THE INVERSE FUNCTION THEOREM FOR LIPSCHITZ MAPS

RALPH HOWARD
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, S.C. 29208, USA
HOWARD@MATH.SC.EDU

Abstract. This is an edited version of a proof, in the from of exercises with detailed hints, of the inverse function that was given to a graduate class in differential equations as homework. The goal of the assignment was not only to present the basics about inverse functions, but to recall/introduce the basics about differential calculus in Banach spaces and to give an example of a nontrivial application of the contraction mapping principle (which is also included as one of the problems) as these were among the basic tools used in the class. Thus the presentation starts more or less from the definition of the derivative of maps between Banach spaces and proves not only the standard version of the inverse function, but also the version for Lipschitz maps.

Contents

1. Introduction. 2
2. Derivatives of maps between Banach Spaces 2
2.1. Bounded linear maps between Banach spaces. 2
2.2. The derivative of maps between Banach spaces 5
3. The Inverse Function Theorem for Lipschitz Maps 9
4. The Inverse Function Theorem 11
5. Implicit Function Theorem: Surjective From 14
5.1. Preliminary results on surjective linear maps. 14
.2. The surjective form of the implicit function theorem for Lipschitz maps. 16
A. Appendix: Contraction Mappings and the Banach Fixed Point Theorem 16
A.1. Some Review 16
A.2. Metric spaces 17
A.3. The Banach Fixed Point Theorem 17

Date: March 1997.

1. Introduction.

There are many motivations for the inverse function theorem. Here is one. Consider a map $f\colon \mathbf{R}^n \to \mathbf{R}^m$ and a point $y \in \mathbf{R}^m$. Then it is interesting to know when the equation $f(x) = y$ can be solve for x. Of course in complete generality there is no reasonable solution to the problem. But assume f is smooth, or at least continuously differentiable, and that $f(x_0) = y_0$. Then it is reasonable to look for conditions on f so that for y sufficiently close to y_0 that $f(x) = y$ has a solution with x close to x_0. This is the problem the inverse function theorem answers. Let $f'(x_0)\colon \mathbf{R}^n \to \mathbf{R}^m$ be the derivative (this is the linear map that best approximates f near x_0 see §2.2 for the exact definition) and assume that $f'(x_0)\colon \mathbf{R}^n \to \mathbf{R}^m$ is onto. Then the implicit function theorem gives us a open neighbor hood V so that for every $y \in V$ the equation $f(x) = y$ has a solution. More than that it allows us to choose a continuously differentiable function $\varphi\colon V \to \mathbf{R}^n$ with $\varphi(y_0) = x_0$ and such that $f(\varphi(y)) = y$ that is not only can we solve the equation $f(x) = y$, we can do it in such a way that the solution $x = \varphi(y)$ is a smooth function of $y \in V$. Showing that $f(x) = y$ can be solved for x is a more or less typical application of the Banach Fixed Point Theorem (i.e contraction mapping principle which is included in an appendix), but is nontrivial enough to be quite interesting. Showing that the solution $x = \varphi(y)$ depends smoothly on y requires more work.

Rather than just work with maps between the Euclidean spaces, we will work with maps between Banach spaces. The definitions and results we need on bounded linear maps and differentiable functions between are given in in the preliminary sections. The work is split into two steps. First a version of the theorem is given for Lipschitz maps (see §3) that are sufficiently close to a linear map. This version gives the existence of an inverse φ which is continuous. This is then used (in §4) to show that if the original map is continuously differentiable then so is the its local inverse φ. One advantage to splitting the proof up in this manner is that if makes it easy to give explicit bounds on the size of the neighborhood where the local inverse is defined.

2. Derivatives of maps between Banach Spaces

2.1. Bounded linear maps between Banach spaces.

Recall that a ***Banach space*** is a normed vector space that is complete (i.e. Cauchy sequences converge) with respect to the metric by the norm. 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

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

Exercise 2.1. 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 $A\colon \mathbf{X}\to\mathbf{Y}$.

Exercise 2.2.

1. 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 $c_1A+c_2B\in\mathcal{B}(\mathbf{X},\mathbf{Y})$ and
$$\|c_1A+c_2B\|\le |c_1|\|A\|+|c_2|\|B\|.$$
2. 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
   (a) For any $x\in\mathbf{X}$ the sequence $\{A_kx\}_{k=1}^\infty$ is a Cauchy sequence and as $\mathbf{Y}$ is a Banach space this implies $\lim_{k\to\infty}A_kx$ exists.
   (b) Define a map $A\colon\mathbf{X}\to\mathbf{Y}$ by $Ax:=\lim_{k\to\infty}A_kx$. Then show A is linear and $|Ax|_{\mathbf{Y}}\le M|x|_{\mathbf{X}}$ for all $x\in\mathbf{X}$. Thus A is bounded.
   (c) Let $\varepsilon>0$ and let N_ε be so that $k,\ell\ge N_\varepsilon$ implies $\|A_k-A_\ell\|\le\varepsilon$ (N_ε exists as $\{A_k\}_{k=1}^\infty$ is Cauchy). Then for any $x\in\mathbf{X}$ and $k\ge N_\varepsilon$ and all $\ell\ge N_\varepsilon$ we have
$$\begin{aligned}|Ax-A_kx|_{\mathbf{Y}}&\le |Ax-A_\ell x|_{\mathbf{Y}}+|(A_\ell-A_k)x|_{\mathbf{Y}}\\&\le |Ax-A_\ell x|_{\mathbf{Y}}+\varepsilon|x|_{\mathbf{X}}\\&\stackrel{\ell\to\infty}{\longrightarrow} 0+\varepsilon|x|_{\mathbf{X}}=\varepsilon|x|_{\mathbf{X}}.\end{aligned}$$
   This implies $\|A-A_k\|\le\varepsilon$ for $k\ge N_\varepsilon$ and thus $\lim_{k\to\infty}A_k=A$. This shows any Cauchy sequence in $\mathcal{B}(\mathbf{X},\mathbf{Y})$ converges.
3. 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 $BA\in\mathcal{B}(X,\mathbf{Z})$ and $\|BA\|\le\|B\|\|A\|$. In particular if $A\in\mathcal{B}(\mathbf{X},\mathbf{X})$ then by induction $\|A^k\|\le\|A\|^k$. □

Remark 2.3. Some inequalities involving norms of linear maps will be used repeatedly in what follows without comment. The inequalities in question are

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

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

As first step in understanding when nonlinear maps have inverses we look at the set of invertible linear maps

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 2.4. *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 $\|A - I_{\mathbf{X}}\|, \|B - I_{\mathbf{X}}\| \leq \rho$ then by what we have just done we have

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

Therefore

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

This completes the proof. □

The next proposition is a somewhat more general version of the last result. It is not required for the proof of the inverse function and can be skipped. However as a good exercise in learning to do calculations with linear operators I recommend you try to prove it on your own.

Proposition 2.5. *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. □

2.2. **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 $A \in B(\mathbf{X}, \mathbf{Y})$ 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.

Exercise 2.6. Show that f is differential at a with $f'(a) = A$ if and only if

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

□

Exercise 2.7. 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 $t \neq 0$

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

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 exercises.

Exercise 2.8. 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

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

□

Exercise 2.9. Let $\mathbf{X} = \mathbf{R}^n$ and $\mathbf{Y} = \mathbf{R}^m$ and let $f\colon \mathbf{R}^n \to \mathbf{R}^m$ be a function which can be 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 $\mathbf{R}^n$ and $\mathbf{R}^n$ 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

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

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 2.5). 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 2.5) so $\lim_{X \to A} X^{-1} = A^{-1}$. Thus

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

The result now follows from Exercise 2.6. □

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 Exercise(2.8))

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

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

$$|\gamma(b) - \gamma(a)|_{\mathbf{Y}} = \left|\int_a^b f'(c(t))c'(t)\,dt\right|_{\mathbf{Y}} \le \int_a^b \|f'(c(t))\||c'(t)|_{\mathbf{X}}\,dt. \tag{2.1}$$

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

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

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 (2.1) implies

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

This completes the proof. □

3. The Inverse Function Theorem for Lipschitz Maps

The results of this section are a little more general than is usually needed in applications, but the proof of the more general results is actually easier in that it abstracts the inequality ((3.1) below) that is central to most proofs of the inverse function. A map $f\colon \mathbf{X} \to \mathbf{Y}$ between metric spaces (see the appendix for basics about metric spaces) in an ***open map*** iff $f[U]$ is an open subset of $\mathbf{Y}$ for all open sets $U \subseteq \mathbf{X}$. A map between f metric spaces is a ***Lipschitz map*** iff there is a constant M so that $d(f(x_2), f(x_1)) \leq Md(x_2, x_1)$ It is a ***Lipeomorphism***[1] iff it is a Lipschitz map that has a Lipschitz inverse. We use the usual notation for open and closed balls, that is for $x_0 \in \mathbf{X}$ and $r > 0$ then the open an closed balls of radius r about x_0 are

$$B(x_0, r) := \{x \in \mathbf{X} : |x - x_0|_{\mathbf{X}} < r\}, \quad \overline{B}(x_0, r) := \{x \in \mathbf{X} : |x - x_0|_{\mathbf{X}} \leq r\}.$$

Theorem 3.1. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be Banach spaces and let* $\overline{B}(x_0, r_0)$ *be a closed ball in* $\mathbf{X}$*. Let* $f\colon \overline{B}(x_0, r_0) \to \mathbf{Y}$ *so that for some invertible linear map* $L\colon \mathbf{X} \to \mathbf{Y}$ *and some* $\rho < 1$

$$|L^{-1}f(x_2) - L^{-1}f(x_1) - (x_2 - x_1)|_{\mathbf{X}} \leq \rho|x_2 - x_1|_{\mathbf{X}} \tag{3.1}$$

holds for all $x_2, x_1 \in \overline{B}(x_0, r_0)$*. Then for all* $x_2, x_1 \in \overline{B}(x_0, r_0)$

$$\frac{(1-\rho)}{\|L^{-1}\|}|x_2 - x_1|_{\mathbf{X}} \leq |f(x_2) - f(x_1)|_{\mathbf{Y}} \leq \|L\|(1+\rho)|x_2 - x_1|_{\mathbf{X}}. \tag{3.2}$$

This f *is injective on* $\overline{B}(x_0, r_0)$*. Also the restriction* $f\big|_{B(x_0,r_0)}\colon B(x_0, r_0) \to \mathbf{Y}$ *of* f *to the open ball* $B(x_0, r_0)$ *is an open map so that* $V := f[B(x_0, r_0)]$

[1] "Lipeomorphism" gets my vote for the ugliest sounding term in mathematics.

is an open set in $\mathbf{Y}$*. Thus* $f\big|_{B(x_0,r_0)}\colon B(x_0,r_0)\to V$ *is a Lipeomorphism and* V *contains the open ball* $B(f(x_0),r_1)$ *where*

$$r_1 := \beta\rho, \quad \text{with} \quad \beta := \frac{(1-\rho)}{\|L^{-1}\|}. \tag{3.3}$$

Finally if $x_1\in B(x_0,r_0)$ *there is the inclusion*

$$B(f(x_1),\beta(r_0-|x_1-x_0|_{\mathbf{X}}))\subseteq V \tag{3.4}$$

(so that for all $y\in B(f(x_1),\beta(r_0-|x|_{\mathbf{X}}))$ *the equation* $f(x)=y$ *will have a solution in* $B(x_1,r_0-|x_1|_{\mathbf{X}})$*.*

Remark 3.2. Most proofs of the inverse function theorem show the existence of solutions to $f(x)=y$ for y close to $f(x_0)$ by using the Banach fixed point theorem to show that for some linear L the map $\Phi_y(x)=x-L^{-1}(f(x)-y)$ has a fixed point (which is then easily see to be also be a solution to $f(x)=y$). (Usually $L=f'(x_0)$.) The condition (3.1) is just that Φ_y is a contraction and as such is a natural condition to assume. □

Proof. For $x_1,x_2\in\overline{B}(x_0,r_0)$ we can use (3.1) and estimate

$$\begin{aligned}|f(x_2)-f(x_1)|_{\mathbf{Y}} &\le \|L\|\|L^{-1}f(x_2)-L^{-1}f(x_1)|_{\mathbf{X}}\\ &\le \|L\|\Big(L^{-1}f(x_2)-L^{-1}f(x_1)-(x_2-x_1)|_{\mathbf{X}}+|x_2-x_1|_{\mathbf{X}}\Big)\\ &\le \|L\|(\rho|x_2-x_1|_{\mathbf{X}}+|x_2-x_1|_{\mathbf{X}})\\ &= \|L\|(1+\rho)|x_2-x_1|_{\mathbf{X}}\end{aligned}$$

which proves the upper bound of (3.2). To prove the lower bound again use (3.1)

$$\begin{aligned}|x_2-x_1|_{\mathbf{X}} &\le |L^{-1}f(x_2)-L^{-1}f(x_1)-(x_2-x_1)|_{\mathbf{X}}\\ &\qquad + |L^{-1}f(x_2)-L^{-1}f(x_1)|_{\mathbf{X}}\\ &\le \rho|x_2-x_1|_{\mathbf{X}}+\|L^{-1}\||f(x_2)-f(x_1)|_{\mathbf{Y}}.\end{aligned}$$

This implies

$$\frac{1-\rho}{\|L^{-1}\|}|x_2-x_1|_{\mathbf{X}}\le|f(x_2)-f(x_1)|_{\mathbf{Y}}$$

which completes the proof of (3.2).

For $y\in\mathbf{Y}$ define $\Phi_y\colon\overline{B}(x_0,r_0)\to\mathbf{X}$ by

$$\Phi_y(x):=x-L^{-1}(f(x)-y).$$

Then

$$\Phi_y(x)=x \quad \text{if and only if} \quad f(x)=y. \tag{3.5}$$

The condition (3.1) implies that for $x_1,x_2\in\overline{B}(x_0,r_0)$

$$|\varphi_y(x_2)-\Phi_y(x_1)|_{\mathbf{X}}=|L^{-1}f(x_2)-L^{-1}f(x_1)-(x_2-x_1)|_{\mathbf{X}}\le\rho|x_2-x_1|_{\mathbf{X}}.$$

Therefore Φ_y is a contraction. If β and r_1 are defined by (3.3) for $x \in \overline{B}(x_0, r_0)$ and $y \in \overline{B}(f(x_0), r_1)$ then $|x - x_0|_{\mathbf{X}} \le r_0$ and $|y - f(x_0)|_{\mathbf{Y}} \le r_1 = \beta r_0 = \frac{(1-\rho)}{\|L^{-1}\|} r_0$. Thus for $x \in \overline{B}(x_0, r_0)$ and $y \in \overline{B}(f(x_0), r_1)$

$$\begin{aligned} |x_0 - \Phi_y(x)|_{\mathbf{X}} &= |x_0 - x - L^{-1}(f(x) - y)|_{\mathbf{X}} \\ &\le |x_0 - x - L^{-1}(f(x) - f(x_0))|_{\mathbf{X}} + |L^{-1}(y - f(x_0))\| \\ &\le \rho|x - x_0|_{\mathbf{X}} + \|L^{-1}\||y - f(x_0)|_{\mathbf{Y}} \\ &\le \rho r_0 + \|L^{-1}\| \frac{(1-\rho)}{\|L^{-1}\|} r_0 \\ &= r_0 \end{aligned}$$

Therefore if $y \in \overline{B}(f(x_0), r_1)$ then $\Phi_y \colon \overline{B}(x_0, r_0) \to \overline{B}(x_0, r_0)$. As Φ_y is a contraction the Banach fixed point theorem (Theorem A.5 in the appendix) Φ_y will have a unique fixed point in $\overline{B}(x_0, r_0)$. By the equivalence (3.5) this implies that for $y \in \overline{B}(f(x_0), r_1)$ there is a unique $x \in \overline{B}(x_0, r_0)$ with $f(x) = y$. This shows $\overline{B}(f(x_0), r_1) \subset f[\overline{B}(x_0, r_0)]$. But if $|f(x) - f(x_0)| < r_1$ then (using the definition of β and (3.2) we have $\beta|x - x_0|_{\mathbf{X}} \le |f(x) - f(x_0)|_{\mathbf{Y}} < r_1$ so that $|x - x_0|_{\mathbf{X}} < r_1/\beta = r_0$. Thus shows that if $y \in B(f(x_0), r_1)$, $x \in \overline{B}(x_0, r_0)$, and $f(x) = y$ then $x \in B(x_0, r_0)$. Therefore $f[B(x_0, r_0)] \subseteq B(f(x_0), r_1)$.

Now for any $x_0 \in B(x_0, r_0)$ we can let $\tilde{x}_0 := x_1$ and $\tilde{r}_0 := r_0 - |x_1 - x_0|_{\mathbf{X}}$ (so that $\overline{B}(\tilde{x}_0, \tilde{r}_0) \subset \overline{B}(x_0, r_0)$). Then the argument we have just given (with x_0 replaced by $\tilde{x}_0$, r_0 replaced by $\tilde{r}_0$ and f by $f\big|_{\overline{B}(\tilde{x}_0, \tilde{r}_0)}$) shows that

$$B(f(f(x_1), \beta(r_0 - |x_1 - x_0|_{\mathbf{X}})) \subseteq f[B(x_1, r_0 - |x_1 - x_0|_{\mathbf{X}})] \subseteq V$$

where $V = f[B(x_0, r_0)]$ as above. This proves (3.4). If $y \in V$ then there is an $x \in B(x_0, r_0)$ with $f(x) = y$. But then $B(y, \beta(r_0 - |x_1 - x_0|_{\mathbf{X}})) \subset V$. Therefore V contains an open ball about any of its points and so V is open.

As $f\big|_{B(x_0,r_0)}$ is injective it has an inverse $\varphi := f\big|_{B(x_0,r_0)} \colon B(x_0, r_0)$. By (3.2) this will satisfy

$$\frac{1}{\|L\|(1+\rho)} |y_2 - y_1|_{\mathbf{Y}} \le |\varphi(y_2) - \varphi(y_1)|_{\mathbf{X}} \le \frac{\|L^{-1}\|}{(1-\rho)} |y_2 - y_1|_{\mathbf{Y}}.$$

Exercise 3.3. Prove these inequalities from the inequalities (3.2). □

Thus both $f\big|_{B(x_0,r_0)}$ and its inverse φ are Lipschitz and so $f\big|_{B(x_0,r_0)}$ is a Lipeomorphism as claimed.

Finally to show that $f\big|_{B(x_0,r_0)}$ is an open map note that in $U \subseteq B(x_0, r_0)$ then $f[U] = \varphi^{-1}[U]$ and as φ is continuous this implies $f[U]$ is open. □

4. The Inverse Function Theorem

Let $\mathbf{X}$, $\mathbf{Y}$ be Banach spaces $U \subset \mathbf{X}$, $V \subset \mathbf{Y}$ be open sets. Then a map $f \colon U \to V$ is a ***diffeomorphism*** if and only if f is continuously differentiable and there is a continuously differentiable map $\varphi \colon V \to U$ so that $\varphi \circ f = \mathrm{Id}_U$ and $f \circ \varphi = \mathrm{Id}_V$. This implies that φ is the inverse of f (that is $\varphi = f^{-1}$). In particular this means that $f\colon U \to V$ bijective. Thus a diffeomorphism is just a continuously differentiable bijection between open sets so that the inverse is also continuously differentiable.

Remark 4.1. A continuously differentiable function f which is injective need not be a diffeomorphism. As an example let $f\colon \mathbf{R} \to \mathbf{R}$ be $f(t) = t^3$. This is smooth and bijective, but the inverse $f^{-1}(s) = \sqrt[3]{s}$ is not differentiable at the origin. $\square$

Theorem 4.2 (Inverse Function Theorem)**.** *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be Banach spaces,* $U \subset \mathbf{X}$ *be open and* $f\colon U \to \mathbf{X}$ *be continuously differentiable. Let* $x_0 \in U$ *be a point where the linear map* $f'(x_0)\colon \mathbf{X} \to \mathbf{Y}$ *is invertible. Then there is an* $r_0 > 0$ *so that if* $V := f[B(x_0, r_0)]$ *is the image of the open ball of radius* r_0 *about* x_0 *then* V *is an open set in* $\mathbf{Y}$ *and* $f\big|_{B(x_0,r_0)}\colon B(x_0, r_0) \to V$ *is a diffeomorphism.*

Remark 4.3. Sometimes it is useful to have explicit estimates for the size of r_0 and the neighborhood V. We state one version of such estimates. (Most of the work in getting this bounds was done in Theorem 3.1.) By hypothesis $x \mapsto f'(x)$ is a continuous map and by Proposition 2.5 the map $x \mapsto f'(x)^{-1}$ is continuous. Thus $x \mapsto (f'(x_0)^{-1}f'(x) - I_{\mathbf{X}})$ is continuous in a neighborhood of x_0 and this map vanishes at $x = x_0$. Now fix ρ with $0 < \rho < 1$, then by continuity there is an $r_0 > 0$ so that $\overline{B}(x_0, r_0) \subseteq U$ and

$$x \in \overline{B}(x_0, r_0) \quad \text{implies} \quad \|f'(x_0)^{-1}f'(x) - I_{\mathbf{X}}\| \le \rho. \tag{4.1}$$

This r_0 will work in the theorem so that if $f\big|_{B(x_0,r_0)}\colon B(x_0, r_0) \to V$ is a diffeomorphism with $V := f[B(x_0, r_0)]$. Let

$$\beta := \frac{(1-\rho)}{\|f'(x_0)^{-1}\|}. \tag{4.2}$$

Then we have the lower bound

$$B(f(x_0), \beta r_0) \subseteq V$$

for the size of V. More generally if $x_1 \in B(x_0, r_0)$ and $\tilde{r} := r_0 - |x - x_0|_{\mathbf{X}}$ then

$$B(f(x_0), \beta\tilde{r}) \subseteq V.$$

Finally in this ball f will satisfy the estimates

$$\beta|x_2 - x_1|_{\mathbf{X}} \le |f(x_2) - f(x_1)|_{\mathbf{Y}} \le \|f'(x_0)\|(1+\rho)|x_2 - x_1|_{\mathbf{X}}. \qquad \square$$

Proof. Fix ρ with $0 < \rho < 1$ and and to simplify set $L := f'(x_0)$. Then as in Remark 4.3 there is an r_0 so that

$$x \in \overline{B}(x_0, r_0) \quad \text{implies} \quad \|L^{-1}f'(x) - I_{\mathbf{X}}\| \le \rho. \tag{4.3}$$

Let $F\colon U \to \mathbf{Y}$ be given by

$$F(x) = L^{-1}f(x) - x.$$

Then the derivative of F is

$$F'(x) = L^{-1}f'(x) - I_{\mathbf{X}}.$$

Thus for $x \in \overline{B}(x_0, r_0)$

$$\|F'(x)\| \le \rho.$$

Now by Proposition 2.13 for $x_1, x_2 \in B(x_0, r_0)$

$$\begin{aligned} |L^{-1}f(x_2) - L^{-1}f(x_1) - (x_2 - x_1)|_{\mathbf{X}} &= |F(x_2) - F(x_1)|_{\mathbf{X}} \\ &\le \rho|x_2 - x_1|_{\mathbf{X}}. \end{aligned} \tag{4.4}$$

By continuity with will also hold for all $x_1, x_2 \in \overline{B}(x_0, r_0)$. Therefore $f\big|_{\overline{B}(x_0,r_0)} : \overline{B}(x_0, r_0) \to \mathbf{Y}$ satisfies the hypothesis of Theorem 3.1 and so we can use the conclusions of that theorem. In particular the image $V := f[B(r_0, r_0)]$ is open and if β is given by (4.2) then for all $x_1, x_2 \in B(x_0, r_0)$

$$\beta|x_2 - x_1|_{\mathbf{X}} \le |f(x_2) - f(x_1)|_{\mathbf{Y}} \le \|L\|(1+\rho)|x_2 - x_1|_{\mathbf{X}}. \tag{4.5}$$

Let $\varphi := f\big|_{B(x_0,r_0)}^{-1} \colon V \to B(x_0, r_0)$ be the inverse of $f\big|_{B(x_0,r_0)}^{-1}$. We now break the rest of the proof up into small steps leaving most of the details to the reader.

Step 1. The map $\varphi\colon V \to B(x_0, r_0)$ satisfies

$$\frac{1}{\|L\|(1+\rho)}|y_2 - y_1|_{\mathbf{Y}} \le |\varphi(y_2) - \varphi(y_1)|_{\mathbf{X}} \le \frac{1}{\beta}|y_2 - y_1|_{\mathbf{Y}}$$

holds for all $y_1, y_2 \in V$. Thus φ is continuous.

Exercise 4.4. Prove this. HINT: Use the inequalities (4.5) □

Step 2. At each point $x \in B(x_0, r_0)$ the derivative $f'(x)\colon \mathbf{X} \to \mathbf{Y}$ is an invertible linear map and

$$\|f'(x)\| \le \frac{\|L\|}{1-\rho}.$$

Exercise 4.5. Prove this. HINT: Write $f'(x) = LL^{-1}f'(x)$. Then $\|I_{\mathbf{X}} - L^{-1}f'(x)\| \le \rho$ by (4.3) and use Proposition 2.4 to conclude that $L^{-1}f'(x)$ is invertible and $\|L^{-1}f'(x)\| \le 1/(1-\rho)$. Then $f'(x)$ is a product of invertible maps. □

Step 3. Let $b \in V$. Then φ is differentiable at b with

$$\varphi'(b) = f'(\varphi(b))^{-1}.$$

Exercise 4.6. Prove this by justifying the following: Let $a = \varphi(b)$ so that $b = f(a)$. As f is differentiable

$$f(x) - f(a) = f'(a)(x - a) + |x - a|_{\mathbf{X}}\varepsilon(x; a)$$

where $\lim_{x\to a} |\varepsilon(x;a)|_{\mathbf{Y}} = 0$. Letting $x = \varphi(y)$ and $a = \varphi(b)$ gives

$$\begin{aligned} y - b &= f(\varphi(y)) - f(\varphi(b)) \\ &= f'(\varphi(b))(\varphi(y) - \varphi(b)) + |\varphi(y) - \varphi(b)|_{\mathbf{X}} \varepsilon(\varphi(y);\varphi(b)). \end{aligned}$$

By step 4 $f'(\varphi(b))$ exists and $\|f'(\varphi(b))^{-1}\| \le \|L\|/(1-\rho)$. So solve for $\varphi(y) - \varphi(b)$ in the last displayed equation to get

$$\begin{aligned} \varphi(y) - \varphi(b) &= f'(\varphi(b))^{-1}(y-b) - |\varphi(y) - \varphi(b)|_{\mathbf{X}} f'(\varphi(b))^{-1}\varepsilon(\varphi(y);\varphi(b)) \\ &= f'(\varphi(b))^{-1}(y-b) \\ &\qquad - |y-b|_{\mathbf{X}} \frac{|\varphi(y) - \varphi(b)|_{\mathbf{X}}}{|y-b|_{\mathbf{X}}} f'(\varphi(b))^{-1}\varepsilon(\varphi(y);\varphi(b)) \\ &= f'(\varphi(b))^{-1}(y-b) - |y-b|_{\mathbf{X}}\varepsilon_1(y;b) \end{aligned}$$

where this defines $\varepsilon_1(x;b)$. But (Step 1)

$$\frac{|\varphi(y) - \varphi(b)|_{\mathbf{X}}}{|y-b|_{\mathbf{Y}}} \le \frac{1}{\beta},$$

so that

$$|\varepsilon_1(y,b)|_{\mathbf{X}} \le \frac{\|f'(\varphi(b))\|}{\beta} |\varepsilon(\varphi(y);\varphi(b))|_{\mathbf{X}}.$$

But φ is continuous and so $\lim_{y\to b} |\varepsilon_1(y;b)|_{\mathbf{X}} = 0$. □

Step 4. The map $y \mapsto \varphi'(y)$ is continuous on the set V.

Exercise 4.7. Prove this. HINT: As $\varphi'(y) = f'(\varphi(y))^{-1}$ the map $y \mapsto \varphi'(y)$ is a composition of the maps $y \mapsto \varphi(y)$, $x \mapsto f'(x)$ and $A \mapsto A^{-1}$ each of which is continuous either by Step 1, the assumption that f is continuously differentiable, or Proposition 2.4. □

Putting the above together completes the proof of the inverse function theorem. □

Exercise 4.8. Prove the statements given in Remark 4.3. HINT: In the proof of Theorem 4.2 the bound (4.4) which allows the results of Theorem 3.1 to be used. □

5. Implicit Function Theorem: Surjective From

5.1. **Preliminary results on surjective linear maps.** Let $\mathbf{X}$, $\mathbf{Y}$ be Banach spaces and $L\colon \mathbf{X} \to \mathbf{Y}$ a bounded linear map. Then L has a ***right inverse*** iff there there is a bounded linear map $S\colon \mathbf{Y} \to \mathbf{X}$ so that $LS = I_{\mathbf{Y}}$. This clearly implies that L is surjective (as if $y \in \mathbf{Y}$ then $y = Lx$ with $x = Sy$). For linear maps L between finite dimensional spaces it is standard and elementary that L is surjective if and only if it has a right inverse. However there are examples of surjective linear maps $L\colon \mathbf{X} \to \mathbf{Y}$ between Banach spaces that do not have right inverses. We do have the following.

Proposition 5.1. *Let $L\colon \mathbf{X} \to \mathbf{Y}$ be a Bounded surjective linear map between Banach spaces. Then L has a right inverse in the following cases:*

1. $\mathbf{Y}$ *is finite dimensional or*
2. $\mathbf{X}$ *is a Hilbert space (that is an inner product space so that the induced norm is complete).*

Exercise 5.2. Prove this. HINT: For the first part let $y_1, \ldots, y_n$ be a basis of $\mathbf{Y}$ and let $x_i \in \mathbf{X}$ so that $Lx_i = y_i$. Then define S to be the unique linear map $S\colon \mathbf{Y} \to X$ so that $Sy_i = x_i$. The second part is unfair unless you know some functional analysis, but here is the idea. Let $\ker L$ be the kernel (also called the null space) of L and let $X_1 := (\ker L)^\perp$ be the orthogonal complement of $\ker L$. Then $L\big|_{\mathbf{X}_1}\colon \mathbf{X} \to \mathbf{Y}$ is bijective and therefore $L\big|_{\mathbf{X}_1}^{-1}\colon \mathbf{Y} \to \mathbf{X}_1$ is a well defined linear map. But then the open mapping theorem implies that $L\big|_{\mathbf{X}_1}^{-1}$ is bounded. Thus $S := L\big|_{\mathbf{X}_1}^{-1}$ gives the desired right inverse. □

Definition 5.3. Let $\mathbf{X}$ be a Banach space and let $\mathbf{X}_0$ and $\mathbf{X}_1$ be a closed linear subspaces of X. Then X is a ***direct sum*** of $\mathbf{X}_0$ and $\mathbf{X}_1$, written as $X = \mathbf{X}_0 \oplus \mathbf{X}_1$, iff $\mathbf{X}_0 \cap \mathbf{X}_1 = \{0\}$ and there is a constant C so that every $x \in \mathbf{X}$ can be written as $x = x_0 + x_1$ with $x_0 \in \mathbf{X}_0$, $x_1 \in \mathbf{X}_1$ and $|x_0|_{\mathbf{X}}, |x_1|_{\mathbf{X}} \le C|x|_{\mathbf{X}}$. □

Definition 5.4. Let X be a Banach space and $\mathbf{X}_0$ a closed linear subspace of $\mathbf{X}$. Then $\mathbf{X}_0$ is ***complimented*** iff there is a closed linear subspace $\mathbf{X}_1$ so that $\mathbf{X}$ is a direct sum $\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_1$. □

Exercise 5.5 (For those who know the open mapping theorem). Let $\mathbf{X}$ be a Banach space and $\mathbf{X}_0$, $\mathbf{X}_1$ closed subspaces of $\mathbf{X}$ such that $\mathbf{X}_0 \cap \mathbf{X}_1 = \{0\}$ and so that for each $x \in X$ there are $x_0 \in \mathbf{X}_0$ and $x_1 \in \mathbf{X}_1$ such that $x = x_0 + x_1$. Then there is a constant $C > 0$ so that if $x = x_0 + x_1$ with $x_0 \in \mathbf{X}_0$ and $x_1 \in \mathbf{X}_1$ then $|x_0|_{\mathbf{X}}, |x_1|_{\mathbf{X}} \le C|x|_{\mathbf{X}}$. HINT: Let $\mathbf{Y} := \mathbf{X}_0 \times \mathbf{X}_1$ and give $\mathbf{Y}$ the norm $|(x_0, x_1)|_{\mathbf{Y}} := |x_0|_{\mathbf{X}} + |x_1|_{\mathbf{X}}$. This norm makes $\mathbf{Y}$ into a Banach space. Define a linear map $T\colon \mathbf{Y} \to \mathbf{X}$ by $T(x_0, x_1) = x_0 + x_1$. This in linear and bijective. Moreover $|T(x_0, x_1)|_{\mathbf{X}} = |x_0 + x_1|_{\mathbf{X}} \le |x_0|_{\mathbf{X}} + |x_1|_{\mathbf{X}} = |(x_0, x_1)|_{\mathbf{Y}}$. Thus T is bounded with $\|T\| = 1$. Therefore by the open mapping theorem the linear map $T^{-1}\colon \mathbf{X} \to \mathbf{Y}$ is bounded. If $x = x_0 + x_1$ with $x_0 \in \mathbf{X}_0$ and $x_1 \in \mathbf{X}_1$ then $T(x_0, x_1) = x$ and so $(x_0 + x_1) = T^{-1}x$. Thus $|x_0|_{\mathbf{X}}, |x_1|_{\mathbf{X}} \le |x_0|_{\mathbf{X}} + |x_1|_{\mathbf{X}} = |(x_0, x_1)|_{\mathbf{Y}} \le \|T^{-1}\||x|_{\mathbf{X}}$. Therefore we can use $C = \|T^{-1}\|$. □

Proposition 5.6. *Let $L\colon \mathbf{X} \to \mathbf{Y}$ be a bounded linear map between Banach spaces and assume that L has a right inverse $S\colon \mathbf{Y} \to \mathbf{X}$. Let $P := SL\colon \mathbf{X} \to \mathbf{X}$. Then*

$$P^2 = P$$

and if

$$X_0 := \ker P = \ker L, \quad \mathbf{X}_1 = \ker(I_{\mathbf{X}} - P)$$

then $\mathbf{X}$ *is a direct sum* $\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_1$*, and* $L_1 := L\big|_{\mathbf{X}_1} \colon \mathbf{X}_1 \to \mathbf{Y}$ *in an bounded invertible linear map with inverse* $S \colon \mathbf{Y} \to \mathbf{X}_1 \subseteq \mathbf{X}$.

Exercise 5.7. Prove this by justifying the following. First $P^2 = SLSL = SI_{\mathbf{Y}}L = SL = P$. Set

$$P_1 := P = SL, \quad P_0 = I_{\mathbf{X}} - P.$$

Then P_0 and P_1 are bounded linear maps $P_i \colon \mathbf{X} \to \mathbf{X}$ that satisfy

$$P_0 + P_1 = I_{\mathbf{X}}, \quad P_0^2 = P_0, \quad P_1^2 = P_1, \quad P_0P_1 = P_1P_0 = 0.$$

These relations and the definition of $\mathbf{X}_0$ and $\mathbf{X}_1$ then imply

$$\mathbf{X}_0 = \operatorname{image} P_0 = \ker P_1, \quad \mathbf{X}_1 = \operatorname{image} P_1 = \ker P_0.$$

(Here $\operatorname{image} P_i = P_i[\mathbf{X}] = \{P_i x : x \in \mathbf{X}\}$ is the image of P_i.) Therefore if $x \in \mathbf{X}$ we have $x = P_1x + P_0x$ where $x_0 = P_1x \in \mathbf{X}_0$ and $x_1 = P_0x \in \mathbf{X}_1$ and $|P_ix|_{\mathbf{X}} \leq C|x|_{\mathbf{X}}$ where $C := \max\{\|P_0\|, \|P_1\|\}$. Also $\mathbf{X}_0 \cap \mathbf{X}_1 = \{0\}$ and so $\mathbf{X} = \mathbf{X}_0 \oplus \mathbf{X}_1$.

Let $y \in \mathbf{Y}$. Then $Sy = SI_{\mathbf{Y}}y = SLSy = Py = P_1y \in \mathbf{X}_1$. Therefore $\operatorname{image} S \subseteq \mathbf{X}_1$. But if $x \in X_1$ then $x = P_1x = Px = SLx \in \operatorname{image} S$. Thus $\operatorname{image} S = X_1$. Also if $x \in X_1$ then $SLx = Px = P_1x = x$ so $SL = I_{\mathbf{X}_1}$. Therefore S is the inverse of $L\big|_{\mathbf{X}_1}$. □

5.2. **The surjective form of the implicit function theorem for Lipschitz maps.**

A. Appendix: Contraction Mappings and the Banach Fixed Point Theorem

In this appendix we give the statement and proof of the Banach Fixed Point Theorem (also called the contraction mapping principle).

A.1. **Some Review.** There are several series that will come up several times and so we show know how to sum them. The first is the geometric series. This is the series

$$\sum_{k=0}^{\infty} ar^k = a + ar + ar^2 + ar^3 + \cdots = \frac{a}{1-r} = \frac{\text{first term}}{1 - \text{ratio}}$$

which converges when the ratio satisfies $|r| < 1$. (This is true for real or complex r.) Thus will often show up as follows:

Proposition A.1. *Let* $\sum_{k=0}^{\infty} a_k$ *be a series or real or complex numbers so that for some* $C > 0$ *and* $r < 1$ *the bound* $|a_k| \leq Ar^k$ *holds. Then* $\sum_{k=0}^{\infty} a_k$ *converges and the sum satisfies* $|\sum_{k=0}^{\infty} a_k| \leq C/(1-r)$. □

It is also good to have and estimate of the error when we only use n terms of the sum. Here is some practice for you:

Exercise A.2. Let $A = \sum_{k=0}^{\infty} a_k$ and $A_n = \sum_{k=0}^{n} a_k$ and assume that the bound $|a_k| \leq Cr^k$ holds for all k and some $r < 1$. Then show

$$|A - A_n| \leq \frac{Cr^{n+1}}{1-r}.$$

A.2. **Metric spaces.** Certain metric spaces will be usefully to us (mostly metric spaces, in fact Banach spaces, of functions. Recall that a ***metric space*** (X, d) is a nonempty set X with a function $d\colon X \times X \to [0, \infty)$ so that

1. $d(x, y) \geq 0$ and $d(x, y) = 0$ iff $x = y$,
2. $d(x, y) = d(y, x)$, (d is symmetric),
3. $d(x, z) \leq d(x, y) + d(y, z)$, (the triangle inequality).

The metric space (X, d) is ***complete*** iff any Cauchy sequence $\{x_k\}_{k=1}^{\infty}$ in X converges. Given a sequence $\{x_k\}_{k=1}^{\infty}$ in a metric space, then by repeated use of the triangle inequality for $m < n$ we can estimate

$$d(x_m, x_n) \leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-2}, x_{n-1}) + d(x_{n-1}, x_n).$$

We can combine this with some of the results above on series to get a dumb looking, but still useful, criterion for a series to be a Cauchy sequence.

Exercise A.3. Let (X, d) be a complete metric space and let $\{x_k\}_{k=0}^{\infty}$ be a sequence in X and assume that for some $C > 0$ and $r < 1$ that $d(x_k, x_{k+1}) \leq Cr^k$. Then show

1. $\{x_k\}_{k=0}^{\infty}$ is a Cauchy sequence and therefore converges to some point x_* of X.
2. For each n the estimate $d(x_n, x_*) \leq Cr^n/(1 - r)$ holds. □

A.3. **The Banach Fixed Point Theorem.** We now give what is one of the more important existence theorems for general non-linear equations. Let (X, d) be a metric space (say that (X, d) is the plane $\mathbf{R}^2$ 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 $X = \mathbf{R}^2$ 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.)

Let (X, d) be a metric space. Then a map $f\colon X \to X$ is a ***contraction*** iff there is a constant $\rho < 1$ so that $d(f(x), f(y)) \leq \rho d(x, y)$. The number ρ is called the ***contraction 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 $g(y_0) = y_0$.

Exercise A.4. 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 $d(a, b) = d(f(a), f(b)) \le \rho d(a, b)$. □

Theorem A.5 (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

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

Remark A.6. 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. □

Exercise A.7. Prove the theorem by doing the following:

1. For the sequence defined above show that $d(x_k, x_{k+1}) \le d(x_0, x_1)\rho^n$.
2. Use exercise A.3 to conclude that $\{x_k\}_{k=0}^{\infty}$ converges to some point x_* of X, and that the estimate (A.1) holds.
3. Show that x_* is a fixed point of f. HINT: Take the limit as $k \to \infty$ of $x_{n+1} = f(x_n)$.
4. Show that x_* is the only fixed point of f. □