Math 711 Course Notes

Dr. Sarah Raynor[1]

[1]These notes were composed by Dr. Robinson last year and edited by his class. I will be editing them for our course as we proceed.

0.1. $C^1([0,1])$**.** Our next example is a modification of $C([0,1])$ allowing for the function to be differentiable:

DEFINITION 0.1.

$$C^1([0,1]) := \{f : [0,1] \to \mathbb{R} : f \textit{ is differentiable on } [0,1] \textit{ and } \frac{df}{dx} \in C([0,1])\},$$

PROPOSITION 0.1. *$C^1([0,1])$ is a real vector space.*

PROOF. Exercise. $\square$

THEOREM 0.1. *Let $d : C^1([0,1]) \times C^1([0,1]) \to \mathbb{R}$ be given by $d(f,g) = ||f-g||_\infty + ||f'-g'||_\infty$. Then $(C^1([0,1]), d)$ is a metric space.*

PROOF. Exercise. $\square$

LEMMA 0.1. *Suppose that f is a continuously differentiable function on $[0,1]$, and $\sup_{x\in[0,1]} |f'(x)| = M$. Then, for all $x, y \in [0,1]$, $|f(x) - f(y)| \leq M|x-y|$.*

PROOF. Let f, x, y be as in the statement of the theorem. By the mean value theorem, there is a c with $x \leq c \leq y$ so that $\frac{f(x)-f(y)}{x-y} = f'(c)$. Therefore,

$$|f(x) - f(y)| = |f'(c)||x-y| \leq M|x-y|.$$

$\square$

PROPOSITION 0.2. *$C^1([0,1])$ is complete with respect to d.*

PROOF. Suppose that (f_n) is a Cauchy sequence in $C^1([0,1])$. Then, since $\|f_n - f_m\|_\infty \leq d(f_n, f_m)$, (f_n) is also Cauchy in $C([0,1])$. Therefore, $\exists f \in C([0,1])$ so that $f_n \to f$ uniformly. Similarly, since $\|f'_n - f'_m\|_\infty \leq d(f_n, f_m)$, $\exists g \in C([0,1])$ so that $f'_n \to g$ uniformly.

Fix $x \in [0,1]$. Consider the difference quotient $D_n(h) = \frac{f_n(x+h)-f_n(x)}{h}$. Then I claim that $(D_n(h))$ is a uniformly Cauchy sequence with respect to h. That is, for every $\epsilon > 0$, there is an $N \in \mathbb{N}$ so that $\forall n, m > N, |D_n(h) - D_m(h)| < \epsilon$. Let $\epsilon > 0$. Because $\|f'_n - f'_m\|_\infty \to 0$, $\exists N \in \mathbb{N}$ such that $\forall n, m > N$, $\sup_{[0,1]} |(f_n - f_m)'| < \epsilon$. By Lemma 0.1, then, for all $x, y \in [0,1]$ it follows that $|(f_n - f_m)(x-y)| < \epsilon|x-y|$. Now, for $n, m > N$ and $h > 0$,

$$\begin{aligned}
|D_n(h) - D_m(h)| &= \left| \frac{f_n(x+h) - f_n(x)}{h} - \frac{f_m(x+h) - f_m(x)}{h} \right| \\
&= \frac{1}{h}|f_n(x+h) - f_m(x+h) - (f_n(x) - f_m(x))| \\
&= \frac{1}{h}|(f_n - f_m)(x+h) - (f_n - f_m)(x)| \\
&< \frac{1}{h}\epsilon|x+h-x| \\
&= \epsilon
\end{aligned}$$

as claimed. Hence the functions $(D_n(h))$ are uniformly Cauchy. Therefore, there is a function $D(h)$ to which they converge. But if we consider a fixed h, then it is clear that $D_n(h)$ must converge (pointwise) to $\frac{f(x+h)-f(x)}{h}$ by the (pointwise) convergence of f_n to f. Hence, by the uniqueness of limits, $D_n(h)$ converges uniformly to $D(h) := \frac{f(x+h)-f(x)}{h}$.

Now, recall that $(f_n'(x))$ is a uniformly Cauchy sequence which converges to some function $g(x)$. We want to show that $\lim_{h\to 0} D(h) = g(x)$. This will show that $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = g(x)$, i.e. f is differentiable and f' is g, the uniform limit of (f_n'), which is what we set out to show.

So, let $\epsilon > 0$, and choose n sufficiently large so that $\|D_n(h) - D(h)\|_\infty < \frac{\epsilon}{3}$. *and* $|f_n'(x) - g(x)| < \frac{\epsilon}{3}$. Then, for this fixed n, choose δ so that if $|h| < \delta$, then $|D_n(h) - f_n'(x)| < \frac{\epsilon}{3}$. Finally, by the triangle inequality, we have that

$$|D(h) - g(x)| \leq |D(h) - D_n(h)| + |D_n(h) - f_n'(x)| + |f_n'(x) - g(x)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,$$

and the claim follows. □

Definition 0.2. *For $f \in C^1([0,1])$, define the $W^{1,2}$-norm of f as*

$$\|f\|_{1,2} := \left(\int_0^1 |f(x)|^2 dx + \int_0^1 |f'(x)|^2 dx \right)^{\frac{1}{2}}. \tag{1}$$

Proposition 0.3. *$\|\cdot\|_{1,2}$ is a norm on $C^1([0,1])$.*

Proof. Exercise. □

Proposition 0.4. *$C^1([0,1])$ is not complete with respect to the $W^{1,2}$-norm.*

Proof. Exercise. □

0.1.1. *Exercises.*

Exercise 0.1. *Prove Proposition 0.1, Theorem 0.1, and Propositions 0.3 and 0.4. Hint: all of the heavy lifting has been done, so be sure to take advantage of what has already been proved.*

CHAPTER 1

Several Important Constructions in Metric Spaces

1. The Completion of a Metric Space

Let (X, d) be a metric space. We have discussed the importance of completeness in a metric space several times. Completeness is of particular interest if we are trying to construct a solution a problem via the convergence of approximate solutions. So what should we do if the metric space with which we want to work is not complete? For example, the L^2 and $W^{1,2}$ norms that were introduced in the previous chapter are both of interest due to their relationship with the energy in physical problems. However, we know that neither norm generates a complete metric space on the spaces $C^k([0,1])$. It is possible, by introducing the Lebesgue integral and a suitable equivalence relation on integrable functions, to explicitly construct the natural, complete metric spaces for working with these norms, but it is a rather involved construction. Instead, we will prove an abstract theorem stating that every metric space can be "completed," and define the desired complete metric spaces based on this construction.

THEOREM 1.1. *Let (X, d) be a metric space. Then there exists a complete metric space $(\tilde{X}, \tilde{d})$, called the* completion *of X, and a natural embedding $i : X \to \tilde{X}$ such that $\forall x, y \in X$, $\tilde{d}(i(x), i(y)) = d(x, y)$. Moreover, $i(X)$ is dense in $\tilde{X}$.*

In order to prove this theorem, we need first to construct an appropriate choice of $\tilde{X}$ and $\tilde{d}$ and then show that they have the desired properties.

DEFINITION 1.1. *Suppose that (p_n) and (q_n) are both Cauchy sequences in X. Then we say that (p_n) is* equivalent *to (q_n), denoted $(p_n) \sim (q_n)$, if $\lim_{n\to\infty} d(p_n, q_n) = 0$.*

LEMMA 1.1. *The relation defined in Definition 1.1 is an equivalence relation.*

PROOF. Let (p_n),(q_n) and (r_n) be arbitrary Cauchy sequences in X. Clearly, for each n $d(p_n, p_n) = 0$, so $\lim_{n\to\infty} d(p_n, p_n) = 0$, so $(p_n) \sim (p_n)$ and the relation is reflexive.

Now, suppose that $(p_n) \sim (q_n)$. Since the metric is symmetric, $d(p_n, q_n) = d(q_n, p_n)$ for every $n \in \mathbb{N}$. Therefore, $\lim_{n\to\infty} d(q_n, p_n) = \lim_{n\to\infty} d(p_n, q_n) = 0$, so $(q_n) \sim (p_n)$ and the relation is symmetric.

Finally, suppose that $(p_n) \sim (q_n)$ and $(q_n) \sim (r_n)$. Then

$$\lim_{n\to\infty} d(p_n, r_n) \le \lim_{n\to\infty} (d(p_n, q_n) + d(q_n, r_n)) = \lim_{n\to\infty} d(p_n, q_n) + \lim_{n\to\infty} d(q_n, r_n) = 0 + 0 = 0,$$

by the triangle inequality, the properties of limits, and the hypotheses. Therefore $(p_n) \sim (r_n)$ and the relation is transitive. □

We are now in a position to define $\tilde{X}$ and $\tilde{d}$:

Definition 1.2. *Define Y to be the collection of all Cauchy sequences in X, and define $\tilde{X}$ to be the quotient of Y by the equivalence relation defined above.*

Lemma 1.2. *Suppose that $(p_n), (\tilde{p}_n), (q_n)$, and $(\tilde{q}_n)$ are Cauchy sequences in X and $(p_n) \sim (\tilde{p}_n)$ and $(q_n) \sim (\tilde{q}_n)$. Then $\lim_{n\to\infty} d(p_n, q_n) = \lim_{n\to\infty} d(\tilde{p}_n, \tilde{q}_n)$.*

Proof. Let $(p_n), (\tilde{p}_n), (q_n)$, and $(\tilde{q}_n)$ satisfy the hypotheses of the lemma. Then

$$\begin{aligned}
\lim_{n\to\infty} d(p_n, q_n) &\leq \lim_{n\to\infty} (d(p_n, \tilde{p}_n) + d(\tilde{p}_n, \tilde{q}_n) + d(\tilde{q}_n, q_n)) \\
&= \lim_{n\to\infty} d(p_n, \tilde{p}_n) + \lim_{n\to\infty} d(\tilde{p}_n, \tilde{q}_n) + \lim_{n\to\infty} d(\tilde{q}_n, q_n) \\
&= 0 + \lim_{n\to\infty} d(\tilde{p}_n, \tilde{q}_n) + 0 \\
&= \lim_{n\to\infty} d(\tilde{p}_n, \tilde{q}_n).
\end{aligned}$$

Therefore, $\lim_{n\to\infty} d(p_n, q_n) \leq \lim_{n\to\infty} d(\tilde{p}_n, \tilde{q}_n)$. By an identical argument, we can also show that $\lim_{n\to\infty} d(\tilde{p}_n, \tilde{q}_n) \leq \lim_{n\to\infty} d(p_n, q_n)$. Hence $\lim_{n\to\infty} d(p_n, q_n) = \lim_{n\to\infty} d(\tilde{p}_n, \tilde{q}_n)$ as claimed. □

Proposition 1.1. *The function $\tilde{d} : \tilde{X} \times \tilde{X} \to \mathbb{R}$ given by $\tilde{d}([(p_n)], [(q_n)]) = \lim_{n\to\infty} d(p_n, q_n)$ is a metric on $\tilde{X}$.*

Proof. First, note that, by the lemma above, the value of $\tilde{d}([(p_n)], [(q_n)])$ is independent of the choice of representatives (p_n) and (q_n) in each equivalence class, so it is a well-defined function on $\tilde{X}$. Also note that it is a nonnegative function by construction, and that, by definition, $\tilde{d}([(p_n)], [(q_n)]) = 0$ if and only if $(p_n) \sim (q_n)$, i.e. $[(p_n)] = [(q_n)]$. Finally, the symmetry and triangle inequality properties are trivial consequences of the same properties on d and the fact that $\tilde{d}$ is well-defined. □

Proposition 1.2. *The metric space $(\tilde{X}, \tilde{d})$ is complete.*

Proof. Suppose that $([p]^m)$ is a Cauchy sequence of elements of $\tilde{X}$. Thus, each $[p]^m$ can be represented by a sequence (p_n^m) in X which is itself Cauchy with respect to d. Choose N_1 so that $\forall j, k > N_1$, $d(p_k^1, p_j^1) < \frac{1}{2}$. and choose some $n_1 > N_1$. Now, suppose that, for $i = 1..I$ a value n_i has been chosen, and then consider the sequence (p_n^{I+1}). Choose N_{I+1} so that, for all $j, k > N_{I+1}$, $d(p_j^{I+1}, p_k^{I+1}) < \frac{1}{2^{I+1}}$, which exists because (p_n^{I+1}) is a Cauchy sequence with respect to n. Then choose $n_{i+1} > \max\{n_I, N_{I+1}\}$.. Now, consider the sequence $(p_{n_m}^m)_{m=1}^\infty$. This is a sequence of elements of X, and I claim that it is Cauchy and therefore its equivalence class represents an element of $\tilde{X}$.

To see that $(p_{n_m}^m)_{m=1}^\infty$ is Cauchy, let $\epsilon > 0$. Choose M so that whenever $m_1, m_2 > M$, then $d([p^{m_1}], [p^{m_2}]) < \frac{\epsilon}{4}$, and also so that $\frac{1}{2^M} < \frac{\epsilon}{4}$. Let $m_1, m_2 > M$. Then,
$d([p^{m_1}], [p^{m_2}]) < \frac{\epsilon}{4}$. Therefore, $\lim_{n\to\infty} d(p_n^{m_1}, p_n^{m_2}) < \frac{\epsilon}{4}$, so there exists an $N \in \mathbb{N}$ so that when $n_0 > N$,

$$|d(p_{n_0}^{m_1}, p_{n_0}^{m_2}) - \lim_{n\to\infty} d(p_n^{m_1}, p_n^{m_2})| < \frac{\epsilon}{4},$$

which implies that for $n_0 > N$, $d(p_{n_0}^{m_1}, p_{n_0}^{m_1}) < \frac{\epsilon}{2}$. Now, fix $n > \max\{N, N_{m_1}, N_{m_2}\}$, and consider the following:

$$d(p_{n_{m_1}}^{m_1}, p_{n_{m_2}}^{m_2}) \leq d(p_{n_{m_1}}^{m_1}, p_n^{m_1}) + d(p_n^{m_1}, p_n^{m_2}) + d(p_n^{m_2}, p_{n_{m_2}}^{m_2}) < \frac{1}{2^{m_1}} + \frac{\epsilon}{2} + \frac{1}{2^{m_2}} < \frac{\epsilon}{4} + \frac{\epsilon}{2} + \frac{\epsilon}{4} = \epsilon,$$

where the first and third terms are controlled by the Cauchy nature of (p^{m_1}), and (p^{m_2}) in X, respectively, and the middle term is controlled by the Cauchy nature of $([(p_n)^m])$ in $\tilde{X}$. Therefore, we may conclude that the sequence $(p^{m_1}_{n_{m_1}})$ is Cauchy in X. For simplicity, let us call this sequence $p = (p_n)$ from here on.

Now, I claim that $([p^m]) \to [p]$ in the $\tilde{X}$ topology. We need to show that $\lim_{m\to\infty} \tilde{d}([p^m], [p]) = 0$, i.e. that $\lim_{m\to\infty} \lim_{n\to\infty} d(p^m_n, p_n) = 0$. Let $\epsilon > 0$. Choose M so that for $n, m > M$, $d(p_n, p_m) < \frac{\epsilon}{2}$, and so that $\frac{1}{2^M} < \frac{\epsilon}{2}$. Now let $m > M$. Since (p^m_n) is Cauchy, we have the N_m found above so that $\forall j, k > N_m$, $d(p^m_j, p^m_k) < \frac{1}{2^m} < \frac{1}{2^M} < \frac{\epsilon}{2}$, and also recall that by construction $n_m > N_m$. Then, for all $n > \max\{M, N_m\}$, we have that

$$d(p^m_n, p_n) \le d(p^m_n, p^m_{n_m}) + d(p^m_{n_m}, p_n) = d(p^m_n, p^m_{n_m}) + d(p_m, p_n) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

by our choice of n and the construction of p. Hence, for all $m > M$, $\lim_{n\to\infty} d(p^m_n, p_n) \le \epsilon$. Thus $\tilde{d}([p^m], [p]) < \epsilon$. Therefore, $\lim_{m\to\infty} \tilde{d}([p^m], [p]) = 0$ as claimed. This means that every Cauchy sequence in $\tilde{X}$ has a limit in $\tilde{X}$, so $\tilde{X}$ is indeed complete. $\square$

Proposition 1.3. *Let $i : X \to \tilde{X}$ be given by $i(x) = [(x, x, \dots, x, \dots)]$. Then i is an* isometry. *That is, for every $x, y \in X$, $\tilde{d}(i(x), i(y)) = d(x, y)$.*

Proof. First note that clearly $(x, x, \dots)$, being a constant sequence, is Cauchy, so it generates a well-defined equivalence class in $\tilde{X}$. Then, compute

$$\tilde{d}(i(x), i(y)) = \lim_{n\to\infty} d(i(x)_n, i(y)_n) = \lim_{n\to\infty} d(x, y) = d(x, y),$$

as claimed. $\square$

Proposition 1.4. *The image $i(X)$ is dense in $\tilde{X}$.*

Proof. Let $[(p_n)]$ be any element of $\tilde{X}$, and let $\epsilon > 0$. Since (p_n) is a Cauchy sequence in X, there is an $N \in \mathbb{N}$ so that, for all $n, k > N$, $d(p_n, p_k) < \epsilon$. Fix a single $k > N$, and let $x = p_k$. Then $i(x) = [(x, x, \dots)]$. Consider

$$\tilde{d}([(p_n)], i(x)) = \lim_{n\to\infty} d(p_n, i(x)_n) = \lim_{n\to\infty} d(p_n, p_k).$$

For n sufficiently large, $d(p_n, p_k) < \epsilon$ by the above calculation, so $\tilde{d}([(p_n)], i(x)) < \epsilon$. Therefore, $i(X)$ is dense in $\tilde{X}$ as claimed. $\square$

This completes the proof of Theorem 1.1.

1.0.2. *Exercises.*

Exercise 1.1. *Suppose that X is a complete metric space. Show that X and $\tilde{X}$ are isometric. (That is, the map i constructed above is surjective.)*

Exercise 1.2. *Compute the completion of the rationals with respect to the standard topology and justify your work.*

Exercise 1.3. *Give several examples of functions in the completion of $C([0,1])$ with respect to the L^2 norm that are not themselves in $C([0,1])$. For each, show an example of a Cauchy sequence whose equivalence class generates the function.*

1.1. The Space L^p.

DEFINITION 1.3. *$L^2([0,1])$ is the completion of $C([0,1])$ with respect to the L^2-metric defined in* (**??**).

DEFINITION 1.4. *$L^p([0,1])$ is the completion of $C([0,1])$ with respect to the L^p-metric defined on Test #1:*

$$\|f\|_p = \left(\int_0^1 |f(x)|^p dx\right)^{\frac{1}{p}}. \tag{2}$$

Our definition of L^p is rather abstract, so let's look at what it means for something to be an element of L^p.

- We normally think of the elements of L^p as being functions.
- Give some examples of elements of L^p:

- Give some examples of elements of L^p that are not in $C([0,1])$:

- What does the equivalence relation in the definition of $\tilde{X}$ mean in this context?
- Give an example of two equivalent functions in L^2. What makes them equivalent?

- Give some examples of elements of L^1 that are not in L^2. What about vice versa? What about, say, L^4?

- If $p \neq q$, how are L^p and L^q related? Do they contain the same functions? Why or why not? Compare and contrast with the situation for l^p and l^q.

  LEMMA 1.3. **(Hölder's Inequality)** *Suppose that $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Let $f \in L^p([0,1])$ and $g \in L^q([0,1])$. Then*

$$\int_0^1 |f(x)g(x)|dx \leq \|f\|_p \|g\|_q.$$

Proof. By Young's Inequality, we have that if $a, b \in \mathbb{R}$ then $|ab| \leq \frac{1}{p}a^p + \frac{1}{q}b^q$. Therefore, for any $x \in [0,1]$, $|f(x)g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q$. Now if f or g is identically zero, the claim is trivial, so we may assume they are not. Let $\tilde{f}(x) = \frac{f(x)}{\|f\|_p}$ and $\tilde{g}(x) = \frac{g(x)}{\|g\|_q}$. By the homogeneity of the L^p norms, these have unit L^p and L^q norms respectively. Then

$$\int \frac{|f(x)g(x)|}{\|f\|_p\|g\|_q}dx = \int |\tilde{f}(x)\tilde{g}(x)|dx \leq \int (\frac{1}{p}|\tilde{f}(x)|^p + \frac{1}{q}|\tilde{g}(x)|^q)dx = \frac{1}{p} + \frac{1}{q} = 1.$$

Thus $\int |f(x)g(x)|dx \leq \|f\|_p\|g\|_q$. □

Proposition 1.5. *Suppose that* $1 \leq p < q < \infty$. *Then* $L^q([0,1]) \subset L^p([0,1])$ *and if* $f \in L^q([0,1])$, *then* $\|f\|_p \leq \|f\|_q$.

Proof. Suppose $f \in L^q([0,1])$. Then there is a sequence (f_n) of continuous functions so that $\|f_n - f\|_q \to 0$. We will show that $\|f_n - f\|_p \to 0$ as well, which will imply that f is in L^p. Now, define $r = \frac{q}{p}$ and $s = \frac{q}{q-p}$. Note that since $q > p$, these numbers are both greater than 1, and also $\frac{1}{r} + \frac{1}{s} = 1$. Then compute:

$$\begin{aligned}\|f_n - f\|_p^p &= \int_0^1 |f_n(x) - f(x)|^p dx \\ &= \int_0^1 1 \cdot |f_n(x) - f(x)|^p dx \\ &\leq \|1\|_s \||f_n(x) - f(x)|^p\|_r,\end{aligned}$$

by Hölder's inequality (Lemma 1.3). Then $\|1\|_s = (\int 1^s dx)^{\frac{1}{s}} = 1$, and

$$\||f_n(x) - f(x)|^p\|_r = (\int_0^1 |f_n(x) - f(x)|^{p\frac{q}{p}} dx)^{\frac{p}{q}} = (\int_0^1 |f_n(x) - f(x)|^q dx)^{\frac{1}{q}p} = \|f_n - f\|_q^p.$$

Taking pth roots, we conclude that $\|f_n - f\|_p \leq \|f_n - f\|_q \to 0$. So, $f \in L^p$ as claimed.

Moreover, note that $\|f\|_p^p \leq \|1\|_s\|f\|_q^p$ by the same Hölder calculation as above, so it follows that $\|f\|_p \leq \|f\|_q$. □

Definition 1.5. *The space* l^p *is defined as* $\{(x_n)_{n=1}^\infty | \sum_{n=1}^\infty |x_n|^p < +\infty\}$, *and equipped with the norm* $\|(x_n)\|_p = (\sum_{n=1}^\infty |x_n|^p)^{\frac{1}{p}}$.

Lemma 1.4. *The space* $(l^p, \|\cdot\|_p)$ *is a Banach space (complete normed vector space).*

Proof. This proof is a combination of the proofs for l^2 and L^p. □

Proposition 1.6. *If* $1 \leq p < q < \infty$, *then* $l^p \subset l^q$.

Proof. Suppose that $(x_n) \in l^p$. Then $\sum_{n=1}^\infty |x_n|^p < +\infty$. Therefore, by the divergence test, $\lim_{n\to\infty} x_n = 0$. Hence there is an $N \in \mathbb{N}$ such that $\forall n > N$

$|x_n| < 1$. Select such an N. Now, consider $\sum_{n=1}^{\infty} |x_n|^q$:

$$\begin{aligned}
\sum_{n=1}^{\infty} |x_n|^q &= \sum_{n=1}^{N} |x_n|^q + \sum_{n=N+1}^{\infty} |x_n|^q \\
&= \sum_{n=1}^{N} |x_n|^q + \sum_{n=N+1}^{\infty} (|x_n|^p |x_n|^{q-p}) \\
&< \sum_{n=1}^{N} |x_n|^q + \sum_{n=N+1}^{\infty} (|x_n|^p 1^{q-p}) \\
&= \sum_{n=1}^{N} |x_n|^q + \sum_{n=N+1}^{\infty} |x_n|^p \\
&< +\infty
\end{aligned}$$

Here, the third line is due to the fact that $q - p > 0$ and $|x_n| < 1$ for all $n \geq N+1$, so $|x_n|^{q-p} < 1^{q-p}$ for all such n. The final line then follows from the fact that the first term on the right-hand side is a finite sum and hence finite, and the second term on the right-hand side is finite by the hypothesis that $(x_n) \in l^p$. Therefore, it follows that $(x_n) \in l^q$ as claimed. $\square$

Notice that the situations in l^p and $L^p([0,1])$ are opposites.

- Compare and contrast the case for $L^p([0,1])$ and $L^p(\mathbb{R})$.

  Suppose that $p < q$. In $\mathbb{R}$, we can construct functions that are in L^p but not L^q **and** vice versa. Therefore there is no containment result like the ones above. This is because the examples from $L^p([0,1])$ and the examples from l^p are both in play (if defined carefully). This was discussed in class and you might want to record the examples here for future reference.

1.1.1. *Exercises.*

EXERCISE 1.4. *Show that* $\|[(f_n)]\|_p := \left(\lim_{n\to\infty} \int_0^1 |f_n(x)|^p dx\right)^{\frac{1}{p}}$ *is a norm on* $L^p([0,1])$ *which generates the metric defined through the completion.*

EXERCISE 1.5. *Consider the square* $S := [0,1] \times [0,1] \subset \mathbb{R}^2$. *Give some examples of functions that are and are not in* $L^2(S)$, *and explain your reasoning.*

EXERCISE 1.6. *Show that* l^p *is a Banach space.*

1.2. The Space $W^{1,2}$. Consider the space $C^1([0,1])$ with the norm defined in Definition 0.3. As discussed above, this space is not complete. But we can compute its completion! Define

$$W^{1,2}([0,1]) := \text{ the completion of } C^1([0,1]) \text{ with respect to } \|\cdot\|_{W^{1,2}}.$$

- What kind of functions can be in $W^{1,2}$?
- Can a function in $W^{1,2}$ be discontinuous?
- Can a function in $W^{1,2}$ have a discontinuous derivative?
- Can a function in $W^{1,2}$ have a cusp?
- How continuous does a function in $W^{1,2}$ have to be?
- How does this compare to $W^{1,p}$?
- What happens if we require more derivatives?
- What happens if we change our domain from $[0,1]$ to $\mathbb{R}$?

1.2.1. *Exercises.*

EXERCISE 1.7. *Show that if* $\|\cdot\|_1$ *and* $\|\cdot\|_2$ *are both norms on a given space* X*, then* $\|\cdot\|_1 + \|\cdot\|_2$ *is also a norm on* X*.*

2. The Contraction Mapping Principle

DEFINITION 2.1. *A point* $x \in X$ *is called a* fixed point *of a function* $f : X \to X$ *if* $f(x) = x$*.*

DEFINITION 2.2. *A function* $f : X \to X$ *is called a* contraction mapping *if there exists a constant* r *with* $0 \leq r < 1$ *such that for all* $x, y \in X$

$$d(f(x), f(y)) \leq r\ d(x, y).$$

THEOREM 2.1 (Contraction Mapping Theorem). *Let* f *be a contraction mapping on a complete nonempty metric space,* X*. Then* f *has a unique fixed point.*

PROOF. Let X be a complete nonempty metric space and $f : X \to X$ be a contraction mapping.

[Uniqueness] Assume $x, y \in X$ are fixed points. This implies $d(x, y) = d(f(x), f(y))$. But since f is a contraction we have $d(x, y) = d(f(x), f(y)) \leq r\ d(x, y)$ for $0 \leq r < 1$. This is impossible if $d(x, y) > 0$. Thus $d(x, y) = 0$. Therefore $x = y$.

[Existence] Let $x_0 \in X$. Define the sequence $\{x_n\}$ recursively by

$$x_{n+1} = f(x_n)$$

for $n \geq 0$. We claim $\{x_n\}$ is a Cauchy sequence:

Let $a = d(x_0, x_1)$. Now consider $d(x_{n+1}, x_n)$. Since f is a contraction mapping, we have

$$d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq r d(x_n, x_{n-1}).$$

By repeating this n times, we see that $d(x_{n+1}, x_n) \leq r^n d(x_0, x_1) = ar^n$. Now if $m < n$, then

$$\begin{aligned}
d(x_m, x_n) &\leq \sum_{j=m}^{n-1} d(x_{j+1}, x_j) \\
&\leq \sum_{j=m}^{n-1} ar^j \\
&= \frac{ar^m - ar^n}{1-r} \\
&= \frac{ar^m(1 - r^{n-m})}{1-r} \\
&\leq \frac{ar^m}{1-r}.
\end{aligned}$$

We used the triangle inequality to obtain the first inequality. For the second inequality, we used the previous calculation that $d(x_{n+1}, x_n) \leq ar^n$. The two subsequent equalities are from the geometric series summation formula. The last inequality holds since $m < n$ and $r < 1$, so we have $1 - r^{n-m} < 1$.

Let $\epsilon > 0$. Choose N large enough such that $\frac{ar^N}{1-r} < \epsilon$. Now for $n \geq m > N$ we have $d(x_m, x_n) \leq \frac{ar^m}{1-r} < \frac{ar^N}{1-r} < \epsilon$. Thus $\{x_n\}$ is a Cauchy sequence.

Now since X is complete and $\{x_n\}$ is Cauchy we know that $\{x_n\}$ converges in X. Let $x = \lim_{n\to\infty} x_n$. Since f is continuous, we have that $f(x) = \lim_{n\to\infty} f(x_n)$. But since $f(x_n) = x_{n+1}$ we have $f(x) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x$. Thus $f(x) = x$. Therefore x is a fixed point. □

This proof actually gives a constructive method to find the fixed point, by iteration of the map.

Corollary 2.1. *Let f be a contraction mapping on a complete nonempty metric space X. If x_0 is any point of X, and $x_{n+1} = f(x_n)$ for $n \geq 0$ then the sequence $\{x_n\}$ converges to the fixed point of f.*

2.0.2. *Exercises.*

Exercise 2.1. *Show that any contraction mapping is continuous.*

2.1. Existence and Uniqueness for Solutions to Differential Equations. In this section we apply the Contraction Mapping Theorem to derive a fundamental result for ordinary differential equations. You may recall from your undergraduate ODE class that you learned to solve *initial value problems* such as

Example 2.1.

$$y'(t) = (y(t))^2, y(0) = 2.$$

If you consider other examples such as

EXAMPLE 2.2.
$$y'(t) = (y(t))^{\frac{1}{2}}, y(0) = 0,$$
then you will discover that this problem has more than one solution.

How can we generalize what we observe in these examples? It would be extremely useful to be able to look at a problem and quickly determine whether or not it has a solution. For applied problems it is also very important to know that the problem has just one solution. Otherwise how do you know which solution is the one that appears in the *real world*?

We begin by noticing that the general initial value problem (IVP), can be stated as
$$(IVP): \quad y'(t) = f(y), y(0) = y_0. \tag{3}$$
From the examples above it is clear that having a continuous f is not enough to guarantee the existence of exactly one solution. We need something that is a little bit better than continuous, such as ...

DEFINITION 2.3. *$f : [a, b] \to \mathbb{R}$ is* Lipschitz continuous *if there is a $k > 0$ so that, $\forall x, y \in [a, b]$, $|f(x) - f(y)| \leq k|x - y|$.*

Note that Lemma 0.1 states that if a function has a bounded derivative then it is Lipschitz. However, Lipschitz functions do not need to have a derivative at every point in their domain.

DEFINITION 2.4. *$f : [a, b] \times \mathbb{R} \to \mathbb{R}$ is* uniformly Lipschitz continuous *if there is a $k > 0$ so that, $\forall t \in [a, b]$, $\forall x, y \in \mathbb{R}$, $|f(t, x) - f(t, y)| \leq k|x - y|$.*

THEOREM 2.2. *If f is uniformly Lipschitz continuous on an open rectangle containing $(0, y_0)$, then there is an $\epsilon > 0$ and a differentiable function $y \in C^1([0, \epsilon])$ such that y is the unique solution of* (3) *on* $[0, \epsilon]$.

The plan is to transform the problem into a fixed point problem and then apply the contraction mapping theorem.

PROOF. First, note that y is a C^1 solution to (3) on $[0, \epsilon]$ if and only if it solves $y(0) = y_0$ and
$$\int_0^t y'(s)ds = \int_0^t f(s, y(s))ds$$
for all $t \in [0, \epsilon]$. This equation was obtained by integrating both sides of the differential equation from 0 to t. Next, we apply the Fundamental Theorem of Calculus and the initial value to get
$$y(t) - y_0 = \int_0^t f(s, y(s))ds.$$
Thus we have a fixed point problem $y = F(y)$, where
$$F(y)(t) = y_0 + \int_0^t f(s, y(s))ds.$$
Rather than seek a fixed point in all of $C([0, 1])$, we will define an appropriate closed ball on which to work. Let
$$B := B_1(y_0) = \{y \in C([0, \epsilon]) : \sup\{|y(t) - y_0| : 0 \leq t \leq \epsilon\} \leq 1\}.$$
Since B is a closed subset of a complete metric space, $C([0, 1])$, it is also complete. (Exercise.) For the moment we have not yet chosen ϵ.

We must check that, for an appropriate choice of ϵ, F is a contraction mapping on B. First we must check that at least F maps B to itself. So, suppose that $y \in B$. Then, forall $t \in [0, \epsilon]$, $|y(t) - y_0| < 1$. Therefore,

$$\begin{aligned}
|F(y)(t) - y_0| &= |y_0 + \int_0^t f(s, y(s))ds - y_0| \\
&\leq \int_0^t |f(s, y(s))|ds \\
&\leq \int_0^t |f(s, y(s)) - f(s, y_0)|ds + \int_0^t |f(s, y_0)|ds \\
&\leq \int_0^t k|y(s) - y_0|ds + \sup_{t\in[0,1]} |f(t, y_0)|\epsilon \\
&\leq k\epsilon + \sup_{t\in[0,1]} |f(t, y_0)|\epsilon
\end{aligned}$$

So, $\sup\{|F(y)(t) - y(0)| : 0 \leq t \leq \epsilon\} \leq \epsilon(k + \sup_{t\in[0,1]} |f(t, y_0)|)$. In order for $F(y)$ to be in B, we need this to be at most 1, and it is clear that we can achieve that for $\epsilon = \frac{1}{2(k+\sup_{t\in[0,1]} |f(t,y_0)|}$. Hence for ϵ sufficiently small, F maps B to itself.

It remains to show that F is a contraction mapping on B. Let $y, z \in B$ and consider

$$\begin{aligned}
|F(y)(t) - F(z)(t)| &= |y_0 + \int_0^t f(s, y(s))ds - (y_0 + \int_0^t f(s, z(s))ds| \\
&\leq \int_0^t |f(s, y(s)) - f(s, z(s))|ds \\
&\leq \int_0^t k|y(s) - z(s)|ds \\
&\leq k\epsilon \sup_{0\leq t\leq\epsilon} |y(t) - z(t)| \\
&\leq \frac{k}{2(k + |f(y_0)|)} \sup_{0\leq t\leq\epsilon} |y(t) - z(t)|.
\end{aligned}$$

Letting $r = \frac{k}{2(k+\sup_{t\in[0,1]} |f(t,y_0)|)}$, which is clearly positive and strictly less than one, we conclude that $\|F(y) - F(z)\|_\infty \leq r\|y - z\|_\infty$ if $y, z \in B$. Therefore F is a contraction mapping on B as claimed.

Then by the Contraction Mapping Principle there is a unique fixed point of F in B. Therefore (3) has a unique solution on $[0, \epsilon]$. □

2.1.1. *Exercises.*

EXERCISE 2.2. *Find a solution to Example 2.1. Notice that you can find infinitely many functions that solve the ODE, but only one which solves the ODE* **and** *satisfies the initial value. Also notice that the solution is only valid on a finite interval, which can be somewhat surprising given that the problem looks* nice.

EXERCISE 2.3. *Find at least two solutions to Example 2.2, both solving the same given initial condition.*

EXERCISE 2.4. *Prove that a closed subset of a complete metric space is complete.*

3. Function Approximation

It is often helpful to know that functions that are not so nice can be approximated by functions that are very nice. Another way to express this is to say that the set of *nice* functions is dense in the set of *not so nice* functions. (How is that for being precise?!)

The term *nice* can mean a lot of things. Being infinitely differentiable is nice. Being analytic, *i.e.* a power series, is nicer. Being a polynomial is perhaps the nicest. In this section we will prove that every function in $C([0,1])$ can be approximated to any degree of accuracy by a polynomial, *i.e.* that polynomials are dense in $C([0,1])$. But before getting to this famous theorem of Weierstrass we will discuss function approximation in general.

Suppose that we are trying to approximate the discontinuous function

$$f(x) = \left\{ \begin{array}{ll} 0 & 0 \leq x < \frac{1}{2} \\ 1 & \frac{1}{2} \leq x \leq 1 \end{array} \right.$$

We would like the approximating function to be a little bit nicer, *i.e.* continuous. For the moment we won't be too picky about the metric being used, we just want to find a nice continuous function that approximates f in an intuitive way. Moreover, we would like to be able to adjust the approximation so that it can be improved, *i.e.* we want to create a sequence of approximations, (f_n), that get better as $n \to \infty$.

We already know one sequence of functions that meets these requirements, namely

$$f_n(x) = \begin{cases} (2x)^n & 0 \leq x < \frac{1}{2} \\ 1 & \frac{1}{2} \leq x \leq 1 \end{cases}.$$

Here is another choice, g_n:

$$g_n(x) = \begin{cases} 0 & 0 \leq x < \frac{1}{2} - \frac{1}{n} \\ \frac{n}{2}(x - \frac{1}{2} + \frac{1}{n}) & \frac{1}{2} - \frac{1}{n} \leq x < \frac{1}{2} + \frac{1}{n} \\ 1 & \frac{1}{2} + \frac{1}{n} \leq x \leq 1 \end{cases}$$

You should sketch some graphs of f, f_n, and g_n. Notice that (g_n) is an approximation of f by piecewise linear functions, which is quite a bit easier (and more natural) to work with than (f_n).

How can we best capture this idea using analysis? First, notice that $g_n(x)$ is the average of $f(x)$ over the interval $[x - \frac{1}{n}, x + \frac{1}{n}]$. This is an important conceptual point. Averaging in analysis is a way of smoothing out irregularities and is most often achieved via integration. Recall from your calculus experience that the average of a function $f(x)$ over an interval $[a, b]$ is given by

$$\frac{1}{b-a} \int_a^b f(x) dx.$$

Thus we can rewrite our formula for $f_n(x)$ as

$$f_n(x) = \frac{n}{2} \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} f(y) dy.$$

Notice that there is a small technical problem with this formulation. If x is close to 1, and we attempt to average over the interval $[x - \frac{1}{n}, x + \frac{1}{n}]$, then it could be that $x + \frac{1}{n} > 1$, so that part of our averaging happens in an area where $f(x)$ is not defined. This problem is easily fixed by *extending* $f(x)$ to a function that has values outside of $[0, 1]$, its original domain. There is more than one way to do this. Most commonly we will extend the function so that it is eventually 0 outside of some larger interval.

There is a second technical problem that, at first, might not seem like a problem. How can all of the averaging information be consolidated into one convenient package inside the integral? Right now the information lies both in the limits of integration and in the constant in front of the integral. If we wanted to modify the averaging process in any way, it is not easy to see how we might do this using the current form. Denote by $\chi_S(x)$ the indicator function of a set S. That is, $\chi_S(x) = 1$ if $x \in S$ and 0 otherwise. Here is a standard way to rewrite the integral...

$$\begin{aligned} \tfrac{n}{2}\int_{x-\frac{1}{n}}^{x+\frac{1}{n}} f(y)dy &= \tfrac{n}{2}\int_{-\infty}^{\infty} \chi_{[x-\frac{1}{n},x+\frac{1}{n}]}f(y)dy \\ &= \tfrac{n}{2}\int_{-\infty}^{\infty} \chi_{[-\frac{1}{n},\frac{1}{n}]}(y-x)f(y)dy \\ &= \int_{-\infty}^{\infty} K_n(y-x)f(y)dy, \end{aligned}$$

where $K_n(y) := \frac{n}{2}\chi_{[-\frac{1}{n},\frac{1}{n}]}$. K_n is often called the *mollifier* or the *convolution kernel.* Notice that it captures all of the information necessary to describe the approximation scheme. The approximation process can be summarized as follows. To approximate a function $f(y)$ near a point x ...

- : (i) Extend f to a function whose domain is all of $\mathbb{R}$.
- : (ii) Create an appropriate kernel $K_n(y)$.
- : (iii) Shift $K_n(y)$ so that it is centered at x, *i.e.* $K_n(y - x)$.
- : (iv) Multiply $f(y)$ by $K_n(y - x)$.
- : (v) Integrate over $\mathbb{R}$.

As a technical detail notice that K_n is even, *i.e.* $K_n(-y) = K_n(y)$, so we can switch $K_n(y - x)$ with $K_n(x - y)$ in the integral without changing anything. This ends up being a technical convenience in later definitions.

The key properties that we abstract from the process above are that

Definition 3.1. *A sequence of functions* (K_n) *are called mollifiers if:*

- : nonnegativity: $K_n \geq 0$,
- : unit area: $\int_{-\infty}^{\infty} K_n(y)dy = 1$, *and*
- : concentration: $K_n(y)$ concentrates *at* 0, i.e. *given any* $\delta > 0$, *we have*

$$\lim_{n\to\infty}\left(\int_{-\infty}^{-\delta} K_n(y)dy + \int_{\delta}^{\infty} K_n(y)dy\right) = 0.$$

We will use the formula

$$f_n(x) := (f * K_n)(x) = \int_{-\infty}^{\infty} f(y)K_n(x-y)dy,$$

to create a sequence of approximations.

3.1. Convolution and Mollification.

DEFINITION 3.2. *Suppose that f, g are functions. The* convolution *of f and g is defined to be*

$$f * g(x) := \int_{-\infty}^{\infty} f(y)g(x-y)dy,$$

as long as this integral is well-defined.

REMARK 3.1. *Given what we know in this course so far, it is difficult to give a technically precise definition of the broadest cases where this convolution operation is well-defined. It will turn out that as long as f is in one of the L^p spaces for some $p \geq 1$ and g is "nice" (e.g. continuous, and compactly supported), then the convolution will be well defined. For the purposes of approximation, this is all that we need. In fact, g does not even need to be quite that nice.*

DEFINITION 3.3. *We say that a function $f : \mathbb{R} \to \mathbb{R}$ is* integrable *if $\int_{-\infty}^{\infty} f(x)dx$ is well-defined and finite.*

DEFINITION 3.4. *A function $g : \mathbb{R} \to \mathbb{R}$ is called* compactly supported *if $\exists M > 0$ such that $g(x) = 0$ for every x such that $|x| > M$.*

LEMMA 3.1. *The convolution of f and g satisfies the following properties:*

(1) *$f * g(x) = g * f(x)$.*
(2) *If f is integrable, g is bounded, and the convolution is well-defined, then $f * g$ is bounded.*
(3) *If f is integrable and g is continuous and compactly supported, then $f*g$ is uniformly continuous.*
(4) *If f is integrable and g is continuously differentiable and compactly supported, then $f * g$ is differentiable and $(f * g)' = f * (g')$.*

Note that the last part of this lemma can be applied repeatedly to obtain higher levels of differentiability for $f * g$ if g is smooth.

PROOF. For part 1, compute, making the u-substitution $u = x-y$, $y = x-u$, $du = -dy$,

$$\begin{aligned} f * g(x) &= \int_{-\infty}^{\infty} f(y)g(x-y)dy \\ &= \int_{\infty}^{-\infty} f(x-u)g(u)(-du) \\ &= \int_{-\infty}^{\infty} f(x-u)g(u)du \\ &= g * f(x) \end{aligned}$$

as claimed.

For part 2, first note that, since f is integrable, $0 \leq \int_{-\infty}^{\infty} |f(x)|dx < +\infty$. Let $M = \int_{-\infty}^{\infty} |f(x)|dx$. Also, since g is continuous with compact support, it is bounded, so there is a $C > 0$ so that, $\forall x \in \mathbb{R}$, $|g(x)| \leq C$. Then, for $x \in \mathbb{R}$,

$$|f * g(x)| \leq \int_{-\infty}^{\infty} |f(y)||g(x-y)|dy \leq C \int_{-\infty}^{\infty} |f(y)|dy \leq CM.$$

Therefore, $f * g$ is bounded.

For part 3, we want to show that $f * g$ is uniformly continuous. Note that if g is continuous and compactly supported, then g is uniformly continuous, so $\forall \epsilon > 0\ \exists \delta > 0$ so that if $|x_1 - x_2| < \delta$ then $|g(x_1) - g(x_2)| < \epsilon$. Let $\epsilon > 0$, and choose $\delta > 0$ so that if $|x_1 - x_2| < \delta$ then $|g(x_1) - g(x_2)| < \frac{\epsilon}{M}$. Then suppose that $x_1, x_2 \in \mathbb{R}$ with $|x_1 - x_2| < \delta$, and note that this implies that, $\forall y \in \mathbb{R}$, $|(x_1 - y) - (x_2 - y)| < \delta$. Then compute:

$$\begin{aligned}
|(f * g)(x_1) - (f * g)(x_2)| &= \left| \int_{-\infty}^{\infty} f(y)g(x_1 - y)dy - \int_{-\infty}^{\infty} f(y)g(x_2 - y)dy \right| \\
&= \left| \int_{-\infty}^{\infty} f(y)(g(x_1 - y) - g(x_2 - y))dy \right| \\
&\leq \int_{-\infty}^{\infty} |f(y)||g(x_1 - y) - g(x_2 - y)|dy \\
&< \int_{-\infty}^{\infty} |f(y)| \frac{\epsilon}{M} dy \\
&= M \frac{\epsilon}{M} = \epsilon.
\end{aligned}$$

Hence it follows that $f * g$ is uniformly continuous.

Finally, for the fourth claim, suppose that g is a continuously differentiable function. Note that this means that g' is a continuous. Let $x \in \mathbb{R}$ and consider the difference quotient $D_h(x) = \frac{f*g(x+h) - f*g(x)}{h}$. We want to show that $\lim_{h \to 0} D_h(x)$ exists and equals $f * (g')(x)$. But

$$\begin{aligned}
D_h(x) &= \frac{f * g(x+h) - f * g(x)}{h} \\
&= \frac{1}{h} \left(\int_{-\infty}^{\infty} f(y)g(x + h - y)dy - \int_{-\infty}^{\infty} f(y)g(x - y)dy \right) \\
&= \frac{1}{h} \left(\int_{-\infty}^{\infty} f(y)[g(x + h - y) - g(x - y)]dy \right) \\
&= \int_{-\infty}^{\infty} f(y) \left[\frac{g(x + h - y) - g(x - y)}{h} \right] dy.
\end{aligned}$$

Therefore, since the integrals are uniformly bounded in h (because g and the derivative of g are uniformly bounded),

$$\begin{aligned}
\lim_{h \to 0} D_h(x) &= \lim_{h \to 0} \int_{-\infty}^{\infty} f(y) \left[\frac{g(x + h - y) - g(x - y)}{h} \right] dy \\
&= \int_{-\infty}^{\infty} f(y) \left[\lim_{h \to 0} \frac{g(x + h - y) - g(x - y)}{h} \right] dy \\
&= \int_{-\infty}^{\infty} f(y)g'(x - y)dy \\
&= f * (g')(x)
\end{aligned}$$

as claimed. $\square$

Based on this lemma, we see that $f*g$ inherits the smoothness properties of the smoothest function (f or g) in the convolution. Suppose that, as in the introduction to this section, we have a function $f : [0,1] \to \mathbb{R}$ that is "not nice." Extend f to $\mathbb{R}$ so that it is compactly supported, and hence integrable. If g is "nice," then $f*g$ will be just as "nice." Our example earlier in the section tells us that, in fact, $f*g$ may be nicer than *both* f and g, since g_n is continuous for each n even though f is not, and $K_n = \frac{n}{2}\chi_{[-\frac{1}{n},\frac{1}{n}]}$ is also not continuous.

Finally, we need to confirm that, if we convolve f with a sequence of averaging kernels K_n, then as $n \to \infty$, $f_n = K_n * f$ does indeed converge to f. For the purposes of the Weierstrass Approximation Theorem, the following lemma will suffice:

Lemma 3.2. *Suppose that f is a bounded, integrable function on $\mathbb{R}$, and that S is a compact subset of $\mathbb{R}$ on which f is continuous. Then, if K_n is a sequence of convolution kernels satisfying Definition 3.1, the functions $f_n := f * K_n$ converge to f uniformly on S.*

Proof. We have that $f_n(x) = \int_{-\infty}^{\infty} f(x-t)K_n(t)dt$ by part 1 of Lemma 3.1. Additionally, because $\int_{-\infty}^{\infty} K_n(x)dx = 1$, we have that $f(x) = \int_{-\infty}^{\infty} f(x)K_n(t)dt$. So,

$$f_n(x) - f(x) = \int_{-\infty}^{\infty} [f(x-t) - f(x)]K_n(t)dt.$$

Let $\epsilon > 0$. We want to show that there is an N so that $\forall n > N$, $\forall x \in S$, $|f_n(x) - f(x)| < \epsilon$. Since S is compact and f is continuous on S, f is uniformly continuous on S. Therefore, there exists a $\delta > 0$ so that, if $x \in S$ and $|t| < \delta$, then $|f(x-t) - f(x)| < \frac{\epsilon}{2}$. Let $M = \sup_{\mathbb{R}} |f(x)|$, which is finite because f is bounded. Since $\lim_{n\to\infty} \left(\int_{-\infty}^{-\delta} K_n(t)dt + \int_{\delta}^{\infty} K_n(t)dt \right) = 0$, there is an $N \in \mathbb{N}$ so that $\forall n > N$, $\int_{-\infty}^{-\delta} K_n(t)dt + \int_{\delta}^{\infty} K_n(t)dt < \frac{\epsilon}{4M}$. Finally, let $n > N$ and compute, for any $x \in S$,

$$\begin{aligned}
|f_n(x) - f(x)| &= \left| \int_{-\infty}^{-\delta} [f(x-t) - f(x)]K_n(t)dt + \int_{-\delta}^{\delta} [f(x-t) - f(x)]K_n(t)dt + \int_{\delta}^{\infty} [f(x-t) - f(x)]K_n(t)dt \right| \\
&\leq \int_{-\infty}^{-\delta} |f(x-t) - f(x)|K_n(t)dt + \int_{-\delta}^{\delta} |f(x-t) - f(x)|K_n(t)dt + \int_{\delta}^{\infty} |f(x-t) - f(x)|K_n(t)dt \\
&= \int_{-\infty}^{-\delta} |f(x-t) - f(x)|K_n(t)dt + \int_{\delta}^{\infty} |f(x-t) - f(x)|K_n(t)dt + \int_{-\delta}^{\delta} |f(x-t) - f(x)|K_n(t)dt \\
&< (2M)\frac{\epsilon}{4M} + \frac{\epsilon}{2} \cdot 1 = \epsilon,
\end{aligned}$$

where we have used our choices of δ and n above, and the fact that $\int_{-\delta}^{\delta} K_n(t)dt \leq \int_{-\infty}^{\infty} |f(x-t) - f(x)|K_n(t)dt = 1$. Since x was arbitrary, it follows that $f_n \to f$ uniformly on S as claimed. □

3.1.1. *Exercises.*

Exercise 3.1. *Suppose that K_n is $\frac{n}{2}\chi_{[-\frac{1}{n},\frac{1}{n}]}(x)$ (the example discussed above). For each of the following functions, draw pictures of f and $f * K_n$ for several values of n.*

-
$$f(x) = \begin{cases} 1 & \frac{1}{2} \leq x \leq 1 \\ 0 & \textit{otherwise} \end{cases}$$

- $$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$
- $$f(x) = \begin{cases} 4(x - \frac{1}{2})^2 & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

(You may want to use Maple to visualize this.)

EXERCISE 3.2. *Show that if f and g are both integrable functions, then $f * g$ is also integrable, and*

$$\int_{-\infty}^{\infty} |f * g(x)| dx \leq \left(\int_{-\infty}^{\infty} |f(x)| dx \right) \left(\int_{-\infty}^{\infty} |g(x)| dx \right).$$

This is a special case of Young's inequality *for convolutions.*

3.2. The Weierstrass Approximation Theorem.

THEOREM 3.1. *(Weierstrass Approximation Theorem) The set of polynomials is dense in $C[0, 1]$.*

In order to prove this theorem, we need to apply our work in the previous section to a well-chosen sequence of convolution kernels which will produce polynomials.

Define

$$q_n(x) = \begin{cases} (1 - \frac{x^2}{4})^n & -2 \leq x \leq 2 \\ 0 & |x| > 2 \end{cases},$$

and $c_n := \int_{-\infty}^{\infty} q_n(x) dx$. Then let $p_n(x) = \frac{1}{c_n} q_n(x)$. Notice that I have not computed c_n, and in fact I never need to while proving the statements below. I would have to compute c_n if I wanted to find explicit formulas for the approximations.

LEMMA 3.3. *The (p_n) form a sequence of convolution kernels satisfying the conditions listed in 3.1.*

PROOF. The nonnegativity and evenness of p_n should be clear from the definition. The unit area condition is guaranteed by our choice of c_n. We just need to check the concentration property. This requires a little bit of calculus. Compute $q_n'(x) = \frac{-2nx}{4} \left(1 - \frac{x^2}{4}\right)^{n-1}$, and $q_n''(x) = n \left(1 - \frac{x^2}{4}\right)^{n-2} [\frac{x^2}{8}(2n - 1) - \frac{1}{2}]$. Notice that this is always greater than or equal to its value at $x = 0$, which is $-\frac{n}{2}$. Therefore the function $q_n(x)$ is greater than or equal to $g_n := 1 - \frac{n}{4} x^2$ for all x in the interval. This parabola is the quadratic Taylor series for q_n at 0. Notice next that $g_n = 0$ for $x = \pm \frac{2}{\sqrt{n}}$. Therefore,

$$c_n = \int_{-2}^{2} q_n(x) dx \geq \int_{-\frac{2}{\sqrt{n}}}^{\frac{2}{\sqrt{n}}} q_n(x) dx \geq \int_{-\frac{2}{\sqrt{n}}}^{\frac{2}{\sqrt{n}}} g_n(x) dx = (x - \frac{nx^3}{12})|_{-\frac{2}{\sqrt{n}}}^{\frac{2}{\sqrt{n}}} = \frac{8}{3\sqrt{n}}.$$

Then, for $\delta > 0$,

$$\begin{aligned}\int_{-\infty}^{-\delta} p_n(x)dx + \int_{\delta}^{\infty} p_n(x)dx &= \frac{2}{c_n}\int_{\delta}^{2} q_n(x)dx \\ &\le \frac{3\sqrt{n}}{4}\int_{\delta}^{2}(1-\frac{\delta^2}{4})^n dx \\ &\le \frac{3\sqrt{n}}{2}(1-\frac{\delta^2}{4})^n.\end{aligned}$$

Set $r = 1 - \frac{\delta^2}{4}$, and note that $r < 1$. Therefore $\lim_{n\to\infty} 3\sqrt{n}r^n = 0$ by an application of L'Hopital's Rule (exercise!), and the concentration property holds as claimed. $\square$

Now that we have constructed p_n, we are ready to prove the theorem:

THEOREM 3.2. *(Weierstrass Approximation Theorem, Version 2) Let $f \in C([0,1])$. Then there is a sequence of polynomials (f_n) on $[0,1]$ so that $f_n \to f$ uniformly on $[0,1]$.*

PROOF. Let $f \in C([0,1])$. We need f to have a domain of $\mathbb{R}$ and to be bounded and piecewise continuous on $\mathbb{R}$. So, extend f to $[-1,2]$ in any fashion which makes f continuous on $[-1,2]$ and $f(-1) = f(2) = 0$. Then extend f to the rest of $\mathbb{R}$ by 0. Clearly, then, f will be bounded and piecewise continuous on $\mathbb{R}$. Note also that $[0,1]$ is a compact subset of $\mathbb{R}$ on which f is continuous.

Let (p_n) be as constructed above. By Lemma 3.3, these are a sequence of mollifiers. Therefore, by Lemma 3.2, the functions $f_n := p_n * f$ converge uniformly to f. It remains only to show that f_n is a polynomial on $[0,1]$ for any n.

Recall that $f_n(x) = \int_{-\infty}^{\infty} f(t)p_n(x-t)dt$. Also, since $f \equiv 0$ outside $(-1,2)$, and $0 \le x \le 1$, the only nonzero contributions to this integral occur when $-2 < x - t < 2$. On this domain, p_n is a polynomial, so $p_n(x-t)$ can be written as $\sum_{i=0}^{k} a_i(x-t)^i$ for some power k and coefficients $\{a_i\}$. By the binomial theorem, this is $\sum_{i=0}^{k}\sum_{j=0}^{i} a_i(-1)^{i-j}\binom{i}{j}x^j t^{i-j}$. So,

$$\begin{aligned}f_n(x) &= \int_{-\infty}^{\infty} f(t)p_n(x-t)dt \\ &= \int_{-1}^{2} f(t)p_n(x-t)dt \\ &= \int_{-1}^{2} f(t)\left[\sum_{i=0}^{k}\sum_{j=0}^{i} a_i(-1)^{i-j}\binom{i}{j}x^j t^{i-j}\right]dt \\ &= \sum_{i=0}^{k}\sum_{j=0}^{i} x^j\left[a_i(-1)^{i-j}\binom{i}{j}\int_{-1}^{2} f(t)t^{i-j}dt\right],\end{aligned}$$

which is clearly a polynomial in x. $\square$

3.2.1. *Exercises.*

EXERCISE 3.3. *Check that* $\lim_{n\to\infty} 3\sqrt{n}r^n = 0$ *if* $r < 1$ *as claimed in the proof above.*

EXERCISE 3.4. *Check that the two versions of the Weierstrass Approximation Theorem are equivalent.*

Exercise 3.5. *Sketch graphs of p_n for $n = 1, 2, 3, ...$*

Exercise 3.6. *Use a computer to compute the first few f_n for each of the following choices of f and confirm that they are polynomials and that they are good uniform approximations of f.*

- $f(x) = x$.
- $f(x) = e^x$.
- $f(x) = \sin(\pi x)$.

Exercise 3.7. *Look up the proof of the Weierstrass Approximation Theorem that uses Bernstein Polynomials. Can you see that this is also just a form of averaging? (You don't have to memorize the proof; you don't have to turn this problem in, it is a reading and thinking assignment.)*

CHAPTER 2

Calculus in Normed Vector Spaces

Having defined metric spaces and normed linear spaces, and spent some time looking at the fundamental topological properties thereon, we are ready to do some calculus. We will begin by discussing differentiability.

1. Differentiability

1.1. Review of Differentiability on $\mathbb{R}^n$.

DEFINITION 1.1. *A function* $f : \mathbb{R} \to \mathbb{R}$ *is* differentiable *at* $x_0 \in \mathbb{R}$ *if* $\lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$ *exists. If so, we define* $f'(x_0) = \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$*.*

DEFINITION 1.2. *The "little o" notation* $o(x - x_0)$ *represents an error term with the property that* $\lim_{x\to x_0} \frac{o(x-x_0)}{(x-x_0)} = 0$.

LEMMA 1.1. *A function* $f : \mathbb{R} \to \mathbb{R}$ *is differentiable at* x_0 *if and only if there is a real number* L *such that* $f(x) = f(x_0) + L(x - x_0) + o(x - x_0)$*.*

REMARK 1.1. *If the number* L *exists, then we say that* L *is the derivative of* f *at* x_0 *and we write* $L = f'(x_0)$*.*

PROOF. We proved this lemma together in class on Monday, March 19. You should record the proof here for your future reference.

$\square$

The alternate definition of differentiability given by this lemma expresses the notion of *local linear approximation* in a natural way, and this is the notion that serves as a defining property for derivatives in more general settings. We recall from multivariable calculus the following definition:

DEFINITION 1.3. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{x}_0$ if there is a vector $\vec{L} \in \mathbb{R}^n$ such that $f(\vec{x}) = f(\vec{x}_0) + \vec{L} \cdot (\vec{x} - \vec{x}_0) + o(\|\vec{x} - \vec{x}_0\|)$. If f is differentiable at x_0, then the function $L(\vec{x}) = \vec{L} \cdot (\vec{x} - \vec{x}_0)$ is called the* local linear approximation *or* tangent plane approximation *to f at $\vec{x}_0$.*

EXAMPLE 1.1. *On $\mathbb{R}^2$, the functions $f(x, y) = |(|x| - |y|)| - |x| - |y|$ and $g(x, y) = \frac{3x^2y}{x^2+y^2}$ are not differentiable at $(0, 0)$ even though $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial g}{\partial x}$, and $\frac{\partial g}{\partial y}$ all exist (and equal 0) at $(0, 0)$.*

1.1.1. *Exercises.*

EXERCISE 1.1. *Construct at least two different examples of differentiable functions on $\mathbb{R}^3$.*

EXERCISE 1.2. *Construct a function $f : \mathbb{R}^3 \to \mathbb{R}$ such that, at $\vec{x}_0 = (0, 0, 0)$, $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial f}{\partial z}$ all exist, but f is not differentiable at x_0.*

EXERCISE 1.3. *Explain, in paragraph form, why the functions in Example 1.1 are not differentiable; make an explicit comparison with Definition 1.3*

1.2. Linear Operators on Normed Linear Spaces. In $\mathbb{R}^n$, we know what it means for a function to be differentiable, because we have a clear idea of what a "local linear approximation" should look like. Before we can define differentiability for functions on a normed linear space, we must first discuss what it means for such a function to be linear.

DEFINITION 1.4. *Let X, Y be normed linear spaces. $L : X \to Y$ is a* linear operator *if*

(1) *$L(cx) = cL(x)$ for all $c \in \mathbb{R}$ and all $x \in X$.*
(2) *$L(x_1 + x_2) = L(x_1) + L(x_2)$ for all $x_1, x_2 \in X$.*

An important question to ask about any new classification of functions is whether or not they are continuous. It turns out that not all linear operators are continuous. This is a surprising statement, given that linear operators are usually considered to be about as *nice* as a function can be.

EXAMPLE 1.2. *Let $X = Y = C^\infty([0, 1])$ = the space of all infinitely differentiable functions on $[0, 1]$. Equip X with the L^2 norm. Then let $L : X \to Y$ be given by $L(f) = \frac{df}{dx}$. Then L is linear but not continuous.*

EXAMPLE 1.3. *Let $X = C^1([0, 1])$, and $Y = C([0, 1])$ and equip* **both** *X and Y with the sup-norm. Then let $L : X \to Y$ be given by $L(f) = \frac{df}{dx}$. Then L is linear but not continuous.*

Note that this is a subtlety that is not encountered in finite dimensions.

DEFINITION 1.5. *Suppose that $L : X \to Y$ is a linear operator as in the previous definition. We say that L is a* bounded *linear operator if there is a constant $c > 0$ such that $||Lx||_Y \leq c||x||_X$ for all $x \in X$.*

LEMMA 1.2. *A linear operator $L : X \to Y$ is bounded if and only if $||L||_{op} := \sup\{||L(x)||_Y : x \in X, ||x||_X = 1\} < +\infty$. If so, then, $\forall x \in X$, $\|Lx\|_Y \leq \|L\|_{op}\|x\|_X$.*

PROOF. Exercise. $\square$

THEOREM 1.1. *Let $\mathcal{B}(X,Y)$ be the set of all bounded linear operators from X to Y. Then $(\mathcal{B}(X,Y), ||\cdot||_{op})$ is a normed linear space. If Y is complete, then so is $\mathcal{B}(X,Y)$.*

PROOF. We define addition and scalar multiplication as follows. If L_1 and L_2 are in $\mathcal{B}(X,Y)$, then $(L_1+L_2) : X \to Y : (L_1+L_2)(x) = L_1(x) + L_2(x)$. If $(\alpha \in \mathbb{R})$, then $\alpha L_1 : X \to Y : (\alpha L_1)(x) = \alpha(L_1(x))$. Notice that for any $x \in X$ with $||x||_X \leq 1$ we have $||(L_1+L_2)(x)||_Y = ||L_1(x)+L_2(x)||_Y \leq ||L_1(x)||_Y + ||L_2(x)||_Y \leq ||L_1||_{op} + ||L_2||_{op}$. Hence (L_1+L_2) is bounded with $||L_1+L_2||_{op} \leq ||L_1||_{op}+||L_2||_{op}$. A similar, and simpler, argument confirms that αL_1 is bounded with $||\alpha L_1||_{op} = |\alpha|||L_1||_{op}$. We have simultaneously checked that $\mathcal{B}(X,Y)$ is closed under addition and scalar multiplication, so it is a linear space, and that $||\cdot||_{op}$ satisfies two important properties of a norm. The other properties are easily checked. Thus $\mathcal{B}(X,Y)$ is a normed linear space.

Let $(L_n) \subset \mathcal{B}(X,Y)$ be a Cauchy sequence. Let $x \in X$. Then

$$||L_n(x) - L_m(x)||_Y = ||(L_n - L_m)x||_Y \leq ||L_n - L_m||_{op}||x||_X.$$

Given $\epsilon > 0$ we can choose $N > 0$ such that $||L_n - L_m||_{op} \leq \frac{\epsilon}{||x||}$ for $n, m > N$, and so $||L_n(x) - L_m(x)||_Y \leq \epsilon$ for $n, m > N$. Hence $(L_n(x))$ is Cauchy in Y. We assumed that Y is complete, so this sequence must converge to some element of Y. Let $L(x) := \lim(L_n(x))$.

Now we must show that $L \in \mathcal{B}(X,Y)$ and that $||L_n - L||_{op} \to 0$. Since (L_n) is Cauchy we know that it is a bounded sequence in $\mathcal{B}(X,Y)$, so assume $||L_n||_{op} \leq K$ for all n. If $||x|| \leq 1$, then $||L_n(x)||_Y \leq K$ for all n, so $||L(x)||_Y = \lim ||L_n(x)|| \leq K$. Hence L is bounded with $||L||_{op} \leq K$. If $||x||_X \leq 1$ we also have $||(L-L_n)(x)||_Y \leq ||(L-L_m)(x)||_Y + ||(L_m-L_n)(x)||_Y$. Let $\epsilon > 0$ be given and choose N such that $||L_n - L_m||_{op} \leq \epsilon$ for all $n, m > N$. Then $||(L-L_n)(x)||_{op} \leq ||(L-L_m)(x)||_Y + \epsilon$ for all $n, m > N$. Let $m \to \infty$ and use the fact that $L_m(x) \to L(x)$ in Y to get $||(L-L_n)(x)|| \leq \epsilon$ for all $n > N$. This statement is true for all $x \in X$ such that $||x||_X \leq 1$, so we can take a supremum of the left hand side of the inequality to get $||L - L_n||_{op} \leq \epsilon$ for all $n > N$. Hence $L_n \to L$ in $\mathcal{B}(X,Y)$. The proof is done. $\square$

LEMMA 1.3. *If $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator, then it is a bounded linear operator.*

Proof: Consider the standard basis $\{e_1, ..., e_n\}$. Then

$$||L(x)|| = ||L(x_1e_1 + ... + x_ne_n)|| \leq \sum_{i=1}^n |x_i|||L(e_i)|| \leq \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n ||L(e_i)||^2\right)^{\frac{1}{2}} = c||x||,$$

where $c = (\sum_{i=1}^n ||L(e_i)||^2)^{\frac{1}{2}}$.

LEMMA 1.4. *A linear operator $L : X \to Y$ is continuous on X if and only if it is bounded.*

PROOF. $\Rightarrow$) Suppose L is continuous on X. Then L is continuous at 0, so $\forall \epsilon > 0$, $\exists \delta > 0$ so that if $x \in X$ with $\|x\| < \delta$, then $\|Lx\| < \epsilon$. Let $\epsilon = 1$ and choose a corresponding δ. Let

$x \in X$ with $\|x\| = 1$. By linearity, $L(\frac{\delta}{2}x) = \frac{\delta}{2}L(x)$. Also, $\|\frac{\delta}{2}x\| < \delta$, so $\|L(\frac{\delta}{2}x)\| < 1$. So $\|L(x)\| < \frac{2}{\delta}$. So, $\sup\{\|Lx\| : \|x\| = 1\} \leq \frac{2}{\delta}$ and L is bounded by Lemma 1.2.

$\Leftarrow$) Suppose L is bounded. Let $\|L\| = M$. Let $\epsilon > 0$ and $x_0 \in X$. Let $\delta = \frac{\epsilon}{M}$ and let $x \in X$ satisfy $\|x - x_0\| < \delta$. Then, using linearity and Lemma 1.2, we obtain

$$\|Lx - Lx_0\| = \|L(x - x_0)\| \leq \|L\|\|x - x_0\| = M\|x - x_0\| < M\delta = M\frac{\epsilon}{M} = \epsilon.$$

Therefore, L is continuous at every x_0 in X. □

1.2.1. *Exercises.*

EXERCISE 1.4. *Prove Lemma 1.2.*

EXERCISE 1.5. *Fill in the missing details in the proof of Theorem 1.1.*

EXERCISE 1.6. *Use Lemma 1.4 to justify the claims in Examples 1.2 and 1.3.*

EXERCISE 1.7. *Prove that a linear operator is continuous at every point in its domain if and only if it is continuous at* 0*.*