Math 118, Spring 2,001

Shlomo Sternberg

January 11, 2001

Contents

Chapter 5

The contraction fixed point theorem

5.1 Metric spaces

Until now we have used the notion of metric quite informally. It is time for a formal definition. For any set X, we let $X \times X$ (called the Cartesian product of X with itself) denote the set of all ordered pairs of elements of X. (More generally, if X and Y are sets, we let $X \times Y$ denote the set of all pairs (x, y) with $x \in$ and $y \in Y$, and is called the Cartesian product of X with Y.)

A **metric** for a set X is a function d from X to the real numbers $\mathbf{R}$,

$$d : X \times X \to \mathbf{R}$$

such that for all $x, y, z \in X$

1. $d(x, y) = d(y, x)$
2. $d(x, z) \leq d(x, y) + d(y, z)$
3. $d(x, x) = 0$
4. If $d(x, y) = 0$ then $x = y$.

The inequality in 2) is known as the **triangle inequality** since if X is the plane and d the usual notion of distance, it says that the length of an edge of a triangle is at most the sum of the lengths of the two other edges. (In the plane, the inequality is strict unless the three points lie on a line.)

Condition 4) is in many ways inessential, and it is often convenient to drop it, especially for the purposes of some proofs. For example, we might want to consider the decimal expansions $.49999\ldots$ and $.50000\ldots$ as different, but as having zero distance from one another. Or we might want to "identify" these two decimal expansions as representing the same point.

A function d which satisfies only conditions 1) - 3) is called a **pseudo-metric**.

A **metric space** is a pair (X, d) where X is a set and d is a metric on X. Almost always, when d is understood, we engage in the abuse of language and speak of "the metric space X".

Similarly for the notion of a **pseudo-metric space**.

In like fashion, we call $d(x, y)$ the **distance** between x and y, the function d being understood.

If r is a positive number and $x \in X$, the (open) **ball of radius** r about x is defined to be the set of points at distance less than r from x and is denoted by $B_r(x)$. In symbols,

$$B_r(x) := \{y|\ d(x, y) < r\}.$$

If r and s are positive real numbers and if x and z are points of a pseudo-metric space X, it is possible that $B_r(x) \cap B_s(z) = \emptyset$. This will certainly be the case if $d(x, z) > r + s$ by virtue of the triangle inequality. Suppose that this intersection is not empty and that

$$w \in B_r(x) \cap B_s(z).$$

If $y \in X$ is such that $d(y, w) < \min[r - d(x, w), s - d(z, w)]$ then the triangle inequality implies that $y \in B_r(x) \cap B_s(z)$. Put another way, if we set $t := \min[r - d(x, w), s - d(z, w)]$ then

$$B_t(w) \subset B_r(x) \cap B_s(z).$$

Put still another way, this says that the intersection of two (open) balls is either empty or is a union of open balls. So if we call a set in X **open** if either it is empty, or is a union of open balls, we conclude that the intersection of any finite number of open sets is open, as is the union of any number of open sets. In technical language, we say that the open balls form a base for a topology on X.

A map $f : X \to Y$ from one pseudo-metric space to another is called **continuous** if the inverse image under f of any open set in Y is an open set in X. Since an open set is a union of balls, this amounts to the condition that the inverse image of an open ball in Y is a union of open balls in X, or, to use the familiar ϵ, δ language, that if $f(x) = y$ then for every $\epsilon > 0$ there exists a $\delta = \delta(x, \epsilon) > 0$ such that

$$f(B_\delta(x)) \subset B_\epsilon(y).$$

Notice that in this definition δ is allowed to depend both on x and on ϵ. The map is called uniformly continuous if we can choose the δ independently of x.

An even stronger condition on a map from one pseudo-metric space to another is the **Lipschitz condition**. A map $f : X \to Y$ from a pseudo-metric space (X, d_X) to a pseudo-metric space (Y, d_Y) is called a **Lipschitz map** with **Lipschitz constant** C if

$$d_Y(f(x_1), f(x_2)) \le C d_X(x_1, x_2) \quad \forall x_1, x_2 \in X.$$

Clearly a Lipschitz map is uniformly continuous.

For example, suppose that A is a fixed subset of a pseudo-metric space X. Define the function $d(A, \cdot)$ from X to $\mathbf{R}$ by

$$d(A, x) := \inf\{d(x, w),\ w \in A\}.$$

The triangle inequality says that

$$d(x, w) \leq d(x, y) + d(y, w)$$

for all w, in particular for $w \in A$, and hence taking lower bounds we conclude that

$$d(A, x) \leq d(x, y) + d(A, y).$$

or

$$d(A, x) - d(A, y) \leq d(x, y).$$

Reversing the roles of x and y then gives

$$|d(A, x) - d(A, y)| \leq d(x, y).$$

Using the standard metric on the real numbers where the distance between a and b is $|a - b|$ this last inequality says that $d(A, \cdot)$ is a Lipschitz map from X to $\mathbf{R}$ with $C = 1$.

A closed set is defined to be a set whose complement is open. Since the inverse image of the complement of a set (under a map f) is the complement of the inverse image, we conclude that the inverse image of a closed set under a continuous map is again closed.

For example, the set consisting of a single point in $\mathbf{R}$ is closed. Since the map $d(A, \cdot)$ is continuous, we conclude that the set

$$\{x | d(A, x) = 0\}$$

consisting of all point at zero distance from A is a closed set. It clearly is a closed set which contains A. Suppose that S is some closed set containing A, and $y \notin S$. Then there is some $r > 0$ such that $B_r(y)$ is contained in the complement of C, which implies that $d(y, w) \geq r$ for all $w \in S$. Thus $\{x | d(A, x) = 0\} \subset S$. In short $\{x | d(A, x) = 0\}$ is a closed set containing A which is contained in all closed sets containing A. This is the definition of the **closure** of a set, which is denoted by $\overline{A}$. We have proved that

$$\overline{A} = \{x | d(A, x) = 0\}.$$

In particular, the closure of the one point set $\{x\}$ consists of all points u such that $d(u, x) = 0$.

Now the relation $d(x, y) = 0$ is an equivalence relation, call it R. (Transitivity being a consequence of the triangle inequality.) This then divides the space X into equivalence classes, where each equivalence class is of the form $\overline{\{x\}}$, the closure of a one point set. If $u \in \overline{\{x\}}$ and $v \in \overline{\{y\}}$ then

$$d(u, v) \leq d(u, x) + d(x, y) + d(y, v) = d(x, y).$$

since $x \in \overline{\{u\}}$ and $y \in \overline{\{v\}}$ we obtain the reverse inequality, and so

$$d(u,v) = d(x,y).$$

In other words, we may define the distance function on the quotient space X/R, i.e. on the space of equivalence classes by

$$d(\overline{\{x\}}, \overline{\{y\}}) := d(u,v), \quad u \in \overline{\{x\}}, v \in \overline{\{y\}}$$

and this does not depend on the choice of u and v. Axioms 1)-3) for a metric space continue to hold, but now

$$d(\overline{\{x\}}, \overline{\{y\}}) = 0 \Rightarrow \overline{\{x\}} = \overline{\{y\}}.$$

In other words, X/R is a *metric* space. Clearly the projection map $x \mapsto \overline{\{x\}}$ is an isometry of X onto X/R. (An isometry is a map which preserves distances.) In particular it is continuous. It is also open.

In short, we have provided a canonical way of passing (via an isometry) from a pseudo-metric space to a metric space by identifying points which are at zero distance from one another.

A subset A of a pseudo-metric space X is called *dense* if its closure is the whole space. From the above construction, the image A/R of A in the quotient space X/R is again dense. We will use this fact in the next section in the following form:

If $f : Y \to X$ is an isometry of Y such that $f(Y)$ is a dense set of X, then f descends to a map F of Y onto a dense set in the metric space X/R.

5.2 Completeness and completion.

The usual notion of convergence and Cauchy sequence go over unchanged to metric spaces or pseudo-metric spaces Y. A sequence $\{y_n\}$ is said to **converge** to the point y if for every $\epsilon > 0$ there exists an $N = N(\epsilon)$ such that

$$d(y_n, y) < \epsilon \quad \forall\, n > N.$$

A sequence $\{y_n\}$ is said to be **Cauchy** if for any $\epsilon > 0$ there exists an $N = N(\epsilon$ such that

$$d(y_n, y_m) < \epsilon \quad \forall\, m, n > N.$$

The triangle inequality implies that every convergent sequence is Cauchy. But not every Cauchy sequence is convergent. For example, we can have a sequence of rational numbers which converge to an irrational number, as in the approximation to the square root of 2. So if we look at the set of rational numbers as a metric space R in its own right, not every Cauchy sequence of rational numbers converges in R. We must "complete" the rational numbers to obtain $\mathbf{R}$, the set of real numbers. We want to discuss this phenomenon in general.

So we say that a (pseudo-)metric space is **complete** if every Cauchy sequence converges. The key result of this section is that we can always "complete" a metric or pseudo-metric space. More precisely, we claim that

Any metric (or pseudo-metric) space can be mapped by a one to one isometry onto a dense subset of a complete metric (or pseudo-metric) space.

By the italicized statement of the preceding section, it is enough to prove this for a pseudo-metric spaces X. Let X_{seq} denote the set of Cauchy sequences in X, and define the distance between the Cauchy sequences $\{x_n\}$ and $\{y_n\}$ to be

$$d(\{x_n\},\{y_n\}) := \lim_{n\to\infty} d(x_n,y_n).$$

It is easy to check that d defines a pseudo-metric on X_{seq}. Let $f: X \to X_{seq}$ be the map sending x to the sequence all of whose elements are x;

$$f(x) = (x,x,x,x,\cdots).$$

It is clear that f is one to one and is an isometry. The image is dense since by definition

$$\lim d(f(x_n),\{x_n\}) = 0.$$

Now since $f(X)$ is dense in X_{seq}, it suffices to show that any Cauchy sequence of points of the form $f(x_n)$ converges to a limit. But such a sequence converges to the element $\{x_n\}$. QED

Of special interest are vector spaces which have metric which is compatible with the vector space properties and which is complete: Let V be a vector space over the real numbers. A **norm** is a real valued function

$$v \mapsto \|v\|$$

on V which satisfies

1. $\|v\| \geq 0$ and > 0 if $v \neq 0$,
2. $\|rv\| = |r|\|v\|$ for any real number r, and
3. $\|v+w\| \leq \|v\| + \|w\| \ \forall \ \ v,w \in V$.

Then $d(v,w) := \|v-w\|$ is a metric on V, which satisfies $d(v+u,w+u) = d(v,w)$ for all $v,w,u \in V$. The ball of radius r about the origin is then the set of all v such that $\|v\| < r$. A vector space equipped with a norm is called a **normed vector space** and if it is complete relative to the metric it is called a **Banach space**.

5.3 The contraction fixed point theorem.

Let X and Y be metric spaces. Recall that a map $f : X \to Y$ is called a **Lipschitz map** or is said to be "Lipschitz continuous", if there is a constant C such that

$$d_Y(f(x_1), f(x_2)) \le C d_X(x_1, x_2), \quad \forall\, x_1, x_2 \in X.$$

If f is a Lipschitz map, we may take the greatest lower bound of the set of all C for which the previous inequality holds. The inequality will continue to hold for this value of C which is known as the Lipschitz constant of f and denoted by $\mathrm{Lip}(f)$.

A map $K : X \to Y$ is called a **contraction** if it is Lipschitz, and its Lipschitz constant satisfies $\mathrm{Lip}(K) < 1$. Suppose $K : X \to X$ is a contraction, and suppose that $Kx_1 = x_1$ and $Kx_2 = x_2$. Then

$$d(x_1, x_2) = d(Kx_1, Kx_2) \le \mathrm{Lip}(K) d(x_1, x_2)$$

which is only possible if $d(x_1, x_2) = 0$, i.e. $x_1 = x_2$. So a contraction can have at most one fixed point. The contraction fixed point theorem asserts that if the metric space X is complete (and non-empty) then such a fixed point exists.

Theorem 5.3.1 *Let X be a non-empty complete metric space and $K : X \to X$ a contraction. Then K has a unique fixed point.*

Proof. Choose any point $x_0 \in X$ and define

$$x_n := K^n x_0$$

so that

$$x_{n+1} = Kx_n, \quad x_n = Kx_{n-1}$$

and therefore

$$d(x_{n+1}, x_n) \le C d(x_n, x_{n-1}), \quad 0 \le C < 1$$

implying that

$$d(x_{n+1}, x_n) \le C^n d(x_1, x_0).$$

Thus for any $m > n$ we have

$$d(x_m, x_n) \le \sum_n^{m-1} d(x_{i+1}, x_i) \le \left(C^n + C^{n+1} + \cdots + C^{m-1}\right) d(x_1, x_0) \le C^n \frac{d(x_1, x_0)}{1 - C}.$$

This says that the sequence $\{x_n\}$ is Cauchy. Since X is complete, it must converge to a limit x, and $Kx = \lim Kx_n = \lim x_{n+1} = x$ so x is a fixed point. We already know that this fixed point is unique. QED

We often encounter mappings which are contractions only near a particular point p. If K does not move p too much we can still conclude the existence of a fixed point, as in the following:

Proposition 5.3.1 *Let D be a closed ball of radius r centered at a point p in a complete metric space X, and suppose $K : D \to X$ is a contraction with Lipschitz constant $C < 1$. Suppose that*

$$d(p, Kp) \leq (1 - C)r.$$

Then K has a unique fixed point in D.

Proof. We simply check that $K : D \to D$ and then apply the preceding theorem with X replaced by D: For any $x \in D$, we have

$$d(Kx, p) \leq d(Kx, Kp) + d(Kp, p) \leq Cd(x, p) + (1-C)r \leq Cr + (1-C)r = r \quad \text{QED.}$$

Proposition 5.3.2 *Let B be an open ball or radius r centered at p in a complete metric space X and let $K : B \to X$ be a contraction with Lipschitz constant $C < 1$. Suppose that*

$$d(p, Kp) < (1 - C)r.$$

Then K has a unique fixed point in B.

Proof. Restrict K to any slightly smaller closed ball centered at p and apply Prop. 5.3.1. QED

Proposition 5.3.3 *Let $K : X \to X$ be a contraction with Lipschitz constant C of a complete metric space. Let x be its (unique) fixed point. Then for any $y \in X$ we have*

$$d(y, x) \leq \frac{d(y, Ky)}{1 - C}.$$

Proof. We may take $x_0 = y$ and follow the proof of Theorem 5.3.1. Alternatively, we may apply Prop. 5.3.1 to the closed ball of radius $d(y, Ky)/(1 - C)$ centered at y. Prop. 5.3.1 implies that the fixed point lies in the ball of radius r centered at y. QED

Prop. 5.3.3 will be of use to us in proving continuous dependence on a parameter in the next section. In the section on iterative function systems for the construction of fractal images, Prop. 5.3.3 becomes the "collage theorem". We might call Prop. 5.3.3 the "abstract collage theorem".

5.4 Dependence on a parameter.

Suppose that the contraction "depends on a parameter s". More precisely, suppose that S is some other metric space and that

$$K : S \times X \to X$$

with

$$d_X(K(s, x_1), K(s, x_2)) \leq Cd_X(x_1, x_2), \quad 0 \leq C < 1, \ \forall s \in S, \ x_1, x_2 \in X. \tag{5.1}$$

(We are assuming that the C in this inequality does not depend on s.) If we hold $s \in S$ fixed, we get a contraction

$$K_s : X \to X, \quad K_s(x) := K(s,x).$$

This contraction has a unique fixed point, call it p_s. We thus obtain a map

$$S \to X, \quad s \mapsto p_s$$

sending each $s \in S$ into the fixed point of K_s.

Proposition 5.4.1 *Suppose that for each fixed $x \in X$, the map*

$$s \mapsto K(s,x)$$

of $S \to X$ is continuous. Then the map

$$s \mapsto p_s$$

is continuous.

Proof. Fix a $t \in S$ and an $\epsilon > 0$. We must find a $\delta > 0$ such that $d_X(p_s, p_t) < \epsilon$ if $d_S(s,t) < \delta$. Our continuity assumption says that we can find a $\delta > 0$ such that

$$d_X(K(s,p_t),p_t) = d_X(K(s,p_t),K(t,p_t) \leq (1-C)\epsilon$$

if $d_S(s,t) < \delta$. This says that K_s moves p_t a distance at most $(1-C)\epsilon$. But then the 'abstract collage theorem", Prop. 5.3.3, says that

$$d_X(p_t,p_s) \leq \epsilon. \quad \text{QED}$$

It is useful to combine Proposition 5.3.1 and 5.4.1 into a theorem:

Theorem 5.4.1 *Let B be an open ball of radius r centered at a point q in a complete metric space. Suppose that $K : S \times B \to X$ (where S is some other metric space) is continuous, satisfies* (5.1) *and*

$$d_X(K(s,q),q) < (1-C)r, \quad \forall\ s \in S.$$

Then for each $s \in S$ there is a unique $p_s \in B$ such that $K(s,p_s) = p_s$, and the map $s \mapsto p_s$ is continuous.

.

5.5 The Lipschitz implicit function theorem

In this section we follow the treatment in [?]. We begin with the inverse function theorem which contains the guts of the argument. We will consider a map $F : B_r(0) \to E$ where $B_r(0)$ is the open ball of radius r about the origin in a Banach space, E, and where $F(0) = 0$. We wish to conclude the existence of an inverse to F, defined on a possible smaller ball by means of the contraction fixed point theorem.

Proposition 5.5.1 *Let $F : B_r(0) \to E$ satisfy $F(0) = 0$ and*

$$\mathrm{Lip}[F - \mathrm{id}] = \lambda < 1. \tag{5.2}$$

Then the ball $B_s(0)$ is contained in the image of F where

$$s = (1 - \lambda)r \tag{5.3}$$

and F has an inverse, G defined on $B_s(0)$ with

$$\mathrm{Lip}[G - \mathrm{id}] \le \frac{\lambda}{1 - \lambda}. \tag{5.4}$$

Proof. Let us set $F = \mathrm{id} + v$ so

$$\mathrm{id} + v : B_r(0) \to E, \quad v(0) = 0, \quad \mathrm{Lip}[v] < \lambda < 1.$$

We want to find a $w : B_s(0) \to E$ with

$$w(0) = 0$$

and

$$(\mathrm{id} + v) \circ (\mathrm{id} + w) = \mathrm{id}.$$

This equation is the same as

$$w = -v \circ (\mathrm{id} + w).$$

Let X be the space of continuous maps of $\overline{B_s(0)} \to E$ satisfying

$$u(0) = 0$$

and

$$\mathrm{Lip}[u] \le \frac{\lambda}{1 - \lambda}.$$

Then X is a complete metric space relative to the sup norm, and, for $x \in \overline{B_s(0)}$ and $u \in X$ we have

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

Thus, if $u \in X$ then

$$u : \overline{B_s} \to \overline{B_r}.$$

If $w_1, w_2 \in X$,

$$\| -v \circ (\mathrm{id} + w_1) + v \circ (\mathrm{id} + w_2) \| \le \lambda \| (\mathrm{id} + w_1) - (\mathrm{id} + w_2) \| = \lambda \| w_1 - w_2 \| .$$

So the map $K : X \to X$

$$K(u) = -v \circ (\mathrm{id} + u)$$

is a contraction. Hence there is a unique fixed point. This proves the proposition.

Now let f be a homeomorphism from an open subset, U of a Banach space, E_1 to an open subset, V of a Banach space, E_2 whose inverse is a Lipschitz map. Suppose that $h : U \to E_2$ is a Lipschitz map satisfying

$$\mathrm{Lip}[h]\mathrm{Lip}[f^{-1}] < 1. \tag{5.5}$$

Let

$$g = f + h. \tag{5.6}$$

We claim that g is open. That is, we claim that if $y = g(x)$, then the image of a neighborhood of x under g contains a neighborhood of $g(x)$. Since f is a homeomorphism, it suffices to establish this for $g \circ f^{-1} = \mathrm{id} + h \circ f^{-1}$. Composing by translations if necessary, we may apply the proposition. QED

We now want to conclude that g is a homeomorphism = continuous with continuous inverse, and, in fact, is Lipschitz:

Proposition 5.5.2 *Let f and g be two continuous maps from a metric space X to a Banach space E. Suppose that f is injective and f^{-1} is injective with Lipschitz constant* $\mathrm{Lip}[f^{-1}]$. *Suppose that g satisfies*

$$\mathrm{Lip}[g - f] < \frac{1}{\mathrm{Lip}[f^{-1}]}.$$

Then g is injective and

$$\mathrm{Lip}[g^{-1}] \le \frac{1}{\frac{1}{\mathrm{Lip}[f^{-1}]} - \mathrm{Lip}[g - f]} = \frac{\mathrm{Lip}[f^{-1}]}{1 - \mathrm{Lip}[g - f]\mathrm{Lip}[f^{-1}]}. \tag{5.7}$$

Proof. By definition,

$$d(x, y) \le \mathrm{Lip}[f^{-1}]\|f(x) - f(y)\|$$

where d denotes the distance in X and $\| \ \|$ denotes the norm in E. We can write this as

$$\|f(x) - f(y)\| \ge \frac{d(x, y)}{\mathrm{Lip}[f^{-1}]}.$$

So

$$\begin{aligned} \|g(x) - g(y)\| &\ge \|f(x) - f(y)\| - \|(g - f)(x) - (g - f)(y)\| \\ &\ge \left(\frac{1}{\mathrm{Lip}[f^{-1}]} - \mathrm{Lip}[g - f]\right) d(x, y). \end{aligned}$$

Dividing by the expression in parenthesis gives the proposition. QED

We can now be a little more precise as to the range covered by g:

Proposition 5.5.3 *Let U be an open subset of a Banach space E_1, and g be a homeomorphism of U onto an open subset of a Banach space, E_2. Let $x \in U$ and suppose that*

$$\overline{B_r(x)} \subset U.$$

If g^{-1} is Lipschitz with

$$\mathrm{Lip}[g^{-1}] < c$$

then

$$\overline{B_{\frac{r}{c}}(g(x))} \subset g\left(\overline{B_r(x)}\right).$$

Proof. By composing with translations, we may assume that $x = 0,\ g(x) = 0$. Let

$$v \in B_{\frac{r}{c}}(0).$$

Let

$$T = T(v) = \sup\{t|[0,t]v \subset g\left(\overline{B_r(0)}\right)\}.$$

We wish to show that $T = 1$. Since $g(B_r(0))$ contains a neighborhood of 0, we know that $T(v) > 0$, and, by definition,

$$(0,T)v \subset g\left(\overline{B_r(0)}\right).$$

By the Lipschitz estimate for g^{-1} we have

$$\|g^{-1}(tv) - g^{-1}(sv)\| \leq c|t-s|\|v\|.$$

This implies that the limit $\lim_{t\to T} g^{-1}(tv)$ exists and

$$\lim_{t\to T} g^{-1}(tv) \in \overline{B_r(0)}.$$

So

$$Tv \in (g\left(\overline{B_r(0)}\right).$$

If $T = T(v) < 1$ we would have

$$\begin{aligned}\|g^{-1}(Tv)\| &\leq \|g^{-1}(Tv) - g^{-1}(0)\| \\ &\leq c\|Tv\| \\ &= cT\|v\| \\ &< c\|v\| \\ &\leq r.\end{aligned}$$

This says that

$$Tv \in g\left(B_r(0)\right).$$

But since g is open, we could find an $\epsilon > 0$ such that $[T, T+\epsilon]v \subset g\left(B_r(0)\right)$ contradicting the definition of T. QED

The above three propositions, taken together, constitute what we might call the inverse function theorem for Lipschitz maps. But the contraction fixed point theorem allows for continuous dependence on parameters, and gives the fixed point as a continuous function of the parameters. So this then yields the implicit function theorem.

The differentiability of the solution, in the case that the implicit function is assumed to be continuously differentiable follows as in Chapter 1.