Real Analysis II

John Loftin*

May 14, 2007

1 Spaces of functions

1.1 Banach spaces

Many natural spaces of functions form infinite-dimensional vector spaces. Examples are the space of polynomials and the space of smooth functions. If we are interested in solving differential equations, then, it is important to understand analysis in infinite-dimensional vector spaces (over $\mathbb{R}$ or $\mathbb{C}$).

First of all, we should recognize the following straightforward fact about finite-dimensional vector spaces:

Homework Problem 1. *Let $x = (x^1, \dots, x^m)$ denote a point in $\mathbb{R}^m$, and let $\{x_n\} = \{(x_n^1, \dots, x_n^m)\}$ be a sequence of points in $\mathbb{R}^m$. Then $x_n \to x$ if and only if $x_n^i \to x^i$ for all $i = 1, \dots, m$.*

(Recall the standard metric on $\mathbb{R}^m$ is given by $|x - y|$, where the norm $|\cdot|$ is given by $|x| = \sqrt{(x^1)^2 + \cdots + (x^m)^2}$.)

Thus for taking limits in $\mathbb{R}^m$, we could even dispense with the notion of the taking limits using the metric on $\mathbb{R}^m$, and simply define the $x_n \to x$ by $x_n^i \to x^i$ for each $i = 1, \dots, m$. This reflects the fact that there is only one natural topology on a finite-dimensional vector space: that given by the standard norm.

For infinite-dimensional vector spaces, say with a countable basis, so that $x = (x^1, x^2, \dots)$, it is possible to define a topology by $x_n \to x$ if and only if each $x_n^i \to x^i$. It turns out that this is not usually the most useful way

*partially supported by NSF Grant DMS0405873

to define limits in infinite-dimensional spaces, however (though a related construction is used in defining the topology of Fréchet spaces).

Finite-dimensional vector spaces are also all complete with respect to their standard norm (in other words, they are all Banach spaces). Given a norm on an infinite dimensional vector space, completeness must be proved, however. There are many examples of Banach function spaces: On a measure space, the L^p spaces of functions are all Banach spaces for $1 \leq p \leq \infty$. Also, on a metric space X, the space of all bounded continuous functions $C^0(X)$ is a measure space under the norm

$$\|f\|_{C^0(X)} = \sup_{x \in X} |f(x)|.$$

The L^p and C^0 form the basis of most other useful Banach spaces, with extensions typically provided by measuring not just the functions themselves, but also their partial derivatives (as in Sobolev and C^k spaces) or their difference quotients (Hölder spaces).

Completeness of a metric space of course means that any Cauchy sequence has a unique limit. More roughly, this means that any sequence that should converge, in that its elements are becoming infinitesimally close to each other, will converge to a limit in the space. As we will see, taking such limits is a powerful way to construct solutions to analytic problems. Unfortunately, many of the most familiar spaces of functions (such as smooth functions) do not have the structure of a Banach space, and so it is difficult to ensure that a given limit of smooth functions is smooth. In fact we have the following theorem, which we state without proof:

Theorem 1. *On $\mathbb{R}^n$ equipped with Lebesgue measure, the space $C_0^\infty(\mathbb{R}^n)$ of smooth functions with compact support is dense in $L^p(\mathbb{R}^n)$ for all $1 \leq p < \infty$. In other words, completion of the space of smooth functions with compact support on $\mathbb{R}^n$ with respect to the L^p norm, is simply the space of all L^p functions for $1 \leq p < \infty$.*

If we are working in L^2, for example, it is possible for the limit of smooth functions to be quite non-smooth: there are many L^2 functions which are discontinuous everywhere. This poses a potential problem if the limit we have produced is supposed to be a solution to a differential equation. In particular, such a limit may be nowhere differentiable. Some of our goals then are to understand (1) how to make sense of taking derivatives of functions which are not classically differentiable (the theory of *distributions* and *weak* *derivatives*), and (2) how to show that a limit function actually has enough derivatives to solve the equation (*bootstrapping*).

Theorem 1 reminds us that the L^p Banach spaces have a very large overlap, which of course includes many more functions than the smooth functions with compact support. In particular, it is often useful to take the point of view that these Banach function spaces are not so much different spaces but different tools to study either the space of all functions or (via the completion process) the space of only very nice functions (e.g., smooth functions of compact support).

In particular, two function spaces which are very closely related to each other are L^∞ and C^0. As we will see below, they have essentially the same norm. First of all, we show that $C^0(X)$ is a Banach space for any metric space X.

1.2 The Banach space C^0

Given a metric space X, define

$$C^0(X) = \{f : X \to \mathbb{R} : f \text{ is continuous and } \sup_X |f| < \infty\}.$$

Define the norm

$$\|f\|_{C^0(X)} = \sup_X |f|.$$

It is straightforward to verify that $\|\cdot\|_{C^0}$ satisfies the requirements for a norm:

- $\|f\|_{C^0} = 0 \iff f \equiv 0$,
- $\|\lambda f\|_{C^0} = |\lambda| \|f\|_{C^0}$,
- $\|f + g\|_{C^0} \leq \|f\|_{C^0} + \|g\|_{C^0}$.

Remark. If $f_i \to f$ in $C^0(X)$, then we say $f_i \to f$ *uniformly* on X, and $C^0(X)$ convergence is the same as *uniform convergence*.

The main thing to check is that the norm gives $C^0(X)$ the structure of a complete metric space:

Proposition 1. *For any metric space X, $C^0(X)$ is a Banach space with norm $\|\cdot\|_{C^0}$.*

Proof. We simply need to check the metric induced on $C^0(X)$ is complete.

Let d denote the metric on X, and consider a Cauchy sequence $\{f_i\} \subset C^0(X)$. In other words, for all $\epsilon > 0$, there is an N so that $n, m > N$ implies $\|f_n - f_m\|_{C^0} < \epsilon$. By the definition of the norm, this is equivalent to $|f_n(x) - f_m(x)| < \epsilon$ for all $x \in X$. Now for each $x \in X$, $\{f_i(x)\} \subset \mathbb{R}$ is a Cauchy sequence, and since $\mathbb{R}$ is complete, there is a limit $f_\infty(x) = \lim_i f_i(x)$.

Now we have produced a limit function f_∞. Now we need to show that $\|f_i - f_\infty\|_{C^0} \to 0$ and $f_\infty \in C^0(X)$. The first statement is straightforward: For all $\epsilon > 0$, there is an N so that for all $n, m > N$, for all $x \in X$,

$$|f_n(x) - f_m(x)| < \epsilon.$$

Now let $m \to \infty$ to see that

$$|f_n(x) - f_\infty(x)| \le \epsilon.$$

So we have that for all $\epsilon > 0$, there is an N so that for all $n > N$, and for all $x \in X$,

$$|f_n(x) - f_\infty(x)| \le \epsilon.$$

Since this is true for all $x \in X$, we have

$$\|f_n - f_\infty\|_{C^0} = \sup_{x \in X} |f_n(x) - f_\infty(x)| \le \epsilon,$$

and so $\|f_i - f_\infty\|_{C^0} \to 0$.

We still need to prove that the limit function f_∞ is continuous. So let $x \in X$ and choose $\epsilon > 0$. Then there is an N so that for $n > N$, $\|f_n - f_\infty\|_{C^0} < \epsilon$. By the previous paragraph and the definition of $\|\cdot\|_{C^0}$,

$$|f_n(x) - f_\infty(x)| < \epsilon \quad \text{and} \quad |f_n(y) - f_\infty(y)| < \epsilon \text{ for all } y \in X.$$

Choose a particular $n > N$ and since f_n is continuous at x, there is a $\delta > 0$ so that $|f_n(x) - f_n(y)| < \epsilon$ for y so that $d(x, y) < \delta$. Then for such y in a δ-ball around x,

$$\begin{aligned} |f_\infty(x) - f_\infty(y)| &= |\,[f_\infty(x) - f_n(x)] + [f_n(x) - f_n(y)] + [f_n(y) - f_\infty(y)]\,| \\ &\le |f_\infty(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f_\infty(y)| \\ &< \epsilon + \epsilon + \epsilon = 3\epsilon. \end{aligned}$$

So we have proved that for all $\epsilon > 0$, $x \in X$, there is a $\delta > 0$ so that $d(x, y) < \delta \Rightarrow |f_\infty(x) - f_\infty(y)| < 3\epsilon$. This proves f_∞ is continuous. $\square$

The last bit of the proof can be remembered as this: *Any uniform limit of continuous functions is continuous.*

Remark. The previous proposition works as well for functions whose range is the complex numbers $\mathbb{C}$, or a vector space $\mathbb{R}^n$, or in fact any Banach space B. The proof is the same. In this last case, we could refer to the Banach space $C^0(X; B)$ as the Banach space of continuous functions from X into B.

Consider an open set $\Omega \subset \mathbb{R}^n$. On Ω, the C^0 norm is essentially the same as the L^∞ norm, but is simpler to define because we can consider functions as elements of C^0, while we need equivalence classes of functions to define L^∞. In fact, more is true. Let Ω inherit the standard metric and Lebesgue measure from $\mathbb{R}^n$. For a measurable function $f : \Omega \to \mathbb{R}$, let $[f]$ be the equivalence class whose members are all functions from $\Omega \to \mathbb{R}$ which agree with f almost everywhere.

Proposition 2. *The map $\Phi\colon C^0(\Omega) \to L^\infty(\Omega)$ given by $\Phi(f) = [f]$ is one-to-one and preserves the norm.*

Proof. First of all, note that it follows immediately from the definitions that for $f \in C^0(\Omega)$, $\Phi(f) \in L^\infty(\Omega)$. Also, we should show that $\|f\|_{C^0} = \|\Phi(f)\|_{L^\infty}$ to show Φ preserves the norm.

The proof hinges on the simple fact that every full-measure subset V of Ω is dense in Ω. (Recall $V \subset \Omega$ has *full measure* if $\Omega \setminus V$ has Lebesgue measure zero.) This fact may be proved as follows: let $V \subset \Omega$ have full measure. Then there is no open ball contained in $\Omega \setminus V$ (since open balls have positive measure). This shows V is dense in Ω. (Question: We need to use Ω is an open subset of $\mathbb{R}^n$ in this paragraph. Where did we use that Ω is open?)

Now we prove the map Φ is injective. So if f and g are in $C^0(\Omega)$, and $[f] = [g]$, then by definition, $f \equiv g$ on a set V of full measure. Let $x \in \Omega$. Since V is dense, there is a sequence $x_n \to x$, $x_n \in V$. Then

$$f(x) = f(\lim_n x_n) = \lim_n f(x_n) = \lim_n g(x_n) = g(\lim_n x_n) = g(x)$$

since f and g are continuous and $f(x_n) = g(x_n)$. So f and g coincide at each point of Ω and so $f = g$ in $C^0(\Omega)$.

Finally, we show that for $f \in C^0(\Omega)$, $\|f\|_{C^0} = \|f\|_{L^\infty}$. In particular, let μ denote Lebesgue measure and compute (recall we often write $\|f\|_{L^\infty}$ instead of the more correct $\|[f]\|_{L^\infty} = \|\Phi(f)\|_{L^\infty}$)

$$\begin{aligned} \|f\|_{L^\infty(\Omega)} &= \inf\{a : |f(x)| \le a \text{ for almost every } x \in \Omega\} \\ &= \inf\{a : \mu\{x : |f(x)| > a\} = 0\}. \end{aligned}$$

But $\mu\{x : |f(x)| > a\} = 0$ implies that $\{x : |f(x)| > a\} = \emptyset$ (Proof: If the set is not empty it is an open subset of Ω since $|f|$ is continuous. The only open subset of Ω with measure zero is the empty set.) So now

$$\begin{aligned}\|f\|_{L^\infty(\Omega)} &= \inf\{a : \mu\{x : |f(x)| > a\} = 0\} \\ &= \inf\{a : \{|f(x)| > a\} = \emptyset\} \\ &= \inf\{a : |f(x)| \le a \text{ for all } x \in \Omega\} \\ &= \sup_{x\in\Omega} |f(x)| = \|f\|_{C^0(\Omega)}.\end{aligned}$$

□

Remark. The previous Proposition is true for any measurable subset Ω of $\mathbb{R}^n$ with the following property: every nonempty open subset of Ω has positive measure.

Remark. The map Φ from $C^0(\Omega)$ to $L^\infty(\Omega)$ is far from being onto. A typical discontinuous function g cannot be changed on a set of measure zero to be continuous. The following homework problem is to show this is the case with the Heaviside function.

Homework Problem 2. *Let $g(x)$ be the Heaviside function on $\mathbb{R}$. In other words, let $g(x) = 0$ if $x < 0$ and $g(x) = 1$ if $x \ge 0$.*

(a) Show there is no function in $C^0(\mathbb{R})$ which is equal to g almost everywhere.

(b) Show that there is no sequence of functions $f_n \in C^0(\mathbb{R})$ which satisfy $f_n \to g$ in $L^\infty(\mathbb{R})$.

Hint for (b)*: Show that if $f_n \to g$ in $L^\infty(\mathbb{R})$, then $\{f_n\}$ is a Cauchy sequence in $C^0(\mathbb{R})$. Then use Proposition 1 and show the resulting limit function $f_\infty \in C^0(\mathbb{R})$ must be equal to g almost everywhere. (This amounts to showing that $\Phi(C^0)$ is a closed subspace of L^∞.) Provide a contradiction.*

1.3 Quantifiers

It is worth taking the time to look in some detail at C^0 convergence, and to compare it to pointwise convergence. By contrast, C^0 convergence is often call *uniform convergence.*

For a metric space X, $f_n \to f$ in $C^0(X)$, if for all $\epsilon > 0$, there is an N so that

$$n > N \quad \Longrightarrow \quad \|f_n - f\|_{C^0(X)} < \epsilon.$$

In other words, for all $\epsilon > 0$, there is an N so that

$$n > N \quad \Longrightarrow \quad \sup_{x \in X} |f_n(x) - f(x)| < \epsilon.$$

So then $f_n \to f$ in $C^0(X)$ implies that for all $\epsilon > 0$, there is an N so that for $x \in X$,

$$n > N \quad \Longrightarrow \quad |f_n(x) - f(x)| < \epsilon.$$

A few easy manipulations imply in fact the following

Lemma 3. *Let X be a metric space and let $f_n \in C^0(X)$. Then $f_n \to f$ in $C^0(X)$ if and only if for every $\epsilon > 0$, there is an $N = N(\epsilon)$ so that for $x \in X$,*

$$n > N \quad \Longrightarrow \quad |f_n(x) - f(x)| < \epsilon.$$

Homework Problem 3. *Prove Lemma 3.*

Since $C^0(X)$ is a Banach space, we know that the limit function $f \in C^0(X)$ as well, and thus the uniform limit of continuous functions is continuous. C^0 convergence is called uniform convergence because the N in Lemma 3 depends only on $\epsilon > 0$ and not on $x \in X$: thus N is *uniform* over all $x \in X$.

We contrast this with pointwise convergence. If f_n are functions on X, then $f_n \to f$ pointwise if for all $\epsilon > 0$ and $x \in X$, there is an $N = N(\epsilon, x)$ so that

$$n > N \quad \Longrightarrow \quad |f_n(x) - f(x)| < \epsilon.$$

The difference between pointwise and uniform convergence is subtle but very important: in pointwise convergence $N = N(\epsilon, x)$ may depend on ϵ and x, while in uniform convergence $N = N(\epsilon)$ only depends on ϵ and is independent of x.

We have belabored this point because it is one of the major issues in analysis: keeping track of which constants, or *quantifiers*, depend on which other quantifiers. (It is even better to have explicit bounds (*estimates*) on the behavior of quantifiers with respect to each other.) Of course it is desirable (though not always possible) to have more uniform dependence of quantifiers, as we see in the following standard example:

We have seen that the uniform limit of continuous functions is continuous. On the other hand, a pointwise limit of continuous functions may be not be:

Example 1. *Consider* $X = [0,1]$ *and* $f_n(x) = x^n$*. Then* $f_n \to f$ *pointwise on* $[0,1]$*, where*

$$f(x) = \begin{cases} 0 & \text{for } x \in [0,1), \\ 1 & \text{for } x = 1. \end{cases}$$

So the pointwise limit f is discontinuous, and thus we see that $f_n \not\to f$ uniformly.

1.4 Derivatives

The theory of derivatives in one variable is fairly straightforward: if a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at p (i.e., $f'(p)$ exists), then f must be continuous at p. For functions of more than one variable, however, consider the following example:

Example 2.

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & \text{for } (x,y) \neq (0,0) \\ 0 & \text{for } (x,y) = (0,0), \end{cases}$$

has first partial derivatives everywhere but is not even continuous at $(0,0)$*.*

Even though f has all its first partial derivatives at $(0,0)$, we do not consider f to be *differentiable* at $(0,0)$. For functions of more than one variable, we introduce the following definition of differentiability, which is stronger than just the existence of all the partial derivatives. Instead of $\mathbb{R}$-valued functions, we consider the slightly more general case of maps from $\mathbb{R}^n$ to $\mathbb{R}^m$. A basic reference is Spivak, *Calculus on Manifolds*, Chapter 2.

Let $\mathcal{O} \subset \mathbb{R}^n$ be a domain, and let $f = (f^1, \dots, f^m) : \mathcal{O} \to \mathbb{R}^m$. Then f is *differentiable* at a point $a \in \mathcal{O}$ if there is a linear map $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ which satisfies

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - Df(a)(h)|}{|h|} = 0,$$

where $h \in \mathbb{R}^n$. $Df(a)$ is called the *derivative*, or *total derivative*, of f at a.

Lemma 4. *In terms of standard bases of* $\mathbb{R}^n$ *and* $\mathbb{R}^m$*,* $Df(a)$ *is written as the Jacobian matrix*

$$Df(a) = \left(\frac{\partial f^i}{\partial x^j}(a) \right), \qquad i = 1, \dots, m, \quad j = 1, \dots, n.$$

In particular, if f is differentiable at a, then all the partial derivatives $\partial f^i/\partial x^j$ exist at a.

Proof. Write $Df(a)$ as the matrix (λ^i_j). Also consider a path $h = (0, \ldots, k, \ldots, 0)$, where $k \to 0$ sits in the j^{th} slot. (In other words, $h^l = \delta^l_j k$, where δ^l_j is the *Kronecker delta*, which is 1 if $l = j$ and 0 otherwise.) We also use *Einstein's summation convention*. In m space, this summation convention requires that any repeated index which appears in both up and down positions—such as the l in the last two lines below—is assumed to be summed from 1 to m. Compute

$$\begin{aligned}
\frac{\partial f^i}{\partial x^j}(a) &= \lim_{k \to 0} \frac{f^i(a^1, \ldots, a^j + k, \ldots, a^n) - f^i(a)}{k} \\
&= \lim_{k \to 0} \frac{[f^i(a^1, \ldots, a^j + k, \ldots, a^n) - f^i(a) - \lambda^i_l h^l] + \lambda^i_l h^l}{k} \\
&= 0 + \lim_{k \to 0} \frac{\lambda^i_l \delta^l_j k}{k} \\
&= \lambda^i_l \delta^l_j = \lambda^i_j.
\end{aligned}$$

The key step, going from the second to the third line, follows from the assumption that f is differentiable at a. $\square$

Another important result with essentially the same proof concerns directional derivatives. For a vector $v = v^j \in \mathbb{R}^n$, The *directional derivative* at a in the direction v is the vector

$$\mathcal{D}_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}.$$

(Note we do not require $\|v\| = 1$ to define the directional derivative.) We have the following lemma:

Lemma 5. *If f is differentiable at a, then the directional derivative $\mathcal{D}_v f(a)$ exists and*

$$\mathcal{D}_v f(a) = v^j \frac{\partial f}{\partial x^j}.$$

As Example 2 above shows, the converse of Lemma 4 is not true without extra assumptions on the partial derivatives. The following proposition gives an easy criterion for a function to be differentiable:

Proposition 6. *If $f = (f^1, \ldots, f^m)$ has continuous first partial derivatives $\partial f^i/\partial x^j$ on a neighborhood of a, then f is differentiable at a.*

Proof. For a component function f^i, write

$$\begin{aligned} f^i(a+h) - f^i(a) &= f^i(a^1+h^1, a^2, \ldots, a^n) - f^i(a^1, a^2, \ldots, a^n) \\ &\quad + f^i(a^1+h^1, a^2+h^2, \ldots, a^n) - f^i(a^1+h^1, a^2, \ldots, a^n) \\ &\quad + \cdots + f^i(a^1+h^1, a^2+h^2, \ldots, a^n+h^n) - \\ &\qquad\quad f^i(a^1+h^1, a^2+h^2, \ldots, a^{n-1}+h^{n-1}, a^n) \end{aligned}$$

Now consider the first term in terms of the function $f^i(x^1, a^2, \ldots, a^n)$ of the first variable x^1 alone. The Mean Value Theorem shows that there is a b^1 between a^1 and $a^1 + h^1$ so that

$$f^i(a^1+h^1, a^2, \ldots, a^n) - f^i(a^1, a^2, \ldots, a^n) = h^1 \frac{\partial f^i}{\partial x^1}(b^1, a^2, \ldots, a^n).$$

Similarly, for all other terms the difference equals

$$h^j \frac{\partial f^i}{\partial x^j}(a^1+h^1, \ldots, a^{j-1}+h^{j-1}, b^j, a^{j+1}, \ldots, a^n)$$

for b^j between a^j and $a^j + h^j$. So if we set $c_j = (a^1 + h^1, \ldots, a^{j-1} + h^{j-1}, b^j, a^{j+1}, \ldots, a^n)$, then we have

$$f^i(a+h) - f^i(a) = \sum_{j=1}^{n} h^j \frac{\partial f^i}{\partial x^j}(c_j),$$

where each $c_j \to a$ as $h \to 0$. So compute

$$\begin{aligned} \lim_{h\to 0} \frac{\left| f^i(a+h) - f^i(a) - \sum_{j=1}^{n} \frac{\partial f^i}{\partial x^j}(a) h^j \right|}{|h|} &= \lim_{h\to 0} \frac{\left| \sum_{j=1}^{n} \left[\frac{\partial f^i}{\partial x^j}(c_j) - \frac{\partial f^i}{\partial x^j}(a) \right] h^j \right|}{|h|} \\ &\le \lim_{h\to 0} \frac{\sum_{j=1}^{n} \left| \frac{\partial f^i}{\partial x^j}(c_j) - \frac{\partial f^i}{\partial x^j}(a) \right| |h^j|}{|h|} \\ &\le \lim_{h\to 0} \sum_{j=1}^{n} \left| \frac{\partial f^i}{\partial x^j}(c_j) - \frac{\partial f^i}{\partial x^j}(a) \right| \\ &= 0 \end{aligned}$$

since each $\partial f^i/\partial x^j$ is assumed to be continuous at a.

So we have proved that each component function f^i is differentiable at a. To show f is differentiable, just note

$$\frac{\left|f(a+h)-f(a)-\dfrac{\partial f}{\partial x^j}(a)h^j\right|}{|h|} \leq \sum_{i=1}^{m} \frac{\left|f^i(a+h)-f^i(a)-\dfrac{\partial f^i}{\partial x^j}(a)h^j\right|}{|h|},$$

which goes to 0 as $h \to 0$. □

Recall a function is (locally) C^1 if its first partial derivatives are continuous. The previous Proposition 6 shows that such functions are differentiable, and Lemma 5 then shows that directional derivatives work as expected for C^1 functions.

Now, for functions f on Ω an open subset of $\mathbb{R}^m$, consider the norm

$$\|f\|_{C^1(\Omega)} = \|f\|_{C^0(\Omega)} + \sum_{i=1}^{m} \left\|\frac{\partial f}{\partial x^i}\right\|_{C^0(\Omega)}$$

and the space

$$C^1(\Omega) = \{f : \Omega \to \mathbb{R} : f,\, \partial_1 f, \ldots, \partial_m f \text{ are bounded and continuous}\}.$$

Similarly, we can consider $\mathbb{R}^p$-valued C^1 functions, the difference being that the functions f, $\partial_i f$ have bounded values in $\mathbb{R}^p$.

Proposition 7. *On any open set $\Omega \subset \mathbb{R}^m$, $C^1(\Omega, \mathbb{R}^p)$ is a Banach space.*

Proof. It is straightforward to check $\|\cdot\|_{C^1}$ is a norm.

Since $\|f\|_{C^1} \geq \|f\|_{C^0}$ and $\|f\|_{C^1} \geq \|\frac{\partial f}{\partial x^j}\|_{C^0}$, then for any Cauchy sequence $\{f_n\}$ in C^1, $\{f_n\}$ and $\{\frac{\partial f_n}{\partial x^j}\}$ are Cauchy sequences in C^0. Therefore, since C^0 is a Banach space, there are uniform limits

$$f_\infty = \lim_n f_n, \qquad g_i = \lim_n \frac{\partial f}{\partial x^i}, \quad i = 1, \ldots, m, \tag{1}$$

and $f_\infty, g_i \in C^0$. Since

$$\|f\|_{C^1} = \|f\|_{C^0} + \sum_{i=1}^{m} \left\|\frac{\partial f}{\partial x^i}\right\|_{C^0},$$

(1) shows it suffices to prove that

$$\frac{\partial f_\infty}{\partial x^i} = g_i, \quad i = 1, \dots, m.$$

As usual, we recognize that integrating has better properties than differentiating. For $x \in \Omega$, choose an $x_0 = x - (0, \dots, k, \dots, 0)$, where the $k > 0$ is in the i^{th} slot. Since Ω is open, we may choose k small enough so that the line segment from x_0 to x is contained in Ω. Compute

$$\begin{aligned} f_\infty(x) &= \lim_n f_n(x) \\ &= \lim_n \left[f_n(x_0) + \int_{y=x^i-k}^{y=x^i} \frac{\partial f_n}{\partial x^i}(x^1, \dots, x^{i-1}, y, x^{i+1}, \dots, x^m)\, dy \right] \\ &= f_\infty(x_0) + \int_{y=x^i-k}^{y=x^i} g_i(x^1, \dots, x^{i-1}, y, x^{i+1}, \dots, x^m)\, dy \qquad (2) \end{aligned}$$

The key step in the computation is the last one: $f_n(x_0) \to f_\infty(x_0)$ is easy, and the integral converges by the Dominated Convergence Theorem: Since $g_i \in C^0$, there is a constant C so that $|g_i| \le C$ on Ω. Moreover, since $\frac{\partial f_n}{\partial x^i} \to g_i$ in C^0, there is an N so that $|\frac{\partial f_n}{\partial x^i} - g_i| \le 1$ for all $n \ge N$. Thus $\frac{\partial f_n}{\partial x^i}$ are all bounded by the integrable function $C + 1$, and the Dominated Convergence Theorem applies.

Now we can differentiate (2) with respect to x^i and we see that $\frac{\partial f_\infty}{\partial x^i} = g_i$ at each $x \in \Omega$. This completes the proof. $\square$

The last part of the proof is of independent interest. We record it as

Proposition 8. *Let f_n be C^1 functions on a domain $\Omega \subset \mathbb{R}^m$. Then if $f_n \to f$ uniformly and $\partial f_n / \partial x^i \to g_i$ uniformly for $i = 1, \dots, m$, then $g_i = \partial f / \partial x^i$.*

Remark. We can also define $C^k(\Omega, \mathbb{R}^p)$ to be the space of all functions $f : \Omega \to \mathbb{R}^p$ so that f and all its partial derivatives up to order k are continuous and bounded. The norm is given by

$$\|f\|_{C^k} = \sum_{|\alpha| \le k} \|\partial_\alpha f\|_{C^0}, \qquad (3)$$

where $\alpha = (\alpha_1, \dots, \alpha_m)$, each $\alpha_i \ge 0$, $|\alpha| = \alpha_1 + \dots + \alpha_m$, and

$$\partial_\alpha f = \frac{\partial^{|\alpha|} f}{(\partial x^1)^{\alpha_1} \cdots (\partial x^m)^{\alpha_m}}$$

(if some $\alpha_i = 0$, then there is no differentiation with respect to x^i).

We can use the same proof as above to conclude that C^k is a Banach space. In particular, we can apply the theorem to $F = (f, f_{,1}, \dots, f_{,n})$ and then relate $\|F\|_{C^1}$ to $\|f\|_{C^2}$ to provide an inductive step.

C^∞ is not a Banach space, as the analog of (3) would involve an infinite sum.

We've used the following problem implicitly a few times above.

Homework Problem 4. *Show that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a point a, then it is continuous at a.*

Homework Problem 5. *Let f be a real-valued function defined on a domain in $\mathbb{R}^2$. Show that if the second mixed partials $f_{,12} = \frac{\partial^2 f}{\partial x^1 \partial x^2}$ and $f_{,21} = \frac{\partial^2 f}{\partial x^2 \partial x^1}$ are continuous in a neighborhood of a point y, then*

$$\frac{\partial^2 f}{\partial x^1 \partial x^2}(y) = \frac{\partial^2 f}{\partial x^2 \partial x^1}(y).$$

Hint: If the two are not equal, assume without loss of generality that the difference $f_{,12} - f_{,21} > 0$ at y. Then it must be positive on a rectangular neighborhood. Integrate this quantity over the rectangular neighborhood, and use Fubini's Theorem and the Fundamental Theorem of Calculus to arrive at a contradiction.

Finally, we introduce the Chain Rule. We need the following lemma first:

Lemma 9. *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then there is a constant $C = C(A)$ so that $|Ax| \le C|x|$ for all $x \in \mathbb{R}^n$.*

Homework Problem 6. *Prove Lemma 9. Hint: write down Ax in terms of the matrix entries of A.*

Proposition 10 (Chain Rule). *Let $g : \mathcal{O} \to \mathbb{R}^n$, $f : \mathcal{U} \to \mathcal{O}$, where $\mathcal{O} \subset \mathbb{R}^m$ and $\mathcal{U} \subset \mathbb{R}^l$ are domains. Assume f is differentiable at $a \in \mathcal{U}$, and g is differentiable at $f(a) \in \mathcal{O}$. Then there is a composition of linear maps*

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a).$$

In terms of partial derivatives, this is equivalent to

$$\frac{\partial g^p}{\partial x^i} = \frac{\partial g^p}{\partial y^j}\frac{\partial y^j}{\partial x^i},$$

where $\{x^i\}$ are coordinates on $\mathbb{R}^l$, $\{y^j\}$ are coordinates on $\mathbb{R}^m$, and we follow the usual rules of Leibniz notation and Einstein summation.

Proof. Let $A = Df(a)$, $B = Dg(f(a))$. Now consider the remainder terms in the definition of differentiable maps. For $h \in \mathbb{R}^l$, $k \in \mathbb{R}^m$,

$$\begin{aligned}
\phi(h) &= f(a+h) - f(a) - A(h), \\
\psi(k) &= g(f(a)+k) - g(f(a)) - B(k), \\
\rho(h) &= (g \circ f)(a+h) - (g \circ f)(a) - (B \circ A)(h).
\end{aligned}$$

Then since f and g are differentiable,

$$\lim_{h \to 0} \frac{|\phi(h)|}{|h|} = 0, \tag{4}$$

$$\lim_{k \to 0} \frac{|\psi(k)|}{|k|} = 0, \tag{5}$$

and we want to show that

$$\lim_{h \to 0} \frac{|\rho(h)|}{|h|} = 0.$$

So compute

$$\begin{aligned}
\rho(h) &= g(f(a+h)) - g(f(a)) - B(A(h)) \\
&= g(f(a+h)) - g(f(a)) - B(f(a+h) - f(a) - \phi(h)) \\
&= [g(f(a+h)) - g(f(a)) - B(f(a+h) - f(a))] + B(\phi(h)) \\
&= \psi(f(a+h) - f(a)) + B(\phi(h))
\end{aligned}$$

So then

$$\frac{|\rho(h)|}{|h|} \leq \frac{|\psi(f(a+h) - f(a))|}{|h|} + \frac{|B(\phi(h))|}{|h|}.$$

$|B(\phi(h))|/|h| \to 0$ as $h \to 0$ by (4) and Lemma 9. On the other hand (5) shows that for all $\epsilon > 0$ there is a δ so that

$$|k| < \delta \quad \Longrightarrow \quad |\psi(k)| \leq \epsilon |k|.$$

Therefore if $|f(a+h) - f(a)| < \delta$ (which can be achieved if $|h| < \gamma$ since f is continuous),

$$\begin{aligned}
\frac{|\psi(f(a+h) - f(a))|}{|h|} &\leq \frac{\epsilon |f(a+h) - f(a)|}{|h|} \\
&\leq \epsilon \left(\frac{|A(h)|}{|h|} + \frac{|\phi(h)|}{|h|} \right)
\end{aligned}$$

Now if we let $h \to 0$, using (4) and Lemma 9,

$$\limsup_{h \to 0} \frac{|\psi(f(a+h) - f(a))|}{|h|} \le C\epsilon.$$

Now we may let $\epsilon \to 0$ to show that $|\rho(h)|/|h| \to 0$ as $h \to 0$. $\square$

1.5 Contraction mappings

Another tool we need is a basic fact about complete metric spaces, the Contraction Mapping Theorem.

A *fixed point* of a map $f : X \to X$ is a point $x \in X$ so that $f(x) = x$. For a metric space X with metric d, a *contraction map* is a map $g : X \to X$ so that there is a constant $\lambda \in (0,1)$ for which

$$d(g(x), g(y)) \le \lambda\, d(x,y) \text{ for all } x, y \in X.$$

Remark. It is important that the constant $\lambda < 1$ is independent of the x and y in X. As we'll see below in a homework exercise, the following theorem is false if we let λ depend on x and y.

Theorem 2 (Contraction Mapping). *Any contraction mapping on a complete metric space has a unique fixed point.*

Proof. As above, denote our metric space by X with metric d, and let $\lambda \in (0,1)$ be the constant for the contraction map g: for all $x, y \in X$, $d(g(x), g(y)) \le \lambda\, d(x,y)$.

First we prove uniqueness. If x and y are fixed points of g (so $g(x) = x$, $g(y) = y$), then

$$d(x,y) = d(g(x), g(y)) \le \lambda\, d(x,y).$$

So $(1-\lambda)d(x,y) \le 0$. Since $\lambda < 1$ and $d(x,y) \ge 0$ (since X is a metric space), we must have $d(x,y) = 0$ and so $x = y$ (again since X is a metric space).

To prove existence of the fixed point, we consider any point $x_0 \in X$, and consider iterates defined inductively by $x_{n+1} = g(x_n)$ for all $n \ge 0$. We claim x_n is a Cauchy sequence and the limit x_∞ of x_n is the fixed point. For

$n \geq m \geq 0$, compute

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x_{n-1}) + \cdots + d(x_{m+1}, x_m) \\ &= d(g(x_{n-1}), g(x_{n-2})) + \cdots + d(g(x_m), g(x_{m-1})) \\ &\leq \lambda\, d(x_{n-1}, x_{n-2}) + \cdots + \lambda\, d(x_m, x_{m-1}) \\ &\leq \lambda^2 d(x_{n-2}, x_{n-3}) + \cdots + \lambda^2 d(x_{m-1}, x_{m-2}) \\ &\leq \lambda^{n-1} d(x_1, x_0) + \cdots + \lambda^m d(x_1, x_0) \\ &= d(x_1, x_0) \sum_{i=m}^{n-1} \lambda^i \leq d(x_1, x_0) \sum_{i=m}^{\infty} \lambda^i = d(x_1, x_0) \frac{\lambda^m}{1-\lambda} \end{aligned}$$

(Note that in this computation, we've used the exact sum of the geometric series, and it is crucial that $\lambda \in (0,1)$: the geometric series diverges for $\lambda \geq 1$.) So if N is a positive integer, then for all $n, m > N$, $d(x_n, x_m) \leq d(x_1, x_0)\lambda^N/(1-\lambda)$, and this last quantity $d(x_1, x_0)\lambda^N/(1-\lambda) \to 0$ as $N \to \infty$. Thus $\{x_n\}$ is a Cauchy sequence which has a limit $x_\infty \in X$ since X is a complete metric space.

Now we prove that x_∞ is a fixed point. Since $x_\infty = \lim_i x_i = \lim_i x_{i+1}$, we have

$$g(x_\infty) = g(\lim_i x_i) = \lim_i g(x_i) = \lim_i x_{i+1} = x_\infty,$$

and so x_∞ is a fixed point. One point to note is that we have interchanged g with lim, which is valid only if g is continuous (this is a homework problem below). □

Homework Problem 7. *Show any contraction map is continuous.*

Homework Problem 8. *Newton's method is an iterative method for finding zeros of differentiable functions. For an initial x_0, we proceed by the recursive definition*

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

Then the limit $\lim x_n$ should produce a zero of the function f.

A differentiable function $f : \mathbb{R} \to \mathbb{R}$ has a nondegenerate zero *at x if $f(x) = 0$ and $f'(x) \neq 0$.*

Assume $f : \mathbb{R} \to \mathbb{R}$ is a locally C^2 function (i.e., f'' is continuous on all of $\mathbb{R}$). Show that every nondegenerate zero x of f has a neighborhood $\mathcal{N}_x$ so that for any initial $x_0 \in \mathcal{N}_x$, Newton's method converges to x. Hints:

(a) *The main point is to exhibit the Newton's method iteration as a contraction map on a complete metric space (recall a closed subset of any complete metric space is complete). You must find an appropriately small neighborhood of x on whose closure Newton's method is a contraction map.*

(b) *You will need the following lemma: For a C^1 function $g\colon \mathbb{R} \to \mathbb{R}$,*

$$y \neq z \in [a,b] \quad \Longrightarrow \quad \frac{|g(y)-g(z)|}{|y-z|} \leq \max_{w\in[a,b]} |g'(w)|.$$

(c) *Show that any fixed point of Newton's method is a zero.*

(d) *Show the zero you have produced via Newton's method must be the original zero x.*

1.6 Differentiating under the Integral

Proposition 11. *Let $f = f(y,x)$ be a locally C^1 real-valued function for $y \in \mathbb{R}^n$, $x \in \mathcal{O}$ an open subset of $\mathbb{R}^m$. Then on a measurable $\Omega \subset\subset \mathcal{O} \subset \mathbb{R}^m$ equipped with Lebesgue measure dx,*

$$\frac{\partial}{\partial y^i}\int_\Omega f(y,x)\,dx = \int_\Omega \frac{\partial f}{\partial y^i}(y,x)\,dx.$$

$\int_\Omega f(y,x)\,dx$ is C^1 as a function of y.

Remark. $\Omega \subset\subset \mathcal{O}$ means that the closure $\bar{\Omega}$ in $\mathbb{R}^m$ is a compact subset of $\mathcal{O}$.

Proof. Compute. Let e_i be the standard i^{th} basis vector on $\mathbb{R}^n$.

$$\begin{aligned}\frac{\partial}{\partial y^i}\int_\Omega f(y,x)\,dx &= \lim_{k\to 0}\frac{1}{k}\left[\int_\Omega f(y+ke_i,x)\,dx - \int_\Omega f(y,x)\,dx\right]\\ &= \lim_{k\to 0}\int_\Omega \frac{f(y+ke_i,x)-f(y,x)}{k}\,dx\end{aligned}$$

Clearly as $k \to 0$, the integrand goes to $\frac{\partial f}{\partial y^i}(y,x)$ pointwise. We need to show that the integrands are bounded in absolute value by a fixed integrable function to use the Dominated Convergence Theorem. This follows from the Mean Value Theorem, which shows that the integrand is equal to

$$\frac{\partial f}{\partial y^i}(\tilde{y},x)$$

for $\tilde{y} = (y^1, \ldots, y^{i-1}, b^i, y^{i+1}, \ldots, y^n)$, b^i between y^i and $y^i + k$. Since f is C^1, $\partial f / \partial y^i$ is continuous, $\bar{\Omega}$ is compact, and $\tilde{y}$ stays in a compact neighborhood of y, then the absolute value of the integrand is bounded by a constant M. Since $\int_\Omega M\, dx < \infty$, the Dominated Convergence Theorem shows that

$$\begin{aligned}
\frac{\partial}{\partial y^i} \int_\Omega f(y, x)\, dx &= \lim_{k \to 0} \int_\Omega \frac{f(y + ke_i, x) - f(y, x)}{k}\, dx \\
&= \int_\Omega \lim_{k \to 0} \frac{f(y + ke_i, x) - f(y, x)}{k}\, dx \\
&= \int_\Omega \frac{\partial f}{\partial y^i}(y, x)\, dx
\end{aligned}$$

To show that $\int_\Omega f(y, x)$ is C^1 as a function of y, note that its partial derivatives

$$g_i(y) = \int_\Omega \frac{\partial f}{\partial y^i}(y, x)\, dx$$

are continuous in y by the Dominated Convergence Theorem again, since if $y \to y_0$, then

$$\begin{aligned}
\lim_{y \to y_0} g_i(y) &= \lim_{y \to y_0} \int_\Omega \frac{\partial f}{\partial y^i}(y, x)\, dx \\
&= \int_\Omega \lim_{y \to y_0} \frac{\partial f}{\partial y^i}(y, x)\, dx \\
&= \int_\Omega \frac{\partial f}{\partial y^i}(y_0, x)\, dx \\
&= g_i(y_0)
\end{aligned}$$

because $\partial f / \partial y^i$ is continuous in y. $\square$

Remark. The last argument also shows that if $f = f(z, x)$ is a continuous function of z and x, and $x \in \Omega$ a compact subset of $\mathbb{R}^n$, then the function

$$z \mapsto \int_\Omega f(z, x)\, dx$$

is continuous.

1.7 The Inverse Function Theorem

We need the following lemma first:

Lemma 12. *If f is a C^1 function from a ball B in $\mathbb{R}^n$ to $\mathbb{R}^m$, which satisfies*

$$\left|\frac{\partial f^i}{\partial x^j}\right| \le C$$

on B, then for $y, z \in B$,

$$|f(y) - f(z)| \le Cmn|y - z|.$$

Proof. If $y, z \in B$, then the line segment $\{ty + (1-t)z : 0 \le t \le 1\}$ between them is also contained in B (see Homework Problem 13 below). Then use the Chain Rule to compute for $i = 1, \dots, m$,

$$\begin{aligned}
|f^i(y) - f^i(z)| &= \left| \int_0^1 \frac{\partial}{\partial t} f^i(ty + (1-t)z)\, dt \right| \\
&= \left| \int_0^1 (y^j - z^j) \frac{\partial f^i}{\partial x^j}(ty + (1-t)z)\, dt \right| \\
&\le Cn|y - z|.
\end{aligned}$$

(Note this argument is essentially the same as the use of the Mean Value Theorem.) Now apply

$$|f(y) - f(z)| \le \sum_{i=1}^m |f^i(y) - f^i(z)|.$$

□

Theorem 3 (Inverse Function Theorem). *Let $f : \mathcal{O} \to \mathcal{U}$ be a C^1 map between domains in $\mathbb{R}^m$. Assume that for $a \in \mathcal{O}$, $Df(a)$ is an invertible matrix (i.e., $\det Df(a) \ne 0$). Then there are neighborhoods $\mathcal{O}' \ni a$ and $\mathcal{U}' \ni f(a)$ so that $f : \mathcal{O}' \to \mathcal{U}'$ is a bijection and f^{-1} is also a C^1 map. For every $b \in \mathcal{O}'$, $D(f^{-1})(f(b)) = (Df(b))^{-1}$.*

Proof. First of all, we may reduce to the case that $a = f(a) = 0$ and $Df(a) = I$ the identity map from $\mathbb{R}^m$ to itself. (This can be achieved by replacing $f(x)$ by $(Df(a))^{-1}(f(x + a) - f(a))$. Then use the Chain Rule and the fact that the derivative of the linear map $(Df(a))^{-1}$ is $(Df(a))^{-1}$ itself.)

Now consider $g(x) = x - f(x)$ and note that $Dg(0) = 0$ the zero linear transformation. Since g is C^1, there is an $r > 0$ so that $|x| < 2r$ implies

$$\left|\frac{\partial g^i}{\partial x^j}(x)\right| < \frac{1}{2m^2}, \qquad \text{for } i, j = 1, \ldots, m. \tag{6}$$

Let $B(r) = \{x \in \mathbb{R}^m : |x| < r\}$. Then Lemma 12 and $g(0) = 0$ imply that $g(\overline{B(r)}) \subset \overline{B(r/2)}$.

Now let $y \in B(r/2)$ and consider

$$g_y(x) = g(x) + y = x - f(x) + y.$$

Then

- $g_y(x) = x$ is equivalent to $f(x) = y$, and so a fixed point of g_y is equivalent to a solution to $f(x) = y$.
- If $x \in \overline{B(r)}$, $|g_y(x)| \leq |g(x)| + |y| \leq r$, and so g_y is a map from the complete metric space $\overline{B(r)}$ to itself.
- Lemma 12 and (6) imply g_y is a contraction map (with $\lambda = 1/2$). In other words, for $x_1, x_2 \in \overline{B(r)}$,

$$|g_y(x_1) - g_y(x_2)| = |g(x_1) - g(x_2)| \leq \tfrac{1}{2}|x_1 - x_2| \tag{7}$$

Therefore, for each $y \in B(r/2)$, there is a unique fixed point x of g_y, which shows there is a unique solution x to $f(x) = y$ in $\overline{B(r)}$.

Now we show $x = f^{-1}(y)$ is continuous: for $x_1, x_2 \in \overline{B(r)}$, we have, by the definition $g = x - f$ and (7)

$$\begin{aligned}
|x_1 - x_2| &\leq |g(x_1) - g(x_2)| + |f(x_1) - f(x_2)| \\
&\leq \tfrac{1}{2}|x_1 - x_2| + |f(x_1) - f(x_2)|, \\
\tfrac{1}{2}|x_1 - x_2| &\leq |f(x_1) - f(x_2)|, \\
|f^{-1}(y_1) - f^{-1}(y_2)| &\leq 2|y_1 - y_2| \qquad (8)
\end{aligned}$$

for $y_i = f(x_i)$. Thus f^{-1} is continuous.

To show f^{-1} is differentiable at y_2 with total derivative $(Df(x_2))^{-1}$, we need to show that

$$\lim_{y_1 \to y_2} \frac{|f^{-1}(y_1) - f^{-1}(y_2) - (Df(x_2))^{-1}(y_1 - y_2)|}{|y_1 - y_2|} = 0.$$

To show this, compute

$$\begin{aligned}
&|f^{-1}(y_1) - f^{-1}(y_2) - (Df(x_2))^{-1}(y_1 - y_2)| \\
&\quad = |x_1 - x_2 - (Df(x_2))^{-1}(f(x_1) - f(x_2))| \\
&\quad = |(Df(x_2))^{-1}[Df(x_2)(x_1 - x_2) + (y_1 - y_2)]| \\
&\quad \leq C|Df(x_2)(x_1 - x_2) + (y_1 - y_2)| \quad \text{(by Lemma 9)} \\
&\quad = C|Df(x_2)(x_1 - x_2) + f(x_1) - f(x_2)| \qquad (9)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{|f^{-1}(y_1) - f^{-1}(y_2) - (Df(x_2))^{-1}(y_1 - y_2)|}{|y_1 - y_2|} \\
&\quad = \frac{|f^{-1}(y_1) - f^{-1}(y_2) - (Df(x_2))^{-1}(y_1 - y_2)|}{|x_1 - x_2|} \cdot \frac{|x_1 - x_2|}{|y_1 - y_2|}
\end{aligned}$$

This goes to zero as $y_1 \to y_2$ by (8) and (9), since f is differentiable at x_2.

Finally we show the total derivative $(Df(x))^{-1}$ is continuous in y. We can think of Df as a map from x to $\mathbb{R}^{m^2}$, which represents the space of $m \times m$ matrices. $Df(x)$ is continuous in x (f is C^1), and thus is continuous in y. The determinant function $\det\colon \mathbb{R}^{m^2} \to \mathbb{R}$ is continuous, since it is a polynomial in the matrix entries. So $\det Df(x)$ is bounded away from zero, by compactness of $\overline{B(r)}$. We are left to prove the continuity of the matrix inverse operation for square matrices with determinant bounded away from 0. This follows from the formula from the inverse in terms of cofactor matrices: Each entry of the inverse matrix $A^{-1} = (a_{ij})^{-1}$ is of the form

$$\frac{(m-1)^{\text{st}}\text{-order polynomial in the } a_{ij}}{\det(a_{ij})}.$$

$\square$

Homework Problem 9. *If, in the Inverse Function Theorem, f is a smooth (C^∞) map, then $f^{-1}\colon \mathcal{U}' \to \mathcal{O}'$, the C^1 local inverse of f, is also C^∞. Hints:*

(a) If $A = A(s)$ is a family of invertible $n \times n$ matrices which depend differentiably on a real parameter s, differentiate the equation $AA^{-1} = I$ to show

$$\frac{d(A^{-1})}{ds} = -A^{-1}\frac{dA}{ds}A^{-1}.$$

(b) *Use the formula for $D(f^{-1})$ to show that f^{-1} is C^∞.*

Hints: It may be helpful to use the following notation. If $f = f(x) = f(x^1, \dots, x^n)$, we may write $(y^1, \dots, y^n) = y = y(x) = f(x)$. And so $f^{-1}(y) = x$ may be written simply as $y = y(x)$. To show f^{-1} is C^2, for example, you should write

$$\frac{\partial^2 (f^{-1})^k}{\partial y^i \partial y^j} = \frac{\partial^2 x^k}{\partial y^i \partial y^j}$$

in terms of (the components of) the first and second derivatives

$$\frac{\partial f}{\partial x^i} = \frac{\partial y}{\partial x^i}, \quad \text{and} \quad \frac{\partial^2 f}{\partial x^i \partial x^j} = \frac{\partial^2 y}{\partial x^i \partial x^j}$$

and verify that the resulting expression is continuous.

Remember to use the Chain Rule, as in e.g.,

$$\frac{\partial}{\partial y^j} = \frac{\partial x^i}{\partial y^j}\frac{\partial}{\partial x^i},$$

and recall that $Df^{-1} = (Df)^{-1}$ can be written as

$$\left(\frac{\partial x^i}{\partial y^j}\right) = \left(\frac{\partial y^k}{\partial x^l}\right)^{-1}.$$

It will also be helpful to use Einstein's summation notation. In particular, the matrix notation used in part (a) is insufficient, as there may be quantities with more than 2 indices which need to be summed.

Theorem 4 (Implicit Function Theorem). *Suppose $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is C^1 in an open set containing (a, b), and assume $f(a, b) = 0$. Assume the $m \times m$ matrix*

$$\left(\frac{\partial f^i}{\partial x^{n+j}}(a, b)\right), \qquad 1 \le i, j \le m$$

is invertible. Then there is an open set $\mathcal{O} \subset \mathbb{R}^n$ containing a and an open set $\mathcal{U} \subset \mathbb{R}^m$ containing b so that for each $x \in \mathcal{O}$, there is a unique $g(x) \in \mathcal{U}$ so that $f(x, g(x)) = 0$. g is locally C^1.

Homework Problem 10. *Prove the Implicit Function Theorem. Hints:*

(a) *Consider* $F: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ *defined by* $F(x, y) = (x, f(x, y))$ *and apply the Inverse Function Theorem to* F.

(b) *Show that, on a suitably small neighborhood,* F^{-1} *is of the form* $F^{-1}(x, y) = (x, p(x, y))$ *for* $p: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$.

(c) *Show that* $g(x) = p(x, 0)$ *satisfies the conditions of the theorem.*

1.8 Lipschitz constants and functions

A closely related concept to the contraction map is the *Lipschitz constant.* A map $f: X \to Y$ has Lipschitz constant

$$L = \sup_{x,x' \in X : x \neq x'} \frac{d_Y(f(x), f(x'))}{d_X(x, x')}.$$

Here of course d_X and d_Y are the metrics on X and Y respectively. An equivalent definition is that L is the smallest constant so that

$$d_Y(f(x), f(x')) \leq L\, d_X(x, x') \quad \text{for all } x, x' \in X.$$

A function with finite Lipschitz constant is called *Lipschitz.* A basic fact is the following:

Lemma 13. *Any Lipschitz function is continuous.*

If $f: X \to X$, then the Lipschitz constant gives a criterion for a mapping to be a contraction mapping:

Lemma 14. $f: X \to X$ *is a contraction map if and only if the Lipschitz constant* L *of* f *is strictly less than 1.*

Idea of proof. The Lipschitz constant is the smallest value of λ for which f is a contraction map. □

If $f: \mathbb{R} \to \mathbb{R}$, then the Lipschitz constant is simply

$$L = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|},$$

which of course is suggestive of the definition of the derivative. In fact, the following is true:

Homework Problem 11. *The Lipschitz constant of a locally C^1 function $f : \mathbb{R} \to \mathbb{R}$ is equal to* $\sup_{x\in\mathbb{R}} |f'(x)|$.

Hint: To show the two quantities are equal, you need to relate the sup of the derivative to the sup of the difference quotients. To relate the derivative $f'(x)$ to difference quotients, use the definition of the derivative. To relate a given difference quotient to a derivative, use the Mean Value Theorem.

The previous problem shows that any differentiable function with bounded derivative is Lipschitz. The converse is false, as we see in the following example.

Example 3. *The function $x \mapsto |x|$ is a Lipschitz function from $\mathbb{R}$ to $\mathbb{R}$. This follows from the observation that for each $x \neq y \in \mathbb{R}$,*

$$\frac{\big||x| - |y|\big|}{|x-y|} \leq 1.$$

(This can be proved using the Triangle Inequality.)

Example 4. *For any constant $\alpha \in (0,1)$, the function from $\mathbb{R}$ to $\mathbb{R}$ $x \mapsto |x|^\alpha$ is not Lipschitz. In particular,*

$$\lim_{x\to 0} \frac{\big||x|^\alpha - |0|^\alpha\big|}{|x-0|} = \lim_{x\to 0} |x|^{\alpha-1} = \infty.$$

In terms of the graph of a function, a function whose graph has a corner (as does $x \mapsto |x|$) is Lipschitz, while a function whose graph has a cusp (as does $x \mapsto |x|^\alpha$) is not Lipschitz.

Another basic fact we establish is this: the conclusion of the Contraction Map Theorem may be false if the Lipschitz constant is equal to 1. An easy example is the map $x \mapsto x+1$ from $\mathbb{R} \to \mathbb{R}$. The Lipschitz constant is obviously 1, and there is no fixed point. A related, but somewhat more surprising fact, is outlined in the following problem:

Homework Problem 12. *Find an example of a differentiable function $f : \mathbb{R} \to \mathbb{R}$ so that for each $x \neq y$,*

$$\frac{|f(x) - f(y)|}{|x-y|} < 1,$$

and yet f has no fixed point. Prove your answer works.

Hint: The point of this problem is that there should be no uniform $L < 1$ which works for all x and y. To construct such a function f, use Problem 11 above. In particular, first construct the derivative f' and then integrate to find f. (You'll need $\sup_x |f'(x)| = 1$; why?) Use the Mean Value Theorem to relate values of f' to difference quotients.

A subset $\mathcal{C}$ of a real vector space is *convex* if every line segment connecting two points in $\mathcal{C}$ is contained in $\mathcal{C}$. More formally, $\mathcal{C}$ is convex if

$$x, y \in \mathcal{C}, \quad t \in [0, 1] \quad \Longrightarrow \quad tx + (1-t)y \in \mathcal{C}.$$

Proposition 15. *Any globally C^1 function from a convex domain $\Omega \subset \mathbb{R}^n$ to $\mathbb{R}^m$ is globally Lipschitz.*

Proof. Lemma 12 above shows that for any $x, y \in \Omega$,

$$|f(x) - f(y)| \le Cnm|x-y|, \quad \text{for } C = \sup \left\{ \left| \frac{\partial f^i}{\partial x^j}(z) \right| : z \in \Omega, i \le n, j \le m \right\}.$$

$C < \infty$ since f is C^1. Thus f is Lipschitz. □

Consider X a locally compact metric space and Y any metric space. Then we say a function $f \colon X \to Y$ is *locally Lipschitz* if f satisfies one of the two following equivalent definitions:

1. f is Lipschitz when restricted to any compact set of X. In other words, if $K \subset X$ is compact, then there is a constant L_K so that

$$x, x' \in K \quad \Longrightarrow \quad d_Y(f(x), f(x')) \le L_K d_X(x, x').$$

2. Each $x \in X$ has a neighborhood on which f is Lipschitz.

We prove these two definitions are equivalent below.

Corollary 16. *On any domain $\Omega \subset \mathbb{R}^n$, any locally C^1 function f is locally Lipschitz.*

Proof. Any ball is convex (see the following homework problem), and so if f is C^1 on a small ball, then it is Lipschitz on the ball by the previous Proposition 15. □

Homework Problem 13. *Show that any ball $B_x(r) = \{y \in \mathbb{R}^n : |y-x| < r\}$ is convex.*

Proposition 17. *Let X be a locally compact metric space and Y be any metric space, then for maps f from X to Y, the two definitions (1) and (2) above are equivalent.*

Proof. To prove (1) $\Longrightarrow$ (2), consider $x \in X$. Since X is locally compact, there is a neighborhood $\mathcal{O}$ of x with compact closure. By the definition of locally Lipschitz, f is Lipschitz when restricted to $\bar{\mathcal{O}}$, and is thus Lipschitz on $\mathcal{O}$ also.

To prove part (2) $\Longrightarrow$ (1), let $K \subset X$ be a compact subset. Given that all points in X have neighborhoods on which f is Lipschitz, we need to prove that f is Lipschitz on K. The set of all neighborhoods of points in K on which f is Lipschitz forms an open cover of K, and thus there is a finite subcover $\mathcal{O}_1, \dots, \mathcal{O}_n$. The set

$$P = K \times K \setminus \left(\bigcup_{i=1}^{n} \mathcal{O}_i \times \mathcal{O}_i \right)$$

is compact, and so the function

$$\frac{d_Y(f(x), f(x'))}{d_X(x, x')},$$

which is continuous on P, attains its maximum M on P.

Consider any $x \neq x' \in K$. Then either $(x, x') \in P$ or $x, x' \in \mathcal{O}_i$ for some $i = 1, \dots, n$. Let L_i be the Lipschitz constant of $f|_{\mathcal{O}_i}$. Choose $L = \max\{M, L_1, \dots, L_n\}$. Then for every $x \neq x' \in K$,

$$\frac{d_Y(f(x), f(x'))}{d_X(x, x')} \leq L$$

and f is Lipschitz on K. $\square$