Differential Topology

Notes for Mathematics 738, Spring 2016.

Ralph Howard
Department of Mathematics
University of South Carolina
Columbia, S.C. 29208, USA
howard@math.sc.edu

Contents

1. A Review of Some Linear Algebra.

Here we recall facts from linear algebra that will motivate some of what we do with nonlinear maps. I assume you know the definition of a vector space, a subspace of a vector space, the dimension of a vector space, and the span of a subset of a vector space. If U and W are subspaces of a vector space then the set

$$U + W = \{u + w : u \in U, w \in W\}$$

is a subspace and is the smallest subspace containing both U and W. The intersection is also a subspace.

Proposition 1.1. *Let V be a finite dimensional real vector space and U and W subspaces of V. Then*

$$\dim(U + W) + \dim(U \cap W) = \dim(U) + \dim(W).$$

Problem 1.1. Prove this. *Hint:* Let $p = \dim U$, $q = \dim W$, and $k = \dim U \cap W$. Let $v_1, \ldots, v_k$ be a basis for $U \cap W$. Expend this set to a basis $v_1, \ldots, v_k, u_{k+1}, \ldots, u_p$ of U and a basis $v_1, \ldots, v_k, w_{k+1}, \ldots, w_q$. Then show $v_1, \ldots, v_k, u_{k+1}, \ldots, u_p, w_{k+1}, \ldots, w_q$ is a basis for $U + W$. $\square$

If V is a vector space over $\mathbb{R}$, and $a, b \in V$ are distinct, then the ***line*** through a and b is the set

$$\{(1-t)a + tb : t \in \mathbb{R}\}.$$

An ***affine subspace*** of V is a subset A, that contains the line through any two of its points. More explicitly A is an affine subspace of V if and only if for all $a_0, a_1 \in A$ and $t \in \mathbb{R}$ we have $(1-t)a_0 + ta_1 \in A$. As we are not assuming here that a_0 and a_1 are distinct this implies that a one element subset of V is is an affine subspace.

If $a_0, a_1, \ldots, a_k$ are in the real vector space V than an ***affine combination*** of these vectors is a sum of the form

$$\sum_{j=0}^{k} t_j a_j \qquad \text{where} \qquad \sum_{j=0}^{k} t_j = 1.$$

Note that an affine combination of two points a_0 and a_1 is of the form $t_0 a_0 + t_1 a_1$ where $t_0 + t_1 = 1$. Letting $t = t_1$ we have $t_0 a_0 + t_1 a_1 = (1-t)a_0 + ta_1$. Thus the set of all affine combinations of two distinct points is just the line through the points.

Proposition 1.2. *A subset of the real vector space V is an affine subspace of V if and only if it is closed under taking affine combinations of its elements.*

Problem 1.2. Prove this. *Hint:* One direction is more or less clear. If a subset is closed under affine combinations, then it is an affine subspace. Conversely if A is an affine subspace of V, then it is closed under taking the affine combination of any two of its elements. Use this as the base case for an induction showing that A is closed under general affine combinations. $\square$

Proposition 1.3. *Let V be a real vector space.*

(a) A linear subspace of V is also an affine subspace of V.
(b) An affine subspace of V that contains the origin is a linear subspace of V.

Problem 1.3. Prove this. *Hint:* For part (b) note that if A is an affine subspace of V with $0 \in A$, then for any $a_1, a_2 \in A$ and $\alpha, \beta \in \mathbb{R}$ we have $\alpha a_1 + \beta a_0 = (1 - \alpha - \beta)0 + \alpha a_1 + \beta a_0$ and thus $\alpha a_1 + \beta a_0$ is an affine combination of 0, a_1, and a_2. $\square$

Proposition 1.4. *Let U be a linear subspace of V and $a \in V$ then the translate*

$$a + U := \{a + u : u \in U\}$$

is an affine subspace of V.

Problem 1.4. Prove this. $\square$

Proposition 1.5. *Let A be an affine subspace of the real vector space V. Then there is a unique linear subspace U of V such that for any vector $a_0 \in A$*

$$A = a_0 + U.$$

(That is the affine subspaces of V are just the translations of the linear subspaces.)

Problem 1.5. Prove this. *Hint:* To start let $a_0 \in A$, set $U = \{a - a_0 : a \in A\}$ and show U is a linear subspace of V. One way to do this is to show U is an affine subspace that contains the origin. $\square$

If A is an affine subspace of the real vector space V and $A = a_0 + U$ where U is a linear subspace of V then define the ***dimension*** of A to be

$$\dim(A) = \dim(U).$$

The following is a very special case of one of the main theorems we will prove this term.

Proposition 1.6. *Let A and B be affine subspaces of the real vector space V and assume $\dim(A) + \dim(B) < \dim(V)$. Then for every $\varepsilon > 0$ there is a vector $v \in V$ with $\|v\| < \varepsilon$ and $A \cap (v + B) = \varnothing$.*

Problem 1.6. Prove this. *Hint:* If $A \cap B = \varnothing$ when we can use $v = 0$. If $A \cap B \neq \varnothing$, let $a \in A \cap B$. Then there are unique linear subspaces U and W such that $A = a + U$ and $B = a + W$. As $\dim(U) + \dim(W) < \dim(V)$ the subspace $U + W$ is not all of V. Let $x \in V$ be a vector that is not in $U + W$ and let $0 \neq t \in \mathbb{R}$. Show that $A \cap (tx + B) = \varnothing$ for all $t \neq 0$. $\square$

This shows that in some cases it is easy to move one affine subspace away from another. In some cases this is not true.

Proposition 1.7. *Let U and W be linear subspaces of the real vector space V with $U + W = V$. Then for all $a, b \in V$ we have $(a + U) \cap (b + W) \neq \varnothing$.*

Problem 1.7. Prove this. $\square$

The following is a basic factor about linear maps.

Proposition 1.8 (Rank plus Nullity Theorem.)**.** *Let $L\colon V \to W$ be a linear map between finite dimensional vector spaces. Then*

$$\dim(\ker(L)) + \dim(\operatorname{Image}(L)) = \dim(V).$$

Problem 1.8. Prove this. *Hint:* Let $k = \dim(\ker(L))$ and choose a basis $v_1, \ldots, v_k$ of $\ker(K)$. Extend this to a basis $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$ of V (where $n = \dim(V)$). Then show $Lv_{k+1}, \ldots, Lv_n$ is a basis of V. $\square$

One of the main ideas in differential topology is to relate the geometry of a map $f\colon M \to N$ between nice spaces so the geometry of preimages $f^{-1}[y] := \{x \in M : f(x) = y\}$ for $y \in N$. One of the main geometric invariants of a space is its dimension. The following is anther special case of a much more general result we will prove later.

Proposition 1.9. *Let $L\colon V \to W$ be a surjective linear map between finite dimensional real vector spaces. Then*

(a) $\dim(W) \leq \dim(V)$
(b) all $y \in W$ *the preimage* $f^{-1}[y]$ *is an affine subspace of* V *of dimension* $\dim(V) - \dim(W)$*.*

Problem 1.9. Prove this. □

Recall that if $f\colon X \to Y$ is a map between sets, then $g\colon Y \to X$ is a ***left inverse*** to f if and only if the composition $g \circ f\colon X \to X$ is the identity map on X. This implies f is injective and g is surjective. Likewise g is a ***right inverse*** if and only if the composition $f \circ g\colon Y \to Y$ is the identity map on Y. In this case f is surjective and g is injective.

Proposition 1.10. *Let* $L\colon V \to W$ *be a linear map between finite dimensional real vector spaces.*
(a) If L *is surjective, then there is a linear map* $S\colon W \to V$ *that is a right inverse to* L*. (Note for* $y \in W$ *that* $x = Sy$ *gives a solution to* $Lx = y$*.)*
(b) If L *is injective, there is a linear map* $T\colon W \to V$ *that is a left inverse to* L*.*

Problem 1.10. Prove this. □

2. The Inverse and Implict Function Theorems.

2.1. **The Banach Fixed Point Theorem.** We now give what is one of the more important existence theorems for general nonlinear equations. Let (X, d) be a metric space (say that (X, d) is the plane $\mathbb{R}^2$ with its usual metric), and let $F\colon X \to X$ be a continuous map. Then given $y \in X$ we would like a method for solving the equation $F(x) = y$ for x. Experience has shown that it is often better to rewrite this equation as $f(x) = x$ for a new function f. (In the case $X = \mathbb{R}^2$ then for any constant let $f(x) = c(F(x) - y) + x$, then the equation $f(x) = x$ has the same solutions as $F(x) = y$.) Thus it turns out that many useful theorems about solving equations are stated as fixed point theorems. I don't have any real insight into why fixed point theorems turn out to easier to handle than other types of theorems on solutions, but this is the case. (In topology there is the Brouwer fixed point theorem which is a very general result about solving n equations in n unknowns.)

We first recall a bit of elementary analysis. Let a, r be real numbers, with $r \neq 1$. To find the sum

$$S = a + ar + \cdots + ar^n = \sum_{k=0}^{n} ar^k.$$

To do this multiple S by $(1 - r)$ to get

$$\begin{aligned}(1-r)S &= S - rS \\ &= (a + ar + \cdots + ar^n) - (ar + ar^2 + \cdots + ar^{n+1}) \\ &= a - ar^{n+1}.\end{aligned}$$

Thus the sum of this finite geometric series is

$$S = \frac{a - ar^{n+1}}{1-r}.$$

We will use this fact several without comment in what follows.

Let (X, d) be a metric space. Then a map $f\colon X \to X$ is a ***contraction*** if and only there is a constant $\rho < 1$ so that $d(f(x), f(y)) \leq \rho d(x, y)$. The number ρ is called the ***c***ontraction factor. It is easy to see that any contraction is continuous.

Let Y be any set and $g\colon Y \to Y$. Then the point $y_0 \in Y$ is a fixed point of g iff $g(y_0) = y_0$.

Problem 2.1. Let (X, d) be a metric space and $f\colon X \to X$ a contraction with contraction factor ρ. Then show that f has at most one fixed point. *Hint:* Assume that $a, b \in X$ are fixed points of f. Then $d(a, b) = d(f(a), f(b)) \leq \rho d(a, b)$. □

Theorem 2.1 (Banach Fixed Point Theorem)**.** *Let (X, d) be a complete metric space and $f\colon X \to X$ be a contraction with contraction factor $\rho < 1$. Then show that f has a unique fixed point x_* in X. This fixed point can be found by starting with any $x_0 \in X$ and defining a sequence $\{x_k\}_{k=0}^{\infty}$ by recursion*

$$x_1 = f(x_0), \quad x_2 = f(x_1), \quad x_3 = f(x_2), \quad \ldots \quad x_{k+1} = f(x_k), \quad \ldots$$

Then $x_* = \lim_{k\to\infty} x_k$. There is also an estimate on the error of using x_n as an approximation to x_*. This is

$$d(x_n, x_*) \leq \frac{d(x_0, x_1)\rho^n}{1-\rho}. \tag{1}$$

Remark 2.2. This result was in Banach's thesis which appeared in published form in 1922. In his thesis and this paper he also introduced "complete normed linear space" which have since been renamed as Banach spaces. The idea of looking at the sequence x_0, $x_1 = f(x_0)$, $x_2 = f(x_1)\ldots$ is an abstraction of an idea of Picard. □

Problem 2.2. Prove the theorem by doing the following:

(a) For the sequence defined above show that $d(x_k, x_{k+1}) \leq d(x_0, x_1)\rho^n$.

(b) Let $m < n$ then by repeated use of the triangle inequality show

$$d(x_m, x_n) \leq \sum_{k=m}^{n-1} d(x_k, x_{k+1}).$$

(c) Use Part (a) and sum a geometric series to show

$$d(x_m, x_n) \leq \frac{d(x_0, x_1)\rho^m - d(x_0, x_1)\rho^n}{1-\rho} \leq \frac{d(x_0, x_1)\rho^m}{1-\rho}.$$

(d) Show the sequence $x_0, x_1, x_2, \ldots$ is Cauchy sequence and therefore converges to some point x_* of X. Use Part (c) to show (1) holds.

(e) Show x_* is a fixed point of f. *Hint:* Take the limit as $k \to \infty$ of the equation $x_{k+1} = f(x_k)$.
(f) Show x_* is the only fixed point of f. □

2.2. **Banach Spaces.** Let $\mathbf{X}$ be a (not necessarily finite dimensional) vector space over the real numbers. Then a ***norm***, $|\cdot|_{\mathbf{X}}$, on $\mathbf{X}$ is a function from $\mathbf{X}$ to the real numbers such that

- (nonnegative) for all $x \in \mathbf{X}$ the inequality $|x|_{\mathbf{X}} \geq 0$ holds and with $|x|_{\mathbf{X}} = 0$ if and only if $x = 0$.
- (triangle inequality) for all $x, y \in \mathbf{X}$
$$|x + y|_{\mathbf{X}} \leq |x|_{\mathbf{X}} + |y|_{\mathbf{X}}$$
- (homogeneity) If $x \in \mathbf{X}$ and $\lambda \in \mathbb{R}$ then
$$|\lambda x|_{\mathbf{X}} = |\lambda||x|_{\mathbf{X}}.$$

As examples let $\mathbb{R}^n$ is the usual vector space of length n column vectors of real numbers. That is, because it is more compatible with matrix operations, we view elements $x \in \mathbb{R}^n$ as column vectors

$$x = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}$$

(The reason for using superscripts rather subscripts will become clear later.) To save space will will sometimes write elements of $\mathbb{R}^n$ as $x = (x^1, x^2, \ldots, x^n)$. For each $p \geq 1$ the function

$$|(x^1, x^2, \ldots, x^n)|_{\ell^p} = \left(|x^1|^p + |x^2|^p + \cdots + |x^n|^p\right)^{\frac{1}{p}}$$

is a norm on $\mathbb{R}^n$. In this case the triangle inequality is a special case of the ***Minkowski inequality*** and is not a trivial inequality. If $\mathbf{X} = C[0,1]$ is the vector space of all continuous functions $f\colon [0,1] \to \mathbb{R}$ and we define

$$|f|_{L^\infty} = \max_{x \in [0,1]} |f(x)|$$

then $|\cdot|_{L^\infty}$ makes $C[0,1]$ into a normed space.

Problem 2.3. Prove that $|\cdot|_{L^\infty}$ does make $C[0,1]$ into a normed vector space. Show this is not finite dimensional by showing that the functions p_n defined by $p_n(t) = t^n$ for $n = 0, 1, 2, \ldots$ are linearly independent. □

Anther example is to let $\mathbf{X} = L^1[0,1]$, that is the Lebesgue integrable functions $f\colon [0,1] \to \mathbb{R}$ and use for the norm

$$|f|_{L^1} = \int_{[0,1]} |f| \, dm$$

where m is Lebesgue measure. (Here I am not being altogether honest, $L^1[0,1]$ is really the vector space of equivalence class of integrable functions were we consider to functions equal if they are equal except on a set of measure zero, that is if they are equal almost everywhere.)

Proposition 2.3. *If* $\mathbf{X}$ *is a normed linear space with norm,* $|\cdot|_{\mathbf{X}}$*, then* $\mathbf{X}$ *is a metric space with distance function*

$$d(x, y) := |x - y|_{\mathbf{X}}.$$

Problem 2.4. Prove this. □

We can now define a ***Banach space***. It is a normed vector space $(\mathbf{X}, |\cdot|_{\mathbf{X}})$ that is complete as a metric space. That is all Cauchy sequences in $\mathbf{X}$ converge. (Recall that a sequence $\langle x_n \rangle_{n=1}^{\infty}$) is ***Cauchy*** if and only if for all $\varepsilon > 0$ there is a positive integer N such that $m, n > N$ implies $|x_m - x_n|_{\mathbf{X}} < \varepsilon$.)

In finite dimensions all normed spaces are complete and thus all finite dimensional normed spaces are Banach spaces. For an example of a normed space that is not a Banach space consider $C[0, 1]$ with the norm

$$|f|_{L^1} = \int_{[0,1]} |f| \, dm.$$

Since the continuous continuous functions are dense in $L^1[0, 1]$, for any discontinuous function f in $L^1[0, 1]$ we choose a sequences $\langle f_n \rangle_{n=1}^{\infty}$ of continuous functions with $\lim_{n\to\infty} |f - f_n| \, dm = 0$. Then $\langle f_n \rangle_{n=1}^{\infty}$ will be Cauchy in $C[0, 1]$ but not convergent in $C[0, 1]$ as the limit function is not in $C[0, 1]$.

Proposition 2.4. *Let* $\mathbf{X}$ *be a finite dimensional normed space with* $e_1, \ldots, e_n$ *a basis of* $\mathbf{X}$*. For any* $x \in \mathbf{X}$ *write* $x = \sum_{j=1}^{n} x_j e_j$ *with* $x_j \in \mathbb{R}$ *and use this basis to define a* $|\cdot|_{\ell^1}$ *by*

$$|x|_{\ell^1} = \sum_{j=1}^{n} |x_j|.$$

Then there are constants $C_1, C_2 > 0$ *such that*

$$C_1 |x|_{\ell^1} \leq |x|_{\mathbf{X}} \leq C_2 |x|_{\ell^1}.$$

Thus on a finite dimensional space all norms define the same topology.

Problem 2.5. Prove this. *Hint:* Let $M = \max_{1 \leq j \leq n} |e_j|_{\mathbf{X}}$. Then show

$$|x|_{\mathbf{X}} \leq M \sum_{j=1}^{n} |x_j| = M |x|_{\ell^1},$$

so $C_2 = M$ works. The inequality just given yields for $x, y \in \mathbf{X}$ that

$$|x - y|_{\mathbf{X}} \leq M |x - y|_{\ell^1}.$$

Therefore $|\cdot|_{\mathbf{X}}$ is continuous with respect to the topology defined by $|\cdot|_{\ell^1}$. The topology defined by $|\cdot|_{\ell^1}$ is the usual topology on a finite dimensional vector space and in this topology all the closed bounded sets are compact. (To be a little more explicit the basis $e_1, \ldots, e_n$ gives a linear isomorphism of $\mathbf{X}$ with $\mathbb{R}^n$ and we use this isomorphism to move the topology of $\mathbb{R}^n$

to $\mathbf{X}$.) Use this fact to show $|\cdot|_{\mathbf{X}}$ achieves it minimum, m, on the set $S := \{x : |x|_{\ell^1} = 1\}$. Then $m > 0$ and

$$m \le |x|_{\mathbf{X}} \qquad \text{whenever} \qquad |x|_{\ell^1} = 1.$$

Now use the homogeneity of the two morns to show

$$m|x|_{\ell^1} \le |x|_{\mathbf{X}}$$

for all $x \in \mathbf{X}$. Set $C_1 = m$ to complete the proof. □

2.3. **Bounded linear maps between Banach spaces.** Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces with norms $|\cdot|_{\mathbf{X}}$ and $|\cdot|_{\mathbf{Y}}$. Then a linear map $A\colon \mathbf{X} \to \mathbf{Y}$ is ***bounded*** iff there is a constant C so that

$$|Ax|_{\mathbf{Y}} \le C|x|_{\mathbf{X}} \quad \text{for all} \quad x \in \mathbf{X}.$$

The best constant C in this inequality is the ***operator norm*** (which we will usually just call the ***norm***) of A and denoted by $\|A\|$. Thus $\|A\|$ is given by

$$\|A\| := \sup_{0 \ne x \in \mathbf{X}} \frac{|Ax|_{\mathbf{Y}}}{|x|_{\mathbf{X}}}.$$

Problem 2.6. Show that a linear map $L\colon \mathbf{X} \to \mathbf{Y}$ is continuous if and only if it is bounded. □

Denote by $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ the set of all bounded linear maps $A\colon \mathbf{X} \to \mathbf{Y}$.

Proposition 2.5. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces with $\mathbf{X}$ finite dimensional. Then every linear map $A\colon \mathbf{X} \to \mathbf{Y}$ is bounded.*

Problem 2.7. Prove this. *Hint:* Let $n = \dim(\mathbf{X})$ and $e_1, \dots, e_n$ a basis of $\mathbf{X}$. If $x \in \mathbf{X}$ write $x = \sum_{j=1}^n x_j e_j$ with $x_j \in \mathbb{R}$. Let $M = \max_{1 \le j \le n} |Ae_j|_{\mathbf{Y}}$. Then, using the notation of Proposition 2.4, show

$$|Ax|_{\mathbf{X}} \le M \sum_{j=1}^n |x_j| \le \frac{M}{C_1} |x|_{\mathbf{X}},$$

which shows that A is bounded. □

Problem 2.8. (a) Show that the norm $\|\cdot\|$ makes $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ into a normed linear space. That is show if $A, B \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ and c_1 and c_2 are real numbers then $c_1 A + c_2 B \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ and

$$\|c_1 A + c_2 B\| \le |c_1| \|A\| + |c_2| \|B\|.$$

(b) Show that the norm $\|\cdot\|$ is complete on $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ and so $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ is a Banach space. *Hint:* This can be done as follows. Let $\{A_k\}_{k=1}^\infty$ be a Cauchy sequence in $\mathbf{X}$. Then $M := \sup_k \|A_k\| < \infty$. Show
  (i) For any $x \in \mathbf{X}$ the sequence $\{A_k x\}_{k=1}^\infty$ is a Cauchy sequence and as $\mathbf{Y}$ is a Banach space this implies $\lim_{k \to \infty} A_k x$ exists.
  (ii) Define a map $A\colon \mathbf{X} \to \mathbf{Y}$ by $Ax := \lim_{k \to \infty} A_k x$. Then show A is linear and $|Ax|_{\mathbf{Y}} \le M|x|_{\mathbf{X}}$ for all $x \in \mathbf{X}$. Thus A is bounded.

(iii) Let $\varepsilon > 0$ and let N_ε be so that $k, \ell \geq N_\varepsilon$ implies $\|A_k - A_\ell\| \leq \varepsilon$ (N_ε exists as $\{A_k\}_{k=1}^\infty$ is Cauchy). Then for any $x \in \mathbf{X}$ and $k \geq N_\varepsilon$ and all $\ell \geq N_\varepsilon$ we have

$$\begin{aligned} |Ax - A_k x|_{\mathbf{Y}} &\leq |Ax - A_\ell x|_{\mathbf{Y}} + |(A_\ell - A_k)x|_{\mathbf{Y}} \\ &\leq |Ax - A_\ell x|_{\mathbf{Y}} + \varepsilon |x|_{\mathbf{X}} \\ &\overset{\ell\to\infty}{\longrightarrow} 0 + \varepsilon|x|_{\mathbf{X}} = \varepsilon|x|_{\mathbf{X}}. \end{aligned}$$

This implies $\|A - A_k\| \leq \varepsilon$ for $k \geq N_\varepsilon$ and thus $\lim_{k\to\infty} A_k = A$. This shows any Cauchy sequence in $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ converges.

(c) If $\mathbf{Z}$ is a third Banach space $A \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ and $B \in \mathcal{B}(\mathbf{Y}, \mathbf{Z}))$ then $BA \in \mathcal{B}(\mathbf{X}, \mathbf{Z})$ and $\|BA\| \leq \|B\|\|A\|$. In particular if $A \in \mathcal{B}(\mathbf{X}, \mathbf{X})$ then by induction $\|A^k\| \leq \|A\|^k$. □

Remark 2.6. Some inequalities involving norms of linear maps will be used repeatedly in what follows without comment. The inequalities in question are

$$|Ax|_{\mathbf{Y}} \leq \|A\||x|_{\mathbf{X}}, \quad \|AB\| \leq \|A\|\,\|B\|, \quad \|A^k\| \leq \|A\|^k.$$

Of course the various forms of the triangle inequality will also be used. This includes the form $|u - v| \geq |u| - |v|$. □

From now on we will use facts about geometric series from the last exercise without comment.

The linear map $A \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ is ***invertible*** iff there is a $B \in \mathcal{B}(\mathbf{Y}, \mathbf{X})$ so that $AB = I_{\mathbf{Y}}$ and $BA = I_{\mathbf{X}}$ (where $I_{\mathbf{X}}$ is the identity map on $\mathbf{X}$). The map B is called the ***inverse*** of A and is denoted by $B = A^{-1}$.

Proposition 2.7. *Let $\mathbf{X}$ be a Banach space and $A \in \mathcal{B}(\mathbf{X}, \mathbf{X})$ with $\|I_{\mathbf{X}} - A\| < 1$. Then the A in invertible and the inverse is given by*

$$A^{-1} = \sum_{k=0}^\infty (I_{\mathbf{X}} - A)^k = I_{\mathbf{X}} + (I_{\mathbf{X}} - A) + (I_{\mathbf{X}} - A)^2 + (I_{\mathbf{X}} - A)^3 + \cdots$$

and satisfies the bound

$$\|A^{-1}\| \leq \frac{1}{1 - \|I_{\mathbf{X}} - A\|}.$$

Moreover if $0 < \rho < 1$ then

$$\|A - I_{\mathbf{X}}\|, \|B - I_{\mathbf{X}}\| \leq \rho \quad \textit{implies} \quad \|A^{-1} - B^{-1}\| \leq \frac{1}{(1-\rho)^2}\|A - B\|$$

Proof. Let $B := \sum_{k=0}^\infty (I_{\mathbf{X}} - A)^k$ then $\|(I_{\mathbf{X}} - A)^k\| \leq \|I_X - A\|^k$ and as $\|I_X - A\| < 1$ the geometric series $\sum_{k=0}^\infty \|I_X - A\|^k$ converges. Therefore by comparison the series defining B converges and

$$\|B\| \leq \sum_{k=0}^\infty \|I_{\mathbf{X}} - A\|^k = \frac{1}{1 - \|I_{\mathbf{X}} - A\|}.$$

Now compute

$$\begin{aligned} AB &= \sum_{k=0}^{\infty} A(I_{\mathbf{X}} - A)^k \\ &= \sum_{k=0}^{\infty} (I_{\mathbf{X}} - (I_{\mathbf{X}} - A))(I_{\mathbf{X}} - A)^k \\ &= \sum_{k=0}^{\infty} (I_{\mathbf{X}} - A)^k - \sum_{k=0}^{\infty} (I_{\mathbf{X}} - A)^{k+1} \\ &= I_{\mathbf{X}}. \end{aligned}$$

A similar calculation shows that $BA = I_{\mathbf{X}}$ (or just note A and B clearly commute). Thus $B = A^{-1}$. (The formula for $B = A^{-1}$ was motivated by the power series $(1-x)^{-1} = \sum_{k=0}^{\infty} x^k$.)

If $\|A - I_{\mathbf{X}}\|, \|B - I_{\mathbf{X}}\| \le \rho$ then by what we have just done

$$\|A^{-1}\|, \|B^{-1}\| \le \frac{1}{1-\rho}$$

Therefore

$$\begin{aligned} \|A^{-1} - B^{-1}\| = \|A^{-1}(B-A)B^{-1}\| &\le \|A^{-1}\|\,\|B^{-1}\|\,\|B - A\| \\ &\le \frac{1}{(1-\rho)^2}\|A - B\|. \end{aligned}$$

This completes the proof. $\square$

The next proposition is a somewhat more general version of the last result. The main point is that the set of invertible operators is an open set.

Proposition 2.8. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and let $A, B \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$. Assume that A is invertible. Then if B satisfies*

$$\|A - B\| < \frac{1}{\|A^{-1}\|}$$

then B is also invertible and

$$\|B^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\|\,\|A - B\|}, \qquad \|B^{-1} - A^{-1}\| \le \frac{\|A^{-1}\|^2\|B - A\|}{1 - \|A^{-1}\|\,\|A - B\|}.$$

Therefore the set of invertible maps from $\mathbf{X}$ to $\mathbf{Y}$ is open in $\mathcal{B}(\mathbf{X}, \mathbf{Y})$ and the map $A \mapsto A^{-1}$ is continuous on this set.

Proof. This is more or less a corollary to the last result. Write $B = A - (A - B) = A(I_{\mathbf{X}} - A^{-1}(A - B))$. But $\|A^{-1}(A-B)\| \le \|A^{-1}\|\,\|A - B\| < 1$ by assumption. Thus the last proposition gives that $I_{\mathbf{X}} - A^{-1}(A - B)$ is invertible and that

$$\|(I_{\mathbf{X}} - A^{-1}(A-B))^{-1}\| \le \frac{1}{1 - \|A^{-1}(A-B)\|} \le \frac{1}{1 - \|A^{-1}\|\,\|(A-B)\|}.$$

Whence $B = A(I_{\mathbf{X}} - A^{-1}(A - B))$ is the product of invertible maps and whence invertible with $B^{-1} = (I_{\mathbf{X}} - A^{-1}(A - B))^{-1}A^{-1}$. Thus

$$\|B^{-1}\| \le \|(I_{\mathbf{X}} - A^{-1}(A - B))^{-1}\| \, \|A^{-1}\| \le \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \, \|(A - B)\|}$$

which gives the required bound on $\|B^{-1}\|$.

Now $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$. Therefore

$$\|A^{-1} - B^{-1}\| \le \|A^{-1}\| \, \|A - B\| \, \|B^{-1}\| \le \frac{\|A^{-1}\|^2 \|B - A\|}{1 - \|A^{-1}\| \, \|A - B\|}.$$

This completes the proof. □

Remark 2.9. When $\mathbf{X}$ is finite dimensional a linear operator $A\colon \mathbf{X} \to \mathbf{X}$ is invertible if and only if $\det(A) \ne 0$. As det is a continuous function the set $\{A : \det(A) \ne 0\}$ is open. Thus in this case that the set of invertible operators is open has an easier proof.

2.4. **The derivative of maps between Banach spaces.** Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces, $U \subseteq \mathbf{X}$ be an open set and $f\colon U \to \mathbf{Y}$ a function. Then f is ***differentiable*** at $a \in U$ if and only if there is a linear map $A \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ so that

$$f(x) - f(a) = A(x - a) + o(|x - a|_{\mathbf{X}}).$$

More explicitly this means these is a function $x \mapsto \varepsilon(x; a)$ from a neighborhood of a in U so that

$$f(x) - f(a) = A(x - a) + |x - a|_{\mathbf{X}} \varepsilon(x; a) \quad \text{where} \quad \lim_{x \to a} |\varepsilon(x, a)|_{\mathbf{Y}} = 0.$$

When f is differentiable at a the linear map A is unique and called the ***derivative*** of f at a. It will be denoted by $A = f'(a)$. Thus for us the derivative is a bounded linear map $f'(a) \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ rather than a number.

Problem 2.9. Show that f is differential at a with $f'(a) = A$ if and only if

$$\lim_{x \to a} \frac{|f(x) - f(a) - A(x - a)|_{\mathbf{Y}}}{|x - a|_{\mathbf{X}}} = 0.$$

□

Problem 2.10. If f is differentiable at a then f is continuous at a. □

To get a feel for what this linear map measures let $v \in \mathbf{X}$, assume that $f\colon U \to \mathbf{Y}$ is differentiable at a and let $c(t) := f(a + tv)$. Then for $t \ne 0$

$$\begin{aligned} \frac{1}{t}(c(t) - c(0)) &= \frac{1}{t}(f(a + tv) - f(a)) = \frac{1}{t}(f'(a)tv + |tv|_{\mathbf{X}} \varepsilon(a + tv; a)) \\ &= f'(a)v + |v|_{\mathbf{X}} \varepsilon(a + tv; a). \end{aligned}$$

But $\lim_{t \to 0} \varepsilon(a + tv; a) = 0$ so this implies that c has a tangent vector at $t = 0$ and that it is given by $c'(0) = f'(a)v$. That is $f'(a)v$ is the "directional derivative" at a of f in the direction v. As a variants on this here are a couple of probs.

Problem 2.11. Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and $U \subset \mathbf{X}$ open. Let $c\colon (a,b) \to U$ be a continuously differentiable map and let $f\colon U \to \mathbf{Y}$ be a map that is differentiable at $c(t_0)$. Then $\gamma(t) := f(c(t))$ is differentiable at t_0 and

$$\frac{d}{dt} f(c(t))\bigg|_{t=t_0} = \gamma'(t_0) = f'(c(t_0))c'(t_0).$$

□

To make this more concrete let us look at some finite dimensional cases.

Problem 2.12. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a differentiable scalar valued function on $\mathbb{R}^n$. Let

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \; e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

be the standard basis of $\mathbb{R}^n$ and write $x \in \mathbb{R}^n$ as

$$x = \sum_{j=1}^n x^j e_j = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}.$$

Then for each $x \in$ the derivative $f'(x)$ is a linear map form $\mathbb{R}^n$ to $\mathbb{R}$. (Linear maps from a vector space to the fields of scalars are called ***linear functionals***.) Show that the matrix of this linear functional is the row vector

$$\left[\frac{df}{dx^1}, \frac{df}{dx^2}, \dots, \frac{df}{dx^n}\right]$$

□

Problem 2.13. Let $\mathbf{X} = \mathbb{R}^n$ and $\mathbf{Y} = \mathbb{R}^m$ and let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function given in components as

$$f(x) = \begin{bmatrix} f^1(x) \\ f^2(x) \\ \vdots \\ f^m(x) \end{bmatrix}.$$

Let $\dfrac{\partial f}{\partial x^i}$ be the partial derivative of f with respect to x^i. That is

$$\frac{\partial f}{\partial x^i} = \begin{bmatrix} \dfrac{\partial f^1}{\partial x^i}(x) \\ \dfrac{\partial f^2}{\partial x^i}(x) \\ \vdots \\ \dfrac{\partial f^m}{\partial x^i}(x) \end{bmatrix}$$

Assume that f is differentiable at $x = a$. Then show that the matrix A of $f'(a)$ in the standard basis of $\mathbb{R}^n$ and $\mathbb{R}^m$ is the matrix with columns $\dfrac{\partial f}{\partial x^1}, \dfrac{\partial f}{\partial x^2}, \dots, \dfrac{\partial f}{\partial x^n}$. That is

$$A = \left[\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2}, \dots, \frac{\partial f}{\partial x^m}\right] = \begin{bmatrix} \dfrac{\partial f^1}{\partial x^1} & \dfrac{\partial f^1}{\partial x^2} & \cdots & \dfrac{\partial f^1}{\partial x^n} \\ \dfrac{\partial f^2}{\partial x^1} & \dfrac{\partial f^2}{\partial x^2} & \cdots & \dfrac{\partial f^2}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f^m}{\partial x^1} & \dfrac{\partial f^m}{\partial x^2} & \cdots & \dfrac{\partial f^m}{\partial x^n} \end{bmatrix}$$

where these are all evaluated at $x = a$. □

The following gives trivial examples of differentiable maps.

Proposition 2.10. *Let $A\colon \mathbf{X} \to \mathbf{Y}$ be a bounded linear map between Banach spaces and $y_0 \in \mathbf{Y}$. Set $f(x) = Ax + y_0$ then f is differentiable at all points $a \in \mathbf{X}$ and $f'(a) = A$ for all a.*

Proof. $f(x) - f(a) = A(x - a)$ so the definition of differentiable is verified with $\varepsilon(x, a) \equiv 0$. □

The following gives a less trivial example.

Proposition 2.11. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and let $U \subset \mathcal{B}(\mathbf{X}, \mathbf{Y})$ be the set of invertible elements (this is an open set by Proposition 2.8). Define a map $f\colon U \to \mathcal{B}(\mathbf{Y}, \mathbf{X})$ by*

$$f(X) = X^{-1}.$$

Then f is differentiable and for $A \in U$ the derivative $f'(A)\colon \mathcal{B}(\mathbf{X}, \mathbf{Y}) \to \mathcal{B}(\mathbf{Y}, \mathbf{X})$ is the linear map whose value on $V \in \mathcal{B}(\mathbf{X}, \mathbf{Y})$ is

$$f'(A)V = -A^{-1}VA^{-1}.$$

Proof. Let $L\colon \mathcal{B}(\mathbf{X}, \mathbf{Y}) \to \mathcal{B}(\mathbf{Y}, \mathbf{X})$ be the linear map $LV := -A^{-1}VA^{-1}$ Then for $X \in U$

$$\begin{aligned} f(X) - f(A) - L(X - A) &= X^{-1} - A^{-1} + A^{-1}(X - A)A^{-1} \\ &= X^{-1}(A - X)A^{-1} + A^{-1}(X - A)A^{-1} \\ &= (-X^{-1} + A^{-1})(X - A)A^{-1} \\ &= X^{-1}(X - A)A^{-1}(X - A)A^{-1}, \end{aligned}$$

so that

$$\|f(X) - f(A) - L((X - A)\| \leq \|X^{-1}\|\, \|A^{-1}\|^2 \|X - A\|^2.$$

The map $X \mapsto X^{-1}$ is continuous (Proposition 2.8) so $\lim_{X\to A} X^{-1} = A^{-1}$. Thus

$$\lim_{X\to A} \frac{\|f(X) - f(A) - L((X-A)\|}{\|X-A\|} \le \lim_{X\to A} \|X^{-1}\|\,\|A^{-1}\|^2\|X-A\| = 0.$$

The result now follows from Exercise 2.9. □

Next we consider the chain rule.

Proposition 2.12. *Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ be Banach spaces $U \subseteq \mathbf{X}$, $V \subseteq \mathbf{Y}$ open sets $f\colon U \to \mathbf{Y}$ and $g\colon V \to \mathbf{Z}$. Let $a \in U$ so that $f(a) \in V$ and assume that f is differentiable at a and g is differentiable at $f(a)$. Then $g \circ f$ is differentiable at a and*

$$(g \circ f)'(a) = g'(f(a))f'(a)$$

Proof. From the definitions $f(x) - f(a) = f'(a)(x-a) + |x-a|_{\mathbf{X}}\varepsilon_1(x;a)$ and $g(y) - g(f(a)) = g'(f(a)) + |y - f(a)|_{\mathbf{Y}}\varepsilon_2(x; f(a))$ where $\lim_{x\to a}\varepsilon_1(x;a) = 0$ and $\lim_{y\to f(a)}\varepsilon_2(y; f(a)) = 0$. Then

$$\begin{aligned}
g(f(x)) - g(f(a)) &= g'(f(a))(f(x) - f(a)) + |f(x) - f(a)|_{\mathbf{Y}}\varepsilon_2(f(x), f(a)) \\
&= g'(f(a))f'(a)(x-a) + g'(f(a))|x-a|_{\mathbf{X}}\varepsilon_1(x,a) \\
&\qquad + |f'(a)(x-a) + |x-a|_{\mathbf{X}}\varepsilon_1(x,a)|_{\mathbf{X}}\varepsilon_2(f(x), f(a)) \\
&= g'(f(a))f'(a)(x-a) + |x-a|_{\mathbf{X}}\bigg(g'(f(a))\varepsilon_1(x,a) \\
&\qquad + \left|f'(a)\frac{x-a}{|x-a|_{\mathbf{X}}} + \varepsilon_1(x,a)\right|_{\mathbf{X}} \varepsilon_2(f(x); f(a))\bigg) \\
&= g'(f(a))f'(a)(x-a) + |x-a|_{\mathbf{X}}\varepsilon_3(x;a),
\end{aligned}$$

where this defines $\varepsilon_3(x,a)$. Then

$$|\varepsilon_3(x,a)|_{\mathbf{X}} \le \|g'(f(a))\||\varepsilon_1(x,a)|_{\mathbf{Y}} + (\|f'(a)\| + |\varepsilon_1(x;a)|_{\mathbf{Y}})|\varepsilon_2(f(x), f(a))|_{\mathbf{Z}}.$$

This (and the continuity of f at a) implies $\lim_{x\to a}\varepsilon_3(x,a) = 0$ which completes the proof. □

Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces, $U \subset \mathbf{X}$ open and $f\colon U \to \mathbf{Y}$. Then f is ***continuously differentiable*** on U iff f is differentiable at each point $x \in U$ and the map $x \mapsto f'(x)$ is a continuous map from U to $\mathcal{B}(\mathbf{X}, \mathbf{Y})$. Or what is the same thing, $f'(x)$ exists for all $x \in U$ and for $a \in U$ we have $\lim_{x\to a}\|f'(x) - f'(a)\| = 0$.

If $c\colon [a,b] \to U$ is a continuously differentiable curve and $f\colon U \to \mathbf{Y}$ is a continuously differentiable map, then $\gamma(t) := f(c(t))$ is a continuously differential curve $\gamma\colon [a,b] \to \mathbf{Y}$ and by the chain rule (or Problem (2.11))

$$\gamma'(t) = f'(c(t))c'(t).$$

Using the fundamental theorem of calculus this gives

$$\gamma(b) = \gamma(a) = \int_a^b \gamma'(t)\,dt = \int_a^b f'(c(t))c'(t)\,dt$$

But

$$|\gamma'(t)|_{\mathbf{Y}} = |f'(c(t))c'(t)|_{\mathbf{Y}} \le \|f'(c(t))\||c'(t)|_{\mathbf{X}}.$$

These can be combined to give

$$(2) \qquad |\gamma(b) - \gamma(a)|_{\mathbf{Y}} = \left| \int_a^b f'(c(t))c'(t)\,dt \right|_{\mathbf{Y}} \le \int_a^b \|f'(c(t))\||c'(t)|_{\mathbf{X}}\,dt.$$

Recall that a U is ***convex*** if and only if when $x_0, x_1 \in U$ and $0 \le t \le 1$, then $(1-t)x_0 + tx_1 \in U$.

Proposition 2.13 (Mean Value inequality)**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and let $U \subset \mathbf{X}$ be open and convex. Assume that $f\colon U \to \mathbf{Y}$ is continuously differentiable and that $\|f'(x)\| \le C$ for all $x \in U$. Then*

$$|f(x_1) - f(x_0)|_{\mathbf{Y}} \le C|x_1 - x_0|_{\mathbf{X}}$$

for all $x_1, x_0 \in U$

Proof. Let $c\colon [0,1] \to U$ be given by $c(t) = (1-t)x_0 + tx_1 = x_0 + t((x_1 - x_0)$ (this curves lies in U as $x_0, x_1 \in U$ and U is convex). Then $c'(t) = (x_1 - x_0)$. Let $\gamma(t) := f(c(t))$. Then $\gamma'(t) = f'(c(t))c'(t) = f'(c(t))(x_1 - x_0)$. Putting this into (2) implies

$$\begin{aligned} |f(x_1) - f(x_0)|_{\mathbf{Y}} = |\gamma(1) - \gamma(0)|_{\mathbf{Y}} &\le \int_0^1 \|f'(c(t))\||x_1 - x_0|_{\mathbf{X}}\,dt \\ &\le \int_0^1 C|x_1 - x_0|_{\mathbf{X}}\,dt = C|x_1 - x_0|_{\mathbf{X}}. \end{aligned}$$

This completes the proof. □

2.5. **Preliminary version of the inverse function theorem.** In a Banach space $\mathbf{X}$ we let

$$B(a,r) := \{x : |x-a|_{\mathbf{X}} < r\}, \qquad \overline{B}(x,r) := \{x : |x-a|_{\mathbf{X}} \le r\}$$

be the open and closed balls of radius r centered at a. In the following theorem and its proof we will always be referring to the same Banach space $\mathbf{X}$, and therefore we simplify notation by shorting $|x|_{\mathbf{X}}$ to $|x|$.

Theorem 2.14. *Let $\mathbf{X}$ be a Banach space and W an open subset of $\mathbf{X}$ that contains 0. Let $f\colon W \to \mathbf{X}$ be a continuously differentiable function with*

$$f(0) = 0, \qquad \textit{and} \qquad f'(0) = I$$

where I is the identity map. Then 0 has an open neighborhood $U \subseteq W$ such that the image $f[U]$ is open and there is a continuously differentiable map $g\colon f[U] \to U$ inverse to $f\big|_U$ and the derivative of g is given by

$$g'(y) = f'(g(y))^{-1}.$$

Problem 2.14. Prove this along the following lines. As motivation note that on some level finding the inverse of f is equivalent to solving $f(x) = y$ for x. We will reduce this to finding the fixed point of a contraction. For $y \in \mathbf{X}$ set

$$\varphi_y(x) := x - f(x) + y.$$

(a) Show

$$f(x) = y \qquad \text{if and only if} \qquad \varphi_y(x) = x.$$

Also show that

$$\varphi_y = \varphi_0 + y$$

and therefore all the φ_y's have the same derivative:

$$\varphi_y'(x) = I - f'(x)$$

where I is the identity map on $\mathbf{X}$.

(b) We would like φ_y to be a contraction. This will not necessarily be true on the entire domain of f. So we need to restrict smaller set. The map $x \mapsto f'(x)$ is continuous by assumption. Use this to show that $x \mapsto \|I - f'(x)\|$ is continuous and therefore there is a $r > 0$ such that

$$x \in \overline{B}(0,r) \qquad \text{implies} \qquad \|I - f'(x)\| \le \frac{1}{2}.$$

(For the continuity of $x \mapsto \|I - f'(x)\|$ note this is the composition of $x \mapsto I - f'(x)$ on W and the map $A \mapsto \|A\|$ on $\mathcal{B}(\mathbf{X}, \mathbf{X})$ and these maps are continuous.) By part (a) this is equivalent to

$$x \in \overline{B}(0,r) \qquad \text{implies} \qquad \|\varphi_y'(x)\| \le \frac{1}{2}.$$

for any $y \in \mathbf{X}$. For future use also note that $\|I - f'(x)\| \le 1/2$ implies (by Proposition 2.7) that for $x \in \overline{B}(0,r)$ the linear map $f'(x)$ has an inverse, $f'(x)^{-1}$ and

$$\|f'(x)^{-1}\| \le 2.$$

(c) Now use the Mean Value Inequality (Proposition 2.13) to show that for all $x_0, x_1 \in \overline{B}(0,r)$ and for any $y \in \mathbf{X}$

$$|\varphi_y(x_1) - \varphi_y(x_0)| \le \frac{1}{2}|x_1 - x_0|.$$

(d) The last part of the problem does not quite show that φ_y is a contraction on $\overline{B}(0,r)$ as φ_y need not map $\overline{B}(0,r)$ into itself. (For example if $|y| \ge 2r$ the $\varphi_y(0) \notin \overline{B}(0,r)$.) This is not hard to fix. Note

$$|\varphi_y(x)| = |\varphi_0(x) + y| \le |\varphi_0(x)| + |y|.$$

Now show if $x \in \overline{B}(0,r)$, then

$$|\varphi_0(x)| = |\varphi_0(x) - \varphi_0(0)| \le \frac{1}{2}|x| \le \frac{r}{2}.$$

Thus

$$x \in \overline{B}(0,r),\ y \in \overline{B}(0,r/2) \qquad \text{implies} \qquad \varphi_y(x) \in \overline{B}(0,r).$$

(e) As $\overline{B}(0,r)$ is a closed subset of a complete metric space, it is itself a complete metric space. Use what has been done so far to show that for all $y \in \overline{B}(0,r/2)$ that $\varphi_y \colon \overline{B}(0,r) \to \overline{B}(0,r)$ is a contraction and therefore by the Banach Fixed Point Theorem (Theorem 2.1) φ_y has a unique fixed point in $\overline{B}(0,r)$ and therefore for all $y \in \overline{B}(0,y)$ the equation $f(x) = y$ has a unique solution with $x \in \overline{B}(0,r)$.

(f) We now deal with an annoying minor point. We are looking an inverse of f on a open neighborhood of 0, but so far we are working with the closed sets $\overline{B}(0,r)$ and $\overline{B}(0,r/2)$. So show that if $y \in B(0,r/2)$ that $f(x) = y$ has a unique solution with $x \in B(0,r)$. *Hint:* If $y \in B(0,r)$, then $|y| < r/2$. If $f(x) = y$ with $f(x)$, then $x = \varphi_y(x)$ and therefore

$$|x| = |\varphi_y(x)| = |\varphi_0(x) + y| \le |\varphi_0(x)| + |y|$$

now proceed as in Part (d).

(g) We have shown the image of f contains $B(0,r/2)$. To get an inverse we also need that f is injective. To do this use that $f(x) = x - \varphi_0(x)$ and therefore use the triangle inequality and the reverse triangle inequality (i.e. $|a+b| \ge |a| - |b|$) to show for all $x_0, x_1 \in B(0,r)$ that

$$|x_1 - x_0| - |\varphi_0(x_1)| \le |f(x_1) - f(x_0)| \le |x_1 - x_0| + |\varphi_0(x_1) - \varphi_0(x_0)|$$

and therefore

$$\frac{1}{2}|x_1 - x_0| \le |f(x_1) - f(x_0)| \le \frac{3}{2}|x_1 - x_0|.$$

This shows that f is injective on $B(0,r)$. (The upper bound on $|f(x_1) - f(x_0)|$ will be useful in showing the inverse of f is continuous.)

(h) Let $U = \left(f\big|_{B(0,r)}\right)^{-1}[B(0,r/2)]$ be the preimage of $B(0,r)$ by the restriction of f to $B(0,r)$. Show that U is open and that $f\big|_U \colon U \to B(0,r/2)$ is a bijection. Therefore there is an inverse $g \colon B(0,r/2) \to U$ to $f\big|_U$.

(i) Now show that g is continuous by showing for $y_0, y_1 \in B(0,r/2)$ that

$$\frac{2}{3}|y_1 - y_0| \le |g(y_1) - g(y_0)| \le 2|y_1 - y_0|.$$

Hint: Let $x_0 = g(y_0)$ and $x_1 = g(y_1)$ and use these values in the inequalities of Part (g).

(j) Show that g is differentiable and its derivative is $g'(y) = f'(g(y))^{-1}$. *Hint:* Let $y, y_0 \in B(0,r/2)$ and let $x = g(y)$ and $x_0 = g(y_0)$. As f is differentiable

$$f(x) - f(x_0) = f'(x - x_0)(x - x_0) + \varepsilon(x, x_0)|x - x_0|$$

with

$$\lim_{x \to x_0} \varepsilon(x, x_0) = 0.$$

Using $f(x) = y$, $f(x_0) = y_0$ show

$$\begin{aligned} g(y) - g(y_0) &= f'(g(y_0))^{-1}(y - y_0) - f^{-1}(g(y_0))\varepsilon(g(y), g(y_0))|g(y) - g(y_0)| \\ &= f'(g(y_0))^{-1}(y - y_0) + \rho(y, y_0) \end{aligned}$$

Where this defines $\rho(y, y_0)$. Rewrite $\rho(y, y_0)$ as

$$\begin{aligned} \rho(y, y_0) &= -f^{-1}(g(y_0))\varepsilon(g(y), g(y_0))\frac{|g(y) - g(y_0)|}{|y - y_0|}|y - y_0| \\ &= \varepsilon_1(y, y_0)|y - y_0| \end{aligned}$$

where

$$\varepsilon_1(y, y_0) = -f^{-1}(g(y_0))\varepsilon(g(y), g(y_0))\frac{|g(y) - g(y_0)|}{|y - y_0|}.$$

To show g is differentiable at y_0 it is enough to show

$$\lim_{y \to y_0} \varepsilon_1(y, y_0) = 0.$$

By Parts (b) and (i) we have

$$\|f'(g(y_0))^{-1}\| \le 2, \qquad \frac{|g(y) - g(y_0)|}{|y - y_0|} \le 2.$$

Use these to show

$$|\varepsilon_1(y, y_0)| \le 4|\varepsilon(g(y), g(y_0))|$$

and use the continuity of g to show

$$\lim_{y \to y_0} \varepsilon(g(y), g(y_0)) = 0.$$

Put these pieces together to conclude $\lim_{y \to y_0} \varepsilon_1(y, y_0) = 0$.

(k) Now that g is differentiable, all that remains is to show it is continuously differentiable. Prove this. *Hint:* From the last part of the problem we have that

$$g'(y) = f'(g(y))^{-1}.$$

This is the composition of three maps. The first is $y \to g(y)$ form $B(0, r/2)$ to U, the second is $x \mapsto f'(x)$ from U to $\mathcal{B}(\mathbf{X}, \mathbf{X})$, and the third is the inverse map $A \mapsto A^{-1}$ on the set of invertible elements of $\mathcal{B}(\mathbf{X}, \mathbf{X})$. All of these maps are continuous (for the continuity of $A \mapsto A^{-1}$ see Proposition 2.8). □

2.6. **Inverse Function Theorem.**

Theorem 2.15. *Let W be an open subset of the Banach space $\mathbf{X}$, and $f\colon W \to \mathbf{Y}$ a continuously differentiable map from W to the Banach space $\mathbf{Y}$. If $a \in W$ and $f'(a)\colon \mathbf{X} \to \mathbf{Y}$ is an invertible linear map, then a has an open neighborhood U and $f(a)$ has an open neighborhood V and there is a continuously differentiable map $g\colon V \to U$ that is the inverse of the restriction $f\big|_U$. To be very explicit*

$$f(g(y)) = y \qquad \text{and} \qquad g(f(x)) = x$$

for all $x \in U$ and $y \in V$.

Problem 2.15. Prove this. *Hint:* This can be reduced to Theorem 2.14. Let $W_0 = W - a := \{w - a : w \in W\}$. This is a neighborhood of 0 in $\mathbf{X}$. Define $f_0 \colon W_0 \to \mathbf{X}$ by

$$f_0(x) = f'(a)\,(f(x+a) - f(a))\,.$$

(a) Show that $f_0(0) = 0$, $f_0'(0) = I$, and

$$f(x) = f'(a)f_0(x-a) + f(a).$$

(b) Now use Theorem 2.14 to show that 0 has open neighborhoods U_0 and $V = f[U_0]$ and a continuously differentiable map $g_0 \colon V_0 \to U_0$ such that g_0 is the inverse of the restriction $f\big|_{U_0}$.

(c) Let $U = U_0 + a$ and $W = W_0 + f(a)$ and

$$g(y) = g_0\left(f'(a)^{-1}(y - f(a))\right).$$

Show

$$g_0(y) = g\left(f'(a)y + f(a)\right) - f(a)$$

and then that

$$f(g(y)) = y \qquad \text{and} \qquad g(f(x)) = x$$

for all $x \in U$ and $y \in V$. Thus g is the required inverse to the restriction of f to U. As g_0 is continuous differentiable we see that g is the composition of continuous differentiable functions and therefore g is continuous differentiable.

(d) To get the formula for $g'(y)$ note that

$$f \circ g(y) = f(g(y)) = y = Iy$$

By the chain rule

$$(f \circ g)'(y) = f'(g(y))g'(y)$$

and the derivative of I is I. Thus $f'(g(y))g'(y) = I$. Now solve this for $g'(y)$ to finish the proof. $\square$

A map f between metric spaces is an ***open map*** if and only if f maps open set to open set. That is if V is an open subset of the domain of f, then the image $f[U]$ is open.

Proposition 2.16. *Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces and U an open subset of $\mathbf{X}$. If $f \colon U \to \mathbf{Y}$ is a continuous differentiable map such that $f'(x)$ is invertible for all $x \in U$, then f is an open map.*

Problem 2.16. Prove this. $\square$

A map $f \colon M \to N$ between metric spaces is a ***proper map*** if and only if for all compact subsets $K \subseteq M$ the preimage $f^{-1}[K]$ is compact.

Problem 2.17. This problem is to give a feel for what it means for a map to be proper. Show that a map $f\colon \mathbb{R}^n \to \mathbb{R}$ is if and only if

$$\lim_{|x|_{\ell^2}\to\infty} |f(x)| = \infty.$$

More generally if $f\colon \mathbb{R}^m \to \mathbb{R}^n$ then f is proper if and only if

$$\lim_{|x|_{\ell^2}\to\infty} |f(x)|_{\ell^2} = \infty.$$

Thus a map between Euclidean spaces is proper exactly when it maps large points to large points. $\square$

Proposition 2.17. *If $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is a proper map, then the image $f[\mathbb{R}^m]$ is closed in $\mathbb{R}^n$.*

Problem 2.18. Prove this. *Hint:* Let y_0 be a point in the closure of $f[\mathbb{R}^n]$. We need to show that y_0 is in $f[\mathbb{R}^m]$. As y_0 is in the closure of $f[\mathbb{R}^n]$ there is a sequence $y_1, y_2, \ldots \in f[\mathbb{R}^m]$ with $\lim_{k\to\infty} y_k = y_0$. Then $K := \{y_0\} \cup \{y_k : k = 1, 2, \ldots\}$ is a compact subset of $\mathbb{R}^n$. Let $x_k \in \mathbb{R}^m$ with $y_k = f(x_k)$ then $\{x_k : k = 1, 2, \ldots\} \subseteq f^{-1}[K]$. Now use that f is proper to show that $x_1, x_2, \ldots$ is a bounded sequence and therefore has a convergent subsequence. $\square$

Recall that a metric space, M, is ***connected*** if and only if the only subsets C of M that are both open and closed are $\varnothing$ and M.

We can now give our first global result.

Proposition 2.18. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable proper map and assume that $f'(x)$ is invertible for all $x \in \mathbb{R}^n$. Then f is surjective.*

Problem 2.19. Prove this. $\square$

Remark 2.19. In the last proposition more can be said. In fact f will be bijective. Showing that f is injective uses that $\mathbb{R}^n$ is simply connected and a some results about covering spaces from algebraic topology. $\square$