NOTES FOR MATH 5210, SPRING 2015

DOMINGO TOLEDO

1. Introduction

These notes are meant to supplement Rudin's book [5]. They contain the notes from the same course in 2013, and other topics covered in 2015. From time to time I will add here some material that is discussed in class but is not covered in [5].

2. The Real Numbers

2.1. **Equivalence Classes of Cauchy Sequences.** There are two standard constructions of the set $\mathbb{R}$ of real numbers from $\mathbb{Q}$, the rational numbers. One is by Dedekind cuts, used in Math 3210 and also described in [5], the other is using equivalence classes of Cauchy sequences of rational numbers. We describe the second construction.

Start with the set $\mathbb{Q}$ of rational numbers, and its structure as an ordered field. We assume that all this structure of $\mathbb{Q}$ is known. See 1.12 to 1.17 of [5] for the definition of ordered field.

Definition 2.1. Let $\{x_n\}$ be a sequence of rational numbers: $x_n \in \mathbb{Q}$ for $n = 1, 2, \ldots$.

(1) $\{x_n\}$ is called a Cauchy sequence if and only if for each $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, there exists a natural number N so that $|x_m - x_n| < \epsilon$ for all $m, n \geq N$.
(2) If $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences in $\mathbb{Q}$, we say that they are equivalent, written as $\{x_n\} \sim \{y_n\}$, if and only if $x_n - y_n \to 0$. Recall that this means that for any $\epsilon \in \mathbb{Q}$, $\epsilon > 0$, there exists a natural number N so that $|x_n - y_n| < \epsilon$ for all $n \geq N$.

It is easy to check that this is an equivalence relation. We denote by $[\{x_n\}]$ the equivalence class of the Cauchy sequence $\{x_n\}$.

Definition 2.2. The set $\mathbb{R}$ of real numbers is the set of equivalence classes of Cauchy sequences in $\mathbb{Q}$:

$$\mathbb{R} = \{[\{x_n\}] : \{x_n\} \text{ is a Cauchy sequence in } \mathbb{Q}\}.$$

Date: February 20, 2015.

2.2. **Basic Properties of $\mathbb{R}$.** It is easy check that the operations and relations listed below are well defined, meaning that they are independent of the choice of representatives in their definition:

(1) There is an injective map $\mathbb{Q} \to \mathbb{R}$ obtained by assigning to the rational number $r \in \mathbb{Q}$ the equivalence class of the constant sequence $\{r, r, r, r, \dots\}$. We consider $\mathbb{Q}$ as a subset of $\mathbb{R}$ by using this identification.
(2) The following operations are well-defined and make $\mathbb{R}$ into a field:
    (a) Addition: define $[\{x_n\}] + [\{y_n\}] = [\{x_n + y_n\}]$.
    (b) Multiplication: define $[\{x_n\}][\{y_n\}] = [\{x_n y_n\}]$.
    (c) Additive identity: $0 = [\{0, 0, 0, \dots\}]$.
    (d) Additive inverse: $-[\{x_n\}] = [\{-x_n\}]$.
    (e) Multiplicative identity: $1 = [\{1, 1, 1, \dots\}]$.
    (f) Multiplicative inverse: If $[\{x_n\}] \neq 0$, then $x_n = 0$ for at most finitely many n. Replace $\{x_n\}$ by an equivalent sequence $\{x'_n\}$ so that $x'_n \neq 0$ for all n and define the multiplicative inverse $1/[\{x_n\}] = [\{1/x'_n\}]$.
(3) There is an order relation defined by $[\{x_n\}] < [\{y_n\}]$ if and only if there exists a natural number N so that $x_n < y_n$ for all $n \geq N$. This is an order relation: if $[\{x_n\}] \neq [\{y_n\}]$, then exactly one of $[\{x_n\}] < [\{y_n\}]$ or $[\{x_n\}] > [\{y_n\}]$ holds.
(4) These operations and order relation make $\mathbb{R}$ into an ordered field, see 1.12 to 1.18 of [5] for the definition and properties of ordered fields.

In the above list of properties, the only ones that require most effort are the multiplicative inverse and the three exclusive possibilities of the order relation. See Homework 1 for details on how to check them.

2.3. **Completeness of $\mathbb{R}$.** It remains to prove that $\mathbb{R}$ is complete, meaning that every Cauchy sequence in $\mathbb{R}$ converges. This is the main point of this construction of $\mathbb{R}$. We proceed to check completeness. In what follows ϵ will always be a rational number, identified with the constant sequence $\{\epsilon, \epsilon, \dots\}$.

Let $\{a_n\}$ be a Cauchy sequence in $\mathbb{R}$. This means, first, that for each $n \in \mathbb{N}$, $a_n = [\{x_{i,n}\}]$, where $\{x_{i,n}\}_{i=1}^{\infty}$ is a Cauchy sequence in $\mathbb{Q}$. In more detail:

- A sequence $\{a_n\}$ of real numbers is, first of all, represented by a *double sequence* $\{x_{i,n}\}_{i,n=1}^{\infty}$ of rational numbers.
- Next, for each n, since a_n is a real number, the sequence $\{x_{i,n}\}$ is a Cauchy sequence of the variable i, in other words

$$\forall \epsilon > 0 \;\; \exists N(\epsilon, n) \in \mathbb{N} \text{ so that } i, j \geq N(\epsilon, n) \Rightarrow |x_{i,n} - x_{j,n}| < \epsilon. \tag{2.1}$$

- The sequence $\{a_n\}$ being a Cauchy sequence in $\mathbb{R}$ means that

$$\forall\epsilon > 0\ \exists M(\epsilon) \text{ s.t. } m, n \geq M(\epsilon) \Rightarrow |a_m - a_n| < \epsilon. \tag{2.2}$$

This means that for each m and n we must have that $|x_{i,m} - x_{i,n}| < \epsilon$ for i sufficiently large, where "sufficiently large" depends on m and n, in other words:

$$\forall\epsilon > 0\ \exists M \text{ s.t. } \forall m, n \geq M\ \exists I \text{ s.t. } i \geq I \Rightarrow |x_{i,m} - x_{i,n}| < \epsilon \tag{2.3}$$

where

$$M = M(\epsilon) \text{ and } I = I(\epsilon, m, n). \tag{2.4}$$

To prove that $\{a_n\}$ converges we must find a single Cauchy sequence $\{y_i\}$ in $\mathbb{Q}$ so that $\{a_n\}$ converges to $[\{y_i\}]$, that is, given any $\epsilon > 0$ we must find $N'(\epsilon)$ and an $I'(\epsilon, n)$ so that

$$n > N'(\epsilon) \text{ and } i > I'(\epsilon, n) \Rightarrow |x_{i,n} - y_i| < \epsilon. \tag{2.5}$$

The only reasonable way to find $\{y_i\}$ would be by a diagonal process. The first guess $y_i = x_{i,i}$ is unlikley to work, because we don't have enough information on $x_{i,i}$. After some trial and error, the following looks like a good choice:

In equation (2.1) take $\epsilon = \frac{1}{n}$, (any sequence of positive rational numbers converging to zero would do) and define

$$\varphi(n) = N(\frac{1}{n}, n) \text{ and } y_i = x_{\varphi(i),i}, \tag{2.6}$$

where $N(\epsilon, n)$ is as in (2.1).

Proposition 2.1. *The sequence* $[\{x_{i,n}\}]$ *converges to the real number* $[\{y_i\}]$ *as* $n \to \infty$

Proof. Given $\epsilon > 0$, consider only values of n so that $\frac{1}{n} < \epsilon/3$, and also so that $n > M(\epsilon/3)$, where M is as in (2.3) and (2.4), in other words, assume that

$$n > \max(3/\epsilon, M(\epsilon/3)).$$

Next consider only values of i can so $i > \varphi(n)$ and also $i > n$. Then by this choice and the choice already made for n we have

$$i > \max(3/\epsilon, M(\epsilon/3), n, \varphi(n))$$

Finally choose

$$j > \max(\varphi(i), \varphi(n), I(\epsilon/3, i, n)).$$

Then for all these values of n, i, j we have

$$|x_{\varphi(i),i} - x_{i,n}| \leq |x_{\varphi(i),i} - x_{j,i}| + |x_{j,i} - x_{j,n}| + |x_{j,n} - x_{i,n}| < \frac{1}{i} + \frac{\epsilon}{3} + \frac{1}{n} < \epsilon.$$

Since in (2.6) we chose $y_i = x_{\varphi(i),i}$, this last inequality means that

$$|y_i - x_{i,n}| < \epsilon \text{ for } n > \max(3/\epsilon, M(\epsilon/3)) \text{ and } i > \max(3/\epsilon, M(\epsilon/3), \varphi(n)),$$

therefore we found that (2.5) holds (with $N'(\epsilon) = \max(3/\epsilon, M(\epsilon/3))$ and $I'(\epsilon, n) = \max(3/\epsilon, M(\epsilon/3), \varphi(n)))$. □

3. Complete Metric Spaces

Recall that a metric space (X, d) is said to be *complete* if every Cauchy sequence $\{x_n\}$ in X has limit in X, that is, there exists $x \in X$ so that $\lim x_n = x$.

3.1. **Examples of Complete Metric Spaces.**

3.1.1. *The real numbers* $\mathbb{R}$. In §2.3 we proved that $\mathbb{R}$ is complete. All other examples will follow from this.

3.1.2. *Euclidean space* $\mathbb{R}^k$. The space $\mathbb{R}^k = \{\mathbf{x} = (x_1, \ldots, x_k) : x_1, \ldots, x_k \in \mathbb{R}\}$ can be given various distance functions, coming from norms:

$$\begin{aligned} ||\mathbf{x}||_1 &= |x_1| + \ldots |x_k| \\ ||\mathbf{x}||_2 &= (x_1^2 + \cdots + x_k^2)^{1/2} \\ ||\mathbf{x}||_\infty &= \max(|x_1|, \ldots, |x_k|), \end{aligned} \tag{3.1}$$

and d_1, d_2, d_∞ will denote the corresponding distance functions $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||$.

Euclidean space usually refers to $\mathbb{R}^k$ with the Euclidean distance d_2, but checking completeness, it is useful to also consider d_∞. The three distances are related by the following inequalities:

$$\begin{aligned} ||\mathbf{x}||_\infty \leq \quad ||\mathbf{x}||_2 &\leq \sqrt{k}||\mathbf{x}||_\infty \\ ||\mathbf{x}||_2 \leq \quad ||\mathbf{x}||_1 &\leq \sqrt{k}||\mathbf{x}||_2 \\ ||\mathbf{x}||_\infty \leq \quad ||\mathbf{x}||_1 &\leq k||\mathbf{x}||_\infty. \end{aligned} \tag{3.2}$$

A sequence $\{\mathbf{x}_n\}$ is Cauchy in one of these three distances if and only if it is Cauchy in any other. For example, to prove completeness in the Euclidean distance d_2, suppose $\{\mathbf{x}_n\}$ is Cauchy in d_2: given $\epsilon > 0$ there is N so that $m, n > N \Rightarrow ||\mathbf{x}_m - \mathbf{x}_n|| < \epsilon$. Then the first half of the top inequality in (3.2) gives $||\mathbf{x}_m - \mathbf{x}_n|| < \epsilon$, for $m, n > N$, in other words, for each $m, n > N$ and for each $i = 1, \ldots, k$,

$$|x_{i,m} - x_{i,n}| \leq \max(|x_{1,m} - x_{1,n}|, \ldots, |x_{k,m} - x_{k,n}|) < \epsilon,$$

where $\mathbf{x}_n = (x_{1,n}, \dots, x_{k,n})$. In other words, for each $i = 1, \dots, k$, the sequence $x_{i,n}$ of ith components of $\mathbf{x}_n$ is a Cauchy sequence in $\mathbb{R}$, so it converges. Call its limit x_i and let $\mathbf{x} = (x_1, \dots, x_k)$. Then, given $\epsilon > 0$, for each i there is N_i so that $n > N_i \Rightarrow |x_{i,n} - x_i| < \epsilon/\sqrt{k}$. Taking $N = \max(N_i)$, we get $||\mathbf{x}_n - \mathbf{x}||_\infty < \epsilon/\sqrt{k}$ for $n > N(\epsilon)$. The second half of the top line of (3.2) gives $||\mathbf{x}_n - \mathbf{x}||_2 < \epsilon$, so $\lim \mathbf{x}_n = \mathbf{x}$ in the Euclidean distance d_2. Thus $(\mathbb{R}^k, d_2)$ is a complete metric space. Notice that what we really proved was that $(\mathbb{R}^k, d_\infty)$ is complete and derived completeness in d_2 from the two inequalities in the top line of (3.2). Completeness in d_1 then follows from either of the next two lines.

3.1.3. *Spaces of continuous functions.* Let (X, d) be a *compact* metric space and let $C(X)$ denote the set of continuous real valued functions on X:

$$C(X) = \{f : f \text{ is a continuous function } f : X \to \mathbb{R}\}, \tag{3.3}$$

and give it a norm and corresponding distance function:

$$||f||_\infty = \max_{x \in X}(|f(x)|), \quad d_\infty(f, g) = ||f - g||_\infty, \quad \text{for all } f, g \in C(X). \tag{3.4}$$

Observe that the definition of the norm and the distance makes sense because X is assumed to be compact, therefore continuos functions $f : X \to \mathbb{R}$ are bounded and attain their maximum.

We need to check that $(C(X), d_\infty)$ is a metric space. It is clear that $d_\infty(f, g) \geq 0$ and $= 0$ if and only if $f = g$, and $d_\infty(f, g) = d_\infty(g, f)$. To check the triangle inequality, given any $f, g, h \in C(X)$, we have for each $x \in X$ the inequality

$$|f(x) - g(x)| \leq |f(x) - h(x)| + |h(x) - g(x)|. \tag{3.5}$$

and, since each term in the right hand side is majorized by the maximum, we have

$$|f(x) - h(x)| + |h(x) - g(x)| \leq \max_{x \in X} |f(x) - h(x)| + \max_{x \in X} |h(x) - g(x)|,$$

so (3.5) gives

$$|f(x) - g(x)| \leq \max_{x \in X} |f(x) - h(x)| + \max_{x \in X} |h(x) - g(x)| = d_\infty(f, h) + d_\infty(h, g),$$

so the maximum of the left hand side, which by definition is $d_\infty(f, g)$, is majorized by the right hand side, which is the triangle inequality for d_∞., so $(C(X), d_\infty)$ is a metric space.

Convergence in d_∞ is actually a familiar notion:

Theorem 3.1. *A sequence $\{f_n\}$ in $C(X)$ converges to $f \in C(X)$ in the metric d_∞ if and only if $f_n(x)$ converges to $f(x)$* uniformly *on X. A sequence $\{f_n\}$ in $(C(X), d_\infty)$ is a Cauchy sequence if and only if the sequence $\{f_n(x)\}$ is* uniformly Cauchy.

See Definition 7.7 of [5] for the definition of uniform convergence; the definition of uniformly Cauchy is implicit in Theorem 7.8.

Proof. Suppose $f_n \to f$ in d_∞. This means that given $\epsilon > 0$ there exists $N(\epsilon)$ so that $n > N(\epsilon) \Rightarrow d_\infty(f_n, f) < \epsilon$, which is equivalent to saying that $n > N(\epsilon) \Rightarrow \max_{x \in X}(|f_n(x) - f(x)|) < \epsilon$, which is equivalent to saying that $n > N(\epsilon) \Rightarrow |f_n(x) - f(x)| < \epsilon$ *for all* $x \in X$, in other words, N is a function of ϵ, independent of x, which is the definition of uniform convergence. This proves the first statement, the proof of the second is similar. □

Theorem 3.2. $(C(X), d_\infty)$ *is a complete metric space.*

Proof. Suppose $\{f_n\}$ is a Cauchy sequence in $(C(X), d_\infty)$. By theorem 3.1, this means that $\{f_n\}$ is uniformly Cauchy, in other words, given $\epsilon > 0$ there is N such that $m, n > N \Rightarrow |f_m(x) - f_n(x)| < \epsilon$ for all $x \in C$. In particular, for each $x \in X$, $\{f_n(x)\}$ is a Cauchy sequence in $\mathbb{R}$, so it converges to a limit that we call $f(x)$.

At this point we have only produced a function $f : X \to \mathbb{R}$ so that $f_n(x) \to f(x)$ pointwise on X and do not even know that $f \in C(X)$. We need more work. First, Theorem 7.8 of Rudin [5], that says that a sequence of continuous functions is uniformly convergent if and only if it is uniformly Cauchy, in particular, the convergence $f_n \to f$ must be uniform (see the argument at the top of page 148 of [5]). Finally, a uniform limit of continuous functions is continuous (see Theorem 7.12 of [5]), so $f \in C(X)$.

□

3.2. **Examples of Incomplete Metric Spaces.** To appreciate the concept of completeness, we also need to have some examples of metric spaces that are not complete. The most familiar examples are of a pair of metric spaces, $X \subset Y$, where the distance on X is restricted from (Y, d), X is not closed in Y. Taking a point $y \in Y \setminus X$ that is a limit point of X we get a sequence $\{x_n\}$ in X converging to y. This is a Cauchy sequence in X that does not converge in X, so X is not complete.

Examples that we have seen of this nature are:

3.2.1. $\mathbb{Q} \subset \mathbb{R}$, $y = \sqrt{2}$. Take any sequence $x_n \in \mathbb{Q}$ converging to $\sqrt{2}$, this is a Cauchy sequence in $\mathbb{Q}$ with no limit in $\mathbb{Q}$.

3.2.2. *The open interval* $(0, 1) \subset \mathbb{R}$, $y = 0$. . The sequence $\{1/n\}$ in $(0, 1)$ converges to $0 \notin (0, 1)$, so is a Cauchy sequence in $(0, 1)$ with no limit in $(0, 1)$.

And there are of course many variations on these examples. They may seem artificial, but all examples have to be like this. It is in fact a theorem that every metric space is a dense subspace of a complete metric space, called its *completion*. For example, the completion of $\mathbb{Q}$ is $\mathbb{R}$ and the completion of $(0,1)$ is $[0,1]$. To give more interesting examples, we have to look at situations where the completion may not be so familiar.

3.2.3. *Incomplete Metrics on $C([a,b])$.* Specialize the construction of §3.1.3 to $X=[a,b]\subset\mathbb{R}$, an interval in the real line ($a<b$). In analogy with the norms and distances (3.1) on $\mathbb{R}^k$, define norms and distances on $C([a,b])$

$$\begin{aligned} ||f||_1 &= \int_a^b |f(x)|dx \\ ||f||_2 &= (\int_a^b f(x)^2 dx)^{1/2} \\ ||f||_\infty &= \max_{x\in[a,b]} |f(x)|, \end{aligned} \tag{3.6}$$

and d_1, d_2, d_∞ will denote the corresponding distance functions $d(f,g) = ||f-g||$, and where d_∞ is the distance previously defined in (3.4). The verification that these are metrics is analogous to the verification for (3.1). In all cases the main issue is the triangle inequality, which has been verified for d_∞, and the verification for d_2 uses the Cauchy-Schwarz inequality for integrals.

The inequalities analogous to (3.2) are:

$$\begin{aligned} ||f||_2 &\le \sqrt{b-a}\ ||f||_\infty \\ ||f||_1 &\le \sqrt{b-a}\ ||f||_2 \\ ||f||_1 &\le (b-a)\ ||f||_\infty. \end{aligned} \tag{3.7}$$

It is easy to prove these inequalities, for instance the last

$$\int_a^b |f(x)|dx \le \int_a^b \max_{x\in[a,b]}(|f(x)|)dx = (b-a)\max_{x\in[a,b]}(f(x)) = (b-a)||f||_\infty.$$

The main difference between (3.2) and (3.7) is that for $C(X)$ we get only half of the inequalities as we got for $\mathbb{R}^k$. This means that the notions of convergence and of Cauchy sequences need not be the same in all the distances. For instance, the last inequality (3.7) says that a sequence f_n is convergent of is Cauchy in d_∞, then the same is true for d_1, but the converse need not be true. In fact, we will see that it is not true:

Theorem 3.3. *The metric space $(C([0,1]), d_1)$ is not complete.*

Proof. The idea is the same as in §3.2.1: Think of $(C([0,1]), d_1)$ as a subspace of a larger space, take f in this larger space but not in $C([0,1])$, and a sequence $f_n\in C([0,1])$ with $f_n\to f$. Let

$$f(x) = x^{-1/2}$$

Then $f : (0,1] \to \mathbb{R}$ is unbounded, cannot be extended continuously to $[0,1]$, but

$$\int_0^1 |f(x)|dx = \lim_{\epsilon\to 0}\int_\epsilon^1 |f(x)|dx = \lim_{\epsilon\to 0} 2x^{1/2}|_\epsilon^1 = 2,$$

so this is a convergent integral.

Define for $n \in \mathbb{N}$, define $f_n \in C([0,1])$ by

$$f_n(x) = \begin{cases} n^{1/2} & \text{if } 0 \le x \le 1/n, \\ x^{-1/2} & \text{if } 1/n \le 1. \end{cases}$$

Then

$$\int_0^1 |f_n(x) - f(x)|dx \le \int_0^{1/n} x^{-1/2}dx = 2/\sqrt{n} \to 0 \text{ as } n \to \infty.$$

By the usual argument using the triangle inequality, this implies that f_n is a Cauchy sequence in the metric d_1. This is a statement that involves only the sequence $f_n \in C([0,1])$ and not the function f.

We can prove that f_n does not converge in d_1 to any function in $C([0,1])$,without using the function f, by arguing as follows. Suppose $f_n \to g \in C([0,1])$ in the distance d_1. Since g is continuous, it is bounded, say $|g(x)| \le C$. Then for $n > C^2$ we have

$$\begin{aligned} d_1(f_n, g) = \int_0^1 |f_n(x) - g(x)|dx &\ge \int_{1/n}^{1/C^2} |f(x) - g(x)|dx \\ \ge \int_{1/n}^{1/C^2} (x^{-1/2} - C)dx &= 1/C + 1/n - 2/\sqrt{n} \to 1/C, \end{aligned}$$

so $d_1(f_n, g)$ cannot approach zero for any $g \in C([0,1])$. $\square$

Similar arguments can be used to prove that $(C([0,1]), d_2)$ is not complete. Use the function $g(x) = x^{-1/4}$ instead of $f(x) = x^{-1/2}$.

3.3. **Non-compact closed and bounded sets in $C(X)$.** We have seen examples of metric spaces where closed and bounded sets are not necessarily compact, for instance in $\mathbb{Q}$. More interesting examples occur in infinite dimensional spaces, for instance, $C([a,b])$.

Here is an example that appears in many places. For $n = 0, 1, 2, \ldots$ define $f_n \in C([0, 2\pi])$ by

$$f_n(x) = \cos(nx). \tag{3.8}$$

The standard addition formulas for the cosine give

$$\begin{aligned} \cos((m+n)x) &= \cos(mx)\cos(nx) - \sin(mx)\sin(nx) \\ \cos((m-n)x) &= \cos(mx)\cos(nx) + \sin(mx)\sin(nx) \end{aligned}$$

we get

$$f_m(x)f_n(x) = \cos(mx)\cos(nx) = (\cos(m+n)x + \cos(m-n)x)/2$$

and integrating

$$\int_0^{2\pi} f_m(x)f_n(x)dx = \begin{cases} 0 & \text{if } m \neq n, \\ \pi & \text{if } m = n. \end{cases} \tag{3.9}$$

If $m \neq n$, we get

$$d_2(f_m, f_n)^2 = \int_0^{2\pi} (f_m(x) - f_n(x))^2 dx =$$

$$\int_0^{2\pi} (f_m(x)^2 - 2f_m(x)f_n(x) + f_m(x)^2)dx \quad = \quad 2\pi,$$

therefore

$$d_2(f_m, f_n) = \sqrt{2\pi} \text{ for } m \neq n, \tag{3.10}$$

so by the first inequality (3.7)

$$d_\infty(f_m, f_n) \geq 1 \text{ for all } m, n \text{ with } m \neq n. \tag{3.11}$$

Let $E = \{f_0, f_1, f_2, \ldots\}$. Since $d_\infty(0, f_n) = ||f_n||_\infty = 1$ for all n, E is a bounded set. Moreover (3.11) shows that E has no limit points. Hence E is closed, is an infinite bounded set without limit points, hence E is not compact, see Theorem 2.37 of [5]. Or just observe that the collection $\{B(f_n, \frac{1}{2})\}$ is an open cover of E that has no finite subcover. This gives a good example of a complete metric space in which closed and bounded sets need not be compact, in contrast to the situation for Euclidean space R^k.

Remark 3.1. For a characterizatioon of the compact subsets of $C(X)$, see the section on equicontinuity, 7.19 to 7.25 of [5].

3.4. **Lipschitz Maps.** Let (X, d_X) and (Y, d_Y) be metric spaces.

Definition 3.1. A map $f : X \to Y$ is called a *Lipschitz map* if there exists a constant $C > 0$ such that

$$d_Y(f(x), f(y)) \leq C\ d_X(x, y) \text{ for all } x, y \in X.$$

The constant C (if it exists) is called a *Lipschitz constant* for f. The infimum of the set of all Lipschitz constants is called *the* Lipschitz constant for f.

Clearly a Lipschitz map is continuous, in fact uniformly continuous: in the $\epsilon - \delta$ definition of continuity, given $\epsilon > 0$, if we let $\delta = \epsilon/C$, then $d_X(x, y) < \delta \Rightarrow d_Y(f(x), f(y)) < \epsilon$. In other words, δ can be chosen as an explicit linear function of ϵ.

Example 3.1. Some examples of Lipschitz maps:

(1) Let $A : \mathbb{R}^k \to \mathbb{R}^l$ be a linear transformation. Define the *norm of A* by

$$||A|| = \max_{||x||=1} ||Ax||, \tag{3.12}$$

this is equivalent to Definition 9.6(c) of [5]. Observe that the definition makes sense because the continuous function $||AX||$ does indeed have a maximum on the compact set $\{||X|| = 1\}$ (the unit sphere). By linearity, this norm has the property that

$$||Ax|| \leq ||A||||x|| \Rightarrow d(Ax, Ay) \leq ||A||d(x, y) \text{ for all } x, y \in \mathbb{R}^k. \tag{3.13}$$

Thus a linear transformation $A : \mathbb{R}^k \to \mathbb{R}^l$ is a Lipschitz map, with Lipschitz constant $||A||$.

(2) Suppose $I \subset \mathbb{R}$ is an interval, and $f : I \to \mathbb{R}$ is differentiable on I and has bounded derivative, that is, there exists $C > 0$ such that $|f'(x)| \leq C$ for all $x \in I$. Then either using the mean value theorem: for some $\xi \in I$,

$$|f(x) - f(y)| = |f'(\xi)(x - y)| \leq C|x - y|,$$

or using the fundamental theorem of calculus:

$$|f(x) - f(y)| = |\int_y^x f'(t)dt| \leq |\int_y^x C dt| = C|x - y|$$

we get that f is Lipschitz on I, and in fact C also works as a Lipschitz constant for f.

(3) The last two examples can be combined as follows: Let $U \subset \mathbb{R}^k$ be an open, convex set, let $f : U \to \mathbb{R}^l$ be a continuously differentiable map, and suppose that it has *bounded derivative* on U meaning

$$\text{There exists } C > 0 \text{ such that } ||d_x f|| \leq C \text{ for all } x \in U, \tag{3.14}$$

where $d_x f : \mathbb{R}^k \to \mathbb{R}^l$ is the differential of f at x and $||d_x f||$ is the norm of the linear transformation $d_x f$ as defined in (3.12). See Chapter 9 of [5] for details on differentiable maps. The notation $\mathbf{f}'(x)$ is used there instead of $d_x f$. The matrix of the linear transformation $d_x f$ in the standard bases for $\mathbb{R}^k$ and $\mathbb{R}^l$ is the Jacobian matrix of f.

We use convexity and the fundamental theorem of caluculs to show that f is Lipschitz, and that C is a Lipschitz constant for f. Let $x, y \in U$. Since U is convex, the line segment

$$\gamma(t) = (1 - t)x + ty, \quad 0 \leq t \leq 1$$

lies in U. Observe that $\gamma(0) = x, \gamma(1) = y$. Then, by the fundamental theorem of calculus and the chain rule

$$f(y) - f(x) = \int_0^1 \frac{d}{dt}(f(\gamma(t))dt = \int_0^1 d_{\gamma(t)} f(\gamma'(t))dt.$$

Taking norms, using $\gamma'(t) = y - x$ and the inequality (3.13) we get

$$\begin{aligned} ||f(y) - f(x)|| &= ||\int_0^1 d_{\gamma(t)} f(y-x)dt|| \leq \\ \int_0^1 ||d_{\gamma(t)} f|| ||y-x||dt &\leq C||y-x|| \end{aligned}$$

in other words

$$d(f(x), f(y)) \leq Cd(x, y),$$

and f is Lipschitz.

Example 3.2. Examples of continuous maps that are not Lipschitz:

(1) $f(x) = x^2$ is not Lipschitz on $\mathbb{R}$, since $|f(x) - f(0)|/|x - 0| = |x|$ is unbounded. This is continuous, but not uniformly continuous on $\mathbb{R}$.
(2) $f(x) = \sqrt{x}$ is uniformly continuous on $[0, \infty)$ but it's not Lipschitz on any interval containing 0, since $|f(x) - f(0)|/|x - 0| = \sqrt{x}/x = 1/\sqrt{x} \to \infty$ as $x \to 0$.

3.5. **Digression: Lebesgue Numbers.** In studiying compact metric spaces, it is very useful to have the following theorem:

Theorem 3.4. *Let X be a compact metric space and let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ be an open cover of X. Then there exists a numbet $\lambda = \lambda(\mathcal{U}) > 0$ with the following property: If $x, y \in X$ are any two points such that $d(x, y) < \lambda$, then there exists an open set $U_\alpha \in \mathcal{U}$ so that $x, y \in U_\alpha$.*

Remark 3.2. A number λ as in the statement of the theorem is called a *Lebesgue number* for $\mathcal{U}$.

Proof. For each $x \in X$ choose $U_\alpha \in \mathcal{U}$ so that $x \in U_\alpha$. Since U_α is open in X, there is $r(x) > 0$ so that $B(x, r(x)) \subset U_\alpha$. Then $\mathcal{B} = \{B(x, r(x))\}_{x \in X}$ and $\mathcal{B}' = \{B(x, r(x)/2)\}_{x \in X}$ are open covers of X with the property that $\mathcal{B}'$ refines $\mathcal{B}$ and $\mathcal{B}'$ refines $\mathcal{U}$ in the sense that each element of $\mathcal{B}'$ is contained in some element of $\mathcal{B}$ and in each element of $\mathcal{B}$ is conained in some element of $\mathcal{U}$.

Since X is compact there is a finite subcollection $\mathcal{B}'' = \{B(x_i, r(x_i)/2\}_{i=1}^n$ that covers X. Let $C = \{x_1, \ldots, x_n\}$ denote the collection of the centers of the balls $B(x_i, r(x_i)/2) \in \mathcal{B}''$. Let $\lambda = \min\{r(x_i)/2 : x_i \in C\}$. Let $x, y \in X$ and suppose $d(x, y) < \lambda$. Then there is an x_i so that $d(x, x_i) < r(x_i)/2$. Then $d(x_i, y) \leq d(x_i, x) + d(x, y) < r(x_i)/2 \; + \; \lambda \leq r(x_i)$. Thus $x, y \in B(x_i, r(x_i))$, which by construction is contained in some U_α. □

A typical application of Lebesgue numbers is the following theorem:

Theorem 3.5. *Let X, Y be metric spaces, with X compact. Suppose $f : X \to Y$ is continuous. Then f is uniformly continuous.*

Proof. Since f is continuous, for each $x \in X$ and each $\epsilon > 0$ there exists $\delta(x, \epsilon) > 0$ with the property that if $y \in X$ and $d_X(x, y) < \delta(x, \epsilon)$, then $d_Y(f(x), f(y)) < \epsilon/2$. Let λ be a Lebesgue number for the cover $\{B(x, \delta(x, \epsilon))\}_{x \in X}$. Then if $x, y \in X$ and $d_X(x, y) < \lambda$, then there exists $z \in X$ so that $x, y \in B(z, \delta(z, \epsilon)$, consequently $d_Y(f(x), f(y)) \leq d_Y(f(x), f(z)) + d_Y(f(z), f(y)) < \epsilon/2 + \epsilon/2 = \epsilon$. □

4. Normed Vector Spaces

The examples of metric spaces in section 3.1 were also vector spaces and the distance function had special features. The unifying concept behind these (and many other) examples is the *normed vector space*:

Definition 4.1. A *normed vector space* (over $\mathbb{R}$) is a real vector space V and a function $N : V \to \mathbb{R}$, called the *norm* (usually write $||\mathbf{x}|| = N(\mathbf{x})$), satisfying the following properties.

(1) For all $\mathbf{x} \in V$, $||\mathbf{x}|| \geq 0$ and $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = 0$.
(2) For all $\mathbf{x} \in V$ and all $a \in \mathbb{R}$, $||a\mathbf{x}|| = |a|\ ||\mathbf{x}||$.
(3) For all $\mathbf{x}, \mathbf{y} \in V$, $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$. (Triangle inequality).

If V is a normed vector space, then $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{y} - \mathbf{x}||$ is a metric on V. This metric is *translation invariant* meaning that for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $d(\mathbf{x} + \mathbf{z}, \mathbf{y} + \mathbf{z}) = d(\mathbf{x}, \mathbf{y})$, in fact, both equal $d(0, \mathbf{y} - \mathbf{x})$. Observe that $||\mathbf{x}|| = d(\mathbf{x}, 0)$.

4.1. **Norms in Euclidean Spaces.**

We have seen in section (3.1.2) examples of norms in the space $\mathbb{R}^n$, namely the norms (3.1) and the comparisons (3.2) which we write again:

$$\begin{aligned} ||\mathbf{x}||_\infty \leq \quad & ||\mathbf{x}||_2 \quad \leq \sqrt{n}||\mathbf{x}||_\infty \\ ||\mathbf{x}||_2 \leq \quad & ||\mathbf{x}||_1 \quad \leq \sqrt{n}||\mathbf{x}||_2 \\ ||\mathbf{x}||_\infty \leq \quad & ||\mathbf{x}||_1 \quad \leq n||\mathbf{x}||_\infty . \end{aligned} \tag{4.1}$$

We used these inequalites to prove that these norms are equivalent in the sense that the notions of convergence, limits, and Cauchy sequences are the same in all these norms. One obvious feature of these inequalities is that the right hand side depends on the dimension n.

We observe, first of all, that these inequalities cannot be improved: these inequalites all become equalities on the vector $\mathbf{x} = (1, 1, \ldots, 1)$. In other words,

$$\lim_{n\to\infty} \left(\sup \left\{ \frac{||\mathbf{x}||_2}{||\mathbf{x}||_\infty} : \mathbf{x} \in \mathbb{R}^n \right\} \right) = \lim_{n\to\infty} \sqrt{n} = \infty \tag{4.2}$$

with similar conclusions for $||\mathbf{x}||_1/||\mathbf{x}||_2$ and $||\mathbf{x}||_1/||\mathbf{x}||_\infty$. On the other hand the reciprocals of these ratios remain bounded. We summarize

$$\sup\left\{\frac{||\mathbf{x}||_r}{||\mathbf{x}||_s} : \mathbf{x} \in \mathbb{R}^n\right\} = 1 \text{ if } r > s \text{ and } \to \infty \text{ as } n \to \infty \text{ if } r < s. \tag{4.3}$$

for $r, s \in \{1, 2, \infty\}$ and using the usual order $1 < 2 < \infty$.

4.1.1. *Some infinite dimensional Euclidean spaces.* We define a single space which contains all the above examples.

Definition 4.2. Let $\mathbb{R}^\infty$ denote the set of all infinite sequences of real numbers that are eventually zero, in other words,

$$\mathbb{R}^\infty = \{\mathbf{x} = (x_1, x_2, \dots) : x_i \in \mathbb{R} \text{ and } x_i = 0 \text{ for all but finitely many } i\}$$

More precisely, for each $\mathbf{x} = (x_1, x_2, \dots) \in \mathbb{R}^\infty$ there exists an $N = N(\mathbf{x}) \in \mathbb{N}$ so that $x_i = 0$ for $i > N$.

Observe that for each n we can regard $\mathbb{R}^n \subset \mathbb{R}^\infty$, namely

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n, 0, 0, \dots) : x_i \in \mathbb{R}\} \subset \mathbb{R}^\infty.$$

and, with this understanding of how $\mathbb{R}^n$ is a subset of $\mathbb{R}^\infty$, we have that

$$\mathbb{R}^\infty = \bigcup_{n \in \mathbb{N}} \mathbb{R}^n.$$

Moreover, the norms $||\mathbf{x}||_1, ||\mathbf{x}||_2, ||\mathbf{x}||_\infty$ all make sense in $\mathbb{R}^\infty$, since the defining expressions are finite sums (for $||\mathbf{x}||_1$ and $||\mathbf{x}||_2$) and the sequence of components, being finite, it is bounded (for $||\mathbf{x}||_\infty$). Or, we could say, each $\mathbf{x} \in \mathbb{R}^\infty$ is in some $\mathbb{R}^n$, we can use the definition of the norm in $\mathbb{R}^n$, and the resulting value is independent of the choice of n.

These norms *are not equivalent.* We have the left-hand inequalities (4.1) hold, but not the others: $||\mathbf{x}||_\infty \le ||\mathbf{x}||_2 \le ||\mathbf{x}||_1$ hold, but (4.3) implies that none of the opposite inequalities can hold.

To give more examples of norms on $\mathbb{R}^n$ we'll first need to prove some inequalities that will be needed to prove the triangle inequality. We take this opportunity to discuss some general consequences of convexity.

4.2. **Convex functions and Jensen's inequality.** Recall that a function $\phi : I \to \mathbb{R}$, where $I \subset \mathbb{R}$ is an interval, is called *convex* if and only if it is continuous and for all $x, y \in I$, we have

$$\phi\Big(\frac{x+y}{2}\Big) \le \frac{\phi(x) + \phi(y)}{2}, \tag{4.4}$$

and ϕ is called *strictly convex* if this inequality is always strict whenever $x \neq y$. For example, a linear function is convex but not strictly convex. Geometrically this inequality means that that the midpoint of every chord spanned by two points on the graph of ϕ lies above the graph. From calculus we know that a twice continuously differentiable function ϕ is convex if and only if $\phi''(x) \geq 0$ for all x.

Often one sees the definition of convexity requiring the stronger inequality that every chord lie, except for its endpoints, above the graph. In other words, for all $x, y \in I$ and for all $t \in [0, 1]$,

$$\phi((1-t)x + ty) \leq (1-t)\phi(x) + t\phi(y). \tag{4.5}$$

One consequence of Jensen's inequality is that for continuous functions ϕ, the two definitions are equivalent.

Theorem 4.1. *Suppose that ϕ is a continuous, convex function. Fix a natural number n and a collection $\mu_1, \mu_2, \cdots \mu_n$ so that $\mu_1+\mu_2+\cdots+\mu_n = 1$. Then for any $(x_1, \cdots, x_n) \in \mathbb{R}^n$ so that so that $\{\sum_{i=1}^n \lambda_i x_i : 0 \leq \lambda_i \leq 1\}$ is contained in the domain of ϕ we have the inequality*

$$\phi(\mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_n x_n) \leq \mu_1\phi(x_1) + \mu_2\phi(x_2) + \cdots + \mu_n\phi(x_n). \tag{4.6}$$

In other words, the expression $\mu_1 x_1+\cdots+\mu_n x_n$ is a weighted average of the numbers $x_1, \ldots, x_n$ relative to the weights $\mu_1, \ldots, \mu_n$. Jensen's theorem is that the value of ϕ at a weighted average of $x_1, \ldots, x_n$ lies below the average, with the same weights, of $\phi(x_1), \ldots, \phi(x_n)$. Thus the defining inequality (4.4) of convexity, which is the same as the simplest case of (4.6) : $n = 2$ and $\mu_1 = \mu_2 = \frac{1}{2}$, implies the same inequality for all other possible weighted averages.

Proof. We follow the proof of (4.6) given in [3]:

(1) Iteration of (4.4) gives (4.6) for the case $n = 2^k$ and $\mu_1 = \mu_2 = \cdots = \mu_n = \frac{1}{n}$. Namely $\phi((x_1 + \cdots + x_{2^k})/2^k)$ can be written as

$$\phi(((x_1 + \cdots + x_{2^{k-1}})/2^{k-1})/2 + ((x_{2^{k-1}+1} + \ldots x_{2^k})/2^{k-1})/2)$$

to which (4.4) can be applied to give the average of two terms of the same form with k replaced by $k-1$, which gives the inductive step for an induction of which (4.4) is the first step.

(2) Prove that (4.6) is true for all μ_i being equal: $\mu_i = \frac{1}{n}$. This is proved as follows: if this statement is true for n then it is true for $n-1$: Assume that for any n numbers $x_1, \ldots, x_n$ we have

$$\phi((x_1 + \cdots + x_n)/n) \leq (\phi(x_1) + \cdots + \phi(x_n))/n. \tag{4.7}$$

Then given any collection of $n-1$ real numbers $x_1, \ldots, x_{n-1}$, apply (4.7) to the n-tuple $x_1, \ldots, x_{n-1}, (x_1 + \ldots x_{n-1})/(n-1)$. Since

$$\frac{x_1 + \cdots + x_{n-1} + (x_1 + \ldots x_{n-1})/(n-1)}{n} = \frac{x_1 + \ldots x_{n-1}}{n-1},$$

applying (4.7) to the left hand side we get that it, and therefore the right hand side, is majorized by

$$\frac{\phi(x_1) + \cdots + \phi(x_{n-1}) + \phi(\frac{x_1+\cdots+x_{n-1}}{n-1})}{n}$$

Multiplying both sides of the resulting inequality by n we get

$$n\phi(\frac{x_1 + \cdots + x_{n-1}}{n-1}) \leq \phi(x_1) + \cdots + \phi(x_{n-1}) + \phi(\frac{x_1 + \cdots + x_{n-1}}{n-1})$$

therefore

$$(n-1)\phi(\frac{x_1 + \cdots + x_{n-1}}{n-1}) \leq \phi(x_1) + \cdots + \phi(x_{n-1}),$$

which is equivalent to (4.7) for n replaced by $n-1$. Since (4.7) is true for $n = 2^k$, all k, it follows that it holds for all natural numbers n.

(3) The inequality (4.6) holds whenever all the μ_i are rational: Let m be a common denominator for $\mu_1, \ldots, \mu_n$, let $\mu_i = \frac{k_i}{m}$, where m and the k_i are natural numbers. Then we can write the weighted average $\sum \mu_i x_i$ as a weighted average with equal weights $\frac{1}{m}$ by repeating the terms x_i as many times as the numerators k_i:

$$\sum_{i=1}^{n} \mu_i x_i = \sum_{j=1}^{m} \frac{1}{m} y_j$$

where $y_1 = \ldots y_{k_1} = x_1$, $y_{k_1+1} = \ldots y_{k_2} = x_2$, etc.

(4) The inequality (4.6) holds for all real weights μ_i: for $i = 1, \ldots, n$ take a sequence $r_{i,j}$ of rational numbers, $0 < r_{i,j} < 1$, so that $\lim_{j\to\infty} r_{i,j} = \mu_i$. With some care, we can choose the $r_{i,j}$ to be weights: $0 < r_{i,j} < 1$ and $r_{1,j} + \ldots r_{n,j} = 1$. We could do this as follows: choose $r_{1,j}, \ldots r_{n-1,j}$ to be weights and also satisfy $r_1, j + \cdots + r_{n-1,j} < 1$. This is possible because $\mu_1 + \cdots + \mu_{n-1} < 1$.

Fix $x_1, \ldots, x_n$. Then (4.6) holds for $r_{i,j}$ and, since ϕ is continuous, each side of (4.6) converges to the corresponding side of (4.6) for the μ_i.

□

4.3. **Applications of Jensen's Inequality.**

4.3.1. *Arithmetic-Geometric Mean.* Let $\phi(x) = e^x$ and let $\mu_i = \frac{1}{n}$. Then Jensens's inequality gives

$$e^{(\frac{x_1+\cdots+x_n}{n})} \leq \frac{e^{x_1} + \ldots e^{x_n}}{n}.$$

If we let $y_i = e^{x_i}$, this inequality reads

$$\left(y_1 \ldots y_n\right)^{\frac{1}{n}} \leq \frac{y_1 + \cdots + y_n}{n}, \quad y_i > 0, \tag{4.8}$$

which is known as the inequality between geometric (left hand side) and arithmetic (right hand side) means. This is best known in the case $n = 2$:

$$\sqrt{ab} \leq \frac{a+b}{2}, \quad a, b > 0. \tag{4.9}$$

4.3.2. *Young's Inequality.* Another variation of the above inequality is with non–equal weigths. Choose two real numbers $p, q > 1$ so that $\frac{1}{p} + \frac{1}{q} = 1$ and again use the convexity of the exponential function to conclude that

$$e^{\frac{x}{p} + \frac{y}{q}} \leq \frac{e^x}{p} + \frac{e^y}{q}.$$

Now let $a = e^{\frac{x}{p}}$ and $b = e^{\frac{y}{q}}$, and we get the inequality

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad \text{if} \quad a, b > 0 \quad \text{and} \quad \frac{1}{p} + \frac{1}{q} = 1. \tag{4.10}$$

4.3.3. *Hölder's Inequality.* This is the inequality, for $x = (x_1, \ldots, x_n), y = (y_1, \ldots y_n) \in \mathbb{R}^n$ and for real numbers $p, q > 1$ satisfying (as above) $\frac{1}{p} + \frac{1}{q} = 1$:

$$|\sum_{i=1}^n x_i y_i| \leq \Big(\sum_{i=1}^n |x_i|^p \Big)^{\frac{1}{p}} \quad \Big(\sum_{i=1}^n |y_i|^q \Big)^{\frac{1}{q}} \tag{4.11}$$

Proof. Let A and B denote the two factors on the right-hand side of this inequality, and observe that we may assume that $A, B > 0$, otherwise the inequality reduces to $0 \leq 0$. Then use the triangle inequality $|\sum x_i y_i| \leq \sum |x_i| \; |y_i|$ and let $a_i = |x_i|/A$, $b_i = |y_i|/B$. Then by the definition of A and B we have

$$\sum a_i^p = 1 \text{ and } \sum b_i^q = 1. \tag{4.12}$$

Next apply the inequality (4.10) to get

$$a_i b_i \leq \frac{a_i^p}{p} + \frac{b_i^q}{q},$$

summing this we get

$$\sum a_i b_i \leq \sum \Big(\frac{a_i^p}{p} + \frac{b_i^q}{q} \Big) = \frac{1}{p} + \frac{1}{q} = 1,$$

where the first equality follows from (4.12) and the second from the definition of p, q. In other words, by the definition of A, B, a_i, b_i:

$$\sum \frac{|x_i|}{A} \frac{|y_i|}{B} \leq 1, \quad \text{equivalently} \quad \sum |x_i| \; |y_i| \leq AB,$$

which gives the desired inequality (4.11). $\square$

4.3.4. *Minkowski's Inequality.* Given a real number $p > 1$, and a vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, let

$$||x||_p = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{\frac{1}{p}} \tag{4.13}$$

which is called the p *norm* of x. Minkowski's inequality justifies calling this a norm, namely, it is the statement of the triangle inequality:

$$||x+y||_p \le ||x||_p + ||y||_p \quad \text{for all} \quad x, y \in \mathbb{R}^n. \tag{4.14}$$

Proof. Write $||x+y||_p^p = \sum |x_i + y_i|^p$ as follows:

$$||x+y||_p^p = \sum |x_i + y_i||x_i + y_i|^{p-1} \le \sum (|x_i| + |y_i|)|x_i + y_i|^{p-1}$$

Split the last sum as the sum of two terms, and apply Hölders inequality (4.11) to each, combine them again, to get

$$||x+y||_p^p \le \Big(\Big(\sum |x_i|^p\Big)\Big)^{\frac{1}{p}} + \Big(\sum |y_i|^p\Big)^{\frac{1}{p}}\Big)\Big(\sum |x_i + y_i|^{(p-1)q}\Big)^{\frac{1}{q}}$$

where q is defined by $\frac{1}{p} + \frac{1}{q} = 1$, which is equivalent to $(p-1)q = p$. Thus the second factor in this last inequality is the same as $||x+y||_p^{\frac{p}{q}}$. Writing this last factor in this form and dividing both sides by it gives

$$||x+y||_p^{p-\frac{p}{q}} \le ||x||_p + ||y||_p.$$

Finally, observe that $p - \frac{p}{q} = 1$ because $\frac{1}{p} + \frac{1}{q} = 1$, therefore this last inequality is Minkowski's inequality (4.14). □

4.4. **Norms on Euclidean Space and Spaces of Sequences.** The last two inequalities can be interpreted in several ways. For fixed n, can define a whole family of norms on $\mathbb{R}^n$, depending on a real number $p \ge 1$, by equation (4.13). Minkowski's inequality tells that these are all norms, they give a one parameter family of norms interpolating the three norms (3.1) for $p = 1, 2, \infty$. Observe the following:

(1) Hölder's inequality (4.11) tells us that these norms come in pairs p, q, where $\frac{1}{p} + \frac{1}{q} = 1$.
(2) Observe that this inequality also holds, and is very easy to prove, in the limiting case $p = 1, q = \infty$.
(3) One pair is exceptional: $p = q = 2$. In this case the 2-norm is the usual Euclidean (or Pythagorean) norm, and Hölder's inequality reduces to the Cauchy-Schwarz inequality. So we can regard the Hölder inequality as a natural generalization of Cauchy-Schwarz.
(4) See Figure 4.1 for a picture of the unit spheres $\{\mathbf{x} : ||\mathbf{x}||_p = 1\}$ of these norms in $\mathbb{R}^2$. Certain features that can be seen from this figure are:

(a) For $r < s$, the unit ball of the r norm is contained in the unit sphere of the s-norm. This translates into the inequality $||\mathbf{x}||_s \leq ||\mathbf{x}||_r$ for $r < s$, and equality holds if and only if all but one of the coordinates of $\mathbf{x}$ vanish.

(b) The exceptional case $p = q = 2$ is more symmetric than the others: it has rotational symmetry, while the others have very little symmetry. More precisely, if A is a 2 by 2 matrix, $||A\mathbf{x}||_p = ||\mathbf{x}||_p$ for all p happens:

 (i) For $p = 2$ if and only if
$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

 (ii) For $p \neq 2$ if and only if
$$A = \begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix} \text{ or } \begin{pmatrix} 0 & \pm 1 \\ \pm 1 & 0 \end{pmatrix}$$

(c) For the extreme cases $p = 1, \infty$ the unit ball is not strictly convex, meaning that there are straight line segments in its boundary. In the other cases $1 < p < \infty$, the unit ball is strictly convex: no straight line segments contained in its boundary.

(d) An equivalent formulation of the last statement is the following: Let
$$E_p(\mathbf{x}) = \{\mathbf{y} : ||\mathbf{x}||_p = ||\mathbf{y}||_p + ||\mathbf{x} - \mathbf{y}||_p\}$$
(the equality set for the triangle inequality). For $1 < p < \infty$, $\mathbf{y} \in E_p(\mathbf{x})$ if and only if $\mathbf{y} = a\mathbf{x}$ for some $a \in \mathbb{R}$, $0 \leq a \leq 1$. In other words, $E_p(\mathbf{x})$ is the straight line segment from the origin to $\mathbf{x}$. For $p = 1, \infty$, $E_p(\mathbf{x})$ is much larger, in fact, it has non-empty interior.

4.4.1. *Equivalence of norms in $\mathbb{R}^n$.*

Theorem 4.2. *Let $||x||$ and $||x||'$ be any two norms on $\mathbb{R}^n$. Then they are equivalent: there exist constants $C_1, C_2 > 0$ so that $C_1||x|| \leq ||x||' \leq C_2||x||$.*

Proof. It is enough to prove that any norm is equivalent to the 1-norm $||x||_1$. Let $||x||$ be a norm. Writing $x = (x_1, \ldots, x_n) = x_1 e_1 + \ldots x_n e_n$, where $e_1, \ldots, e_n$ are the standard basis vectors, the triangle inequality gives
$$||x|| \leq |x_1|\ ||e_1|| + \ldots |x_n|\ ||e_n|| \leq C||x||_1, \quad \text{where} \quad C = \max\{||e_i||\}.$$
This gives one of the two desired inequalities, and shows that the funcion $||x||$ is *continuous* with respect to $||x||_1$. Note that a norm is always continuous with respect to the distance it defines, what may not be clear a-priori is that it is continuous with respect to another norm.

Now, since $||x||$ is continuous with respect to the usual norm (or the usual topology), and the unit sphere $S = \{x \in \mathbb{R}^n : ||x||_1 = 1\}$ is *compact*, $||x||$

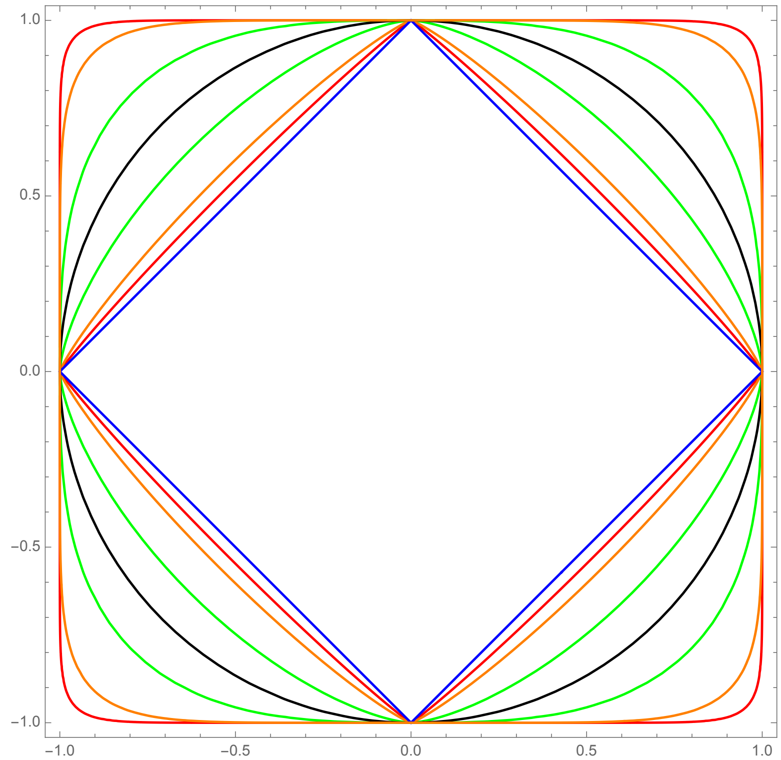

FIGURE 4.1. The Unit Spheres of the l^p-norms

has a minimum on S. Let M be this minimum value of $||x||$ for $x \in S$. Then $M > 0$ since $||x|| = 0$ implies $x = 0$ and $0 \notin S$. Then for all $x \neq 0$ we have $||\frac{x}{||x||_1}|| \geq M$ or $||x|| \geq M||x||_1$ as desired. □

4.4.2. *Inequivalent norms in infinite dimensions.* Let's go back to the space $\mathbb{R}^\infty$, where we saw in section (4.1.1) that the norms $||x||_1, ||x||_2, ||x||_\infty$ were defined, and, by the same reasoning, we see that the norms $||x||_p$ are defined for all real $p \geq 1$. These norms are *not equivalent.* We have seen that $||x||_s \leq ||x||_r$ whenever $r < s$, but not in the opposite direction. Let's look at what goes wrong with the proof of Theorem 4.2. Let's take $r = 1$ and $s > 1$. We know that it is true (even though the argument in the proof of Theorem 4.2 doesn't work) that $||x||_s \leq ||x||_1$. Thus we know, in particular, that $||x||_s$ is continuous in $\mathbb{R}^\infty$ with respect to $||x||_1$. But it is not bounded away from zero on the unit sphere of the 1-norm. Let $x_n = (\frac{1}{n}, \dots, \frac{1}{n}, 0, 0, \dots)$ (thus x_n has precisely n nonzero entries, all equal to $\frac{1}{n}$). Then $||x_n||_1 = 1$, so it is on the unit sphere, but for any $s > 1$

$$||x_n||_s = \left(\sum_{i=1}^{n} \frac{1}{n^s}\right)^{\frac{1}{s}} = \left(\frac{n}{n^s}\right)^{\frac{1}{s}} = \frac{1}{n^{\frac{s-1}{s}}} \to 0.$$

Therefore s-norm is not bounded away from zero on the unit sphere of the 1-norm, which gives us another proof that these norms are not equivalent on the infinite dimensional space $\mathbb{R}^\infty$, as we have already seen in section (4.1.1).

4.4.3. *More examples of incomplete metric spaces.* The spaces $\mathbb{R}^\infty$ with norm $||\mathbf{x}||_s$, $1 \le s \le \infty$ provide more examples of incomplete metric spaces. Take, for concreteness, $k = 1$ and consider the sequence $\{\mathbf{x}_n\}$ in $\mathbb{R}^\infty$ defined by

$$\mathbf{x}_n = (1, \frac{1}{2^2}, \frac{1}{3^2}, \dots, \frac{1}{n^2}, 0, 0, \dots). \tag{4.15}$$

Then $\{\mathbf{x}_n\}$ is a Cauchy sequence, since for $m < n$,

$$||\mathbf{x}_m - \mathbf{x}_n||_1 = \sum_{i=m+1}^{n} \frac{1}{i^2}, \tag{4.16}$$

which is the "tail end" of the convergent series $\sum \frac{1}{i^2}$, hence given any $\epsilon > 0$ there is an N so that $\sum_{m+1}^{n} \frac{1}{i^2} < \epsilon$ whenever $n > m > N$. But $\{\mathbf{x}_n\}$ is not convergent: given any fixed $\mathbf{y} \in \mathbb{R}^\infty$, since there is some fixed k so that $\mathbf{y} = (y-, \dots, y_k, 0, 0, \dots)$, we see that for $n > k$, $||\mathbf{x}_n - \mathbf{y}||_1 \ge \frac{1}{(k+1)^2}$, thus $\mathbf{x}_n\}$ cannot have limit $\mathbf{y}$. Since this holds for any $\mathbf{y} \in \mathbb{R}^\infty$, $\{\mathbf{x}_n\}$ cannot have a limit in $\mathbb{R}^\infty$ using the 1-norm.

Since $\sum \frac{1}{i^{2s}}$ converges, the same $\mathbf{x}_n$ can be used to show that $\mathbb{R}^\infty$ is not complete in any of the s-norms for $1 < s < \infty$. For the ∞-norm the inequality (4.16) is replaced by $||\mathbf{x}_m - \mathbf{x}_n||_\infty = \frac{1}{(m+1)^2}$, so again we get a Cauchy sequence. Since the same lower estimate for $||\mathbf{x}_n - \mathbf{y}||_\infty$ holds, this is not a convergent sequence in the ∞-norm.

4.4.4. *Spaces of infinite sequences.* We have seen these inequivalent norms on $\mathbb{R}^\infty$, and we have seen that they are not complete. One natural way to try to get complete spaces would be to allow infinite sequences that satisfy the summability conditions that make the p-norm finite. In other words, define first a space of arbitrary infinite sequences:

$$\mathbb{R}^\omega = \{(x_1, x_2, \dots) : x_i \in \mathbb{R}\} \tag{4.17}$$

and then take subspaces (denoted l^p) that satisfy summability conditions:

$$l^p = \{x \in \mathbb{R}^\omega : \sum_{i=1}^{\infty} |x_i|^p < \infty\}. \tag{4.18}$$

On l^p we can define a norm $||x||_p$ by

$$||x||_p = \Big(\sum_{i=1}^{\infty} |x_i|^p \Big)^{\frac{1}{p}} \tag{4.19}$$

We will later use some of these spaces, in particular l^2.

4.4.5. *Norms on function spaces.* We have seen in sections (3.1.3) and (3.2.3) examples of norms on spaces of continuous functions. Most of what we have said about norms in finite dimensional spaces or spaces of sequences can also be said for spaces of functions, replacing sums by integrals. This is true of Jensen's inequality, therefore the Hölder and Minkowski inequalities hold for integrals. This means, in particular, that the p-norms can be defined. We should see more examples later, particularly of applications of the $1, 2$ and ∞-norms on spaces of functions.

5. The Contraction Mapping Theorem

Definition 5.1. Let (X, d) be a metric space. A map $f : X \to X$ is called a *contraction* if there exists a constant $C < 1$ so that $d(f(x), f(y)) \leq C\, d(x, y)$ for all $x, y \in X$.

In other words, a contraction is a Lipschitz map with Lipschitz constant $C < 1$.

Theorem 5.1. *Let (X, d) be a complete metric space, and let $f : X \to X$ be a contraction. Then f has a unique fixed point. That is, there is a unique $x \in X$ so that $f(x) = x$.*

Proof. This is Theorem 9.23 of [5]. □

5.1. Newton's Method.

One example of how Theorem 5.1 could be applied is to Newton's method for solving an equation $f(x) = 0$. Of course Newton's method is a couple of centuries older than Theorem 5.1 and gives better results. But it is still interesting to see how it is a fixed point theorem.

Recall that the method finds a solution of $f(x) = 0$ by starting from some guess x_1 and improving it by drawing the tangent line to f at $(x_1, f(x_1))$ and letting x_2 be the intersection of this tangent line with the x-axis. In other words, x_2 is the solution of $f(x_1) + f'(x_1)(x - x_1) = 0$, or

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

and this is the first step of an iteration

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{5.1}$$

which should converge to a solution.

The first thing that is evident from this formula is that this could only work for a solution x with $f'(x) \neq 0$. The next thing that is evident is that the solution is a fixed point, and the iteration method is the same as used in the proof of Theorem 5.1. This suggests we should define

$$N(x) = x - \frac{f(x)}{f'(x)} \tag{5.2}$$

It is clear that $N(x) = x$ if and only if $f(x) = 0$, and (5.1) is the iteration $x_{i+1} = N(x_i)$ used to find the fixed point of a contraction.

To see if N is a contraction, we compute its derivative:

$$N'(x) = \frac{f(x)f''(x)}{f'(x)^2}, \tag{5.3}$$

and use the reasoning of Example 3.1 (2): a bound on N' gives Lipschitz constant, so we want intervals where $|N'(x)| \leq C < 1$. Assume that f is twice continuously differentiable, and that x_0 is a zero of f where $f'(x_0) \neq 0$. Then we can choose constants $c_1, c_2, c_3, c_4 > 0$ so that $0 < c_2 \leq |f'(x)| \leq c_3$ and $|f''(x)| \leq c_4$ for $|x - x_0| \leq c_1$. (First choose a positive constant $c_2 < |f'(x_0)|$, since f' is continuos we can find c_1 so that $c_2 \leq |f'(x)|$ for $|x - x_0| \leq c_1$. Then $|f'|$ and $|f''|$ are bounded on $|x - x_0| \leq c_1$, choose some upper bounds c_3 and c_4 for these quantities.)

Since $|f'(x)| \leq c_3$ and $f(x_0) = 0$, we get

$$|f(x)| \leq c_3|x - x_0| \text{ on } |x - x_0| \leq c_1 \tag{5.4}$$

This gives the estimates

$$|N'(x)| \leq c_5|x - x_0| \text{ on } |x - x_0| \leq c_1, \text{ where } c_5 = \frac{c_3c_4}{c_2^2}. \tag{5.5}$$

Choosing $c_6 < c_1$ and $< 1/c_5$, we get that for $|x - x_0| \leq c_6$

$$|N'(x)| \leq C < 1, \text{ where } C = c_5c_6, \tag{5.6}$$

and by (2) of Example 3.1 we get that N is a contraction on $|x - x_0| \leq c_6$. This means, in particular, that if we start the iteration (5.1) sufficiently close to x_0 it will converge to x_0.

But the situation is really much better. The estimate (5.5) gives a much stronger estimate

$$|N(x) - x_0| = |N(x) - N(x_0)| \leq c_5|x - x_0|^2, \tag{5.7}$$

called quadratic convergence. In practice it means that the number of accurate digits doubles in each iteration (rather than increase by one as in a geometric series).

To see some concrete examples, Figure 5.1 shows the graph of the Newton map N for $f(x) = x^2 - 2$, and the line $y = x$. The fixed points of N are the points where the two meet. The derivative at the fixed points $(\pm\sqrt{2})$ are visibly less than one in absolute value, so they are clearly attractors. In this case the iteration always converges to one of the fixed points in a predictable way.

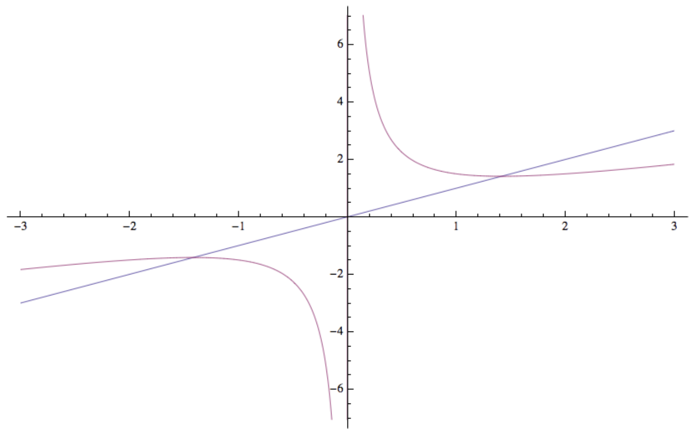

FIGURE 5.1. The Newton Map for $x^2 - 2$

Figure 5.2 shows the Newton map for the cubic polynomial $f(x) = x^3 - x$ with three real roots, $0, \pm 1$. Again visibly $|N'| < 1$ at the fixed points. The convergence to a fixed point its not predictable. It is worth experimenting with a computer by taking several initial values, say, in $(0.4, 0.6)$, and seeing which of the three fixed points it converges to.

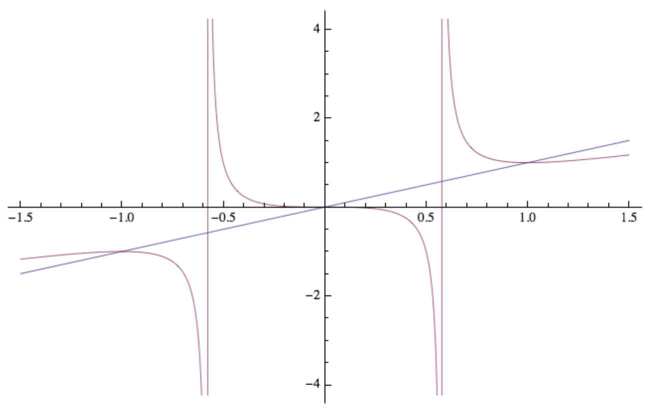

FIGURE 5.2. The Newton Map for $x^3 - x$

One experiment, starting at 0.4, 0.45, 0.455, 0.5, 0.6 produced the following list:

{0.4, -0.2461, 0.0364, -0.0000972 }

{0.45, -0.464, 0.567, -10.156, -6.79, -4.56, -3.09, -2.13, -1.53, -1.19, -1.038}

{0.455, -0.497, 0.951, 1.004, 1.00002, 1., 1.}

{0.5, -1. ,-1.}

$\{0.6, 5.4, 3.64, 2.490, 1.754, 1.3117, 1.0846, 1.00897, 1.00012, 1., 1.\}$

In other words, starting at 0.4 converges very rapidly to 0, while 0.45 converges more slowly to -1, 0.455 leads more rapidly to 1, 0.5 rapidly to -1, 0.6 slowly to 1.

This alternation of convergence, and the drastic difference in the speed of convergence (notice that starting at 0.45 we get to -10 before approaching -1, while starting at 0.5 we quickly approach -1, and starting at 0.4 we quickly approach 0) is hard to understand looking just at real numbers. The pictures become more clear (and more appealing) if we use the complex plane. This leads to the subject of *complex dynamics*.

We quickly illustrate with this example: consider the *complex* polynomial $f(z) = z^3 - z$ and its *complex Newton map* $N(z) = z - f(z)/f'(z) = 2z^3/(3z^2 - 1)$. If we iterate N most points converge to one of the roots $0, \pm 1$. The collection of these points form an open set, called he *Fatou set*. Its complement is a closed set, called the *Julia set*. The way that the Julia set divides the complex plane into the regions of attraction of the three fixed points of N (in other words, into the connected components of the Fatou set) is not an obvious one: The Fatou set has infinitely many connected components, and the number of iterations needed to get close to the limit can vary widely within each component.

A picture of what happens is in Figure 5.3. Note that the colors represent the number of iterations needed, with the color scale explained next to the picture. There are three large regions, each converging to one of the 3 fixed points, as expected. From the distance the separating set looks like two curves, and the picture can be explained by looking at the level sets of $|N'|$ in Figure 5.4.

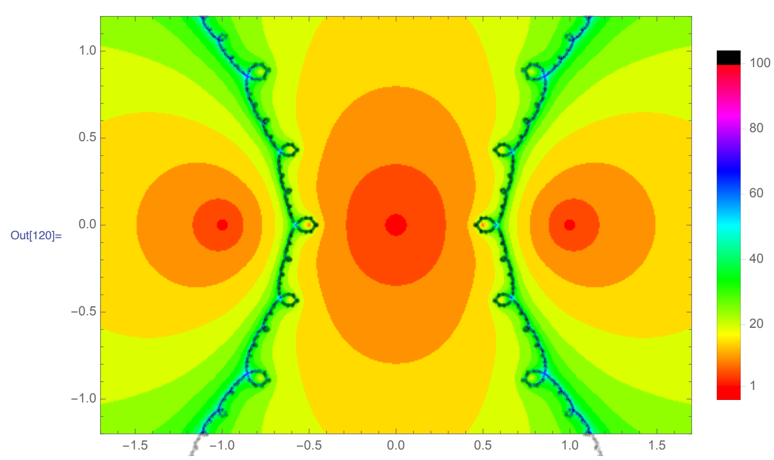

FIGURE 5.3. Complex Dynamics of the Newton map of $z^3 - z$

The three blue regions are the sets where $|N'(z)| < 0.5$, while the white regions are sets where $|N'(z)| > 3.5$. The intermediae level curves represent increases of 0.5 units.

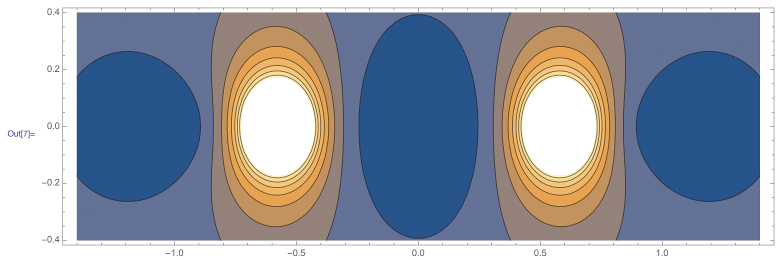

FIGURE 5.4. Level sets of $|N'|$

A closer look at Figure 5.3 shows that the Julia set is much more complicated than what Figure 5.4 may suggest. In particular the Julia set encloses other compoinents of the Fatour set, which contain more divisions, and so on. For example, Figure 5.5 shows what happens in a neighborhood of the interval $[0.4, 0.6]$.

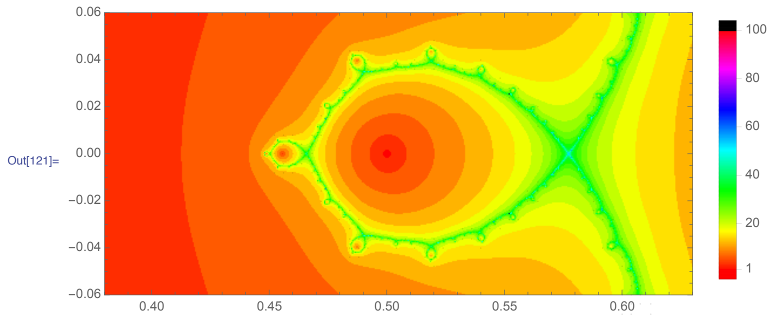

FIGURE 5.5. Interval $[0.4, 0.6]$

This explains why convergence is so fast starting at 0.4, since this is inside a red region, of fastest conergence, to the fixed point 0. And visibly there are 4 connected components that the interval $[0.4, 0.6]$ crosses. But we found 5 changes of behavior in this interval. To see why there should be one more change around 0.455, let's blow up the interval $[0.445, 0.470]$, as shown in Figure 5.6

We see from the figure that there is another, much smaller picture, but similar to the larger one, showing another connected component of the Fatou set. In other words, the Julia set is a *fractal*.

5.2. **Picard's Method.** A more substantial use of the contraction mapping theorem is Picard's method for solving the initial value problem of a first

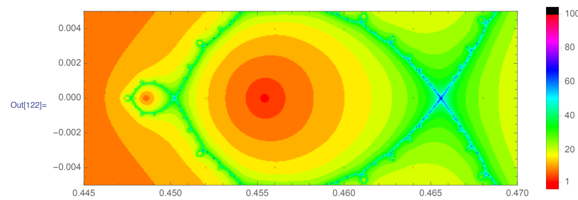

FIGURE 5.6. Interval $[0.445, 0.470]$

order differential equation. More details can be found in many books on ODE's. We roughly follow §31 of [1].

Start from a function $f(t, x)$ defined on an open subset $U \subset \mathbb{R}^2$. Assume, for simplicity, that U is an open rectangle, possibly infinite, that is, $U = (a, b) \times (c, d)$ for $-\infty \leq a < b \leq \infty$ and $-\infty \leq c < d \leq \infty$. We think of the first coordinate t as time and the second coordinate x as position, and $f(t, x)$ as a "slope" at (t, x), since $f(t, x)$ defines a differential equation

$$\begin{aligned} \frac{dx}{dt} &= f(t, x(t)) \\ x(0) &= x_0 \end{aligned} \tag{5.8}$$

where $(0, x_0) \in U$, for a function $x(t)$. We are also given an initial condition $x(0) =$ some given number x_0. (Could have used an initial condition $x(t_0) = x_0$ for any fixed value t_0). We have to assume that $(0, x_0) \in U$, the set where f is defined. The tangent line to the curve $(t, x(t))$ has slope $f(t, x(t))$ and the curve passes through $(0, x_0)$.

The purpose of this section is to prove, by what is known as Picard's method, the following version of the existence and uniqueness theorem for first order differential equations:

Theorem 5.2. *Suppose that $f : U \to \mathbb{R}$ is continuous and satisfies a local, time - independent Lipschitz conditon on U, meaning that on every closed sub-rectangle $R = [a, b] \times [c, d] \subset U$, $a, b, c, d \in \mathbb{R}$, there is a constant $c_R > 0$ so that $|f(t, x) - f(t, y)| \leq c_R\ |x - y|$ for all x, y, t so that (t, x) and (t, y) are in U.*

Then there exist numbers $a, b > 0$ so that the rectangle $[-a, a] \times [x_0 - b, x_0 + b] \subset U$ and so that (5.8) has a unique solution with graph contained in this rectangle, that is, $x(t)$ is defined for $|t| < a$ and satisfies $|x(t) - x_0| < b$ for $|t| < a$. Moreover, $x(t)$ is a continuously differentiable function of t.

Remark 5.1. One sufficient condition on $f : U \to \mathbb{R}$ to satisfy the assumptions of the theorem is that it be continuously differentiable. Then, on each closed rectangle $R \subset U$ there is a bound c_R for $\partial f/\partial x$, in other words, $|\partial f/\partial x(t, s)| \leq c_R$ for all $(x, t) \in R$. We chose to emphasize the Lipschitz condiition in x of $f(t, x)$ because it is, in some sense, the optimal condition for both existence and uniqueness of solutions. Example 5.3 below shows that this condition is not always needed for existence, but it is essential for *uniqueness*

Remark 5.2. Note that the theorem only gives *local* existence of solutions of (5.8), meaning that, even if $U = \mathbb{R}^2$ and the equation (5.8) is defined for all t, the solution need only exist for $|t| < a$, This is not a flaw in the statement, it is the way it really is, as example 5.2 shows.

Example 5.1. Let $f(t, x) = x$, then the solution of (5.8) is $x(t) = x_0 e^t$.

Example 5.2. Let $f(t, x) = x^2$, and take $x_0 = 1$. Then the solution of (5.8) is $x(t) = \frac{1}{1-t}$. Observe that, even though $f(t, x)$ is defined for all (t, x), the solution $x(t)$ is defined only on $(-\infty, 1)$. This shows that the existence theorem can only be local, justifying Remark 5.2. In this example the number a in the statement of the existence theorem cannot be larger than one. In other words, the solution "blows up in finite time".

Example 5.3. Let $f(t, x) = x^{1/2}$. Then f does not satisfy a Lipschitz condition in x at $x = 0$, see Example 3.2. The initial value problem $dx/dt = x^{1/2}$, $x(0) = 0$ has solution $x(t) = t^2/4$ (obtained by the usual method of separation of variables), but also has the solution $x(t) = 0$, both infinitely differentiable. Moreover, it has two additional solutions, both continuously differentiable but not twice differentiable, by patching together one of the above solutions for $t \leq 0$ with the other for $t \geq 0$,

In the same way, $dx/dt = x^{2/3}$, $x(0) = 0$ has four solutions: $x(t) = t^3/27$, $x(t) = 0$, both infinitely differentiable, and two twice continuously differentiable (but not three times differentiable) solutions obtained by patching these two as above. And so on with $f(t, x) = x^{(n-1)/n}$.

The first step of Picard's method is to convert (5.8) to an integral equation: The fundamental theorem of calculus gives

$$x(t) - x(0) = \int_0^t \frac{dx}{d\tau} d\tau = \int_0^t f(\tau, x(\tau)) d\tau$$

So $x(t)$ satisfies (5.8) if and only if it satisfies

$$x(t) = x_0 + \int_0^t f(\tau, x(\tau)) d\tau. \tag{5.9}$$

which is a fixed-point problem for the function $x(t)$: if we define

$$(Px)(t) = x_0 + \int_0^t f(\tau, x(\tau)) d\tau. \tag{5.10}$$

then $x = x(t)$ satisfies (5.8) if and only if

$$Px = x. \tag{5.11}$$

Exercise 1. Apply the iteration procedure of the proof of Theorem 5.1 to $f(t, x) = x$ and $x_0 = 1$ (see Example 5.1): $x_{i+1} = P(x_i)$ with $x_1 = 1$. Check that you get the partial sums of the power series for e^t. Then take $x_1 = 100$ and see what you get from the iteration.

The challenge in solving (5.11) is to find the right metric space (X, d) of functions so that $P : X \to X$ is a contraction.

To do this, calculations are easier if we let

$$\psi(t) = x(t) - x_0 \tag{5.12}$$

so that $\psi(0) = 0$ and (5.11) becomes

$$Q\psi = \psi, \text{ where } Q\psi = P(x_0 + \psi) - x_0. \tag{5.13}$$

Explicitly

$$Q\psi(t) = \int_0^t f(\tau, x_0 + \psi(\tau))d\tau. \tag{5.14}$$

To solve (5.14), introduce constants $c_1, \dots c_4$ as follows:

(1) Choose c_1 and c_2 so that the rectangle

$$R = \{(t, x) : |t| \le c_1, |x - x_0| \le c_2\} \subset U. \tag{5.15}$$

(2) Choose c_3 so that $|f(t, x)| \le c_3$ for all $(t, x) \in R$,
(3) Choose c_4 to be the Lipschitz constant for R: $|f(t, x) - f(t, y)| \le c_4|x - y|$ for all t, x, y with $(t, x), (t, y) \in R$.

If $|\psi(t)| \le c_2$ for $|t| \le c_1$ the following integral is defined and satisfies the estimate

$$|Q\psi(t)| \le \int_0^{|t|} |f(\tau, x_0 + \psi(\tau))|d\tau \le \int_0^{|t|} c_3 d\tau = c_3|t|. \tag{5.16}$$

If $|\psi_1(t)|, |\psi_2(t)| \le c_2$ for $|t| \le c_1$, then the following integral is defined and satisfies the estimate

$$(5.17)\quad |Q\psi_1(t) - Q\psi_2(t)| \le \int_0^{|t|} |f(\tau, x_0 + \psi_1(\tau)) - f(\tau, x_0 + \psi_2(\tau))|d\tau \le$$
$$\int_0^{|t|} c_4|\psi_1(\tau) - \psi_2(\tau)|d\tau \le c_4\ |t| \max_{|t| \le c_1}(|\psi_1(t) - \psi_2(t)|).$$

With these estimates we will get that Q is a contraction on a suitable space of functions. Choose $a > 0$ so that

$$a < \min(c_1, c_2/c_3, 1/c_4). \tag{5.18}$$

and define a space

$$X = \{\psi \in C([-a, a]) : |\psi(t)| \le c_3|t| \text{ for all } t \in [-a, a]\}, \tag{5.19}$$

with the d_∞ metric.

Theorem 5.3. *If $\psi \in X$, then $Q\psi$ is defined and $Q\psi \in X$, so $Q : X \to X$. The space X is closed in $(C([-a,a]), d_\infty)$, so it is complete. There is a constant $C < 1$ such that $d_\infty(Q\psi_1, Q\psi_2) \leq C\, d_\infty(\psi_1, \psi_2)$ for all $\psi_1, \psi_2 \in X$. Therefore there exists a unique $\psi \in X$ so that $Q\psi = \psi$.*

Proof. If $\psi \in X$, then (5.18) gives

(1) $|\psi(t)| \leq c_3 a < c_3(c_2/c_3) = c_2$, and $a < c_1$, $(t, x_0 + \psi(t)) \in R$ (the rectangle (5.15)) for all $t \in [-a, a]$, so $Q\psi$ is defined.
(2) (5.16) gives that $Q\psi \in X$, so $Q : X \to X$.
(3) (5.17) gives that $d_\infty(Q\psi_1, Q\psi_2) \leq C d_\infty(\psi_1, \psi_2)$ for all $\psi_1, \psi_2 \in X$, where $C = a\, c_4 < 1$ by the choice (5.18) of a.

Therefore $Q : X \to X$ and is a contraction.

Finally, to see that X is closed in $(C([-a,a]), d_\infty)$, observe that if $\psi_n \in X$ and $\psi_n \to \psi$ uniformly on $[-a, a]$, then $\psi \in X$ (Proof: if $|\psi_n(t)| \leq c_3|t|$ holds for all n and t and, given $\epsilon > 0$ there is $N(\epsilon)$ so that $|\psi_n(t) - \psi(t)| < \epsilon$ for all $n > N$ and all t, then $|\psi(t)| \leq |\psi(t) - \psi_n(t)| + |\psi_n(t)| < \epsilon + c_3|t|$ holds for all t and all $\epsilon > 0$, so $|\psi(t)| \leq c_3|t|$.)

Since a closed subset of a complete metric space is complete, we can apply the contraction mapping theorem to conclude that (5.13) has a unique solution in X. □

Proof of Theorem 5.2. We have seen that the initial value problem (5.8) is equivalent to the integral equation (5.9), which in turn is equivalent to (5.13). By Theorem 5.3 (5.13) has a unique solution in the space X, therefore (5.9) has a unique solution in the space $x_0 + X$, therefore a unique continuous solution. Since $x(\tau)$ is continuous, the right hand side of (5.9) is visibly continuously differentiable, so we conclude that $x(t)$ must actually be continuously differentiable and therefore the unique solution of (5.8). □

6. The Inverse Function Theorem

Let $U \subset \mathbb{R}^k$ be an open set and let $f : U \to \mathbb{R}^n$ be *differentiable*. Recall that this means that for every $x \in U$ there exists a linear transformation $d_x f : \mathbb{R}^k \to \mathbb{R}^n$, called the *differential* of f at x, so that for all h with $x + h \in U$,

$$f(x+h) - f(x) = (d_x f)(h) + \epsilon(x,h) \text{ where } \frac{||\epsilon(x,h)||}{||h||} \to 0 \text{ as } h \to 0. \tag{6.1}$$

The matrix of the linear transformation $d_x f$ in the standard bases for $\mathbb{R}^k$ and $\mathbb{R}^n$ is the *Jacobian matrix* of f

$$\begin{pmatrix} f^1_{x_1} & \cdots & f^1_{x_k} \\ \cdots & \cdots & \cdots \\ f^n_{x_1} & \cdots & f^n_{x_k} \end{pmatrix} \tag{6.2}$$

where $f(x) = (f^1(x_1 \dots, x_k), f^2(x_1, \dots x_k), \dots, f^n(x_1 \dots, x_k))$ and $f^i_{x_j} = \partial f^i / \partial x_j$.

Let $L(\mathbb{R}^k, \mathbb{R}^n)$ denote the vector space of linear transformations from $\mathbb{R}^k$ to $\mathbb{R}^n$, a kn-dimensional vector space isomorphic to the space of k by n matrices, in turn isomorphic to $\mathbb{R}^{kn}$. We say that f is *continuously differentiable* in U, or *of class* C^1 in U, if f is differentiable at every $x \in U$ and the map $df : U \to L(\mathbb{R}^k, \mathbb{R}^n)$ is continuous. This is equivalent to saying that all partial derivatives $\partial f^i / \partial x_j$ exist and are continuous in U.

The purpose of this section is to prove the following theorem, known as the *Inverse Function Theorem*:

Theorem 6.1. *Let $U \subset \mathbb{R}^k$ be an open set, let $f : U \to \mathbb{R}^k$ be of class C^1 (continuously differentiable). Let $x_0 \in U$ and suppose that $d_{x_0} f : \mathbb{R}^k \to \mathbb{R}^k$ is invertible. Then there exist neighborhoods $N(x_0) \subset U$ of x_0 and $N(y_0)$ of $y_0 = f(x_0)$ so that $f(N(x_0)) = N(y_0)$ and the restriction of f to $N(x_0)$, denoted $f|_{N(x_0)} : N(x_0) \to N(y_0)$ is bijective, so it has an inverse. This inverse map $(f|_{N(x_0)})^{-1} : N(y_0) \to N(x_0)$ is also of class C^1.*

In this statement, by a neighbothood of a point we mean an open set containing the point (not necessarily a ball). The proof will take the rest of the section. It roughly follows the proof of the inverse function theorem for Banach spaces given in §5 of [2].

6.1. **The one-dimensional case.** This theorem is quite simple if $k = 1$. If $f'(x_0) \neq 0$, say $f'(x_0) > 0$, then $f'(x) > 0$ in some open interval I containing x_0, thus f is strictly increasing on I, in particular injective on I. Choose a, b so that $[a, b] \subset I$ and $x_0 \in (a, b)$. If $\alpha = f(a)$ and $\beta = f(b)$, then, given any $y \in (\alpha, \beta)$, by the intermediate value theorem there exists an $x \in (a, b)$ so that $f(x) = y$, and we have already seen that x is unique. So $f^{-1} : (\alpha, \beta) \to (a, b)$ is defined (namely, $f^{-1}(y) = x$). Moreover, since f maps open intervals onto open intertvals, it is an open map, so f^{-1} is continuous.

To see that f^{-1} is of class C^1, use the mean value theorem

$$f(x_2) - f(x_1) = f'(\xi)(x_2 - x_1) \text{ for some } \xi \in (x_1, x_2). \tag{6.3}$$

Letting $y_i = f(x_i)$, this is equivalent to

$$f^{-1}(y_2) - f^{-1}(y_1) = \frac{1}{f'(\xi)}(y_2 - y_1),$$

thus, letting $y_2 \to y_1$ we get that f^{-1} is differentiable at y_1 and $(f^{-1})'(y_1) = 1/f'(x_1) = 1/f'(f^{-1}(y_1))$. Since f^{-1} and f' are continuous, so is $(f^{-1})'$, and the proof is complete.

Note the two ingredients in this proof:

- The Mean Value Theorem (6.3) used to prove injectivity (this is how you prove positive derivative implies strictly increasing) and to prove that the inverse, once known to exist, was of class C^1.
- The Intermediate Value Theorem used to prove that the image of an interval is a whole interval.

We will need to find similar ingredients in higher dimensions. These wil be:

- Lemma 6.2 below to replace the Mean Value Theorem.
- The contraction mapping theorem in §6.5 to replace the Intermediate Value Theorem.

6.2. **Linear Transformations.** We will need some facts about linear transformations $A \in L(\mathbb{R}^k, \mathbb{R}^n)$. Recall first of all the definition of the norm:

$$||A|| = \max\{||Ax|| : ||x|| = 1\} = \sup\{||Ax||/||x|| : x \neq 0\} \tag{6.4}$$

which, by the first definition, is clearly well-defined, and by the equivalent second characterization, gives us the basic estimate

$$||Ax|| \leq ||A||||x|| \text{ for all } x \in \mathbb{R}^k. \tag{6.5}$$

If we compose with a linear transformation $B \in L(\mathbb{R}^m, \mathbb{R}^k)$, from $||ABx|| \leq ||A||||Bx|| \leq ||A||||B||||x||$ we get the inequality

$$||AB|| \leq ||A||||B|| \tag{6.6}$$

In particular, if $A \in L(\mathbb{R}^k, \mathbb{R}^k)$ is invertible, from this inequality and $||I|| = 1$, we get $1 \leq ||A||||A^{-1}||$ or $||A^{-1}|| \geq ||A||^{-1}$.

Furthermore, if we use $||x|| = ||A^{-1}Ax|| \leq ||A^{-1}||\ ||Ax||$ we get $||Ax|| \geq (||A^{-1}||^{-1})||x||$, which means that an invertible linear transformation is bi-Lipschitz:

$$(||A^{-1}||^{-1})||x|| \leq ||Ax|| \leq ||A||||x|| \text{ for all } x \in \mathbb{R}^k. \tag{6.7}$$

As a side remark, let's give a formula for $||A||$:

$$||A|| = \sqrt{\lambda_{\max}} \text{ where } \lambda_{\max} = \text{ largest eigenvalue of } A^tA. \tag{6.8}$$

Here A^t denotes the transpose matrix of A. The matrix A^tA is then a symmetric, positive semi-definite matrix, hence it is diagonalizable and its eigenvalues are non-negative. Thus $\lambda_{\max}$ makes sense, and is non-negative, so we can take its square root.

To prove the formula (6.8), write x as a column vector, then $||x||^2 = x^tx$, $||Ax||^2 = (Ax)^t(Ax) = x^tA^tAx = x^t(A^tA)x$. Changing to an orthonormal basis of eigenvectors for A^tA with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$, and letting $y_1, \ldots y_k$ be the components of x in this basis, we get

$$||Ax||^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots \lambda_k y_k^2,$$

which clearly has maximum value λ_1 on the unit sphere $y_1^2 + \cdots + y_k^2 = 1$, attained at $y_1 = 1$, all other $y_i = 0$. Thus $||Ax||$ has maximum value $\sqrt{\lambda_1} = \sqrt{\lambda_{\max}}$.

Finally we will need one other fact about linear transformations:

Lemma 6.1. *Let $G = \{A \in L(\mathbb{R}^k, \mathbb{R}^k) : A$ is invertible$\} \subset L(\mathbb{R}^k, \mathbb{R}^k)$. Then G is an open set and the map $G \to G$ that takes A to A^{-1} is continuous.*

Proof. $A \in G$ if and only if $\det(A) \neq 0$, and $\det : L(\mathbb{R}^k, \mathbb{R}^k) \to \mathbb{R}$ is continuous, so $\det^{-1}(\mathbb{R} \setminus \{0\})$ is open. If $A \in G$, there is an explicit formula for A^{-1} as the transpose of the matrix of cofactors divided by $\det(A)$, which shows that A^{-1} is continuous. □

6.3. **A Characterization of C^1-Maps.** Let $U \subset \mathbb{R}^k$ be an open, convex set, and let $f : U \to \mathbb{R}^n$ be a function.

Lemma 6.2. *Let $f : U \to \mathbb{R}^n$ be as above. Then f is of class C^1 if and only if there exists a continuous function $A : U \times U \to L(\mathbb{R}^k, \mathbb{R}^n)$ such that for all $x_1, x_2 \in U$,*

$$f(x_2) - f(x_1) = A(x_1, x_2)(x_2 - x_1). \tag{6.9}$$

Moreover, if $A : U \times U \to L(\mathbb{R}^k, \mathbb{R}^n)$ with the stated properties exists, then we must have that $A(x, x) = d_x f$ holds for all $x \in U$.

Proof. Suppose f is C^1. The method used in Example (3) of subsection 3.4 allows us to find A. Let $x_1, x_2 \in U$, and let $\gamma_{x_1,x_2}(t) = (1-t)x_1 + tx_2$ be the straight line segment from x_1 to x_2, which lies in U by convexity. This is a continuous map $U \times U \times [0, 1] \to U$. Then the fundamental theorem of calculus and the chain rule give

$$f(x_2) - f(x_1) = \int_0^1 \frac{d}{dt} f(\gamma_{x_1,x_2}(t))dt = \int_0^1 (d_{\gamma_{x_1,x_2}(t)} f)(x_2 - x_1)dt.$$

Thus, if we let

$$A(x_1, x_2) = \int_0^1 d_{\gamma_{x_1,x_2}(t)} f dt,$$

then $A(x_1, x_2)$ satisfies (6.9). Observe that this formula gives $A(x, x) = d_x f$, since $\gamma_{x,x}$ is a constant path.

Converselry, suppose that $A : U \times U \to L(\mathbb{R}^k, \mathbb{R}^n)$ satisfying (6.9) exists. Then

$$f(x + h) - f(x) = A(x, x + h)h = A(x, x)h + (A(x + h, x) - A(x, x))h$$

that is,

$$f(x + h) - f(x) = A(x, x)h + \epsilon(x, h)$$

where $\epsilon(x, h) = (A(x + h, x) - A(x, x))h$. Then $||\epsilon(x, h)||/||h|| \leq (||A(x + h, x) - A(x, x)||\ ||h||)/||h|| = ||A(x + h, x) - A(x, x)|| \to 0$ by the continuity of A. Thus $A(x,x)$ satisfies the defining equation (6.1) of d_xf. Therefore f is differentiable, and $d_xf = A(x,x)$ is continuous, so f is C^1.

□

Remark 6.1. Note that the convexity of U was not used at all in proving that the existence of A implies that f is of class C^1. It was only used in the opposite direction, mostly for convenience. So this hypothesis can be safely ignored.

6.4. **Proof of the Inverse Function Theorem.** Let $f : U \to \mathbb{R}^k$ be as in the statement of Theorem 6.1. By Lemma 6.2, there exists $A : U \times U \to L(\mathbb{R}^k, \mathbb{R}^k)$ satisfying (6.9). Since $d_{x_0}f = A(x_0, x_0)$ is invertible, by Lemma 6.1 there exists a neighborhood N of (x_0, x_0) in $U \times U$ so that $A(x_1, x_2)$ is invertible for all $(x_1, x_2) \in N$. Moreover, again by Lemma 6.1, $A^{-1} : N \to L(\mathbb{R}^k, \mathbb{R}^k)$ is continuous.

Since $||A^{-1}||$ is continuous on N, for any $\epsilon > 0$ so that $\bar{B}(x_0, \epsilon) \times \bar{B}(x_0, \epsilon) \subset N$, there is a constant C such that $||A^{-1}(x_1, x_2)|| \leq C$ for all $(x_1, x_2) \in \bar{B}(x_0, \epsilon) \times \bar{B}(x_0, \epsilon)$. Multiplying both sides of (6.9) by $A^{-1}(x_1, x_2)$ we get

$$A^{-1}(x_1, x_2)(f(x_2) - f(x_1)) = x_2 - x_1 \tag{6.10}$$

therefore

$$||x_2 - x_1|| \leq ||A^{-1}(x_1, x_2)||\ ||f(x_2) - f(x_1)|| \leq C||f(x_2) = f(x_1)||. \tag{6.11}$$

In other words,

$$||f(x_2) - f(x_1)|| \geq C^{-1}||x_2 - x_1|| \text{ for all } x_1, x_2 \in B(x_0, \epsilon). \tag{6.12}$$

This immediately implies that f is injective on $B(x_0, \epsilon)$, since $f(x_1) = f(x_2)$ gives $||x_2 - x_1|| = 0$. Moreover, rewriting (6.12) for $y_1, y_2 \in f(B(x_0, \epsilon))$ as

$$||y_2 - y_1|| \leq C^{-1}||f^{-1}(y_2) - f^{-1}(y_1)|| \tag{6.13}$$

we see that $f^{-1} : f(B(x_0, \epsilon)) \to B(x_0, \epsilon)$ is Lipschitz with Lipschitz constant C, in particular it is continuous.

Suppose that $f(B(x_0, \epsilon))$ contains a neighborhood V of y_0. Then f^{-1} is of class C^1 on V, because equation (6.10) reads

$$A^{-1}(f^{-1}(y_1), f^{-1}(y_2))(y_2 - y_1) = f^{-1}(y_2) - f^{-1}(y_1). \tag{6.14}$$

Since we know that f^{-1} is continuous, this is equation (6.9) for f^{-1}, thus, by Lemma 6.2, f^{-1} is of class C^1. This finishes the first part of the outline in §6.1. We have seen that the first ingredient, the mean value theorem, is replaced by Lemma 6.2. It only remains to find the second ingredient:

6.5. **Proof that $f(B(x_0, \epsilon))$ Contains a Neighborhood of y_0.** By translating in the domain and range we may assume that $x_0 = 0$ and $y_0 = f(x_0) = 0$, (Replace f by $f(x + x_0) - y_0$.) Then, by composing with $(d_0f)^{-1}$, we may assume that $d_0 f = I$. (Replace f by $(d_0 f)^{-1} \circ f$.) In other words, we may suppose that

$$f(x) = x + \phi(x) \text{ where } \phi : B(0,\epsilon) \to \mathbb{R}^k,\ \phi(0) = 0,\ d_0\phi = 0. \tag{6.15}$$

To find a function $x = g(y)$ that solves $f(x) = y$ for x, given the way we have re-written f, it is is reasonable to look for g of the form

$$g(y) = y + \psi(y) \text{ where } \psi(0) = 0 \tag{6.16}$$

Then, setting $x = y + \psi(y)$ in (6.15), we see that ψ satisfies

$$(y + \psi(y)) + \phi(y + \psi(y)) = y$$

which is the same as

$$-\phi(y + \psi(y)) = \psi(y). \tag{6.17}$$

This converts our problem into a fixed-point problem, as follows. Define, for suitably small δ to be determined shortly, a map

$$F : \bar{B}(0,\delta) \times \bar{B}(0,\delta) \to \mathbb{R}^k \text{ by } F(y,z) = -\phi(y+z) \tag{6.18}$$

and, for each y, let

$$F_y(z) = F(y,z) = -\phi(y+z), \tag{6.19}$$

then (6.17) says that $\psi(y)$ is a fixed point of F_y.

For (6.18) to make sense we need, first of all, $\delta < \epsilon/2$ so that $y+z \in B(0,\epsilon)$ when $y, z \in \bar{B}(0,\delta)$. Moreover we need:

- For all $y, z \in \bar{B}(0,\delta)$, $\phi(y+z) \in \bar{B}(0,\delta)$ so that, for each $y \in \bar{B}(0,\delta)$, $F_y : \bar{B}(0,\delta) \to \bar{B}(0,\delta)$.
- $F_y : \bar{B}(0,\delta) \to \bar{B}(0,\delta)$ is a contraction: there exists a constant $C > 0$ so that for all $z_1, z_2 \in \bar{B}(0,\delta)$, $||F_y(z_1) - F_y(z_2)|| = ||\phi(y+z_1) - \phi(y+z_2)|| \le C||z_1 - z_2||$

These statements follow easily from the properties of ϕ. First, from $\phi(0) = 0$, $d_0\phi = 0$ and the definition of differentiability, (6.1) becomes

$$||\phi(x)|| \le \epsilon(x)(x) \text{ where } ||\epsilon(x)|| \to 0 \text{ as } ||x|| \to 0, \tag{6.20}$$

and, since ϕ is of class C^1 and $d_0\phi = 0$, there exists $\eta > 0$ so that $||d_x\phi|| < 1/2$ if $||x|| < \eta$.

Choose δ so that $||\epsilon(x)|| < 1/2$ in (6.20), thus $||\phi(y+z)|| < \frac{1}{2}||y+z|| < \delta$, and also so that $\delta < \eta/2$. Then, for $||y||, ||z_1||, ||z_2|| < \delta$, and $0 \le t \le 1$, $||\gamma(t)|| = ||(1-t)(y+z_1) + t(y+z_2)|| = ||(1-t)z_1 + tz_2|| < 2\delta < \eta$ and

$$||\phi(y+z_1) - \phi(y+z_2)|| = ||\Big(\int_0^1 d_{\gamma(t)}\phi dt\Big)(z_1 - z_2)|| \le \frac{1}{2}||z_1 - z_2||, \tag{6.21}$$

Thus we get that for each $y \in \bar{B}(0,\delta)$ the map $F_y : \bar{B}(0,\delta) \to \bar{B}(0,\delta)$ defined in (6.19) is a contraction, so it has a unique fixed point $\psi(y)$. Looking at (6.16), we have found our inverse map g and proved that $f(B(0,\epsilon))$ contains $B(0,\delta)$. This completes the poof of the Inverse Function Theorem.

7. Inner Product Spaces

In §3.1, 3.2 we saw examples of normed vector spaces. The ones with the subscript 2 have a more special structure, called an *inner product space*.

Definition 7.1. (1) A *real inner product space* is a real vector space V together with a function $V \times V \to \mathbb{R}$, called the *inner product*. Its value on x, y is denoted $< x, y >$ and it satisfies:
   (a) It is *symmetric*: $< x, y >=< y, x >$ for all $x, y \in V$.
   (b) It is *bilinear*, meaning that it is linear in each variable: for all $x, y, z \in V$ and all $a, b \in \mathbb{R}$, we have

$$< ax + by, z >= a < x, z > +b < y, z >,$$

   and by symmetry linear in the second variable.
   (c) It is *positive definite*: for all $x \in V$ we have $< x, x >\geq 0$, and $< x, x >= 0$ only if $x = 0$.

(2) A *complex inner product space* is a complex vector space V together with a function $V \times V \to \mathbb{C}$, called the *inner product*. Its value on $x, y \in V$ is denoted by $< x, y >$ and it satisfies:
   (a) For all $x, y \in V$, $< y, x >= \overline{< x, y >}$, where the bar denotes the complex conjugate.
   (b) It is *sesquilinear* (= one and a half linear) meaning that it is complex linear in the first variable: for all $x, y, z \in V$ and all $a, b \in \mathbb{C}$, we have

$$< ax + by, z >= a < x, z > +b < y, z >,$$

   and by (a) it is *conjugate linear* in the second variable:

$$< z, ax + by >= \overline{a} < z, x > +\overline{b} < z, y > .$$

7.1. **Examples of Real Inner Product spaces.**

7.1.1. *Euclidean Space.* Just as in §3.1 let $V = \mathbb{R}^n$ with $< x, y >$ the usual dot-product:

$$< x, y >= x \cdot y = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

The properties of an inner product are easily verified. In particular,

$$< x, x >= x_1^2 + \cdots + x_n^2 \geq 0, \quad \text{and } = 0 \text{ if and only if } x = 0.$$

7.1.2. L^2 *- Space of Integrable Functions.* Just as in §3.2, let $[a, b] \subset \mathbb{R}$ be an interval, and let $R[a, b]$ denote the vector space of real, Riemann-integrable functions on $[a, b]$. If $f, g \in R[a, b]$, so is fg, and we can define $< f, g >$ by

$$< f, g >= \int_a^b f(x)g(x)dx.$$

This is called the L^2 - inner product. The properties of an inner product are easily verified. In paraticular

$$< f, f >= \int_a^b f(x)^2 dx \geq 0 \text{ and } = 0 \text{ if and only if } f(x) = 0 \text{ for all } x.$$

7.2. **Examples of Complex Inner Product Spaces.**

7.2.1. *Complex Euclidean Space.* Let $\mathbb{C}^n$ denote the space of all n-tuples $z = (z_1, \ldots, z_n)$ of complex numbers, with pointwise addition, componentwise scalar multiplication. Define the inner product $< z, w >$ by

$$< z, w >= z_1\overline{w_1} + \cdots + z_n\overline{w_n}.$$

The properties of a complex inner product are easily verified. For example,

$$< w, z >= w_1\overline{z_1} + \ldots w_n\overline{z_n} = \overline{z_1\overline{w_1} + \cdots + z_n\overline{w_n}} = \overline{< z, w >}$$

and

$$< z, z >= z_1\overline{z_1} + \cdots + z_n\overline{z_n} = |z_1|^2 + \cdots + |z_n|^2 \geq 0,$$

and $= 0$ if and only if $z = 0$. Observe that for this last property to be true we need the complex conjugates in one of the variables, hence $< z, w >$ is complex "sesquilinear" rather than complex bilinear.

7.2.2. L^2 *- Space of Complex Integrable Functions.* In analogy with (7.1.2), let $[a, b] \subset \mathbb{R}$ and let $R_{\mathbb{C}}[a, b]$ denote the space of Riemann-integrable complex functions on $[a, b]$, meaning that the real and imaginary parts of f are both Riemann integrable. If $f \in R_{\mathbb{C}}[a, b]$, so is $\overline{f}$, and if $f, g \in R_{\mathbb{C}}[a, b]$, so is fg. Therefore it makes sense to define $< f, g >$ by

$$< f, g >= \int_a^b f(x)\overline{g(x)}dx.$$

This is also called the L^2 - inner product. The properties of an inner product are also easily checked. For instance,

$$< g, f >= \int_a^b g(x)\overline{f(x)}dx = \overline{\int_a^b f(x)\overline{g(x)}dx} = \overline{< f, g >},$$

and

$$< f, f >= \int_a^b f(x)\overline{f(x)}dx = \int_a^b |f(x)|^2 dx \geq 0,$$

and $= 0$ if and only if $f(x) = 0$ for all x. Thus it is positive definite, thanks to the introduction of the complex cojugate in one of the variables. Again we see that we have to give up complex bilinearity if we want the inner product to be positive definite.

7.2.3. *L^2-Space of Periodic Functions.* Let $R_{\mathbb{C}}(S^1)$ denote the space of Riemann-integrable functions $f : \mathbb{R} \to \mathbb{R}$ which are periodic of period 2π: $f(x+2\pi) = f(x)$ for all $x \in \mathbb{R}$. Here S^1 denotes the unit circle in $\mathbb{C}$. Define an inner product $< f, g >$ by

$$< f, g >= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx. \tag{7.1}$$

This is the L^2 - inner product used in Chapter 8 of [5] to study Fourier series. In particular, the functions $\{e^{inx}\}$ form an *orthonormal set* meaning that

$$< e^{imx}, e^{inx} >= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{imx} e^{-inx} dx = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{otherwise.} \end{cases}$$

and the *Fourier Coefficients* of f are the numbers $c_n =< f, e^{inx} >$.

7.3. **The Cauchy-Schwarz Inequality.** The Cauchy-Schwarz inequality is a formal consequence of the properties of an inner product. We have seen this before for a real inner product space $(V, < , >)$. Namely, for any $x, y \in V$, since for all t we have that $< x + ty, x + ty >\geq 0$, expanding the left-hand side we get $< x, x > +2t < x, y > +t^2 < y, y >\geq 0$. Thus this quadratic polynomial in t must have discriminant ≤ 0, in oher words

$$(2 < x, y >)^2 - 4 < x, x >< y, y >\leq 0,$$

which is the same as the Cauchy-Schwarz inequality

$$< x, y >^2 \leq < x, x >< y, y > .$$

From this we get that the function $||x|| = \sqrt{< x, x >}$ is a norm, namely it satisfies the triangle inequality $||x+y|| \leq ||x|| + ||y||$, since, squaring, we get

$$||x + y||^2 =< x + y, x + y >=< x, x > +2 < x, y > + < y, y >,$$

which, by the Cauchy-Schwarz inequality and the definition of the norm is

$$\leq ||x||^2 + 2||x|| \; ||y|| + ||y||^2 = (||x| \; + ||y||)^2,$$

which, taking square roots, is the triangle inequality.

Similarly, for a complex inner product space, using the fact that for all $x, y \in V$ and for all $t \in \mathbb{R}$, the polynomial $< x + ty, x + ty > \geq 0$, or $< x, x > +t(< x, y > + < y, x >) + t^2 < y, y > \geq 0$, or $< x, x > +2t\Re(< x, y >) + t^2 < y, y > \geq 0$. Taking the discriminant, we obtain the following form of the Cauchy-Schwarz inequality:

$$(\Re(< x, y >))^2 \leq < x, x >< y, y >$$

which is exactly what is needed to derive the triangle inequality for the norm $||x|| = \sqrt{< x, x >}$ on the complex inner product space $(V, < , >)$.

We leave it as an exercise to use the stronger fact that $< x + ty, x + ty > \geq 0$ for all *complex* t to derive the stronger form of the Cauchy-Schwarz inequality for a complex inner product space:

$$| < x, y > |^2 \ \leq < x, x >< y, y > .$$

As a hint, write down $< x + ty, x + ty >=< x, x > +t < y, x > +\bar{t} < x, y > +|t|^2 < y, y >$. This is of the form $at\bar{t} + bt + \bar{b}\bar{t} + c$ where $a, c \in \mathbb{R}$. Find the value of t that minimizes this expression (answer:$t = -\bar{b}/a$), put this value of t back into the expresiion and use the fact that it gives a number ≥ 0. The resulting inequality is Cauchy-Schwarz.

A few remarks are in order.

Remark 7.1. Every inner product space $(V, < \ , \ >)$, real or complex, becomes a metric space in the standard and familiar way by defining $d(x, y) = ||x-y||$. In the notation of §3.1 and 3.2, what we now call $||x||$ is $||x||_2$, the L^2-norm. The resulting distance is complete for the examples of §3.1. It is also complete for the complex Euclidean space $\mathbb{C}^n$ of (7.2.1). But it is *not* complete for the infinite-dimensional spaces of functions in (7.1.2),(7.2.2) and (7.2.3). The reason is as in the proof of Theorem 3.3: the space of Riemann-integrable functions is too small, we need to allow unbounded functions that are L^2-limits of Riemann-integrable functions. We will see that the right space will be the space $L^2[a, b]$ of Lebesgue-measurable functions with finite L^2-norm.

Remark 7.2. A related remark is the comparison of norms. For the finite dimensional examples there are comparisons in all directions, as in in the inequalities (3.2). But for the infinite-dimensional spaces of functions there are only the inequalities (3.7). For instance, Theorem 3.3 shows that there can be no inequality of the form $||f||_\infty \leq C||f||_1$ becuase $||f||_\infty$ is complete, $||f||_1$ is not complete, and there is the third inequality (3.7) that bounds $||f||_1 \leq (b-a)||f||_\infty$. So, if we had an inequality $||f||_\infty \leq C||f||_1$, a Cauchy seqence in $|| \ \ ||_1$ would be Cauchy in $|| \ \ ||_\infty$, which is complete, so it would converge in $|| \ \ ||_\infty$, hence in $|| \ \ ||_1$ by the last inequality of (3.7).

Remark 7.3. Along the same lines, we see that there can be no inequality of the form $||f||_2 \leq C||f||_1$ because the function $f(x) = x^{-1/2}$ on $(0, 1]$ used in the proof of Theorem 3.3 has finite L^1-norm but infinite L^2-norm. But, if we restrict ourselves to Riemann integrable funcitons, which, by definition, are bounded, then there is the inequality $||f||_2 \leq (||f||_\infty ||f||_1)^{1/2}$.

7.4. **Orthonormal Sets and Fourier Series.** We have already introduced in (7.2.3) the L^2 inner product on the space of periodic functions, which is used in [5] to talk about convergence of Fourier series. In fact, Parseval's Theorem, 8.16 of [5], says that the Fourier series of f converges in L^2-norm to f, in the sense that $||f - s_N(f)|| \to 0$.

The procedure for finding Fourier coefficients and series is a general procedure for complex (or real) inner product spaces. If $(V, < \ , \ >)$ is an inner product space, a set $\{e_1, e_2, \dots\}$, finite or infinite, is called an *orthonotmal set* if

$$< e_m, e_n >= \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{otherwise.} \end{cases} \tag{7.2}$$

If $x \in V$ is a linear combination $x = \Sigma_{m=1}^N c_m e_m$, then each coefficient c_n is easily found by $< x, e_n >=< \Sigma_m c_m e_m e_n >= \Sigma_m c_m < e_m, e_n >= c_n$ by (7.2). This is exactly how the Fourier coefficients of f are found by using the inner product (7.1) and the orthonormal set $\{e^{inx} : n \in \mathbb{Z}\}$.

The following theorem contains theorems 8.11 and 8.12 of [5]

Theorem 7.1. *Let $(V, < \ , \ >)$ be a (real or complex) inner product space, let $S = \{e_n\}$ be an orthonormal set, and, for each N, let $E_N \subset V$ be the linear subspace of V spanned by $e_1, \dots, e_N$. Let $x \in V$ and, for each n so that $e_n \in S$, define $c_n =< x, e_n >$. Then*

$$s_N = \Sigma_{n=1}^N c_n e_n \tag{7.3}$$

is the point in E_N closest to x. In other words, $||x - s_N|| \leq ||x - y||$ for all $y \in E_N$, with equality if and only if $y = s_N$. Moreover, $||s_N||^2 \leq ||x||^2$.

Proof. Let $z = x - s_N$. Observe that z is perpendicular to E_N, because, for $n = 1, \dots, N$, $< z, e_n >=< x - s_N, e_n >=< x, e_n > - < s_N, e_n >= c_n - c_n = 0$. Therefore, if $y \in E_N$, then $x - y = (x - s_N) + (s_N - y) = z + w$, where $w = s_N - y \in E_N$. By the Pythagorean theorem,

$$||x - y||^2 = ||z||^2 + ||w||^2 = ||x - s_N||^2 + ||s_N - y||^2. \tag{7.4}$$

The formal calculation is: $||x - y||^2 =< z + w, z + w > =< z, z > + 2\Re < z, w > + < w, w > =< z, z > + < w, w >$ because $w \in E_N$ and z has zero inner product with any vector in E_N.

Thus (7.4) gives $||x - y|| \geq ||x - s_N||$, with equality if and only if $||s_N - y|| = 0$, that is, $y = s_N$. Finally, since $x = z + s_N$ and $< z, s_N >= 0$ because $s_N \in E_N$, the same formal Pythagorean argument gives $||x||^2 = ||z||^2 + ||s_N||^2 \geq ||s_N||^2$, thus the last assertion. □

To interpret Parseval's theorem, 8.16 of [5] in the same spirit, let's use the following terminology:

Definition 7.2. Let V and W be inner product spaces (real or complex) and let $A; V \to W$ be a linear transformation. A is called an *isometry* if for all $x, y \in V$ we have $< Ax, Ay >_W =< x, y >_V$.

Example 7.1. Let V be the space of $R_{\mathbb{C}}(S^1)$ of periodic functions as in (7.2.3) with the L^2-inner product, and let W be the space of all doubly infinite sequences $\{c_m\}_{-\infty}^{\infty}$ with the property that $\Sigma_{-\infty}^{\infty} |c_m|^2 < \infty$, and with inner product

$$< \{c_m\}, \{\gamma_m\} >= \Sigma_{-\infty}^{\infty} c_m \overline{\gamma_m}, \tag{7.5}$$

which converges by the Cauchy-Schwarz inequality. Then Parseval's Theorem says:

(1) If $f \in V$ and $\{c_m\}$ is the sequence of Fourier coefficients of f, in other words,
$$f(x) \sim \ \Sigma c_n e^{inx},$$
then the sequence $\{c_n\} \in W$, in other words,
$$\Sigma_{-\infty}^{\infty}|c_n|^2 < \infty$$

(2) The linear transformation that assigns to f its sequence of Fourier coefficients is an isometry from V to W. In particular
$$\Sigma_{-\infty}^{\infty}|c_n|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)|^2 dx$$

(3) The Fourier series of f converges to f in L^2:
$$\lim_{N\to\infty} ||f - s_N(f)||_2 \ = \ 0.$$

Observe that we are not saying that the isometry is surjective. we will need to enlarge the class of functions in V beyond the Riemann integrable ones to be able to get surjectivity.

8. The L^p-spaces

Fix an interval $[a,b] \subset \mathbb{R}$ and a real number $p \geq 1$. A measurable function f on E is *of class* L^p if the Lebesgue integral $\int_a^b |f|^p < \infty$. Call two functions f, g of class L^p equivalent, written $f \sim g$, if $f = g$ a.e., that is, if $m(\{x : f(x) \neq g(x)\}) = 0$. The space $L^p[a,b]$ is the set of these equivalence classes:

$$L^p[a,b] = \{f : [a,b] \to \mathbb{R} \text{ measurable} : \int_a^b |f|^p < \infty\}/\sim \tag{8.1}$$

Remark 8.1. In practice we usually think of the elements of $L^p[a,b]$ as functions, we usually have a specific representative in mind. For example, if f is a continuous function, we take this function as the representative of its equivalence class. In many discussions we define various operations on $L^p[a,b]$ by taking representatives, we do so without further comments provided that it is obvious that the definition is independent of the representatrive. This will always be the case when defining a quantity that depends just on an integral, since the value of an integral depends just on the value of the function a. e.. The next definition illustrates this point;

Define the L^p norm on $L^p[a,b]$ by

$$||f||_p = (\int_a^b |f|^p)^{1/p} \tag{8.2}$$

It is clear that this satisfies two of the three defining properties of a norm: First, for any real number a, $||af||_p = |a|\ ||f||_p$. Second, $||f||_p \geq 0$, and if $||f||_p = 0$, then $\int_a^b |f|^p = 0$, hence $f = 0$ a.e. (see [4], Exercise 4.3), therefore f represents 0 in $L^p[a,b]$. It is for this reason that we define the elements of $L^p[a,b]$ to be equivalence classes of functions, rather than the functions themselves. Had we used the functions, we would get that any function f which vanishes a.e. (for example, the characteristic function of a set of measure 0) to have zero norm.

More difficult is to verify the third and most important property of a norm: the triangle inequality $||f+g||_p \leq ||f||_p + ||g||_p$. This is needed to verity that $L^p[a,b]$ is closed under addition, hence is a vector space, and that it is a normed vector space. This is the content of the Minkowski inequality, see §6.2 of [4] or Exercise 6.10 of [5] for a proof.

We will only look at the cases $p = 1, 2$. For these cases we have verified the triangle inequality in §3.2.3. There we used the Riemann integral in defining the norms and checking the triangle inequality, but the proofs are identical. In §3.2.3 we looked at the norms only on the spaces $C[a,b]$ of continuous functions, and proved that these spaces are not complete. Now we have the following theorems:

Theorem 8.1. *The spaces $L^1[a,b]$, $L^2[a,b]$ are complete.*

Proof. See §6.3 of [4], or Theorem 11.42 of [5]. □

Remark 8.2. The space $L^1[a,b]$ is an example of a *Banach space*: a complete normed space. The space $L^2[a,b]$ is an example of a *Hilbert space*: a complete inner product space. Every Hilbert space is a Banach space, but not conversely. The inner product on $L^2[a,b]$ is

$$< f, g >= \int_a^b f(x)g(x)dx \tag{8.3}$$

which makes sense because, if $f, g \in L^2[a,b]$, then the Cauchy-Schwarz inequality gives that fg is Lebesgue integrable:

$$| \int_a^b fg \mid \leq (\int_a^b f^2)^{1/2} (\int_a^b g^2)^{1/2}$$

and the L^2-norm is $||f||_2 =< f, f >^{1/2}$. In both cases, completeness is in the distance given by the norm: If $\{f_n\}$ is a Cauchy sequence in L^p, $p = 1$ or 2, meaning that for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that $m, n > N \Rightarrow ||f_m - f_n||_p < \epsilon$, there exists $f \in L^p[a,b]$ so that $||f_n - f||_p \to 0$.

We also have the following theorem:

Theorem 8.2. *The continuous functions $C[a,b]$ are dense in $L^p[a,b]$, $p = 1, 2$, In other words, given any $f \in L^p[a,b]$ and $\epsilon > 0$ there exists $h \in C[a,b]$ so that $||f - h||_p < \epsilon$.*

Proof. For $p = 2$ see the proof of Theorem 11.38 of [5]. For $p = 1$ the argument is similar. $\square$

Let us recall that any metric space (X, d_X) has a *completion*. This means that there is a *complete* metric space (Y, d_Y) and an *isometry* $\Phi : X \to Y$ so that $\Phi(X)$ is dense in Y. Recall that Φ being an isometry means that $d_Y(\Phi(x), \Phi(y)) = d_X(x, y)$ for all $x, y \in X$. See, for example, exercise 7.24 of [5] for a construction of Y from X. It is easy to see that any two completions of X are isometric to each other.

Using this language, if we go back to §3.2.3, the last two theorems can be phrased as follows: For $p = 1, 2$, $L^p[a, b]$ is the completion of $(C[a, b], d_p)$, where $d_p(f, g) = ||f - g||_p$.

8.1. **Fourier Series of L^2-functions.** We can now complete the discussion of Fourier series started in §7.4. There we developed general properties of orthonormal sets in inner product spaces and representations of elements of such a space in terms of linear combinations of elements of an orthonormal set. In Example 7.1 we stated Parseval's theorem for the class of Riemann integrable functions on $[-\pi, \pi]$. The optimal way to state the theorem is for $L^2[-\pi, \pi]$. This means, replace the space V of Example 7.1 by $L^2[-\pi, \pi]$ with the same inner product of Equation (7.1). Then any $f \in L^2[-\pi, \pi]$ has Fourier coefficients $\{c_n\}$ belonging to the space W of doubly infinite sequences with the inner product of Equation (7.5) (usually denoted $l^2(\mathbb{Z})$). Parseval's theorem asserts:

(1) $||f - s_N(f)||_2 \to 0$ as $N \to \infty$, where $s_N(f) = \sum_{-N}^{N} c_n e^{inx}$.
(2) The linear transformation
$$L^2[-\pi, \pi] \to l^2(\mathbb{Z})$$
that assigns to f the sequence $\{c_n\}_{n\in\mathbb{Z}}$ of its Fourier coefficients is an isometric bijection.

The proof of the first statement is as in the proof of Theorem 8.16 of [5], where the first step in that proof, L^2-density of continuous functions in the Riemann integrable ones, is replaced by the stronger density statement of Theorem 8.2, otherwise the proof is the same. The second statement has been proved before except for the proof of surjectivity. But this is an immediate consequence of the completeness of $L^2[-\pi, \pi]$ (Theorem 8.1): If $\{c_n\}_{n\in\mathbb{Z}} \in l^2(\mathbb{Z})$, then $||\sum_{-\infty}^{\infty} c_n e^{inx}||_2^2 \leq \sum_{-\infty}^{\infty} |c_n|^2 < \infty$, which implies that $\sum_{-\infty}^{\infty} c_n e^{inx}$ converges in $L^2[-\pi, \pi]$, and the Fourier series of this limit is $\sum_{-\infty}^{\infty} c_n e^{inx}$. See the Theorems 11.43 and 11.45 of [5] for more details.

References

[1] V. I. Arnold, *Ordinary Differential Equations*, Third Edition, Springer.

[2] R. S. Hamilton, The Inverse Function Theorem of Nash and Moser, Bulletin (New Series) of the American Mathematical Society, **7**, (1982), 65–222.
[3] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press.
[4] H. L. Royden, *Real Analysis*, Third Edition, McMillan.
[5] W. Rudin, *Principles of Mathematical Analysis*, Third Edition, Mc Graw - Hill.