_L\}$. So if K was contained in a finite subcover, then there would be N with $K \subseteq U_N$. But this would imply that for $q \in K$, $d(p,q) > 1/N$, and so $K \cap B(p;1/N)$ would be empty, contradiction. □

Proposition 1.113. *Let (X,d) be a metric space, $K \subseteq X$ a compact set and let $S \subseteq K$ be a closed subset. Then S is compact.*

Proof. Let $\{U_\alpha\}_{\alpha \in A}$ be an open cover of S and let $V = S^c$ so that V is open. Check that $\{V, U_\alpha : \alpha \in A\}$ is an open cover of K. So finitely many of these sets cover K and the same finite collection covers S since S is a subset of K. □

The next few results give a better understanding of the structure of compact sets.

Proposition 1.114. *Let (X,d) be a metric space, $K \subseteq X$ a compact set and $S \subseteq K$ an infinite subset. Then S has a cluster point in K.*

Proof. We prove the contrapositive. Suppose $S \subseteq K$ is a set with no cluster point in K. Then for each point $p \in K$, since it is not a cluster point of S there would exist an $r > 0$ such that $B(p;r)$ contains only finitely many points from S. The set of all such balls is an open cover of K. Now because K is compact, some finite collection of these balls cover K. But each of these balls has only finitely many points from S so their union contains only finitely many points from S. Since their union contains K, it contains all of S and so S has only finitely many points. □

Definition 1.115. *Let (X,d) be a metric space, $K \subseteq X$. Then K is called* **sequentially compact** *if every sequence in K has a subsequence that converges to a point in K.*

In some texts a set is said to have the **Bolzano-Weierstrass property** if and only if it is sequentially compact.

Proposition 1.116. *Let (X,d) be a metric space, $K \subseteq X$. If K is compact, then K is sequentially compact.*

Proof. Let $\{p_n\} \subseteq K$ be a sequence. We do two cases.

Case 1. The sequence has only finitely many points. In this case there must be some p so that $p_n = p$ for infinitely many n. Let n_1 be the first such integer, n_2 the second, etc., this defines a subsequence $\{p_{n_k}\}$ with $p_{n_k} = p$ for all k. Clearly this subsequence converges to p.

Case 2. The sequence has infinitely many values. In this case if we let S be the set of points in the sequence, then S is an infinite set and so has a cluster point p in K. We will construct a subsequence that converges to p. First, let n_1 be the smallest integer so that $p_{n_1} \in B(p;1)$ and $p_{n_1} \neq p$. Now let n_2 be the smallest integer strictly larger than n_1 so that $p_{n_2} \in B(p;1/2)$ and $p_{n_2} \neq p$. Continuing in this way we obtain a subsequence $\{p_{n_k}\}$ such that $d(p, p_{n_k}) < 1/k$, and hence this subsequence converges to p. ☐

Definition 1.117. *Let (X,d) be a metric space, $K \subseteq X$ and let $\epsilon > 0$. A subset $E \subseteq K$ is called an* **ϵ-net for** *K provided that given any $p \in K$ there is $q \in E$, such that $d(p,q) < \epsilon$. The subset K is called* **totally bounded** *if for each $\epsilon > 0$ there is an ϵ-net for K consisting of finitely many points.*

Example 1.118. *Let $K = [0,1]$ for each $\epsilon > 0$, let N be the largest integer so that $N\epsilon \leq 1$. Then $\{0, \epsilon, 2\epsilon, ..., N\epsilon\}$ is an ϵ-net for K.*

Proposition 1.119. *Let (X,d) be a metric space and $K \subseteq X$. If K is sequentially compact, then K is totally bounded and complete.*

Proof. First we show that K is complete. To see this let $\{p_n\} \subseteq K$ be a Cauchy sequence. Since K is sequentially compact, there is a $p \in K$ and a subsequence so that $\lim_k p_{n_k} = p$. We now show that $\lim_n p_n = p$. Given $\epsilon > 0$, there is N so that $n, m > N$ implies that $d(p_n, p_m) < \epsilon/2$. But we can pick a large k so that $n_k > N$ and $d(p, p_{n_k}) < \epsilon/2$. Hence for $n > N$, $d(p, p_n) \leq d(p, p_{n_k}) + d(p_{n_k}, p_n) < \epsilon$.

Now to show that K sequentially compact implies K totally bounded, we prove the contrapositive. So assume that K is not totally bounded. Then there exists some $\epsilon > 0$ for which we can find no finite ϵ-net. Pick any $p_1 \in K$, since this point is not an ϵ-net, there must be $p_2 \in K$ such that $d(p_2, p_1) > \epsilon$. Since $\{p_1, p_2\}$ can not be an ϵ-net, there is $p_3 \in K$ such that $d(p_3, p_2) > \epsilon$ and $d(p_3, p_1) > \epsilon$. Continuing in this way, we obtain a sequence $\{p_n\}$ in K

with $d(p_n, p_m) > \epsilon$ for all $n \neq m$. But for a sequence to converge to a point, it must be Cauchy. Since no subsequence of this sequence could be Cauchy, no subsequence converges. Hence, K is not sequentially compact. $\square$

Now we come to the main theorem.

Theorem 1.120. *Let (X,d) be a metric space and $K \subseteq X$. Then the following are equivalent:*

1. *K is compact,*
2. *K is sequentially compact,*
3. *K is totally bounded and complete.*

Proof. We have already shown that 1) implies 2) and that 2) implies 3). So it will be enough to show that 3) implies 1). To this end assume that 3) holds but that K is not compact. Then there is some open cover $\{U_\alpha\}$ of K with no finite subcover.

For $\epsilon = 1/2$ take a finite 1/2-net, $\{p_1, ..., p_L\}$, So that $K \subseteq \cup_{j=1}^{L} B(p_j; 1/2)$. Hence, $K = (K \cap B^-(p_1; 1/2)) \cup \cdots \cup (K \cap B^-(p_L; 1/2)$. This writes K as the union of L closed sets. Now if each of these sets could be covered by a finite collection of sets in $\{U_\alpha\}$ then K would have a finite subcover. So one of these sets cannot have a finite subcover. Pick one and call it K_1. Then $K_1 \subseteq K$ is closed and since it is the intersection with a closed ball of radius 1/2 no two points in K_1 can be distance greater than 1 apart.

Now take a finite 1/4-net, use it to write K_1 as a finite union of sets of radius 1/4, and arguing as before pick one which cannot be covered by a finite collection of the U's. This defines a closed subset $K_2 \subseteq K_1$ with no two points greater than 1/2 apart. Proceed inductively to define a sequence of closed subsets, $K \supseteq K_1 \subseteq K_2 \supseteq \ldots$, so that no two points in K_n are distance greater than $1/2^{n-1}$ apart and each K_n cannot be covered by a finite collection of the U's.

If we pick a point $p_n \in K_n$, then the sequence is Cauchy since for $n, m > N$, $d(p_n, p_m) \leq 1/2^{N-1}$. Since K is complete, there is $p \in K$ with $\lim_n p_n = p$. Also, since $p_n \in K_m$ for all $n \geq m$, this implies that the limit is in K_m, i.e., $p \in K_m$ for all m.

Now since $\{U_\alpha\}$ is an open cover of K, there is some α_0, with $p \in U_{\alpha_0}$ and since this set is open, there is an $r > 0$, so that $B(p : r) \subseteq U_{\alpha_0}$. But since no two points in K_n are more than distance $1/2^{n-1}$ apart, if we pick n, so that $1/2^{n-1} < r$, then $K_n \subseteq B(p; r) \subseteq U_{\alpha_0}$. This contradicts the fact that K_n is not contained in a finite union of the U's. This contradiction completes the proof of the theorem. $\square$

Theorem 1.121 (Heine-Borel). *Let $(\mathbb{R}^k, d)$ denote k-dimensional Euclidean spaceand let $K \subseteq \mathbb{R}^k$. Then K is compact if and only if K is closed and bounded.*

Proof. We have already shown that compact sets are closed. Every compact set is bounded by Problem 1.121.

If K is closed and bounded, then there is some $M > 0$ so that $K \subseteq [-M, +M] \times [-M, +M] \times \cdots \times [-M, +M]$. If we can show that this k-dimensional cube is totally bounded, then it will be compact and since K is a closed subset of this compact set, it will follow that K is compact.

To see that the cube is totally bounded, divide each interval into $2N$ equal length subintervals of length M/N. This divides the large cube up into a finite number of smaller cubes. Each of these smaller cubes is contained in a Euclidean ball of radius $\sqrt{k}\frac{M}{2N}$. Given any $\epsilon > 0$, pick N large enough that $\sqrt{k}\frac{M}{2N} < \epsilon$, then these balls define a finite ϵ-net. Thus, the big cube is totally bounded. □

Theorem 1.122 (Bolzano-Weierstrass). *Let $(\mathbb{R}^k, d)$ be k-dimensional space. If $\{p_n\} \subseteq \mathbb{R}^k$ is a bounded sequence, then it has a convergent subsequence.*

Proof. Since the sequence is bounded it is contained in a closed ball of some finite radius. This set is compact by Heine-Borel, hence sequentially compact. □

Historically, the Bolzano-Weierstrass theorem was proved before Heine-Borel. The original proof used a "divide and conquer" strategy.

Definition 1.123. *Let (X,d) be a metric space and let $Y \subseteq X$. A subset $S \subseteq Y$ is called* **dense** *provided that $\overline{S} \cap Y = Y$. The set Y is called* **separable** *if there is a sequence $S = \{p_n\} \subseteq Y$ that is dense.*

Proposition 1.124. *Let (X,d) be a metric space. If $K \subseteq X$ is compact, then K is separable.*

Proof. For each $n \in \mathbb{N}$, let S_n be a finite $1/n$-net for K. Now we write the elements in $S = \cup_{n\in\mathbb{N}} S_n$ as a sequence by numbering them beginning with S_1, then S_2, etc.

Now we claim that the sequence $S = \{p_n\}$ is dense. To see this, given any $p \in K$ and $r > 0$, if we take $1/n < r$ then since S_n is a $1/n$-net for K, we have that $B(p;r) \cap S_n \neq \emptyset$. Hence, $B(p;r) \cap S \neq \emptyset$. This shows that $p \in \overline{S}$. Thus, $K \subseteq \overline{S} \cap K \subseteq K$ and so S is dense. □

Problem 1.125. *Let (X, d) be a metric space and $K \subseteq X$ a compact subset. Prove that K is bounded.*

Problem 1.126. *Let (X, d) be a metric space. Let $K_i \subseteq X, i = 1, ..., n$ be a finite collection of compact subsets. Prove that $K_1 \cup \cdots \cup K_n$ is compact.*

We now look at a few applications of compact sets. First we need an inequality sometimes called the *reverse triangle inequality.*

Proposition 1.127. *Let (X, d) be a metric space and let $x, y, z \in X$. Then $|d(x, y) - d(x, z)| \leq d(y, z)$.*

Proof. We have that $d(x, y) - d(x, z) \leq (d(x, z) + d(z, y)) - d(x, z) = d(y, z)$. On the other hand, $-(d(x, y) - d(x, z)) = d(x, z) - d(x, y) \leq (d(x, y) + d(y, z)) - d(x, y) = d(y, z)$. Since this number and its negative are both less than $d(y, z)$ we have that $|d(x, y) - d(x, z)| \leq d(y, z)$. $\square$

Proposition 1.128. *Let (X, d) be a metric space, let $\{q_n\} \subseteq X$ be a convergent sequence with limit q and let $p \in X$. Then $d(p, q) = \lim_n d(p, q_n)$.*

Proof. Given $\epsilon > 0$, there is a N so that $n > N$, implies that $d(q, q_n) < \epsilon$. Hence, by the reverse triangle inequality, for $n > N$, we have $|d(p, q) - d(p, q_n)| \leq d(q, q_n) < \epsilon$. This shows that $d(p, q) = \lim_n d(p, q_n)$. $\square$

Definition 1.129. *Let (X, d) be a metric space, let $S \subseteq X$ and let $p \in X$. Then the* **distance from** p **to S** *is the number*

$$dist(p, S) = \inf\{d(p, q) : q \in S\}.$$

In general, the distance from a point to a set is really an infinum and not a minimum. That is there is not generally a point in $q_0 \in S$ so that $dist(p, S) = d(p, q_0)$. For example, in $\mathbb{R}$ with the usual metric, if we take $S = [0, 1)$ and $p = 3$, then $dist(p, S) = 2$, but there is no point $q_0 \in S$, with $d(p, q_0) = |p - q_0| = 2$.

However, for compact sets the distance is always attained as the following result shows.

Proposition 1.130. *Let (X, d) be a metric space, let $K \subseteq X$ and let $p \in X$. If K is compact, then there is a point $q_0 \in K$, such that $dist(p, K) = d(p, q_0)$.*

Proof. Let $\{q_n\}$ be a sequence in K so that $dist(p, K) = \lim_n d(p, q_n)$. Since K is compact, this sequence has a subsequence that converge to a point in K. Let $\{q_{n_k}\}$ be a convergent subsequence with limit $q_0 \in K$. Then by the above proposition, we have that $d(p, q_0) = \lim_k d(p, q_{n_k})$. Since $\{d(p, q_n)\}$ is a convergent sequence of reals, the subsequence has the same limit. Hence, $d(p, q_0) = \lim_n d(p, q_n) = dist(p, K)$. $\square$

Proposition 1.131. *Let* $(\mathbb{R}^k, d)$ *be* k*-dimesnional Euclidean space, let* $C \subseteq \mathbb{R}^k$*, and let* $p \in \mathbb{R}^k$*. If* C *is closed, then there is a point* $q_0 \in C$ *with* $dist(p, C) = d(p, q_0)$*.*

Proof. Pick any point $q_1 \in C$ and let $r = d(p, q_1)$. Then $dist(p, C) \leq d(p, q_1)$ and so, $dist(p, C) = \inf\{d(p, q) : q \in C, d(p, q) \leq r\}$. This shows that if we let $K = C \cap B^-(p; r)$, then $dist(p, C) = dist(p, K)$. Now since K is closed and bounded, it is a compact set by Heine-Borel. Thus, by the last result, there is $q_0 \in K($ and hence, $q_0 \in C)$ with $dist(p, C) = dist(p, K) = d(p, q_0)$. $\square$

Chapter 2

Finite and Infinite Sets, Countability

Later we will need a little familiarity with infinite sets and the fact that they can have "different sizes". So we discuss these ideas now.

One formal way to think of the statement that a set S has n elements, is that there is a function $f : \{1, 2, ..., n\} \to S$ that is one-to-one and onto. Usually, we write $f(j) = s_j$ and write $S = \{s_1, s_2, ..., s_n\}$ to show that it has n elements.

A set S is **finite** when there is some natural number n and a one-to-one onto function $f : \{1, 2, ..., n\} \to S$.

A set S is called **countably infinite** when there is a one-to-one onto function $f : \mathbb{N} \to S$. This can be thought of as saying that a set is countable infinite precisely when the elements of S can be listed in sequence of $S = \{s_1, s_2, ...\}$ with no repetitions, i.e., $i \neq j$ implies that $s_i \neq s_j$.

A set is **countable** when it is either finite or countably infinite. It is not hard to see that a set S is countable if and only if there is a function $f : \mathbb{N} \to S$ that is onto.

There are many sets that are countable, some are a little surprising.

For example, $\mathbb{Z}$ is countable, since I could "list" the elements as,

$$\{0, 1, -1, 2, -2, ...\}.$$

Formally, $f : \mathbb{N} \to \mathbb{Z}$ is the function, $f(2k - 1) = -k$, for $k = 1, 2, ...$ and $f(2k) = k$, for $k = 1, 2, ...$

Also, even though $\mathbb{N} \times \mathbb{N}$ seems to be much "larger" than $\mathbb{N}$ it is also countable. To see this first I'll write $\mathbb{N} \times \mathbb{N}$ as a union of sets, each of which is finte. For each natrual number $n \geq 2$, let $S_n = \{(i, j) \in \mathbb{N} \times \mathbb{N} : i + j = n\}$.

Thus, $S_2 = \{(1,1)\}$, $S_3 = \{(1,2),(2,1)\}$, $S_4 = \{(1,3),(2,2),(3,1)\}$. It is not hard to see that S_n has $n-1$ elements. Now to list these points, we will first list the points in S_2 then S_3 etc. Within each set, we will list the elements in order of their first coordinate, so for example, $(1,2)$ comes before $(2,1)$. This time the formula for the one-to-one onto function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$, is difficult to write down. But equivalently, we could define its inverse, i.e., a one-to-one onto function $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. Now if $i+j = n$ then all the elements of the sets $S_1, ..., S_{n-1}$ come before. These sets have $1 + 2 + ... + (n-2) = \frac{(n-2)(n-1)}{2}$ elements, then the order after that is determined by whether $i = 1, 2, ...$ Thus, the function g is defined by

$$g((i,j)) = \frac{(i+j-2)(i+j-1)}{2} + i$$

and we have shown that $\mathbb{N} \times \mathbb{N}$ is countable.

Now since $\mathbb{N} \times \mathbb{N}$ is countable $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ is also countable! To see how to do this, first define $h : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ by $h((i,j,k)) = (g((i,j)), k)$. Since g is one-to-one and onto $\mathbb{N}$, it is easy to see that h is one-to-one and onto $\mathbb{N} \times \mathbb{N}$. Now if we compose with g then we will have $g \circ h : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is one-to-one and onto!

In a similar way one sees that doing the Cartesian product of $\mathbb{N}$ with itself any finite number of times yields a countable set.

These facts make it easy to see that some other sets are countable. For example to see that, $\mathbb{Q} = \{p/q : p, q \in \mathbb{Z}, q \neq 0\}$ is countable. First, the map from $\mathbb{Z} \times (\mathbb{Z} \backslash \{0\}) \to \mathbb{Q}$ defined by $(p,q) \to p/q$ is onto(but not one-to-one). We know that $\mathbb{Z}$ is countable and $\mathbb{Z} \backslash \{0\}$ is countable. So there is a one-to-one onto map from $\mathbb{N} \times \mathbb{N} \to \mathbb{Z} \times (\mathbb{Z} \backslash \{0\})$, and hence, a one-to-one onto map from $\mathbb{N}$ to $\mathbb{Z} \times (\mathbb{Z} \backslash \{0\})$. Composing with the above map gives an onto map from $\mathbb{N}$ onto $\mathbb{Q}$. Hence, $\mathbb{Q}$ is countable.

With so many countable sets it is a little surprising that there are sets so big that their elements cannot be "listed", i.e., that are not countable. Such sets are called **uncountable.**

One example of an uncountable set is the set $\mathbb{R}$. To prove that $\mathbb{R}$ is uncountable it will be enough to prove that the real numbers in the interval $(0,1)$ is uncountable.

Suppose that the real numbers in $(0,1)$ was a countable set, then we could list them all, let r_n denote the n-th real number, so that $\{r_n : n \in \mathbb{N}\} = (0,1)$. Now each real number has a decimal expansion. Recall that $0.4000... = 0.399999...$. We shall only allow the first type of decimal expansion. When we disallow infinite tails of nines, then the decimal expansion is unique. Now let

$$r_1 = 0.a_{1,1}a_{1,2}a_{1,3}.....$$

$$r_2 = 0.a_{2,1}a_{2,2}a_{2,3}.....$$

$$etc.$$

where each of the $a_{i,j}$ is an integer between 0 and 9. Note that $a_{i,j}$ is the j-th decimal entry of the i-th number.

Now define a sequence of integers b_n, as follows: if $0 \leq a_{n,n} \leq 4$ let $b_n = a_{n,n} + 1$, and if $5 \leq a_{n,n} \leq 9$ let $b_n = a_{n,n} - 1$.

Note that $1 \leq b_n \leq 8$. Hence, there is a real number $r, 0 < r < 1$, where r has decimal expansion,

$$r = 0.b_1b_2....$$

Note that the way that we have defined the b_n's, there are no nines. Since $b_1 \neq a_{1,1}$, $r \neq r_1$ and since $b_2 \neq a_{2,2}$, $r \neq r_2$, etc. Thus, r was not included in the list! Contradiction.

This method of proof–of listing things, somehow getting a doubly indexed array and then altering the (n, n)-entries is often referred to as **Cantor's diagonalization method.**

Chapter 3

Continuous Functions

Continuity plays an important role for functions on the real line. Intuitively, a function is continuous provided that it sends "close" points to "close" points. When we say "close" we naturally have a notion of distance between points, both in the domain and the range. Thus, we see that continuity is really a property for functions between metric spaces. In this chapter, we define what we mean by continuous functions between metric spaces, then study the properties of continuous functions. We will see that our notions of open, closed and compact sets all play an important role.

Definition 3.1. *Let (X, d) and (Y, ρ) be metric spaces and let $p_0 \in X$. We say that a function $f : X \to Y$ is* **continuous at** p_0 *provided that for every $\epsilon > 0$ there is $\delta > 0$ such that whenever $p \in X$ and $d(p_0, p) < \delta$ then $\rho(f(p_0), f(p)) < \epsilon$.*

Note that in this definition the value of δ really depends on the ϵ for this reason some authors write $\delta(\epsilon)$.

Two quick examples.

Example 3.2. *Let $X = Y = \mathbb{R}$ both endowed with the usual metric and let $f : \mathbb{R} \to \mathbb{R}$ be defined by*

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0. \end{cases}$$

Then f is not continuous at $p_0 = 0$.

To see that f is not continuous at 0, take $\epsilon = 1/2$. For any $\delta > 0$, the point $p = \delta/2$ satisfies $d(p_0, p) = |p_0 - p| = |0 - \delta/2| = \delta/2$, but

$\rho(f(p_0), f(p)) = |f(p_0) - f(p)| = 0 - 1| > \epsilon$. Thus, every possible value of δ fails to meet the critieria.

Note that the domain of the function is *really* important when trying to decide continuity. For this same formula, if we made the domain instead just $X = [-1, 0]$, then f would be continuous at 0, since now the function would be the function $f(x) = 0$, for every $x \in X$.

Example 3.3. *Let $X = Y = \mathbb{R}$ with the usual metric, let $f(x) = x^2$ and let $p_0 = 3$. Then f is continuous at p_0.*

To see this given $\epsilon > 0$, we take $\delta = \min\{1, \epsilon/7\}$. Then when $d(p_0, p) = |3-p| < \delta$ we know that $|3-p| < 1$ and so $2 < p < 4$. Hence, $d(f(p_0), f(p)) = |3^2 - p^2| = |3 - p||3 + p| < (3 + 4)|3 - p| < 7\delta \leq \epsilon$.

If we wanted to prove that this function was continuous at $p_0 = 5$, then we could take $\delta = \min\{1, \epsilon/11\}$, and the same argument would work.

Definition 3.4. *Let (X, d) and (Y, ρ) be metric spaces. We say that a function $f : X \to Y$ is* **continuous** *provided that f is continuous at every point in X.*

Thus, f is continuous provided that for each $p_0 \in X$ and $\epsilon > 0$ there is $\delta > 0$, such that $p \in X$ and $d(p_0, p) < \delta$ implies that $\rho(f(p_0), f(p)) < \epsilon$. In this case δ depends on both ϵ and the point p_0. For this reason some author's write $\delta(p_0, \epsilon)$.

Example 3.5. *Let $X = Y = \mathbb{R}$ with the usual metric. Then $f(p) = p^2$ is continuous.*

Given $p_0 \in \mathbb{R}$ and $\epsilon > 0$ let $\delta = \min\{1, \frac{\epsilon}{2|p_0|+1}$. Then $d(p_0, p) < \delta$ implies that $|p| < |p_0| + 1$ and hence $\rho(f(p_0), f(p)) = |p_0^2 - p^2| = |p_0 + p||p_0 - p| \leq (|p_0| + |p|)\delta < (2|p_0| + 1)\delta \leq \epsilon$.

In this example, the δ that we picked depended on both p_0 and ϵ. When it can be chosen independent of p_0, the function is called *uniformly continuous.* That is:

Definition 3.6. *Let (X, d) and (Y, ρ) be metric spaces. We say that a function $f : X \to Y$ is* **uniformly continuous** *provided that for every $\epsilon > 0$ there is $\delta > 0$ such that whenever $p, q \in X$ and $d(p, q) < \delta$ then $\rho(f(p), f(q)) < \epsilon$.*

The following is immediate:

Proposition 3.7. *Let (X, d) and (Y, ρ) be metric spaces and let $f : X \to Y$. If f is uniformly continuous, then f is continuous.*

Example 3.8. *Let $X = Y = \mathbb{R}$ with the usual metric and let $f(p) = 5p+7$, then f is uniformly continuous.*

Given $\epsilon > 0$, let $\delta = \epsilon/5$, then when $d(p,q) = |p-q| < \delta$ we have that $\rho(f(p), f(q)) = |f(p) - f(q)| = 5|p-q| < 5\delta = \epsilon$.

Example 3.9. *Let $X = Y = \mathbb{R}$ with the usual metric and let $f(p) = p^2$. Then f is not uniformly continuous.*

To prove this it will be enough to take $\epsilon = 1$, and show that for this value of ϵ, we can find no corresponding $\delta > 0$. By way of contradiction, suppose that there was a δ such that $|p-q| < \delta$ implied that $|f(p) - f(q)| < 1$. Set $p_n = n$, $q_n = n + \delta/2$. Then $|p_n - q_n| < \delta$, but $|f(p_n) - f(q_n)| = q_n^2 - p_n^2 = n\delta + \delta^2/4 > n\delta > 1$, for n sufficiently large, in fact for $n > \delta^{-1}$.

Example 3.10. *Let $X = [-M, +M]$, $Y = \mathbb{R}$ with usual metrics and let $f : X \to Y$ with $f(p) = p^2$. Then f is uniformly continuous.*

To see this, given $\epsilon > 0$, set $\delta = \frac{\epsilon}{2M}$. then $|p-q| < \delta$ implies that $|p^2 - q^2| = |p+q||p-q| \leq 2M|p-q| < 2M\delta = \epsilon$.

We now look at a characterization of continuity in terms of open and closed sets.

Recall that if $f : X \to Y$ and $S \subseteq Y$, then the *preimage of S* is the subset of X given by

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

It is important to realize that to make this definition, we do not need for the function f to have an inverse function.

Theorem 3.11. *Let (X,d) and (Y,ρ) be metric spaces and let $f : X \to Y$. Then the following are equivalent:*

1. *f is continuous,*
2. *for every open set $U \subseteq Y$, the set $f^{-1}(U)$ is an open set in X,*
3. *for every closed set $C \subseteq Y$, the set $f^{-1}(C)$ is a closed subset of X.*

Proof. First we show that 2) and 3) are equivalent. To see this note that $(f^{-1}(S))^c = \{x \in X : f(x) \notin S\} = f^{-1}(S^c)$. So if say 2) holds and C is a closed set, then C^c is an open set. Hence, $(f^{-1}(C))^c = f^{-1}(C^c)$ is open, which implies that $f^{-1}(C)$ is closed. The proof that 3) implies 2) is similar.

Now we prove that 1) implies 2). Let $U \subseteq Y$ be open. We must prove that $f^{-1}(U)$ is open. Since we will be talking about balls in two different metric spaces, we will use a subscript to help indicate which metric. Let $p_0 \in f^{-1}(U)$, so that $f(p_0) \in U$. Since U is open there is $\epsilon > 0$, so that $B_\rho(f(p_0); \epsilon) \subseteq U$. Because f is continuous at p_0, there is a $\delta > 0$, so that $d(p_0, p) < \delta$ implies that $\rho(f(p_0), f(p)) < \epsilon$.

But this last statement shows that $B_d(p_0; \delta) \subseteq f^{-1}(U)$, because $p \in B_d(p_0; \delta)$ implies that $d(p_0, p) < \delta$, implies that $\rho(f(p_0), f(p)) < \epsilon$, implies that $f(p) \in B_\rho(f(p_0); \epsilon)$, implies that $f(p) \in U$.

Hence, $f^{-1}(U)$ is open in X.

Conversely, assume that 2) holds and we will prove 1). Given $p_0 \in X$ and an $\epsilon > 0$, we have that $U = b_\rho(f(p_0); \epsilon)$ is open. Hence, by assumption, $f^{-1}(U)$ is open and $p_0 \in f^{-1}(U)$. Thus, we may pick $\delta > 0$ so that $B_d(p_0; \delta) \subseteq f^{-1}(U)$. But this implies that if $d(p_0, p) < \delta$, then $p \in f^{-1}(U)$, which implies that $f(p) \in U = B_\rho(f(p_0); \epsilon)$, which implies that $\rho(f(p_0), f(p)) < \epsilon$.

Hence, f is continuous at p_0, and since this was an arbitrary point, we have that f is continuous at every point in X. $\square$

Proposition 3.12. *Let (X, d),(Y, ρ) and (Z, γ) be metric spaces and let $f : X \to Y$ and $g : Y \to Z$ be continuous functions. Then $g \circ f : X \to Z$ is continuous.*

Proof. First note that if $S \subseteq Z$, then $x \in (g \circ f)^{-1}(S)$ iff $(g \circ f)(x) \in S$ iff $g(f(x)) \in S$ iff $f(x) \in g^{-1}(S)$ iff $x \in f^{-1}(g^{-1}(S))$. Thus, we have that $(g \circ f)^{-1}(S) = f^{-1}(g^{-1}(S))$.

We use the above theorem. If $U \subseteq Z$ is open in Z, then $g^{-1}(U)$ is open in Y and hence $f^{-1}(g^{-1}(U))$ is open in X. Thus, when U is open in Z, then $(g \circ f)^{-1}(U)$ is open in X. Hence, $g \circ f$ is continuous. $\square$

Problem 3.13. *Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^3$. Prove that f is continuous.*

Problem 3.14. *Prove that $f(x) = x^3$ is not uniformly continuous on $\mathbb{R}$.*

Problem 3.15. *Prove that $f : [-M, +M] \to \mathbb{R}$ with $f(x) = x^3$ is uniformly continuous.*

Problem 3.16. *Let (X, d) and (Y, ρ) be metric spaces. A function $f : X \to Y$ is called Lipschitz continuous provided that there is a constant $K > 0$, so that $\rho(f(p), f(q)) \leq Kd(p, q)$. Prove that every Lipschitz continuous function is uniformly continuous.*

Given a function $f : X \to Y$ and $S \subseteq X$, the *image of S* is the subset of Y given by $f(S) = \{f(p) : p \in S\}$.

Problem 3.17. *Let (X,d) and (Y,ρ) be metric spaces and $f : X \to Y$ a function. Prove that f is continuous at p_0 if and only if for every $\epsilon > 0$, there exists a $\delta > 0$ so that the image of $B_d(p_0;\delta)$ is contained in $B_\rho(f(p_0);\epsilon)$.*

3.1 Functions into Euclidean space

Let (X,d) be a metric space. Given functions $f_i : X \to \mathbb{R}$, $i = 1, ..., k$ we can define a function $F : X \to \mathbb{R}^k$ by setting $F(x) = (f_1(x), ..., f_k(x))$. Conversely, any function $F : X \to \mathbb{R}^k$ is readily seen to be of this form. We call the functions $f_1, ..., f_k$ the *component functions* or, more simply, the *components* of F.

To avoid possibly confusion, we shall let d_2 denote the Euclidean metric on $\mathbb{R}^k$.

Theorem 3.18. *Let (X,d) be a metric space and let $(\mathbb{R}^k, d_2)$ be Euclidean space. A function $F : X \to \mathbb{R}^k$ is continuous at a point $p_0 \in X$ if and only if each of its component functions $f_i : X \to \mathbb{R}$ is continuous at p_0 for every $i = 1, ..., k$. The function F is continuous if and only if each of its component functions is continuous.*

Proof. First assume that F is continuous at p_0. Given $\epsilon > 0$, there exists $\delta > 0$ such that when $d(p_0,p) < \delta$, then

$$d_2(F(p_0), F(p)) = \sqrt{(f_1(p_0) - f_1(p)^2 + \cdots + (f_k(p_0) - f_k(p))^2} < \epsilon.$$

For each i, since $|f_i(p_0) - f_i(p)| < d_2(F(p_0), F(p))$, we see that $d(p_0,p) < \delta$ implies that $|f_i(p_0) - f_i(p)| < \epsilon$. Thus, f_i is continuous at p_0 for each i.

Conversely, assume that each f_i is continuous at p_0 and let $\epsilon > 0$ be given. For each i, there exists $\delta_i > 0$, so that $d(p_0,p) < \delta_i$ implies that $|f_i(p_0) - f_i(p)| < \frac{\epsilon}{\sqrt{k}}$. Now let $\delta = \min\{\delta_1, ..., \delta_k\}$. Then for $d(p_0,p) < \delta$ we have that $d_2(F(p_0), F(p)) < \epsilon$.

The second equivalence follows from the first and the fact that continuous means continuous at every point. □

Theorem 3.19. *Let (X,d) be a metric space, let $p_0 \in X$, let $\mathbb{R}$ have the usual metric and let $f, g : X \to \mathbb{R}$ be functions. If f and g are both continuous at p_0, then:*

1. *$f + g$ is continuous at p_0,*
2. *fg is continuous at p_0,*

3. *for any constant c, cf is continuous at p_0,*

4. *if $g(p) \neq 0$ for all p, then $\frac{f}{g}$ is continuous at p_0.*

Proof. To prove the first statement, given $\epsilon > 0$, there is $\delta_1 > 0$, so that $d(p_0, p) < \delta_1$ implies that $f(p_0) - f(p)| < \epsilon/2$. Also there is $\delta_2 > 0$, so that $d(p_0, p) < \delta_2$ implies that $|g(p_0) - g(p)| < \epsilon/2$. Let $\delta = \min\{\delta_1, \delta_2\}$.

To prove the second, given $\epsilon > 0$, first pick $\delta_1 > 0$ and $\delta_2 > 0$, so that $d(p_0, p) < \delta_1$ implies $|f(p_0) - f(p)| < 1$ and $d(p_0, p) < \delta_2$ implies that $|g(p_0) - g(p)| < 1$. Now pick $\delta_3 > 0$ and $\delta_4 > 0$ so that $d(p_0, p) < \delta_3$ implies that $|f(p_0) - f(p)| < \frac{\epsilon}{2(|g(p_0)|+1)}$ and $d(p_0, p) < \delta_4$ implies that $|g(p_0) - g(p)| < \frac{\epsilon}{2(|f(p_0)|+1)}$. Let $\delta = \min\{\delta_1, ..., \delta_4\}$, then for $d(p_0, p) < \delta$ we have that

$$|f(p_0)g(p_0) - f(p)g(p)| \leq |f(p_0) - f(p)||g(p_0)| + |g(p_0) - g(p)||f(p)| < \frac{\epsilon}{2(|g(p_))| + 1)}|g(p_0)| + \frac{\epsilon}{2(|f(p_0)| + 1)}(|f(p_0)| + 1) < \epsilon.$$

Statement 3) follows from 2) and the fact constants are continuous.

Statement 4) will follow from 2), if we prove that the function $\frac{1}{g}$ is continuous at p_0. Given $\epsilon > 0$, we first pick $\delta_1 > 0$, so that $d(p_0, p) < \delta_1$ implies that $|g(p_0) - g(p)| < \frac{|g(p_0)|}{2}$. This guarantees that $|g(p)| > \frac{|g(p_0)|}{2}$. We now pick $\delta_2 > 0$ so that $d(p_0, p) < \delta_2$ implies that $|g(p_0) - g(p)| < \frac{|g(p_0)|^2\epsilon}{2}$. Now if we let $\delta = \min\{\delta_1, \delta_2\}$, then when $d(p_0, p) < \delta$, we have

$$|\frac{1}{g(p_0)} - \frac{1}{g(p)}| = \frac{|g(p) - g(p_0)|}{|g(p_0)g(p)|} \leq \frac{|g(p) - g(p_0)|}{|g(p_0)|^2/2} < \frac{|g(p_0)|^2\epsilon}{2}\frac{2}{|g(p_0)|^2} = \epsilon.$$

□

Corollary 3.20. *Let (X, d) be a metric space, let $\mathbb{R}$ have the usual metric and let $f, g : X \to \mathbb{R}$ be functions. If f and g are both continuous, then:*

1. *$f + g$ is continuous,*

2. *fg is continuous,*

3. *for any constant c, cf is continuous,*

4. *if $g(p) \neq 0$ for all p, then $\frac{f}{g}$ is continuous.*

Corollary 3.21. *Let (X, d) be a metric space, let $p_0 \in X$,and let $F, G : X \to \mathbb{R}^k$. If F, G are both continuous at p_0(respectively, both continuous), then*

1. *$F + G$ is continous at p_0(respectively, continuous),*
2. *cF is continuous at p_0(respectively, continuous),*
3. *$F \cdot G$ is continuous at p_0(respectively, continuous).*

Proof. We only prove 3), the rest are similar. If we let $f_1, ..., f_k$ and $g_1, .., g_k$ be the component functions for F and G, then the fact that F and G are continuous at p_0 implies that $f_1, ..., f_k, g_1, ..., g_k$ are all continuous at p_0. By the above result about continuity of products, each $f_i g_i$ is continuous at p_0, and so by the result about continuity of sums, $F \cdot G = f_1 g_1 + \cdots + f_k g_k$ is continuous at p_0.

The results about continuity on all of X then follow since they are true at each point. □

3.2 Continuity of Some Basic Functions

Proposition 3.22. *Every polynomial defines a continuous function on $\mathbb{R}$.*

Proof. The proof is by induction on the degree of the polynomial. First note that $_1(x) = x$ is not only continuous but uniformly continuous since we can pick $\delta = \epsilon$. Applying Corollary 3.19.2 with $f(x) = g(x) = x$, we get that the function $p_2(x) = x^2$ is continuous. Now assume that we have shown that $p_n(x) = x^n$ is continuous. Then applying Corollary 3.20.2 again, we have that $p_n(x)p_1(x) = x^{n+1} = p_{n+1}(x)$ is continuous.

Since p_n is continuous, applying Corollary 3.20.3 any function of the form $a_n x^n$, with $a_n \in \mathbb{R}$ a constant is continuous. Since constants are continuous, applying Corollary 3.20.1 to $f(x) = a_1 x$ and $g(x) = a_0$, we get that any first degree polynomial is continuous. Now assume that we have proved that all polynomials of degree n are continuous and we are given an $(n + 1)$-st degree polynomial, $q(x) = a_{n+1}x^{n+1} + q_n(x)$ where q_n is an n-th degree polynomial. By our inductive hypothesis q_n is continuous and by our earlier work, $a_{n+1}x^{n+1}$ is continuous. Thus, applying Corollary 3.20.1, gives that q is continuous.

Thus, all polynomials are continuous. □

Proposition 3.23. *Let p, q be polynomials and let $E = \{x \in \mathbb{R} : q(x) \neq 0\}$. Then $r : E \to \mathbb{R}$ defined by $r(x) = p(x)/q(x)$ is continuous.*

Proof. We know that p and q are continuous on $\mathbb{R}$ by the above result and hence they are continuous on E. The result now follows by applying Corollary 3.20.4 with $X = E$. □

We now prove continuity of our favorite trigonometric functions. We will need a few important facts and inequalities:

$$cos(x+y) = cos(x)cos(y) - sin(x)sin(y), sin(x+y) = sin(x)cos(y) + cos(x)sin(y),$$
$$|sin(x)| \leq |x|, |1 - cos(x)| \leq |x|.$$

The first two equalities are the double angle formulas. The two inequalities follow by examining the unit circle and recalling that x in radian measure is the length of the arc.

Proposition 3.24. *The functions, $sin, cos : \mathbb{R} \to \mathbb{R}$ are both Lipschitz continuous with constant 2.*

Proof. We have that

$$|sin(x) - sin(y)| = |sin((x-y)+y) - sin(y)| =$$
$$|sin(x-y)cos(y) + cos(x-y)sin(y) - sin(y)| \leq$$
$$|sin(x-y)cos(y)| + |(cos(x-y)-1)sin(y)| \leq$$
$$|sin(x-y)| + |cos(x-y) - 1| \leq 2|x-y|.$$

Thus, $d(sin(x), sin(y)) \leq 2d(x,y)$. □

Recall that by Problem 3.16, we have that the functions sin and cos are not only continuous on $\mathbb{R}$ but are uniformly continuous on $\mathbb{R}$. Applying Corollary 3.20.4, we have that the functions $tan, cot, sec, cosec$ are continuous wherever their denominators do not vanish.

We now look at some multivariable functions. We formally define functions, $x_i : \mathbb{R}^k \to \mathbb{R}$ by $x_i((a_1, ..., a_k)) = a_i$. Thus, x_i is the function that yields the i-th coordinate of a point. These are called the **coordinate functions of a point.** By a **polynomial in the coordinate functions,** we mean any function $p : \mathbb{R}^k \to \mathbb{R}$ that can be expressed as a finite sum of products of constants times products of coordinate functions of points. For example, $6x_1^2x_3^5 + 2x_2^3 + 7$ is a polynomial in the coordinate functions.

Proposition 3.25. *Every polynomial in the coordinate functions defines a continuous function from the Euclidean space $\mathbb{R}^k$ to $\mathbb{R}$.*

Proof. Given $p = (a_1, ..., a_k)$ and $q = (b_1, ..., b_k)$ two points in $\mathbb{R}^k$, we have that $d(x_i(p), x_i(q)) = |x_i(p) - x_i(q)| = |a_i - b_i| \leq d_2(p, q)$. This shows that each coordinate function is a Lipschitz continuous function from $(\mathbb{R}^k, d_2)$ to $\mathbb{R}$. The remainder of the proof follows as in the case of ordinary polynomials. First, one does induction to show that every monomial function is continuous and then do an induction on the number of terms in the polynomial. □

Problem 3.26. *Prove that the function* $g : [0, +\infty) \to \mathbb{R}$ *defined by* $g(x) = \sqrt{x}$ *is continuous.*

Problem 3.27. *Prove or disprove that the above function is uniformly continuous.*

3.3 Continuity and Limits

In this section we generalize the concept of limit and establish the connections between limits and continuity. The first result gives a sequential test for continuity.

Theorem 3.28. *Let* (X, d) *and* (Y, ρ) *be metric spaces, let* $f : X \to Y$ *be a function and let* $p_0 \in X$*. Then* f *is continuous at* p_0 *if and only if whenever* $\{p_n\} \subseteq X$ *is a sequence that converges to* p_0*, the sequence* $\{f(p_n)\} \subseteq Y$ *converges to* $f(p_0)$*.*

Proof. Assume that f is continuous at p_0 and let $\epsilon > 0$. Then there exists $\delta > 0$, such that $d(p_0, p) < \delta$ implies that $\rho(f(p_0), f(p)) < \epsilon$. Since $\lim_n p_n = p_o$, there is N so that $n > N$ implies that $d(p_0, p) < \delta$. Hence, if $n > N$, then $\rho(f(p_0), f(p_n)) < \epsilon$. This proves that $\lim_n f(p_n) = f(p_0)$.

To prove the converse, we show its contrapositive. So assume that f is not continuous at p_0. Then there exists $\epsilon > 0$, for which we can obtain no δ that satisfies the definition of continuity. The fact that $\delta = 1/n$ fails to satisfy the definition, means that there exists a point p_n, with $d(p_0, p_n) < 1/n$, but $\rho(f(p_0), f(p_n)) \geq \epsilon$.

In this manner we obtain a sequence $\{p_n\}$ with $d(p_0, p_n) < 1/n$ and hence $\lim_n p_n = p_0$. Yet $\rho(f(p_0), f(p_n)) \geq \epsilon$ for every n, and so $f(p_0)$ cannot be the limit of the sequence $\{f(p_n)\}$. □

Corollary 3.29. *Let* (X, d) *and* (Y, ρ) *be metric spaces and let* $f : X \to Y$*. Then* f *is continuous if and only if whenever* $\{p_n\} \subseteq X$ *is a convergent sequence we have* $\lim_n f(p_n) = f(\lim_n p_n)$*.*

There is another notion of limit, familiar from calculus, that is also related to continuity.

Definition 3.30. *Let (X, d) and (Y, ρ) be metric spaces, let $p_0 \in X$ be a cluster point, let $f : X\backslash\{p_0\} \to Y$ be a function and let $q_0 \in Y$. We write*

$$\lim_{p \to p_0} f(p) = q_0$$

provided that for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever $d(p_0, p) < \delta$ and $p \neq p_0$, then $\rho(q_0, f(p)) < \epsilon$.

Note that we need p_0 to be a cluster point to guarantee that the set of p's satisfying $d(p_0, p) < \delta$ and $p \neq p_0$, is non-empty. Often, when using this definition, the function f is actually defined on all of X, in which case we simply ignore its value at p_0.

We leave the proof of the following result as an exercise.

Theorem 3.31. *Let (X, d) and (Y, ρ) be metric spaces, let $f : X \to Y$ be a function and let $p_0 \in X$ be a cluster point. Then f is continuous at p_0 if and only if $\lim_{p \to p_0} f(p) = f(p_0)$.*

Lemma 3.32. *Let (X, d) and (Y, ρ) be metric spaces and let $f : X \to Y$ be any function. If $p_0 \in X$ is not a cluster point, then f is continuous at p_0.*

Proof. Since p_0 is not a cluster point, there exists $r > 0$ so that $d(p_0, p) < r$ implies that $p = p_0$. Thus, given any $\epsilon > 0$, if we let $\delta = r$, then $d(p_0, p) < \delta$ implies that $\rho(f(p_0), f(p)) = 0 < \epsilon$. □

Corollary 3.33. *Let (X, d) and (Y, ρ) be metric spaces, and let $f : X \to Y$ be a function. Then f is continuous if and only if for every cluster point $p_0 \in X$, we have that $\lim_{p \to p_0} f(p) = f(p_0)$.*

Problem 3.34. *Prove Theorem 3.31.*

Problem 3.35. *Prove Corollary 3.33.*

Problem 3.36. *Let (X, d) and (Y, ρ) be metric spaces, let $f : X \to Y$ be a function and let $p_0 \in X$ be a cluster point. Prove that $\lim_{p \to p_0} f(p) = q_0$ if and only if for every sequence of points $\{p_n\} \subseteq X$ such that $p_n \neq p_0$ and $\lim_n p_n = p_0$, we have $\lim_n f(p_n) = q_0$.*

3.4 Continuous Functions and Compact Sets

If we let $C = \{(x, y) \in \mathbb{R}^2 : xy = 1\}$ then it is not hard to see that C is a closed subset of $\mathbb{R}^2$. Also we know that the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f((x, y)) = x$ is a continuous function, it is the first coordinate function. But the image $f(C) = \{x \in \mathbb{R} : x \neq 0\}$ is not a closed subset of $\mathbb{R}$. Thus, the continuous image of a closed set need not be a closed set. The story is quite different for compact sets.

Theorem 3.37. *Let (X, d) and (Y, ρ) be metric spaces, and let $f : X \to Y$ be a continuous function. If $K \subseteq X$ is compact, then its image $f(K) \subseteq Y$ is compact.*

Proof. Let $\{U_\alpha\}_{\alpha \in A}$ be an open cover of $f(K)$, then each of the sets $f^{-1}(U_\alpha)$ is open and $\{f^{-1}(U_\alpha)\}_{\alpha \in A}$ is an open cover of K. Since K is compact, there is a finite subset $F \subseteq A$ such that $\{f^{-1}(U_\alpha)\}_{\alpha \in F}$ covers K.

Note that $f(f^{-1}(U_\alpha)) = U_\alpha$. Hence, $\{U_\alpha\}_{\alpha \in F}$ covers $f(K)$. Since every open cover of $f(K)$ has a finite subcover, $f(K)$ is compact. □

Corollary 3.38. *Let (X, d) and (Y, ρ) be metric spaces, and let $f : X \to Y$ be a continuous function. If X is compact, then $f(X)$ is a bounded subset of Y.*

Thus, when X is non-empty, compact and $f : X \to \mathbb{R}$ is continuous, then there exists M such that $|f(x)| \leq M$ for every $x \in X$. So in particular $\sup\{f(x) : x \in X\}$ and $\inf\{f(x) : x \in X\}$ exists. The following shows that not only does the supremum and infimum exist but that they are attained.

Corollary 3.39. *Let (X, d) be a non-empty compact metric space and let $f : X \to \mathbb{R}$ be continuous. Then there are points $x_m, x_M \in X$, such that for any $x \in X$, $f(x_m) \leq f(x) \leq f(x_M)$. That is, $f(x_m) = \inf\{f(x) : x \in X\}$ and $f(x_M) = \sup\{f(x) : x \in X\}$.*

Proof. The proof is similar to the proof of Proposition 1.126. Choose a sequence of points $\{x_n\}$ such that $\lim_n f(x_n) = \inf\{f(x) : x \in X\}$. Then since X is sequentially compact, there is a subsequence $\{x_{n_k}\}$ that converges to some point x_m. Since f is continuous, $f(x_m) = \lim_k f(x_{n_k}) = \inf\{f(x) : x \in X\}$. The proof for the supremum is similar. □

The above result gives the proof that whenever $f : [a, b] \to \mathbb{R}$ is continuous, then f attains its maximum and minimum value. First, by Heine-Borel the interval $[a, b]$ is compact, now apply the above result. Note that $(0, 1)$ is a bounded interval, $f(x) = 1/x$ is continuous on this set but is not even bounded.

Theorem 3.40. *Let (X, d) and (Y, ρ) be metric spaces, and let $f : X \to Y$ be a continuous function. If X is compact, then f is uniformly continuous.*

Proof. Given $\epsilon > 0$, since f is continuous at p, there is a $\delta(p) > 0$, such that $d(p, q) < \delta(p)$ implies that $\rho(f(p), f(q)) < \epsilon/2$.

The collection of sets $\{B(p; \frac{\delta(p)}{2})\}_{p \in X}$ is an open cover of X. Hence, there exists a finite collection of points, $\{p_1, ..., p_n\}$ so that

$$X \subseteq B(p_1; \frac{\delta(p_1)}{2}) \cup \cdots \cup B(p_n; \frac{\delta(p_n)}{2}).$$

Let $\delta = \frac{1}{2} \min\{\delta(p_1), ..., \delta(p_n)\}$.

Now let $p, q \in X$ be any points with $d(p, q) < \delta$. Because the above balls are a cover, there exists an i, so that $p \in B(p_i; \frac{\delta(p_i)}{2})$. Hence, $\rho(f(p_i), f(p)) < \epsilon/2$. Also, $d(p_i, q) \leq d(p_i, p) + d(p, q) < \frac{\delta(p_i)}{2} + \delta \leq \delta(p_i)$ and so, $\rho(f(p_i), f(q)) < \epsilon/2$. Thus, we have that $\rho(f(p), f(q)) \leq \rho(f(p), f(p_i)) + \rho(f(p_i), f(q)) < \epsilon$. □

Thus, for example any rational function $r(x) = p(x)/q(x)$, such that $q(x) \neq 0$ on the set $[a, b]$ will be uniformly continuous.

Problem 3.41. *Let (X, d) be a metric space, fix $p \in X$ and define $f : X \to \mathbb{R}$ by $f(x) = d(p, x)$. Prove that f is continuous. Use this fact to give another proof of Proposition 1.130.*

3.5 Connected Sets and the Intermediate Value Theorem

We will see in this section that the intermediate value theorem from calculus is really a consequence of the fact that an interval of real numbers is a connected set. First we need to define this concept.

Definition 3.42. *A metric space (X, d) is* **connected** *if the only subsets of X that are both open and closed are X and the empty set. A subset S of X is called* **connected** *provided that the subspace (S, d) is a connected metric space. If S is not connected then we say that S is* **disconnected** *or* **separated.**

Proposition 3.43. *A metric space (X, d) is disconnected if and only if X can be written as a union of two disjoint, non-empty open sets.*

Proof. We have that (X, d) is disconnected if and only if there is a subset A with $A \neq X$ and A non-empty and A is open and closed. Let $B = A^c$, then since A is closed B is open and since $A \neq X$, B is non-empty. Thus, $X = A \cup B$ expresses X as a disjoint union of two non-empty open sets.

Conversely, if $X = A \cup B$ with A and B disjoint, non-empty and open. Then necessarily $A = B^c$ and so A is both open and closed. Since neither set was empty, $A \neq X$ and A is non-empty. □

Example 3.44. *If (X, d) is a discrete metric space with two or more points, then X is disconnected since $X = \{p_0\} \cup \{p_0\}^c$ expresses X as a disjoint union of two non-empty open sets.*

We now come to perhaps the most important example of a connected space. By an *interval* in $\mathbb{R}$ we mean either an open interval, closed interval, or half-open interval. The endpoints can be either an actual number or $+\infty$ or $-\infty$.

Theorem 3.45. *Let $I \subseteq \mathbb{R}$ be an interval or all of $\mathbb{R}$. Then I is a connected set.*

Proof. Suppose not, then we could write $I = A \cup B$ where A and B are both non-empty and open in the metric space (I, d) where d is the usual metric. Let $a \in A$ and $b \in B$. Without loss of generality, we can assume that $a < b$(otherwise just change the names of A and B). Let $A_1 = A \cap [a, b]$, and $B_1 = B \cap [a, b]$. Then A_1 and B_1 are disjoint, non-empty open sets in the topological space $([a, b], d)$. Since A_1 is closed and bounded, it is compact and hence $c = \sup\{x : x \in A_1\}$ is an element of A_1. Hence, $c < b$. But since A_1 is open, there is an open interval centered at c that is still in A_1. This open interval contains points in A_1 that are larger than c. This contradiction completes the proof. □

Now we come to a general version of the Intermediate Value Theorem.

Theorem 3.46 (Intermediate Value Theorem for Metric Spaces)**.** *Let (X, d) be a connected metric space and let $f : X \to \mathbb{R}$ be a continuous function. If $x_0, x_1 \in X$ with $f(x_0) < L < f(x_1)$, then there is $x_2 \in X$ with $f(x_2) = L$.*

Proof. Suppose not. Then we would have that $X = f^{-1}((-\infty, L)) \cup f^{-1}((L, +\infty))$. Since f is continuous, both of these sets are open, with x_0 in the first set and x_1 in the second set. Thus, X is written as a disjoint union of two non-empty open sets. This makes each of these sets open and closed, contradiction. □

Corollary 3.47 (Intermediate Value Theorem). *Let $I \subseteq \mathbb{R}$ be an interval or the whole real line and let $f : I \to \mathbb{R}$ be continuous. If $x_0, x_1 \in I$ and $f(x_0) < L < f(x_1)$, then there is x_2 between x_0 and x_1 with $f(x_2) = L$.*

Proof. If $x_0 < x_1$, apply the theorem to the connected space $[x_0, x_1]$, if $x_1 < x_0$ apply the theorem to the connected space $[x_1, x_0]$. □

Theorem 3.48. *Let (X, d) and (Y, ρ) be metric spaces and let $f : X \to Y$ be continuous. If X is connected, then $f(X) \subseteq Y$ is a connected subset.*

Proof. Let $S = f(X)$ and let (S, ρ) be the subspace. Then $f : X \to S$ is also continuous. If we could write $S = A \cup B$ as a disjoint union of two non-empty open sets, then we would have that $X = f^{-1}(A) \cup f^{-1}(B)$ expresses X as a disjoint union of non-empty open sets. Hence, S must be connected. □

Definition 3.49. *A metric space (X, d) is called* **pathwise connected** *provided that for any two points, $a, b \in X$ there exists a continuous function $f : [0, 1] \to X$ such that $f(0) = a$ and $f(1) = b$.*

Intuitively, a space is pathwise connected if and only if you can draw a "curve" between any two points with no breaks in the curve.

Problem 3.50. *Prove that if (X, d) is pathwise connected, then (X, d) is connected.*

Example 3.51. *The subset of the plane defined by*

$$X = \{(x, sin(\frac{1}{x})) : 0 < x \leq 1\} \cup \{(0, y) : -1 \leq y \leq +1\}$$

is a space that is connected, but not pathwise connected.

Problem 3.52. *Let $X = \{(x, y) : a \leq x \leq b, c \leq y \leq d\} \subseteq \mathbb{R}^2$. Prove that X is pathwise connected.*

Definition 3.53. *Let $g : [a, b] \to \mathbb{R}$. By the* **graph of g** *we mean the set $G = \{(x, g(x)) : a \leq x \leq b\} \subseteq \mathbb{R}^2$.*

Problem 3.54. *Let $g : [a, b] \to \mathbb{R}$. Prove that g is continuous if and only if the graph of g is a pathwise connected subset of the plane.*

This last problem is often used in calculus to intuitively describe continuous functions as those whose graph can be drawn without lifting your pencil.

Chapter 4

The Contraction Mapping Principle

At this point we would like to present an important result that uses the concepts that we have developed and together with some of its applications. Logically, this material should be done later, since it uses facts from calculus about derivatives and integrals which we have not yet discussed. But we find that a little practical math at this stage helps to motivate the rest of the course.

Definition 4.1. *Let (X, d) be a metric space. A function $f : X \to X$ is called a* **contraction mapping** *provided that there is $r, 0 < r < 1$, so that $d(f(x), f(y)) \leq rd(x, y)$ for every $x, y \in X$.*

Note that saying that f is a contraction mapping is the same as saying that it is Lipschitz continuous with constant $r < 1$. It is important to note that when we say that f is a contraction mapping that is stronger than just saying that $d(f(x), f(y)) < d(x, y)$, since we need the r.

Given a function $f : X \to X$ any point satisfying $f(x_0) = x_0$ is called a **fixed point** of the function.

Theorem 4.2 (Contraction Mapping Principle)**.** *Let (X, d) be a complete metric space and let $f : X \to X$ be a contraction mapping. Then:*

1. *there exists a unique point $x_0 \in X$, such that $f(x_0) = x_0$,*

2. *if $x_1 \in X$ is any point and we define a sequence inductively, by setting $x_{n+1} = f(x_n)$, then $\lim_n x_n = x_o$,*

3. *for this sequence, we have that $d(x_0, x_n) \leq \frac{d(x_2,x_1)r^{n-1}}{1-r}$.*

Proof. First, we show that the inductively defined sequence does converge. To see this, since X is complete, it will be enough to show that the sequence is Cauchy.

Let $A = d(x_2, x_1)$. Then we have that $d(x_3, x_2) = d(f(x_2), f(x_1)) \leq rd(x_2, x_1) = rA$. Similarly, $d(x_4, x_3) = d(f(x_3), f(x_2)) \leq rd(x_3, x_2) \leq r^2A$. By induction, we prove that $d(x_{n+1}, x_n) \leq r^{n-1}A$.

Now, if $m > n$, then $d(x_m, x_n) \leq d(x_m, x_{m-1}) + \cdots + d(x_{n+1}, x_n) \leq \sum_{k=n}^{m-1} r^{k-1}A \leq \frac{Ar^{n-1}}{1-r}$. Given any $\epsilon > 0$, we may choose an integer N such that $\frac{Ar^N}{1-r} < \epsilon$. Then if $m, n > N$, we have that $d(x_m, x_n) < \epsilon$. Thus, $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy.

Let $x_0 = \lim_n x_n$. Then since f is continuous, $f(x_0) = \lim_n f(x_n) = \lim_n x_{n+1} = x_0$. to a point x_0 satisfying $f(x_0) = x_0$.

Now if we fix any n, then $d(x_0, x_n) = \lim_m d(x_m, x_n) \leq \frac{Ar^{n-1}}{1-r}$, by the above estimate, which proves 3).

Thus, we know that there is a point x_0, with $f(x_0) = x_0$ and that this sequence converges to one such point. To complete the proof of the theorem it will be enough to show that if $f(x_0) = x_0$ and $f(p_0) = p_0$, then $p_0 = x_0$. To see this last fact, note that $d(p_0, x_0) = d(f(p_0), f(x_0)) \leq rd(p_0, x_0)$. Since $r < 1$, this implies that $d(p_0, x_0) = 0$. □

Thus, the contraction mapping principle not only guarantees us the existence of a fixed point, but shows that there is a unique fixed point and gives us a method for approximating the fixed point, together with an estimate of how close the sequence x_n is to the fixed point! This is a remarkable amount of information.

Here's a typical application of this theorem. Let $f(x) = cos(x)$ and note that since $1 < \pi/2$ for $0 \leq x \leq 1$, we have that $0 \leq f(x) \leq 1$. Thus, $f : [0, 1] \to [0, 1]$ and is continuous. Also, by the mean value theorem, for $0 \leq x \leq y \leq 1$, there is $c, x \leq c \leq y$ with

$$f(y) - f(x) = f'(c)(y - x) = -sin(c)(y - x).$$

Hence, $|f(y) - f(x)| \leq sin(c)|y - x| \leq sin(1)|y - x|$. Thus, $f(x) = cos(x)$ is a contraction mapping with $r = sin(1) < 1$.

Using the fact that $[0, 1]$ is complete, by the contraction mapping principle, there is a unique point, $0 \leq x_0 \leq 1$, such that $cos(x_0) = x_0$. Moreover, we can obtain this point(or at least approximate it) by choosing any number $x_1, 0 \leq x_1 \leq 1$ and forming the inductive sequence $x_{n+1} = cos(x_n)$.

The third part of the theorem gives us an estimate of the distance between our "approximate" fixed point x_n and the true fixed point. In particular, if we pick $x_1 = 0$, then $x_2 = cos(x_1) = 1$, so $|x_2 - x_1| = 1$. Hence, $|x_0 - x_n| \leq \frac{sin(1)^{n-1}}{1-sin(1)}$.

Problem 4.3. *Let $f(x) = \frac{x}{2} + \frac{1}{x}$. Use basic calculus to show that $f : [1, 2] \to [1, 2]$ and use the mean value theorem to show that f is a contraction. Show that the unique fixed point of f is $\sqrt{2}$. Using $x_1 = \frac{3}{2}$, give an estimate on $|\sqrt{2} - x_n|$.*

4.1 Application: Newton's Method

Newton's method gives an iterative method for approximating the solution to an equation of the form $f(x) = 0$, when f is differentiable.

The idea of the iteration is to choose any x_1 and then get an "improved" estimate to the solution by tracing the tangent line to the graph at the point $(x_1, f(x_1))$ until it intersects the x-axis and letting this determine the point x_2. The formula that one obtains is

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

Newton's method consists of repeating this formula iteratively to generate a sequence of points

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

that, hopefully, converge to the zero of the function.

Note that if we set $g(x) = x - \frac{f(x)}{f'(x)}$, then $f(x) = 0$ if and only if $g(x) = x$. Thus, finding a zero of f is equivalent to finding a fixed point of g. Moreover, the above iteration is simply computing $x_{n+1} = g(x_n)$.

Thus, if we can construct an interval $[a, b]$ such that $g : [a, b] \to [a, b]$ and such that g is a contraction mapping on $[a, b]$ then we will have a criterion for convergence of Newton's method.

The details are below.

Theorem 4.4 (Newton's Method)**.** *Let f be a twice continuously differentiable function and assume that there is a point x_0 with $f(x_0) = 0$ and $f'(x_0) \neq 0$. Then there is $M > 0$ so that for any x_1 with $x_0 - M \leq x_1 \leq x_0 + M$ the sequence of points obtained by Newton's method starting at x_1 converges to x_0.*

Proof. Fix an $r, 0 < r < 1$. Let g be as above, then $g'(x) = \frac{f(x)f''(x)}{(f'(x))^2}$, which is continuous in a neighborhood of x_0. Since $g'(x_0) = 0$, there is some $M > 0$, so that $|x - x_0| \leq M$ implies that $|g'(x)| \leq r$. Hence, for any x, in the interval $I = [x_0 - M, x_0 + M]$, we have that $|g(x) - x_0| = |g(x) - g(x_0)| = |g'(c)(x - x_0)| \leq rM < M$. Hence, $x \in I$ implies that $g(x) \in I$.

Thus, $g : I \to I$ and for any $x, y \in I$, $|g(x) - g(y)| = |g'(c)(x - y)| \leq r|x - y|$, since $c \in I$. Hence, we may apply the contraction mapping principle to g to obtain the desired result. □

There are two difficulties in using the above theorem. First, the interval is centered at x_0, which is a value that we are trying to obtain! The second lies in having enough detailed information about g to explicitly find M. However, since the proof of Newton's method uses the contraction mapping principle, in some cases it is possible to say a great deal.

Problem 4.5. *Let $f(x) = x^2 - 5$. Show that a zero of this equation lies somewhere in the interval $[2, 3]$. Calculate $g(x)$ for this function and prove that $g : [2, 3] \to [2, 3]$ and $|g'(x)| \leq 1/2$ for $x \in [2, 3]$. Prove that if we perform Newton's method with $x_1 = 2$, then $|x_n - \sqrt{5}| \leq (1/2)^n$.*

Problem 4.6. *Let $f(x) = x - cos(x)$, so that $x_0 = cos(x_0)$, the problem that we studied earlier. Compute the corresponding g for this f and find a concrete interval $I = [a, b]$ with $0 \leq a < b \leq 1$ such that $g : I \to I$ is a contraction mapping. How does this r compare to the earlier r from the last section?*

4.2 Application: Solution of ODE's

In this section we show how the contraction mapping principle can be used to deduce that solutions exist and are unique for some very complicated ordinary differential equations.

Starting with a continuous function $h(x, y)$ and an intial value y_0, we wish to solve

$$\frac{dy}{dx} = h(x, y), y(a) = y_0.$$

That is, we seek a function $f(x)$ so that $f'(x) = h(x, f(x))$ for $a < x < b$ and $f(a) = y_0$. Such a problem is often called an *initial value problem(IVP).*

For example, when $h(x, y) = x^2y^3$, then we are trying to solve, $\frac{dy}{dx} = x^2y^3$ which can be done by separation of variables. But our function could be $h(x, y) = sin(xy)$, in which case the differential equation becomes $\frac{dy}{dx} = sin(xy)$, which cannot be solved by elementary means.

For this application, we will use a number of things that we have not yet developed fully. But again, we stress that we are seeking motivation.

First note that by the fundamental theorem a continuous function $f : [a,b] \to \mathbb{R}$ satisfies the IVP if and only if for $a \le x \le b$,

$$f(x) = \int_a^x f'(t)dt + y_0 = \int_a^x h(t, f(t))dt + y_0.$$

This is often called the *integral form* of the IVP.

Now let $C([a,b])$ denote the set of all real-valued continuous functions on the interval $[a,b]$. Note that if we are given any $f \in C([a,b])$ and we set

$$g(x) = \int_a^x h(t, f(t))dt + y_0,$$

then g is also a continuous function on $[a,b]$.

Thus, we can define a map $\Phi : C([a,b]) \to C([a,b])$ by letting $\Phi(f) = g$, where f and g are as above. We see that solving our IVP is the same as finding a fixed point of the map Φ! Also, we are now in a situation, where by a "point" in our space $C([a,b])$, we mean a function.

This is starting to look like an application of the contraction mapping principle. For this we would first need a metric d on the set $C([a,b])$(again points in this metric space are functions!) so that $(C([a,b]), d)$ is a complete metric space and then we would need Φ to be a contraction mapping.

It turns out that there is such a metric on $C([a,b])$ and that for many functions h the corresponding map Φ is a contraction mapping.

First for the metric. Given any two functions $f, g \in C([a,b])$, we set

$$d(f,g) = \sup\{|f(x) - g(x)| : a \le x \le b\}.$$

This is the example that was introduced in the our first section on metric spaces. Note that since f, g are continuous functions on the compact metric space $[a,b]$, we have that the continuous function $f - g$ is bounded. Hence, the supremum is finite. Also, it is clear that $d(f,g) = 0$ if and only if $f(x) = g(x)$ for all x, i.e., if and only if $f = g$. Note that $d(f,g) = d(g,f)$. Finally, if $f, g, h \in C([a,b])$, then

$$\begin{aligned} d(f,g) = \sup\{|f(x) - h(x) + h(x) - g(x)| : a \le x \le b\} \le \\ \sup\{|f(x) - h(x)| + |h(x) - g(x)| : a \le x \le b\} \le \\ \sup\{|f(x) - h(x)| : a \le x \le b\} + \sup\{|h(x) - g(x)| : a \le x \le b\} = \\ d(f,h) + d(h,g). \end{aligned}$$

Thus, we see that d is indeed a metric on $C([a,b])$. To apply the contraction mapping principle, we need this metric space to be complete. This fact is shown by the following theorem.

Theorem 4.7. *The metric space* $(C([a,b]),d)$ *is complete.*

Proof. We need to prove that if $\{f_n\} \subseteq C([a,b])$ is a Cauchy sequence in the d metric, then there is a continuous function f, to which it converges. First, note that is we fix $x, a \leq x \leq b$, then we have that

$$|f_n(x) - f_m(x)| \leq \sup\{|f_n(t) - f_m(t)| : a \leq t \leq b\} = d(f_n, f_m).$$

Thus, for each fixed x, the sequence of real numbers $\{f_n(x)\}$ is Cauchy and hence converges to a real number.

Thus, we may define a function by setting, $f(x) = \lim_n f_n(x)$. Doing this for each $a \leq x \leq b$, defines $f : [a,b] \to \mathbb{R}$.

We now wish to prove that f is a continuous function, so that f defines a point in the metric space $C([a,b])$ and then prove that the sequence $\{f_n\}$ converges to this point.

First to see that f is continuous at some point x_0, given $\epsilon > 0$, pick N so that $n, m > N$ implies that $d(f_n, f_m) < \epsilon/3$. Now fix an $n > N$, then we have that

$$|f(x) - f_n(x)| = \lim_m |f_m(x) - f_n(x)| \leq \epsilon/3.$$

Using that f_n is continuous, pick $\delta > 0$, so that $|x_0 - x| < \delta$ implies that $|f_n(x_0) - f_n(x)| < \epsilon/3$. Then for $|x_0 - x| < \delta$ we have that

$$\begin{aligned} |f(x_0) - f(x)| = |f(x_0) - f_n(x_0) + f_n(x_0) - f_n(x) + f_n(x) - f(x)| \leq \\ |f(x_0) - f_n(x_0)| + |f_n(x_0) - f_n(x)| + |f_n(x) - f(x)| < \epsilon. \end{aligned}$$

Thus, we have that f is continuous at x_0 and since x_0 was arbitrary, f is continuous.

Finally, we must show that $\lim_n d(f, f_n) = 0$. But to see this, given $\epsilon > 0$, take N as before and recall that we have already shown, that $|f(x) - f_n(x)| \leq \epsilon/3$, for any $n > N$. Hence, for $n > N$, we have that $d(f, f_n) \leq \sup\{|f(x) - f_n(x)| : a \leq x \leq b\} \leq \epsilon/3 < \epsilon$. □

We now have all the tools at our disposal needed to solve some IVP's.

Theorem 4.8. *Let* $h(x,y)$ *be a continuous function on* $[a,b] \times \mathbb{R}$ *and assume that* $|h(x,y_1) - h(x,y_2)| \leq K|y_1 - y_2|$ *with* $r = K(b-a) < 1$. *Then for any* y_0, *the initial value problem,* $f'(x) = h(x, f(x))$, $f(a) = y_0$ *has a unique solution on the interval* $[a,b]$.

Proof. Consider the mapping $\Phi : C([a,b]) \to C([a,b])$ defined by

$$\Phi(f)(x) = \int_a^x h(t, f(t))dt + y_0.$$

Given $f, g \in C([a,b])$, we have that

$$|\Phi(f)(x) - \Phi(g)(x)| = |\int_a^x h(t, f(t)) - h(t, g(t))dt| \leq$$
$$\int_a^x |h(t, f(t)) - h(t, g(t))|dt \leq \int_a^x K|f(t) - g(t)|dt \leq$$
$$\int_a^x Kd(f,g)dt = K(x-a)d(f,g) \leq rd(f,g).$$

Thus, Φ is a contraction mapping in the metric d. Since $(C([a,b]), d)$ is a complete metric space, by the contraction mapping principle, there exists a unique $f \in C([a,b])$, such that $\Phi(f) = f$. Earlier, we saw that an f was a solution to the IVP if and only if $\Phi(f) = f$. Thus, the solution not only exists but it is unique. □

The contraction mapping principle not only gives us the existence and uniqueness of solutions to the IVP, but it also gives us a method for approximating solutions and very nice bounds on the error of the approximation.

The above theorem is far from the most useful, because the conditions on $h(x,y)$ are too restrictive for most applications. For example, $h(x,y) = x^3y^2$ doesn't satisfy our hypothesis. For this reason you will seldom see it in a textbook. But it is does have the advantage of being the simplest to prove and we believe that it illustrates the key guiding principles of the proofs of more complicated results.

Problem 4.9. *Consider the IVP on the interval* $[0, \pi/2]$ *given by* $\frac{dy}{dx} = \frac{sin(xy)}{5}$, $y(0) = 1$. *Prove that the corresponding map* Φ *is a contraction mapping with* $r = \frac{\pi^2}{10}$ *(if you get a better bound, that is good!). Let* f *denote the unique solution to this IVP and let* $f_{n+1} = \Phi(f_n)$ *denote the sequence that one obtains starting with* f_1 *equal to the constant function* 1. *Give an estimate on* $d(f, f_n)$.

Chapter 5

Riemann and Riemann-Stieltjes Integration

In this chapter we develop the theory of the Riemann integral, which is the type of integration used in your calculus courses and we also introduce Riemann-Stieltjes integration which is widely used in probability, statistics and financial mathematics.

Given a closed interval $[a, b]$ by a **partition** of $[a, b]$ we mean a set $\mathcal{P} = \{x_0.x_1, , ..., x_{n-1}, x_n\}$ with $a = x_0 < x_1 < ... < x_n = b$. The **norm** or **width** of the partition is

$$\|\mathcal{P}\| = \max\{x_i - x_{i-1} : 1 \leq i \leq n\}.$$

Given two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ we say that $\mathcal{P}_2$ is a **refinement of** $\mathcal{P}_1$ or $\mathcal{P}_2$ **refines** $\mathcal{P}_1$ provided that as sets $\mathcal{P}_1 \subseteq \mathcal{P}_2$. Note that if $\mathcal{P}_2$ refines $\mathcal{P}_1$, then $\|\mathcal{P}_2\| \leq \|\mathcal{P}_1\|$.

Given a bounded function $f : [a, b] \to \mathbb{R}$ and a partition $\mathcal{P} = \{x_0, ..., x_n\}$ of $[a, b]$ for $i = 1, ..., n$, we set

$$M_i = \sup\{f(x) : x_{i-1} \leq x \leq x_i\}$$

and

$$m_i = \inf\{f(x) : x_{i-1} \leq x \leq x_i\}.$$

The **upper Riemann sum of f** given the partition $\mathcal{P}$ is the real number,

$$U(f, \mathcal{P}) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}),$$

and the **lower Riemann sum of f** is

$$L(f,\mathcal{P}) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}).$$

Note that if we hadn't assumed that f is a bounded function then some of the numbers M_i or m_i would have been infinite. This is the one reason that we can only define Riemann integrals for bounded functions.

By a **general Riemann sum of f given** $\mathcal{P}$, we mean a sum of the form

$$\sum_{i=1}^{n} f(x_i')(x_i - x_{i-1}),$$

where x_i' is any choice of points satisfying, $x_{i-1} \le x_i' \le x_i$, for $i = 1, ..., n$. Since $m_i \le f(x_i') \le M_i$, the upper and lower Riemann give an upper and lower bound for general Riemann sums, i.e.,

$$L(f,\mathcal{P}) \le \sum_{i=1}^{n} f(x_i')(x_i - x_{i-1}) \le U(f,\mathcal{P}).$$

In fact, since we can choose the points x_i' so that $f(x_i')$ is arbitrarily close to M_i, we see that $U(f,\mathcal{P})$ is actually the supremum of all general Riemann sums of f given $\mathcal{P}$. Similarly, by choosing the points x_i' so that $f(x_i')$ is arbitrarily close to m_i, we see that $L(f,\mathcal{P})$ is the infimum of all general Riemann sums.

Thus, if we want all general Riemann sums of a function to be "close" to a value that we wish to think of as the "integral of f", then it will be enough to study the "extreme" cases of the upper and lower sums.

Proposition 5.1. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function and let $\mathcal{P}_1$ and $\mathcal{P}_2$ be partitions of $[a,b]$ with $\mathcal{P}_2$ a refinement of $\mathcal{P}_1$. Then*

$$L(f,\mathcal{P}_1) \le L(f,\mathcal{P}_2) \le U(f,\mathcal{P}_2) \le U(f,\mathcal{P}_1).$$

Proof. We have already observed that for any partition, $L(f,\mathcal{P}) \le U(f,\mathcal{P})$. So we only need to consider the other two inequalities.

It is enough to prove these inequalities in the case that $\mathcal{P}_2$ is obtained from $\mathcal{P}_1$ by adding just one extra point. Because if we do this then $\mathcal{P}_2$ can be obtained from $\mathcal{P}_1$ by successively adding a point at a time and in each case the inequalities "move" in the right direction.

So let us set $\mathcal{P}_1 = \{x_0, ..., x_n\}$ so that for some j, $\mathcal{P}_2 = \{x_0, ..., x_{j-1}, \hat{x}, x_j, ..., x_n\}$ where $x_{j-1} < \hat{x} < x_j$.

We start with the case of the upper sums. Note that most of the terms in the sum for $U(f,\mathcal{P}_1)$ and $U(f,\mathcal{P}_2)$ will be equal, except that the one term $M_j(x_j - x_{j-1})$ in the sum for $U(f,\mathcal{P}_1)$ will be replaced by two terms in the sum for $U(f,\mathcal{P}_2)$. These two new terms are $M_j'(\hat{x} - x_{j-1}) + M_j''(x_j - \hat{x})$, where

$$M_j' = \sup\{f(x) : x_{j-1} \le x \le \hat{x}\} \text{ and } M_j'' = \sup\{f(x) : \hat{x} \le x \le x_j\}.$$

Since $[x_{j-1}, \hat{x}] \subseteq [x_{j-1}, x_j]$ and $[\hat{x}, x_j] \subseteq [x_{j-1}, x_j]$, we have that $M_j' \le M_j$ and $M_j'' \le M_j$. Hence, the two terms appearing in $U(f,\mathcal{P}_2)$, satisfy

$$M_j'(\hat{x} - x_{j-1}) + M_j''(x_j - \hat{x}) \le M_j(\hat{x} - x_{j-1}) + M_j(x_j - \hat{x}) = M_j(x_j - x_{j-1}),$$

which is the one term appearing in $U(f,\mathcal{P}_1)$. Thus, $U(f,\mathcal{P}_2) \le U(f,\mathcal{P}_1)$.

The proof for the lower sums is similar, except that since we are dealing with infima, we will have that

$$m_j' = \inf\{f(x) : x_{j-1} \le x \le \hat{x}\} \ge m_j \text{ and } m_j'' = \inf\{f(x) : \hat{x} \le x \le x_j\} \ge m_j.$$

$\square$

Definition 5.2. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then the* **upper Riemann integral of f** *is the number*

$$\overline{\int_a^b} f(x)dx = \inf\{U(f,\mathcal{P}) : \mathcal{P} \text{ a partition of } [a,b]\}.$$

The **lower Riemann integral of f** *is the number*

$$\underline{\int_a^b} f(x)dx = \sup\{L(f,\mathcal{P}) : \mathcal{P} \text{ a partition of } [a,b]\}.$$

We say that f is **Riemann integrable** *when these two numbers are equal and in this case we define the* **Riemann integral of f** *to be*

$$\int_a^b f(x)dx = \overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx.$$

To help cement these definitions, let us consider the function, $f : [a,b] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & x \text{ rational} \\ 0 & x \text{ irrational,} \end{cases}$$

then for any partition $\mathcal{P}$ we will have that $M_i = 1$ and $m_i = 0$ for every i. Hence, $U(f, \mathcal{P}) = (b - a)$ and $L(f, \mathcal{P}) = 0$. Thus,

$$\overline{\int_a^b} f(x)dx = b - a \text{ and } \underline{\int_a^b} f(x)dx = 0.$$

In particular, f is not Riemann integrable.

The following helps to explain the terms “upper” and “lower”.

Proposition 5.3. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function and let $\mathcal{P}_1$ and $\mathcal{P}_2$ be any two partitions of $[a, b]$. Then $L(f, \mathcal{P}_1) \leq U(f, \mathcal{P}_2)$ and $\underline{\int_a^b} f(x)dx \leq \overline{\int_a^b} f(x)dx$.*

Proof. Let $\mathcal{P}_3$ be the partition of $[a, b]$ that as a set is $\mathcal{P}_3 = \mathcal{P}_1 \cup \mathcal{P}_2$. Then $\mathcal{P}_3$ is a refinement of $\mathcal{P}_1$ and of $\mathcal{P}_2$. So by the above result,

$$L(f, \mathcal{P}_1) \leq L(f, \mathcal{P}_3) \leq U(f, \mathcal{P}_3) \leq U(f, \mathcal{P}_2).$$

Now we have that

$$\underline{\int_a^b} f(x)dx = \sup\{L(f, \mathcal{P}_1) : \mathcal{P}_1 \text{ is a partition of } [a, b]\} \leq U(f, \mathcal{P}_2),$$

and so

$$\underline{\int_a^b} f(x)dx \leq \inf\{U(f, \mathcal{P}_2) : \mathcal{P}_2 \text{ is a partition of } [a, b]\} = \overline{\int_a^b} f(x)dx.$$

□

The following gives an important means of determining if a function is Riemann integrable.

Theorem 5.4. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then f is Riemann integrable if and only if for every $\epsilon > 0$ there exists a partition $\mathcal{P}$ such that $U(f, \mathcal{P}) - L(f, \mathcal{P}) < \epsilon$.*

Proof. Assuming that the latter condition is met, for any $\epsilon > 0$ and $\mathcal{P}$ as above, we have

$$L(f, \mathcal{P}) \leq \underline{\int_a^b} f(x)dx \leq \overline{\int_a^b} f(x)dx \leq U(f, \mathcal{P}),$$

which implies that

$$0 \leq \overline{\int_a^b} f(x)dx - \underline{\int_a^b} f(x)dx \leq U(f,\mathcal{P}) - L(f,\mathcal{P}) < \epsilon.$$

Since $\epsilon > 0$ was arbitrary, the lower and upper Riemann integrals must be equal.

Conversely, assume that f is Riemann integrable and that $\epsilon > 0$ is given. Since the upper integral is an infimum, we may choose a partition $\mathcal{P}_1$ so that

$$\overline{\int_a^b} f(x)dx \leq U(f,\mathcal{P}_1) < \overline{\int_a^b} f(x)dx + \epsilon/2.$$

Similarly, we may choose a partition $\mathcal{P}_2$ so that

$$\underline{\int_a^b} f(x)dx - \epsilon/2 < L(f,\mathcal{P}_2) \leq \underline{\int_a^b} f(x)dx.$$

Let $\mathcal{P}_3 = \mathcal{P}_1 \cup \mathcal{P}_2$, then

$$U(f,\mathcal{P}_3) - L(f,\mathcal{P}_3) \leq U(f,\mathcal{P}_1) - L(f,\mathcal{P}_2) < \left(\overline{\int_a^b} f(x)dx + \epsilon/2\right) - \left(\underline{\int_a^b} f(x)dx - \epsilon/2\right) = \epsilon$$

since the upper and lower integrals are equal. Thus, we have found a partition satisfying our criteria. □

Definition 5.5. *We say that a bounded function $f : [a,b] \to \mathbb{R}$ satisfies the* **Riemann integrability criterion** *provided that for every $\epsilon > 0$, there exists a partition $\mathcal{P}$, such that $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \epsilon$.*

Thus, the above theorem says that a bounded function is Riemann integrable if and only if it satisfies the Riemann integrability criterion.

Theorem 5.6. *Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then f is Riemann integrable.*

Proof. Since f is continuous and $[a,b]$ is compact, we have that f is bounded and f is uniformly continuous. We will prove that f meets the Riemann integrability criterion.

Given any $\epsilon > 0$, since f is uniformly continuous, there is a $\delta > 0$, so that whenever $|x-y| < \delta$ then $|f(x) - f(y)| < \epsilon/(b-a)$. Now let $\mathcal{P} = \{a =$

$x_0, ..., x_n = b\}$ be any partition of $[a, b]$ with $\|\mathcal{P}\| < \delta$. Then for each i, we have that $x_{i-1} \leq z, y \leq x_i$ implies that $|f(z) - f(y)| < \epsilon/(b - a)$. Since f is continuous, we also have that there are points z_i and y_i in $[x_{i-1}, x_i]$ with $M_i = f(z_i)$ and $m_i = f(y_i)$. Hence, $0 \leq M_i - m_i < \epsilon/(b - a)$. This implies that

$$0 \leq U(f, \mathcal{P}) - L(f, \mathcal{P}) = \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) < \sum_{i=1}^{n} [\epsilon/(b-a)](x_i - x_{i-1}) = \epsilon.$$

Thus, the Riemann integrability criterion is met by f. □

For the following problem, we consider the interval $[0, 1]$ and let $\mathcal{P}_n = \{0, 1/n, ..., (n-1)/n, 1\}$ denote the partition on $[0, 1]$ into n subintervals each of length $1/n$. Thus, $\|\mathcal{P}_n\| = 1/n$. We will also use the summation formula,

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}.$$

Problem 5.7. *Let $f(x) = x$. Compute $U(f, \mathcal{P}_n)$ and $L(f, \mathcal{P}_n)$. Use these formulas and the Riemann integrability criterion to prove that f is Riemann integrable on $[0, 1]$ and to prove that $\int_0^1 x dx = 1/2$.*

Problem 5.8. *Let $a \leq c < d \leq b$ and let $f : [a, b] \to \mathbb{R}$ be the function defined by*

$$f(x) = \begin{cases} 0 & a \leq x \leq c \\ 1 & c < x < d \\ 0 & d \leq x \leq b \end{cases}.$$

Prove that f is Riemann integrable on $[a, b]$ and that $\int_a^b f(x)dx = (d - c)$.

5.1 The Riemann-Stieltjes Integral

The Riemann-Stieltjes integral is a slight generalization of the Riemann integral. The new ingredient in Riemann-Stieltjes integration is a function,

$$\alpha : [a, b] \to \mathbb{R}$$

that is increasing, i.e., $x \leq y$ implies that $\alpha(x) \leq \alpha(y)$. It is best to think of α as a function that measures a new "length" of subintervals by setting the length of a subinterval $[x_{i-1}, x_i]$ equal to $\alpha(x_i) - \alpha(x_{i-1})$. One case where this concept arises is if we imagine that we have a piece of wire of varying density stretched from a to b and $\alpha(x_i) - \alpha(x_{i-1})$ represents the weight of the section of wire from x_{i-1} to x_i.

Given a bounded function $f : [a, b] \to \mathbb{R}$ the Riemann-Stieltjes integral is denoted

$$\int_a^b f d\alpha,$$

and it is designed to also define a " signed area" under the graph of f but now if we want the area of a rectangle to be the length of the base times the height, then a rectangle from x_{i-1} to x_i of height h should have area

$$h(\alpha(x_i) - \alpha(x_{i-1})).$$

Thus, given a bounded function f an increasing function α, a partition $\mathcal{P} = \{a = x_0, ..., x_n = b\}$, the numbers $M_i = \sup\{f(x) : x_{i-1} \leq x \leq x_i\}$ and $m_i = \inf\{f(x) : x_{i-1} \leq x \leq x_i\}$, we are led to define the **upper Riemann-Stieltjes sum** as

$$U(f, \mathcal{P}, \alpha) = \sum_{i=1}^{n} M_i(\alpha(x_i) - \alpha(x_{i-1}))$$

and the **lower Riemann-Stieltjes sum** as

$$L(f, \mathcal{P}, \alpha) = \sum_{i=1}^{n} m_i(\alpha(x_i) - \alpha(x_{i-1})).$$

The **upper Riemann-Stieltjes integral of f with respect to α** is then defined to be

$$\overline{\int_a^b} f d\alpha = \inf\{U(f, \mathcal{P}, \alpha) : \mathcal{P} \text{ is a partition of } [a, b]\}.$$

Similarly, the **lower Riemann-Stieltjes integral of f with respect to α** is defined to be

$$\underline{\int_a^b} f d\alpha = \sup\{L(f, \mathcal{P}, \alpha) : \mathcal{P} \text{ is a partition of } [a, b]\}.$$

When

$$\overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha,$$

then we say that f **is Riemann-Stieltjes integrable with respect to** α and we let

$$\int_a^b f d\alpha$$

denote this common value.

We repeat the key facts about Riemann-Stieltjes integration below. Since the proofs are almost identical to the corresponding proofs in the case of Riemann integration, we omit the details.

Proposition 5.9. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function, let $\alpha : [a, b] \to \mathbb{R}$ be an increasing function and let $\mathcal{P}_1$ and $\mathcal{P}_2$ be partitions of $[a, b]$ with $\mathcal{P}_2$ a refinement of $\mathcal{P}_1$. Then*

$$L(f, \mathcal{P}_1, \alpha) \leq L(f, \mathcal{P}_2, \alpha) \leq U(f, \mathcal{P}_2, \alpha) \leq U(f, \mathcal{P}_1, \alpha).$$

Proposition 5.10. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function, let $\alpha : [a, b] \to \mathbb{R}$ be increasing and let $\mathcal{P}_1$ and $\mathcal{P}_2$ be any two partitions of $[a, b]$. Then $L(f, \mathcal{P}_1, \alpha) \leq U(f, \mathcal{P}_2, \alpha)$ and $\underline{\int}_a^b f(x) d\alpha \leq \overline{\int}_a^b f(x) d\alpha$.*

Theorem 5.11. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function and let $\alpha : [a, b] \to \mathbb{R}$ be increasing. Then f is Riemann-Stieltjes integrable with respect to α if and only if for every $\epsilon > 0$ there exists a partition $\mathcal{P}$ such that $U(f, \mathcal{P}, \alpha) - L(f, \mathcal{P}, \alpha) < \epsilon$.*

Definition 5.12. *Given an increasing function $\alpha : [a, b] \to \mathbb{R}$, we say that a bounded function $f : [a, b] \to \mathbb{R}$ satisfies the* **Riemann-Stieltjes integrability criterion with respect to** α *provided that for every $\epsilon > 0$, there exists a partition $\mathcal{P}$, such that $U(f, \mathcal{P}, \alpha) - L(f, \mathcal{P}, \alpha) < \epsilon$.*

Thus, the above theorem says that a bounded function is Riemann-Stieltjes integrable with respect to α if and only if it satisfies the Riemann-Stieltjes integrability criterion with respect to α.

Theorem 5.13. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function and let $\alpha : [a, b] \to \mathbb{R}$ be an increasing function. Then f is Riemann-Stieltjes integrable with respect to α.*

One of the main motivations for Riemann-Stieltjes integration comes from the concept of a *cumulative distribution function of a random variable.* To understand the Riemann-Stieltjes integral, one need not understand any probability theory, but we introduce these ideas from probability here, via an example, in order to motivate the desire for the Riemann-Stieltjes integral.

For a *probability space,* suppose that we consider the flip of a "biased" coin, so tht the probability of heads(H) is p and the probability of tails(T) is $1-p$ with $0 < p < 1$. When $p = 1/2$, we think of the coin as a "fair" coin. A *random variable* would then be a function that assigned a real number to each of these possible outcomes. For example, we could define a random variable X by setting $X(H) = 1$, $X(T) = 3$. If we let $Prob(X = x)$ denote the probability that X is equal to the real number x, then we would have $Prob(X = 1) = p$ and $Prob(X = 3) = 1 - p$. The *cumulative distribution function of* X*,* is the function $Prob(X \leq x)$. So in our case,

$$Prob(X \leq x) = \begin{cases} 0 & x < 1 \\ p & 1 \leq x < 3 \\ 1 & 3 \leq x \end{cases} .$$

So our cumulative distribution function has two "jump" discontinuities, the first of size p ocurring at 1 and the second of size $1 - p$ occurring at 3.

More generally, if one has a probability space, a random variable X, and we let $\alpha(x) = Prob(X \leq x)$, then α will be an increasing function, i.e., one for which we can consider Riemann-Stieltjes integrals. Much work in probability and its applications, e.g., financial math, is concerned with computing these Riemann-Stieltjes integrals for various functions f in the case that α is the cumulative distribution function of a random variable.

Returning to our example, if for any c we let

$$J_c(x) = \begin{cases} 0 & x < c \\ 1 & c \leq x \end{cases} ,$$

then we can also write $Prob(X \leq x) = pJ_1(x) + (1-p)J_3(x)$.

These simple "jump" functions are useful for expressing the cumulative distributions of many random variables. For example, if we consider one roll of a "fair" die, so that each side has probability $1/6$ of facing up and we let X be the random variable that simply gives the number that is facing up, then

$$Prob(X \leq x) = 1/6[J_1(x) + J_2(x) + J_3(x) + J_4(x) + J_5(x) + J_6(x)].$$

For this reason we want to take a careful look at what Riemann-Stieltjes integration is when $\alpha(x) = hJ_c(x)$, for some real number c and $h > 0$.

Definition 5.14. *Let* $f : [a, b] \to \mathbb{R}$ *and let* $a < c \leq b$*. If there is a number* L *so that for every* $\epsilon > 0$*, there exists* $\delta > 0$*, so that when* $c - \delta < x < c$*,* *we have* $|f(x) - L| < \epsilon$, *then we say that* **L is the limit from the left of** f **at** c. *We write* $\lim_{x\to c^-} f(x) = L$ *and say that* f **is continuous from the left at** c *provided that* $\lim_{x\to c^-} f(x) = f(c)$. *When* $a \le c < b$, *then the* **limit from the right of** f **at** c *is defined similarly and is denoted* $\lim_{x\to c^+} f(x)$. *We say that* f **is continuous from the right at** c *provided that* $\lim_{x\to c^+} f(x) = f(c)$.

Theorem 5.15. *Let* $a < c \le b$, *let* $h > 0$, *let* $\alpha(x) = hJ_c(x)$, *and let* $f : [a,b] \to \mathbb{R}$ *be a bounded function. Then* f *is Riemann-Stieltjes integrable with respect to* α *if and only if* f *is continuous from the left at* c. *In this case,*

$$\int_a^b f d\alpha = hf(c).$$

Proof. First we assume that f is continuous from the left at c and prove that f satisfies the Riemann-Stieltjes integrability criterion. So let $\epsilon > 0$, be given. Since $\lim_{x\to c^-} f(x) = f(c)$, there is $\delta > 0$, so that $c - \delta < x < c$ implies that $|f(x) - f(c)| < \frac{\epsilon}{3h}$.

Now consider the partition, $\mathcal{P} = \{a = x_0, x_1 = c - \delta/2, x_3 = c, x_4 = b\}$. Then we have that

$$U(f,\mathcal{P},\alpha) - L(f,\mathcal{P},\alpha) = \sum_{i=1}^{4}(M_i - m_i)(\alpha(x_i) - \alpha(x_{i-1}) = (M_2 - m_2)h,$$

since $\alpha(x_i) - \alpha(x_{i-1}) = 0$ when $i \ne 2$. Now for $x_1 = c - \delta/2 \le x \le c = x_2$, we have that $|f(x) - f(c)| < \frac{\epsilon}{2h}$, so that $f(c) - \frac{\epsilon}{3h} \le f(x) \le f(c) + \frac{\epsilon}{3h}$. Hence, $f(c) - \frac{\epsilon}{3h} \le m_2 \le M_2 \le f(c) + \frac{\epsilon}{3h}$, which implies that $0 \le M_2 - m_2 \le 2\frac{\epsilon}{3h}$. Thus, $U(f,\mathcal{P},\alpha) - L(f,\mathcal{P},\alpha) \le (M_2 - m_2)h \le \frac{2\epsilon}{3} < \epsilon$, and we have shown that the Riemann-Stieltjes integrability criterion is met.

Also, note that these inequalities show that

$$(f(c) - \frac{\epsilon}{2h})h \le m_i h = L(f,\mathcal{P},\alpha) \le \int_a^b f d\alpha \le U(f,\mathcal{P},\alpha) \le (f(c) + \frac{\epsilon}{2h})h$$

which shows that $|f(c)h - \int_a^b f d\alpha| \le \frac{\epsilon}{2} < \epsilon$. So it follows that $\int_a^b f d\alpha = f(c)h$.

Now we must prove that if f satisfies the Riemann-Stieltjes integrability criterion, then f is continuous from the left at c, i.e., $\lim_{x\to c^-} f(x) = f(c)$.

To this end fix $\epsilon > 0$ and pick a partition $\mathcal{P}$ so that $U(f,\mathcal{P},\alpha) - L(f,\mathcal{P},\alpha) < h\epsilon$. Let $\mathcal{P}_2$ be the partition that we obtain from $\mathcal{P}$ by including the point c. Since $\mathcal{P}_2$ is a refinement of $\mathcal{P}$, we will have that

$$L(f,\mathcal{P},\alpha) \le L(f,\mathcal{P}_2,\alpha) \le U(f,\mathcal{P}_2,\alpha) \le U(f,\mathcal{P},\alpha),$$

and hence, $U(f, \mathcal{P}_2, \alpha) - L(f, \mathcal{P}_2, \alpha) \leq U(f, \mathcal{P}, \alpha) - L(f, \mathcal{P}, \alpha) < h\epsilon$. Now our partition has the form $\mathcal{P}_2 = \{a = x_0 < ... < x_{j-1} < x_j = c < ... < x_n = b\}$, for some j. Since $\alpha(x_i) - \alpha(x_{i-1}) = 0$ for all $i \neq j$, we have that $(M_j - m_j)h = U(f, \mathcal{P}_2, \alpha) - L(f, \mathcal{P}_2, \alpha) < h\epsilon$ and $M_j - m_j < \epsilon$. Thus, for $x_{j-1} \leq x \leq c$, we have that $m_j \leq f(c) \leq M_j$ and $m_j \leq f(x) \leq M_j$, so that $|f(x) - f(c)| \leq M_j - m_j < \epsilon$.

Thus, if we let $\delta = c - x_{j-1}$, then for $x_{j-1} = c - \delta < x < c$, we have that $|f(x) - f(c)| < \epsilon$. Since $\epsilon > 0$ was arbitrary, $\lim_{x \to c^-} f(x) = f(c)$. $\square$

In probability, two events are called independent if the probability of both occurring is the product of the probabilities of each occurring. For example, suppose that we flip a biased coin with $Prob(H) = p, Prob(T) = 1-p$ twice, so that the possible outcomes are $\{HH, HT, TH, TT\}$. If we assume that the flips are independent then the outcomes would have probabilities,

$$\begin{gathered} Prob(HH) = Prob(H)Prob(H) = p^2, \\ Prob(HT) = Prob(TH) = Prob(H)Prob(T) = p(1-p), \\ Prob(TT) = Prob(T)Prob(T) = (1-p)^2. \end{gathered}$$

Problem 5.16. *Suppose that we flip our biased coin twice, assume that the flips are independent and let our random variable be the sum of our earlier random variable, so that* $X(HH) = X(H) + X(H) = 2, X(HT) = X(TH) = 1 + 3, X(TT) = 3 + 3$. *Find the cumulative distribution function,* $\alpha(t) = Prob(X \leq t)$.

Problem 5.17. *Suppose that we flip a fair coin and roll a fair die and asume that these events are independent, so that each possible outcome has probbility* $(1/2)(1/6) = 1/12$. *Let* X *be the random variable that adds* $+1$ *to the number on the die when a* H *is flipped and* $+3$ *to the number on the die when a* T *is flipped. Find the cumulative distribution function of* X.

Problem 5.18. *Let* $\alpha : [a, b] \to \mathbb{R}$ *be an increasing function and let* $f : [a, b] \to \mathbb{R}$ *be a bounded function. If* $m = \inf\{f(x) : a \leq x \leq b\}$ *and* $M = \sup\{f(x) : a \leq x \leq b\}$, *then* $m(\alpha(b) - \alpha(a)) \leq \underline{\int}_a^b f d\alpha \leq \overline{\int}_a^b f d\alpha \leq M(\alpha(b) - \alpha(a))$.

Problem 5.19. *Write out the proof of Theorem 5.13. You may use the earlier stated results in your proof without proving them.*

5.2 Properties of the Riemann-Stieltjes Integral

Before trying to compute many integrals, it will be helpful to have proven some basic properties of these integrals.

Theorem 5.20. *Let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function, let $f, g : [a,b] \to \mathbb{R}$ be bounded functions that are Riemann-Stieltjes integrable with respect to α and let $c \in \mathbb{R}$ be a constant. Then:*

1. *cf is Riemann-Stieltjes integrable with respect to α and*
$$\int_a^b cf d\alpha = c\int_a^b f d\alpha,$$
2. *$f+g$ is Riemann-Stieltjes integrable with respect to α and*
$$\int_a^b (f+g)d\alpha = \int_a^b f d\alpha + \int_a^b g d\alpha,$$
3. *fg is Riemann-Stieltjes integrable with respect to α.*

Proof. To prove 1), first consider the case that $c \geq 0$. Then we have that for any partition, $L(cf, \mathcal{P}, \alpha) = cL(f, \mathcal{P}, \alpha)$ and $U(cf, \mathcal{P}, \alpha) = cU(f, \mathcal{P}, \alpha)$. Thus,

$$\underline{\int_a^b} cf d\alpha = \sup\{L(cf, \mathcal{P}, \alpha)\} =$$
$$\sup\{cL(f, \mathcal{P}, \alpha)\} = c\underline{\int_a^b} f d\alpha = c\int_a^b f d\alpha =$$
$$c\overline{\int_a^b} f d\alpha = \inf\{cU(f, \mathcal{P}, \alpha\} =$$
$$\inf\{U(cf, \mathcal{P}, \alpha)\} = \overline{\int_a^b} f d\alpha,$$

which shows that the upper and lower integrals are equal for cf, so that cf is Riemann-Stieltjes integrable with respect to α and shows that they are both equal to $c\int_a^b f d\alpha$. Thus, 1) is proven in the case that $c \geq 0$.

When $c < 0$, recall that for any bounded set S of real numbers, $\sup\{cs : s \in S\} = c\inf\{s : s \in S\}$ and $\inf\{cs : s \in S\} = c\sup\{s : s \in S\}$. Thus, for any interval, $\sup\{cf(x) : x_{i-1} \leq x \leq x_i\} = c\inf\{f(x) : x_{i-1} \leq x \leq x_i\}$

and $\inf\{cf(x) : x_{i-1} \leq x \leq x_i\} = c\sup\{f(x) : x_{i-1} \leq x \leq x_i\}$. From this it follows that $L(cf, \mathcal{P}, \alpha) = cU(f, \mathcal{P}, \alpha)$ and $U(cf, \mathcal{P}, \alpha) = cL(f, \mathcal{P}, \alpha)$.

Thus,

$$\begin{aligned}
\underline{\int_a^b} cf d\alpha = \sup\{L(cf, \mathcal{P}, \alpha)\} = \\
\sup\{cU(f, \mathcal{P}, \alpha)\} = c\inf\{U(f, \mathcal{P}, \alpha)\} = \\
c\overline{\int_a^b} f d\alpha = c\int_a^b f d\alpha = c\underline{\int_a^b} f d\alpha = \\
c\sup\{L(f, \mathcal{P}, \alpha)\} = \inf\{cL(f, \mathcal{P}, \alpha)\} = \\
\inf\{U(cf, \mathcal{P}, \alpha)\} = \overline{\int_a^b} f d\alpha,
\end{aligned}$$

and again we have that cf is Riemann-Stieltjes integrable with respect to α and the equality of the integrals.

Now we prove 2). Note that for any interval, we have that

$$\sup\{f(x)+g(x) : x_{i-1} \leq x \leq x_i\} \leq \sup\{f(x) : x_{i-1} \leq x \leq x_i\}+\sup\{g(x) : x_{i-1} \leq x \leq x_i\}$$

from which it follows that for any partition, $U(f + g, \mathcal{P}, \alpha) \leq U(f, \mathcal{P}, \alpha) + U(g, \mathcal{P}, \alpha)$. Similarly,

$$\inf\{f(x)+g(x) : x_{i-1} \leq x \leq x_i\} \geq \inf\{f(x) : x_{i-1} \leq x \leq x_i\}+\inf\{g(x) : x_{i-1} \leq x \leq x_i\}$$

and it follows that $L(f + g, \mathcal{P}, \alpha) \geq L(f, \mathcal{P}, \alpha) + L(g, \mathcal{P}, \alpha)$.

Given any $\epsilon > 0$, choose partitions $\mathcal{P}_1, \mathcal{P}_2$ so that $\overline{\int_a^b} f d\alpha \leq U(f, \mathcal{P}_1, \alpha) < \overline{\int_a^b} f d\alpha + \epsilon/2$ and $\overline{\int_a^b} g d\alpha \leq U(g, \mathcal{P}_2, \alpha) < \overline{\int_a^b} g d\alpha + \epsilon/2$. If we let $\mathcal{P}_3 = \mathcal{P}_1 \cup \mathcal{P}_2$ denote their common refinement, then

$$\begin{aligned}
\overline{\int_a^b} (f + g) d\alpha \leq U(f + g, \mathcal{P}_3, \alpha) \leq U(f, \mathcal{P}_3, \alpha) + U(g, \mathcal{P}_3, \alpha) \leq \\
U(f, \mathcal{P}_1, \alpha) + U(g, \mathcal{P}_2, \alpha) < \\
\overline{\int_a^b} f d\alpha + \overline{\int_a^b} g d\alpha + \epsilon.
\end{aligned}$$

since $\epsilon > 0$ was arbitrary, we have that

$$\overline{\int_a^b} (f + g) d\alpha \leq \overline{\int_a^b} f d\alpha + \overline{\int_a^b} g d\alpha.$$

A similar calculation shows that

$$\underline{\int_a^b} (f+g)d\alpha \geq \underline{\int_a^b} f d\alpha + \underline{\int_a^b} g d\alpha.$$

Since the upper bound and lower bound are both equal to $\int_a^b f d\alpha + \int_a^b g d\alpha$ we have that these inequalities must be equalities and 2) follows.

Before proving 3) we first make a number of observations to reduce the work. First suppose that we only prove that if a function h is Riemann-Stieltjes integrable with respect to α, then h^2 is Riemann-Stieltjes integrable. Then, since f and g are both integrable, by part 2), $f+g$ is integrable and hence we would have $(f+g)^2 = f^2 + 2fg + g^2$, f^2 and g^2 are all integrable. Applying 1) and 2) again, we would have that $(f+g)^2 - f^2 - g^2 = 2fg$ is integrable and hence by 1), fg is integrable. Thus, it will be enough to prove that h integrable implies that h^2 is integrable.

Next, we claim that to show that squares of integrable functions are integrable, it will be enough to show that squares of non-negative integrable are integrable. To see this, given any h that is integrable, we know that h is bounded, so there is some constant $c > 0$ so that $h + c \geq 0$. Since constants are continuous functions and continuous functions are integrable, $h+c$ is integrable. Thus, if we know that squares of non-negative integrable functions are integrable, then we will have that $(h+c)^2 = h^2 + 2ch + c^2$ is integrable. But $2ch+c^2$ is integrable by 1) and 2), so $h^2 = (h+c)^2 - 2ch - c^2$ is integrable.

Thus, it remains to show that if $h \geq 0$ is Riemann-Stieltjes integrable with respect to α, then h^2 is also Riemann-Stieltjes integrable with respect to α. To this end we show that h^2 satisfies the Riemann-Stieltjes integrability criterion. So let $\epsilon > 0$, be given and let $K = \sup\{h(x) : a \leq x \leq b\}$. If $h = 0$, then the result is clearly true, so we may assume that $h \neq 0$ and so $K > 0$.

Since h is Riemann-Stieltjes integrable, it satisfies the criterion and hence there exists a partition $\mathcal{P}$, so that $U(h, \mathcal{P}, \alpha) - L(h, \mathcal{P}, \alpha) < \epsilon/2K$. Let $\mathcal{P} = \{a = x_0, ..., x_n = b\}$, let $M_i = \sup\{h(x) : x_{i-1} \leq x \leq x_i\}$ and let $m_i = \inf\{h(x) : x_{i-1} \leq x \leq x_i\}$. Since $h(x) \geq 0$, we have that $\sup\{h(x)^2 : x_{i-1} \leq x \leq x_i\} = M_i^2$ and $\inf\{h(x)^2 : x_{i-1} \leq x \leq x_i\} = m_i^2$. Thus, we have that

$$U(h^2,\mathcal{P},\alpha)-L(h^2,\mathcal{P},\alpha)=\sum_{i=1}^{n}(M_i^2-m_i^2)(\alpha(x_i)-\alpha(x_{i-1}))=$$
$$\sum_{i=1}^{n}(M_i-m_i)(M_i+m_i)(\alpha(x_i)-\alpha(x_{i-1}))\leq 2K\sum_{i=1}(M_i-m_i)(\alpha(x_i)-\alpha(x_{i-1}))=$$
$$2K(U(h,\mathcal{P},\alpha)-L(h,\mathcal{P},\alpha))<\epsilon$$

and we have shown that h^2 satisfies the Riemann-Stieltjes integrability criterion with respect to α. □

Proposition 5.21. *Let $\alpha_1,\alpha_2:[a,b]\to\mathbb{R}$ be increasing functions, let $c>0$ be a constant and let $f:[a,b]\to\mathbb{R}$ be a bounded function. Then*

$$\underline{\int_a^b} fd(c\alpha_1+\alpha_2)=c\underline{\int_a^b} fd\alpha_1+\underline{\int_a^b} fd\alpha_2$$

and

$$\overline{\int_a^b} fd(c\alpha_1+\alpha_2)=c\overline{\int_a^b} fd\alpha_1+\overline{\int_a^b} fd\alpha_2.$$

Proof. First note that for any partition $\mathcal{P}=\{a=x_0,...,x_n=b\}$, we have that

$$U(f,\mathcal{P},c\alpha_1+\alpha_2)=\sum_{i=1}^{n}M_i[(c\alpha_1(x_i)+\alpha_2(x_i))-(c\alpha_1(x_{i-1})+\alpha_2(x_{i-1}))]=$$
$$cU(f,\mathcal{P},\alpha_1)+U(f,\mathcal{P},\alpha_2).$$

Thus,

$$\overline{\int_a^b} fd(c\alpha_1+\alpha_2)=\inf\{cU(f,\mathcal{P},\alpha_1)+U(f,\mathcal{P},\alpha_2)\}\geq$$
$$\inf\{cU(f,\mathcal{P},\alpha_1)\}+\inf\{U(f,\mathcal{P},\alpha_2)\}=$$
$$c\overline{\int_a^b} fd\alpha_1+\overline{\int_a^b} fd\alpha_2.$$

On the other hand, given any $\epsilon>0$, there is a partition $\mathcal{P}_1$ so that

$$U(f,\mathcal{P}_1,\alpha_1)<\overline{\int_a^b} fd\alpha_1+\epsilon/2c$$

and a partition $\mathcal{P}_2$ so that

$$U(f, \mathcal{P}_2, \alpha_2) < \overline{\int_a^b} f d\alpha_2 + \epsilon/2.$$

Let $\mathcal{P}_3 = \mathcal{P}_1 \cup \mathcal{P}_2$, then

$$\begin{aligned}\overline{\int_a^b} fd(c\alpha_1 + \alpha_2) \leq U(f, \mathcal{P}_3, c\alpha_1 + \alpha_2) = \\ cU(f, \mathcal{P}_3, \alpha_1) + U(f, \mathcal{P}_3, \alpha_2) \leq \\ cU(f, \mathcal{P}_1, \alpha_1) + U(f, \mathcal{P}_2, \alpha_2) < \\ c[\overline{\int_a^b} fd\alpha_1 + \epsilon/2c] + [\overline{\int_a^b} fd\alpha_2 + \epsilon/2].\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, we have that

$$\overline{\int_a^b} fd(c\alpha_1 + \alpha_2) \leq c\overline{\int_a^b} fd\alpha_1 + \overline{\int_a^b} fd\alpha_2$$

and these two inequalties for the upper Riemann-Stieltjes, prove equality of these integrals. The proof for the lower Riemann-Stieltjes integrals is similar. $\square$

Proposition 5.22. *Let $\alpha_1, \alpha_2 : [a, b] \to \mathbb{R}$ be increasing, let $c > 0$ and let $f : [a, b] \to \mathbb{R}$ be a bounded function. If f is Riemann-Stieltjes integrable with respect to α_1 and α_2, then f is Riemann-Stieltjes integrable with respect to $c\alpha_1 + \alpha_2$ and*

$$\int_a^b fd(c\alpha_1 + \alpha_2) = c\int_a^b fd\alpha_1 + \int_a^b fd\alpha_2.$$

Proof. The previous result shows that the upper and lower Riemann-Stieltjes of f with respect to $c\alpha_1 + \alpha_2$ are equal and are given by the above formula. $\square$

Theorem 5.23. *Let $a < c_1 < .... < c_n \leq b$, let $h_i > 0, i = 1, ..., n$ and let $\alpha = h_1 J_{c_1} + \cdots h_n J_{c_n}$. If $f : [a, b] \to \mathbb{R}$ is a bounded function that is continuous from the left at each $c_i, i = 1, ..., n$, then f is Riemann-Stieltjes integrable with respect to α and*

$$\int_a^b fd\alpha = h_1 f(c_1) + \cdots h_n f(c_n).$$

Proof. Apply Theorem 5.15 together with the above result. □

Problem 5.24. *Let $a < 1$ and $3 < b$ and let α be the cumulative distribution function for the random variable on page 69 for one flip of a biased coin. Compute $\mu = \int_a^b t d\alpha(t)$, and $\sigma^2 = \int_a^b (t-\mu)^2 d\alpha(t)$.*

This two values are known as the "mean" and "variance" of the random variable. The square root of the variance, σ, is called the "standard deviation".

Problem 5.25. *Let $a < 2$ and $6 < b$. Compute the mean and variance of the random variable of Problem 5.16.*

Problem 5.26. *Let $a < 2$ and $9 < b$. Compute the mean and variance of the random variable of Problem 5.17.*

Problem 5.27. *Let $\alpha : [-1,3] \to \mathbb{R}$ be defined by $\alpha(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x < 1 \\ 1+x & 1 \leq x < 2 \\ 7 & 2 \leq x \leq 3 \end{cases}$. Compute $\int_{-1}^3 t d\alpha(t)$.*

Problem 5.28. *Let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function, let $a < c < b$ and let $f : [a,b] \to \mathbb{R}$ be a bounded function. Prove:*

- $\overline{\int}_a^b f d\alpha = \overline{\int}_a^c f d\alpha + \overline{\int}_c^b f d\alpha$,
- $\underline{\int}_a^b f d\alpha = \underline{\int}_a^c f d\alpha + \underline{\int}_c^b f d\alpha$

Problem 5.29. *Let $\alpha : [a,b] \to \mathbb{R}$, let $a < c < b$ and let $f : [a,b] \to \mathbb{R}$ be Riemann-Stieltjes integrable. Prove that f is Riemann-Stieltjes integrable on $[a,c]$ and on $[c,b]$ and that*

$$\int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha.$$

5.3 The Fundamental Theorem of Calculus

Before proceeding it will be useful to have a few more facts about the Riemann-Stieltjes integral.

Theorem 5.30 (Mean Value theorem for Integrals). *Let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function and let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then there exists $c, a \le c \le b$ such that*

$$\int_a^b f d\alpha = f(c)(\alpha(b) - \alpha(c)).$$

Proof. When $\alpha(b) - \alpha(a) = 0$, then both sides of the equation are 0 and any point c will suffice. So assume that $\alpha(b) - \alpha(a) \ne 0$.

By Problem 5.18,

$$m \le \frac{\int_a^b f d\alpha}{\alpha(b) - \alpha(a)} \le M,$$

where $m = f(x_0)$ and $M = f(x_1)$ are the minimum and maximum of f on $[a,b]$. Thus, by the intermediate value theorem, there is a point c between x_0 and x_1 such that $f(c)$ is equal to this fraction. □

Proposition 5.31. *Let $a \le c < d \le b$, let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function and let $f : [a,b] \to \mathbb{R}$ be a bounded function that is Riemann-Stieltjes integrable with respect to α on $[a,b]$, then f is Riemann-Stieltjes integrable with respect to α on $[c,d]$.*

Proof. Given $\epsilon > 0$ there is a partition $\mathcal{P}$ of $[a,b]$ with $U(f,\mathcal{P},\alpha) - L(f,\mathcal{P},\alpha) < \epsilon$. Let $\mathcal{P}_1 = \mathcal{P} \cup \{c,d\}$, then $\mathcal{P}_1$ is a refinement of $\mathcal{P}$ and so $U(f,\mathcal{P}_1,\alpha) - L(f,\mathcal{P}_1,\alpha) \le U(f,\mathcal{P},\alpha) - L(f,\mathcal{P},\alpha) < \epsilon$. Now if we only consider the points in $\mathcal{P}_1$ that lie in $[c,d]$ then that will be a partition of $[c,d]$ and the terms in the above sum corresponding to those subintervals will be even smaller. Thus, we obtain a partition of $[c,d]$ that satisfies the condition in the Riemann-Stieltjes integrability criterion. □

In the case that $\alpha(x) = x$ the following theorem reduces to the classic fundamental theorem of calculus.

Theorem 5.32 (Fundamental Theorem of Calculus). *Let $\alpha : [a,b] \to \mathbb{R}$ be an increasing function that is differentiable on (a,b) and let $f : [a,b] \to \mathbb{R}$ be continuous. If $F(x) = \int_a^x f(t)d\alpha(t)$, then F is differentiable on (a,b) and $F'(x) = f(x)\alpha'(x)$.*

Proof. Let $a < x < x+h < b$, then by the mean value theorem, there is $c, x \le c \le x+h$ so that

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)d\alpha(t) = \frac{f(c)(\alpha(x+h) - \alpha(x)}{h}.$$

By choosing h sufficiently small, $|f(c) - f(x)|$ and $|\frac{\alpha(x+h)-\alpha(x)}{h} - \alpha'(x)|$ can be made arbitrarily small. The case of $h < 0$ is similar. $\square$

Corollary 5.33. *Let $\alpha : [a, b] \to \mathbb{R}$ be an increasing, continuous function that is differentiable on (a, b) and let $f : [a, b] \to \mathbb{R}$ be continuous. If $F : [a, b] \to \mathbb{R}$ is a continuous function that is differentiable on (a, b) and $F'(x) = f(x)\alpha'(x)$, then $\int_a^b f d\alpha = F(b) - F(a)$.*

Proof. Let $G(x) = \int_a^x f d\alpha + F(a)$, then for $a < x < b$, we have that $G'(x) = F'(x)$ and $G(a) = F(a)$. Hence, $G(x) = F(x)$ for $a < x < b$. Since $|G(b) - G(x)| = |\int_x^b f d\alpha| \leq K(\alpha(b) - \alpha(x))$, where $K = \sup\{|f(x)| : a \leq x \leq b\}$, we see that $G(b) = \lim_{x \to b^-} G(x) = \lim_{x \to b^-} F(x) = F(b)$, since F is continuous. Therefore, $F(b) = G(b) = \int_a^b f d\alpha + F(a)$ and the formula follows. $\square$

Problem 5.34. *Let $\alpha : [0, 2] \to \mathbb{R}$ be defined by $\alpha(x) = \begin{cases} x & 0 \leq x < 2 \\ 5 & x = 2 \end{cases}$ and let $f(x) = x^m$. Calculate $\int_0^2 f d\alpha$. Show that if we let F be continuous on $[0, 2]$ with $F'(x) = f(x)\alpha'(x)$ for $0 < x < 2$, then $F(2) - F(0) \neq \int_0^2 f d\alpha$.*