v: 2019-01-17

REAL ANALYSIS LECTURE NOTES

RASUL SHAFIKOV

2. Inverse Function Theorem and Friends.

2.1. Inverse Function theorem.

Lemma 2.1 (Contraction Lemma). *Let (X, d) be a complete metric space, and $\phi : X \to X$ a contraction, i.e., a map satisfying for some $c < 1$,*

$$d(\phi(x), \phi(y)) \le c\, d(x, y), \quad x, y \in X.$$

Then there exists a unique fixed point p of ϕ, i.e., $p \in X$ such that $\phi(p) = p$.

Proof. Pick any $x_0 \in X$, and define $\{x_n\}$ inductively by setting $x_{n+1} = \phi(x_n)$, $n = 0, 1, \ldots$. Then for $n > 0$ we have

$$d(x_{n+1}, x_n) = d(\phi(x_n), \phi(x_{n-1})) \le c\, d(x_n, x_{n-1}).$$

This gives the following relation

$$d(x_{n+1}, x_n) \le c^n d(x_1, x_0), \quad n = 0, 1, 2, \ldots$$

If $n < m$, then

$$d(x_n, x_m) \le \sum_{i=n+1}^{m} d(x_i, x_{i-1}) \le (c^n + c^{n+1} \cdots + c^{m-1}) d(x_1, x_0) \le \frac{c^n}{1-c}\, d(x_1, x_0).$$

Thus, $\{x_n\}$ is a Cauchy sequences which converges to some point p by completeness of X. Since ϕ is a contractions, it is continuous, and $\phi(p) = \lim_{n\to\infty} \phi(x_n) = \lim_{n\to\infty} x_{n+1} = p$.

The uniqueness of p is trivial. □

Definition 2.2. *A map $f : \mathbb{R}^n \to \mathbb{R}^m$ is called Lipschitz continuous on $\Omega \subset \mathbb{R}^n$ if there is a constant $C > 0$ such that*

$$|f(x) - f(y)| \le C|x - y|, \quad x, y \in \Omega.$$

Such C is called a Lipschitz constant for f.

Lemma 2.3. *Let $\Omega \subset \mathbb{R}^n$ be a domain and $f : \Omega \to \mathbb{R}^m$ be a map of class $C^1(\Omega)$. Then f is Lipschitz continuous on any compact convex subset $B \subset \Omega$.*

Proof. Let $M = \sup_{x\in B} ||Df(x)||$. Let $a, b \in B$. Since B is convex, the straight line segment

$$\{x = a + t(b - a),\ t \in [0, 1]\}$$

connecting a and b is contained in B. By the Fundamental Theorem of Calculus and the Chain Rule we have for each component of f,

$$f_i(b) - f_i(a) = \int_0^1 \frac{d}{dt} f_i(a + t(b - a))dt = \int_0^1 Df_i(a + t(b - a))(b - a)dt.$$

Hence,

$$|f(b) - f(a)|^2 = \sum_{i=1}^{m} |f_i(b) - f_i(a)|^2 \le \sum_{i=1}^{m} \left(\int_0^1 |Df_i(a + t(b - a))|\ |b - a| dt \right)^2 \le n(M|b - a|)^2.$$

From this the assertion follows. □

A C^k-smooth map $f : \Omega \to \Omega'$ between open sets in $\mathbb{R}^n$ is called a (C^k-)*diffeomorphism* if $f^{-1} : \Omega' \to \Omega$ is well defined and C^k-smooth. In general, the inverse of a smooth map, if exists, is not necessarily smooth (but always continuous!). For example, the function $f(x) = x^3$ is C^∞-smooth on $\mathbb{R}$, and has a continuous inverse $f^{-1}(x) = \sqrt[3]{x}$. However, f^{-1} is not differentiable at the origin (note that $f'(0) = 0$). The situation is different if Df is invertible.

Theorem 2.4 (Inverse Function Theorem)**.** *Suppose $U, V \subset \mathbb{R}^n$ are open subsets, $f : U \to V$ is of class $C^k(\Omega)$ and $f'(p)$ is nonsingular (invertible) for some $p \in U$. Then there exist connected neighbourhoods $U_0 \subset U$ of p and $V_0 \subset V$ of $f(p)$ such that $f|_{U_0} : U_0 \to V_0$ is a C^k-diffeomorphism.*

Proof. We may replace f with $f_1(x) = f(x+p) - f(p)$. The map f_1 is smooth and satisfies $f_1(0) = 0$ and $Df(p) = Df_1(0)$. We may further replace f_1 with $f_2 = Df_1(0)^{-1} \circ f_1$. The map f_2 is smooth, $f_2(0) = 0$, and $Df_2(0) = Id$, the identity map. Hence, we may assume that f is defined in a neighbourhood U of the origin, $f(0) = 0$ and $Df(0) = Id$.

Set $h(x) = x - f(x)$. Then $Dh(0) = 0$, and so for any $\varepsilon > 0$ there exists $\delta > 0$ such that $||Dh(x)|| \le \varepsilon$ for $x \in B(0, \delta) = \{x \in \mathbb{R}^n : |x| \le \delta\}$. By Lemma 2.3 we may $\delta > 0$ such that

$$\big|h(x') - h(x)\big| \le \frac{1}{2}|x' - x|, \quad \forall x, x' \in B(0, \delta). \tag{1}$$

Then $|x' - x| \le |f(x') - f(x)| + |h(x') - h(x)| \le |f(x') - f(x)| + \frac{1}{2}|x - x'|$, and so

$$|x' - x| \le 2|f(x') - f(x)|, \quad x, x' \in B(0, \delta). \tag{2}$$

This shows, in particular, that f is injective on $B(0, \delta)$. For an arbitrary $y \in B(0, \delta/2)$ we show that there exists a unique $x \in B(0, \delta)$ such that $f(x) = y$. Let $g(x) = y + h(x) = y + x - f(x)$, so $g(x) = x$ if and only if $f(x) = y$. If $|x| \le \delta$, then

$$|g(x)| \le |y| + |h(x)| \le \frac{\delta}{2} + \frac{1}{2}|x| \le \delta, \tag{3}$$

so g maps $B(0, \delta)$ to itself. By (1), $|g(x) - g(x')| = |h(x) - h(x')| \le \frac{1}{2}|x - x'|$, hence g is a contraction, and by Lemma 2.1, g has a unique fixed point $x \in B(0, \delta)$. By (3), $|x| = |g(x)| < \delta$, so $x \in B(0, \delta)$ as claimed.

Let $U_1 = B(0, \delta) \cap f^{-1}(B(0, \delta/2))$. Then $U_1 \subset \mathbb{R}^n$ is open, and $f : U_1 \to B(0, \delta/2)$ is bijective, so f^{-1} exists. Estimate (2) shows that f^{-1} is continuous. Let U_0 be a connected component of U_1 containing the origin, and $V_0 = f(U_0)$. Then $f : U_0 \to V_0$ is a homeomorphism.

To show that $f : U_0 \to V_0$ is a diffeomorphism it remains to show that $f^{-1} \in C^1(V_0)$. Let $b = f(a)$ for some $a \in U_0$, $b \in V_0$, and set

$$R(v) = f(a + v) - f(a) - Df(a)v,$$

and

$$S(h) = f^{-1}(b + h) - f^{-1}(b) - Df(a)^{-1}h.$$

Let

$$v(h) = f^{-1}(b + h) + f^{-1}(b) = f^{-1}(b + h) - a.$$

Then $h = f(a + v(h)) - f(a)$, and so

$$S(h) = v(h) - Df(a)^{-1}h = Df(a)^{-1}\left[Df(a)v(h) + f(a) - f(a + v(h))\right] = -Df(a)^{-1}R(v(h)).$$

If there exist constants $C, c > 0$ such that

$$c|h| \le |v(h)| \le C|h|, \tag{4}$$

then

$$\frac{|S(h)|}{|h|} \leq ||Df(a)^{-1}|| \frac{|R(v(h))|}{|h|} \leq ||Df(a)^{-1}|| \frac{|R(v(h))|}{|v(h)|}\frac{|v(h)|}{|h|} \leq C||Df(a)^{-1}|| \frac{|R(v(h))|}{|v(h)|}.$$

The expression on the right converges to zero as $h \to 0$ by differentiability of f. This proves that f^{-1} is differentiable at b. It remains to show (4). We have

$$v(h) = Df(a)^{-1}Df(a)v(h) = Df(a)^{-1}\left[f(a+v(h)) - f(a) - R(v(h))\right] = Df(a)^{-1}(h - R(v(h))),$$

and so

$$|v(h)| \leq ||Df(a)^{-1}||\,|h| + ||Df(a)^{-1}||\,|R(v(h))|.$$

Since $|R(v)|/|v| \to 0$ as $|v| \to 0$ by differentiability of f, there exists $\delta_1 > 0$ such that

$$|R(v)| \leq |v|/(2||Df(a)^{-1}||), \quad \text{for } |v| \leq \delta_1 \tag{5}$$

By continuity of f^{-1}, there exists $\delta_2 > 0$ such that $|h| < \delta_2$ implies $|v(h)| \leq \delta_1$, and therefore,

$$|v(h)| \leq 2||Df(a)^{-1}||\,|h|,$$

whenever $|h| \leq \delta_2$ which gives half of (4). For the other half, consider

$$h = f(a + v(h)) - f(a) = Df(a)v(h) + R(v(h)).$$

Therefore, in view of (5) for $|h| < \delta_2$,

$$|h| \leq ||Df(a)||\,|v(h)| + |R(v(h))| \leq \left(||Df(a)|| + \frac{1}{2||Df(a)^{-1}||}\right)|v(h)|.$$

By Theorem 1.5 the partial derivatives of f^{-1} are defined at each point $y \in V_0$. Observe that the formula $Df^{-1}(y) = Df(f^{-1}(y))^{-1}$ implies that the map Df^{-1} from V_0 into the space of invertible $n \times n$ matrices can be written in the form

$$V_0 \xrightarrow{f^{-1}} U_0 \xrightarrow{Df} GL(n,\mathbb{R}) \xrightarrow{\iota} GL(n,\mathbb{R}),$$

where $\iota : GL(n,\mathbb{R}) \to GL(n,\mathbb{R})$ is the matrix inversion map. It follows from Cramer's rule that ι is a smooth map of the matrix components. Thus the partial derivatives of f^{-1} are continuous, and so f^{-1} is of class C^1. To prove that $f^{-1} \in C^k(V_0)$ assume by induction that we have shown that f^{-1} is of class C^{k-1}. Because Df^{-1} is a composition of C^{k-1}-smooth functions, it is itself C^{k-1}-smooth, which implies that the partial derivatives of f^{-1} are of class C^{k-1}, so f^{-1} is C^k-smooth. This completes the proof. □

Example 2.1 (Spherical coordinates)**.** Consider the map $f : (\rho, \phi, \theta) \to (x, y, z)$ given by

$$\begin{aligned} x &= \rho \sin\phi \cos\theta \\ y &= \rho \sin\phi \sin\theta \\ z &= \rho \cos\phi \end{aligned}$$

A computation shows that the differential of this map equals $\rho^2 \sin\phi$. Hence, by the Inverse Function theorem, f is a local diffeomorphism from $\{\rho > 0, \theta \in \mathbb{R},\ 0 < \phi < \pi\}$ to $\mathbb{R}^3$. By choosing a domain U where f is injective we conclude that the map $f : U \to f(U)$ is a diffeomorphism.

This choice of coordinates can be generalized to arbitrary dimension. Consider the map

$$\Phi : (r, \theta_1, ..., \theta_{n-1}) \mapsto (x_1, ..., x_n)$$

defined on the domain

$$U = (0,\infty) \times (0,\pi) \times ... \times (0,\pi) \times (0, 2\pi) \subset \mathbb{R}^n$$

by the equations

$$
\begin{aligned}
x_1 &= r\cos\theta_1,\\
x_2 &= r\sin\theta_1\cos\theta_2,\\
x_3 &= r\sin\theta_1\sin\theta_2\cos\theta_3,\\
&\dots\\
x_{n-1} &= r\sin\theta_1\sin\theta_2\dots\sin\theta_{n-2}\cos\theta_{n-1},\\
x_n &= r\sin\theta_1\sin\theta_2\dots\sin\theta_{n-1}.
\end{aligned}
$$

By the Inverse Function theorem, Φ is a diffeomorphism since its differential satisfies

$$\det D\Phi = r^{n-1}(\sin\theta_1)^{n-2}...\sin\theta_{n-2},$$

which does not vanish on U. A diffeomorphism that is used to simplify considerations or calculations is usually called a (local) change of coordinates. $\diamond$

The rank of a map $f:\mathbb{R}^n\to\mathbb{R}^m$ at a point x is defined as the rank of the differential $Df(x)$ (viewed as an $n\times m$ matrix), which is the same as $\dim Df(x)(\mathbb{R}^n)$. The following theorem can be viewed as a generalization of the Inverse Function theorem.

Theorem 2.5 (Rank theorem)**.** *Suppose $U\subset\mathbb{R}^m$ and $V\subset\mathbb{R}^n$ are open sets and $f:U\to V$ is a smooth map with constant rank k. For any point $p\in U$, there exist a connected neighbourhood $U_1\subset U$, a change of coordinates (i.e., a diffeomorphism) $\phi:U_1\to U_0$, $\phi(p)=0$ and connected neighbourhood $V_1\subset V$ with a change of coordinates $\psi:V_1\to V_0$, $\psi(f(p))=0$, such that*

$$\psi\circ f\circ\phi^{-1}(x_1,\dots,x_k,x_{k+1},\dots,x_m)=(x_1,\dots,x_k,0,\dots,0).$$

Here U_0 and V_0 can be assumed to be connected open neighbourhoods of the origin in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively.

Proof. Since $Df(p)$ has rank k, there exists a $k\times k$ minor with nonzero determinant. By reodering the coordinates, we may assume that it is the upper left minor, $(\frac{\partial f_i}{\partial x_j})$ for $i,j=1,\dots,k$. After translation we may assume that $p=0$, and $f(0)=0$. Let $(x,y)\in\mathbb{R}^k\times\mathbb{R}^{m-k}$, $(v,w)\in\mathbb{R}^k\times\mathbb{R}^{n-k}$ be the coordinates. If we write $f(x,y)=(Q(x,y),R(x,y))$ for some smooth maps $Q:U\to\mathbb{R}^k$, $R:U\to\mathbb{R}^{n-k}$, then $\left(\frac{\partial Q_i}{\partial x_j}\right)_{1\le i,j\le k}$ is nonsingular at the origin. Define $\phi(x,y)=(Q(x,y),y)$. Then

$$D\phi(0)=\begin{pmatrix}\frac{\partial Q_i}{\partial x_j}(0) & \frac{\partial Q_i}{\partial y_j}(0)\\ 0 & I_{m-k}\end{pmatrix}$$

is nonsingular. By the Inverse Function theorem there are connected neighbourhoods U_1 and U_0 of the origin in $\mathbb{R}^m$ such that $\phi:U_1\to U_0$ is a diffeomorphism. Writing the inverse map $\phi^{-1}(x,y)=(A(x,y),B(x,y))$, $A:U_0\to\mathbb{R}^k$, $B:U_0\to\mathbb{R}^{m-k}$, we have

$$(x,y)=\phi(A(x,y),B(x,y))=(Q(A(x,y),B(x,y)),B(x,y)).$$

It follows that $B(x,y)=y$, and so $\phi^{-1}(x,y)=(A(x,y),y)$, $Q(A(x,y),y)=x$, and therefore,

$$f\circ\phi^{-1}(x,y)=(x,\tilde R(x,y)),\quad \tilde R(x,y)=R(A(x,y),y).$$

The Jacobian matrix of this map at an arbitrary point $(x,y)\in U_0$ is

$$D(f\circ\phi^{-1})(x,y)=\begin{pmatrix}I_k & 0\\ \frac{\partial\tilde R_i}{\partial x_j} & \frac{\partial\tilde R_i}{\partial y_j}\end{pmatrix}.$$

Since composing with a diffeomorphism does not change the rank of a map, this matrix has rank equal to k everywhere on U_0. Since the first k columns are obviously independent, the rank can be k only if the partial derivatives $\frac{\partial \tilde{R}_i}{\partial y_j}$ vanish identically on U_0, which implies that $\tilde{R}$ is independent of variables y. Thus, setting $S(x) = \tilde{R}(x, 0)$, we have

$$f \circ \phi^{-1}(x, y) = (x, S(x)). \tag{6}$$

Let $V_1 = \{(v, w) \in V : (v, 0) \in U_0\}$, which is a neighbourhood of the origin. The map $\psi(v, w) = (v, w - S(v))$ is a diffeomorphism from V_1 onto its image, which can be seen by observing that $\psi^{-1}(s, t) = (s, t + S(s))$. It follows from (6) that

$$\psi \circ f \circ \phi^{-1}(x, y) = \psi(x, S(x)) = (x, S(x) - S(x)) = (x, 0).$$

□

For a domain $\Omega \subset \mathbb{R}^n$, a smooth map $f : \Omega \to \mathbb{R}^m$ is called an *immersion* if $Df(x)$ is injective for all $x \in \Omega$ (i.e., $Df(x)$ has a trivial kernel for all x), and a *submersion* if $Df(x)$ is surjective for all $x \in \Omega$. Clearly $n \le m$ is a necessary condition for f to be an immersion, while $n \ge m$ is required for a submersion. These are important examples of maps of constant rank. The Rank theorem is a powerful tool for the study of such maps. For example, let us show that if $f : \mathbb{R}^m \to \mathbb{R}^n$ is an injective map of constant rank, then it is an immersion. Indeed, if f is not an immersion, then the rank k of f is less than m. By the Rank theorem in a neighbourhood of any point there is a local change of coordinates such that f becomes

$$f(x_1, \dots x_k, x_{k+1}, \dots, x_m) = (x_1, \dots, x_k, 0, \dots, 0).$$

It follows that $f(0, \dots, 0, \varepsilon) = f(0)$ for ε small, which contradicts injectivity of f.

Another useful consequence of the Inverse Function theorem is the following theorem which gives conditions under which a level set of a smooth map is locally the graph of a smooth function.

Theorem 2.6 (Implicit Function Theorem). *Let $U \subset \mathbb{R}^n \times \mathbb{R}^k$ be an open set, and let $(x, y) = (x_1, \dots, x_n, y_1, \dots, y_k)$ denote the standard coordinates on U. Suppose $\Phi : U \to \mathbb{R}^k$ is a smooth map, $(a, b) \in U$, and $c = \Phi(a, b)$. If the $k \times k$ matrix*

$$\left(\frac{\partial \Phi^i}{\partial y^j}(a, b) \right)$$

is nonsingular, then there exist neighbourhoods $V_0 \subset \mathbb{R}^n$ of a and $W_0 \subset \mathbb{R}^k$ of b, and a smooth map $f : V_0 \to W_0$ such that $\Phi^{-1}(c) \cap V_0 \times W_0$ is the graph of f, i.e., $\Phi(x, y) = c$ for $(x, y) \in V_0 \times W_0$ if and only if $y = f(x)$.

Proof. Consider the map $\Psi : U \to \mathbb{R}^n \times \mathbb{R}^k$ defined by $\Psi(x, y) = (x, \Phi(x, y))$. Its differential at (a, b) is

$$D\Psi(a, b) = \begin{pmatrix} I_n & 0 \\ \\ \frac{\partial \Phi_i}{\partial x_j}(a, b) & \frac{\partial \Phi_i}{\partial y_j}(a, b) \end{pmatrix},$$

which is nonsingular by hypothesis. Thus by the Inverse Function theorem there exist connected open neighbourhoods U_0 of (a, b) and Y_0 of (a, c) such that $\Psi : U_0 \to Y_0$ is a diffeomorphism. Shrinking U_0 and Y_0 if necessary, we may assume that $U_0 = V \times W$ is a product neighbourhood. The inverse map has the form (why?)

$$\Psi^{-1}(x, y) = (x, B(x, y))$$

for some smooth map $B : Y_0 \to W$. Let $V_0 = \{x \in V : (x, c) \in Y_0\}$ and $W_0 = W$, and define $f : V_0 \to W_0$ by $f(x) = B(x, c)$. Comparing y components in the relation $(x, c) = \Psi \circ \Psi^{-1}(x, c)$ yields

$$c = \Phi(x, B(x, c)) = \Phi(x, f(x)),$$

whenever $x \in V_0$ so the graph of f is contained in $\Phi^{-1}(c)$. Conversely suppose $(x, y) \in V_0 \times W_0$ and $\Phi(x, y) = c$. Then $\Psi(x, y) = (x, \Phi(x, y)) = (x, c)$, so

$$(x, y) = \Psi^{-1}(x, c) = (x, B(x, c)) = (x, f(x)),$$

which implies that $y = f(x)$. $\square$