Math 598 Feb 11, 2005[1]
Geometry and Topology II
Spring 2005, PSU

Lecture Notes 6

2.5 The inverse function theorem

Recall that if $f\colon M \to N$ is a diffeomorphism, then df_p is nonsingular at all $p \in M$ (by the chain rule and the observation that $f \circ f^{-1}$ is the identity function on M). The main aim of this section is to prove a converse of this phenomenon:

Theorem 1 (The Inverse Function Theorem). *Let $f\colon M \to N$ be a smooth map, and $\dim(M) = \dim(N)$. Suppose that df_p is nonsingular at some $p \in M$. Then f is a* local diffeomorphism *at p, i.e., there exists an open neighborhood U of p such that*

1. *f is one-to-one on U.*
2. *$f(U)$ is open in N.*
3. *$f^{-1}\colon f(U) \to U$ is smooth.*

In particular, $d(f^{-1})_{f(p)} = (df_p)^{-1}$.

A simple fact which is applied a number of times in the proof of the above theorem is

Lemma 2. *Let $f\colon M \to N$, and $g\colon N \to L$ be diffeomorphisms, and set $h := g \circ f$. If any two of the mappings f, g, h are diffeomorphisms, then so is the third.* □

In particular, the above lemma implies

Proposition 3. *If Theorem 1 is true in the case of $M = \mathbf{R}^n = N$, then, it is true in general.*

[1]Last revised: February 13, 2005

Proof. Suppose that Theorem 1 is true in the case that $M = \mathbf{R}^n = N$, and let $f\colon M \to N$ be a smooth map with df_p nonsingular at some $p \in M$. By definition, there exist local charts (U, ϕ) of M and (V, ψ) of N, centered at p and $f(p)$ respectively, such that $\tilde{f} := \phi^{-1} \circ f \circ \psi$ is smooth. Since ϕ and ψ are diffeomorphisms, $d\phi_p$ and $d\psi_{f(p)}$ are nonsingular. Consequently, by the chain rule, $d\tilde{f}_o$ is nonsingular, and is thus a local diffeomorphism. More explicitly, there exists open neighborhoods A and B of the origin o of $\mathbf{R}^n$ such that $\tilde{f}\colon A \to B$ is a diffeomorphism. Since $\phi\colon \phi^{-1}(A) \to A$ is also a diffeomorphism, it follows that $\phi \circ \tilde{f}\colon \phi^{-1}(A) \to B$ is a diffeomorphism. But $\phi \circ \tilde{f} = f \circ \psi$. So $f \circ \psi\colon \phi^{-1}(A) \to B$ is a diffeomorphism. Finally, since $\psi\colon \psi^{-1}(B) \to B$ is a diffeomorphism, it follows, by the above lemma, that $f\colon \phi^{-1}(A) \to \psi^{-1}(B)$ is a diffeomorphism. □

So it remains to prove Theorem 1 in the case that $M = \mathbf{R}^n = N$. To this end we need the following fact. Recall that a metric space is said to be complete provided that every Cauchy sequence of that space converges.

Lemma 4 (The contraction Lemma). *Let (X, d) be a complete metric space, and $0 \le \lambda < 1$. Suppose that there exists mapping $f\colon X \to X$ such that $d(f(x_1), (x_2)) \le \lambda d(x_1, x_2)$, for all x_1, $x_2 \in X$. Then there exists a* unique *point $x \in X$ such that $f(x) = x$.*

Proof. Pick a point $x_0 \in X$ and set $x_n := f^n(x)$, for $n \ge 1$. We claim that $\{x_n\}$ is a Cauchy sequence. To this end note that

$$d(x_n, x_{n+m}) = d(f^n(x_0), f^n(x_m)) \le \lambda^n d(x_0, x_m).$$

Further, by the triangle inequality

$$\begin{aligned} d(x_0, x_m) &\le d(x_0, x_1) + d(x_1, x_2) + \cdots + d(x_{m-1}, x_m) \\ &\le (1 + \lambda + \lambda^2 + \cdots + \lambda^m) d(x_0, x_1) \\ &\le \frac{1}{1-\lambda} d(x_0, x_1). \end{aligned}$$

So, setting $K := d(x_0, x_1)/(1-\lambda)$, we have

$$d(x_n, x_{n+m}) \le \lambda^n K.$$

Since K does not depend on m or n, the last inequality shows that $\{x_n\}$ is a Cauchy sequence, and therefore, since X is complete, it has a limit point, say x_∞. Now note that, since $d\colon X \times X \to \mathbf{R}$ is continuous (why?),

$$d(x_\infty, f(x_\infty)) = \lim_{n \to \infty} d(x_n, f(x_n)) = 0.$$

Thus X_∞ is a fixed point of f. Finally, note that if a and b are fixed points of f, then

$$d(a,b) = d(f(a), f(b)) \le \lambda d(a,b),$$

which, since $\lambda < 1$, implies that $d(a,b) = 0$. So f has a unique fixed point. □

Exercise 5. Does the previous lemma remain valid if the condition that $d(f(x_1),(x_2)) \le \lambda d(x_1,x_2)$ is weakened to $d(f(x_1),(x_2)) \le d(x_1,x_2)$?

Next we recall

Lemma 6 (The mean value theorem). *Let $f\colon \mathbf{R}^n \to \mathbf{R}$ be a C^1 functions. Then for every p, $q \in \mathbf{R}^n$ there exists a point s on the line segment connecting p and q such that*

$$f(p) - f(q) = Df(s)(p-q) = \sum_{i=1}^{n} D_i f(s_i)(p^i - q^i).$$

□

Exercise 7. Prove the last lemma by using the mean value theorem for functions of one variable an the chain rule. (*Hint:* Parametrize the segment joining p and q by $tq + (1-t)p$, $0 \le t \le 1$).

The above lemma implies:

Proposition 8. *Let $f\colon \mathbf{R}^n \to \mathbf{R}^m$ be a C^1 function, U be a convex open neighborhood of o in $\mathbf{R}^n$, and set*

$$K := \sup\left\{ \left|D_j f^i(p)\right| \,\middle|\, 1 \le i \le m, 1 \le j \le n, p \in U \right\}$$

Then, for every p, $q \in U$,

$$\|f(p) - f(q)\| \le \sqrt{mn}\, K \|p - q\|$$

Proof. First note that

$$\|f(p) - f(q)\|^2 \;=\; \sum_{i=1}^{m} \big(f^i(p) - f^i(q)\big)^2.$$

Secondly, by the mean value theorem (Lemma 6), there exists, for every i a point s_i on the line segment connecting p and q such that

$$f^i(p) - f^i(q) = Df^i(s_i)(p-q) = \sum_{j=1}^{n} D_j f^i(s_j)(p^j - q^j).$$

Since U is convex, $s_i \in U$, and, therefore, by the Cauchy-Schwartz inequality

$$|f^i(p) - f^i(q)| \leq \sqrt{\sum_{j=1}^{n} D_j f^i(s_j)^2} \sqrt{\sum_{j=1}^{n} (p^j - q^j)^2} \leq \sqrt{n} K \|p - q\|.$$

So we conclude that

$$\|f(p) - f(q)\|^2 \leq m\, n\, K^2 \|p - q\|^2.$$

□

Finally, we recall the following basic fact

Lemma 9. *Let $f\colon \mathbf{R}^n \to \mathbf{R}^m$, and $p \in \mathbf{R}^n$. Suppose there exists a linear transformation $A\colon \mathbf{R}^n \to \mathbf{R}^m$ such that*

$$f(x) - f(p) = A(p - x) + r(x, p)$$

where $r\colon \mathbf{R}^2 \to \mathbf{R}$ is a function satisfying

$$\lim_{x \to p} \frac{r(x,p)}{\|x - p\|} = 0.$$

Then all the partial derivatives of f exist at p, and A is given by the jacobian matrix $Df(p) := (D_1 f(p), \ldots, D_n f(p))$ whose columns are the partial derivatives of f. In particular, A is unique. Conversely, if all the partial derivative $D_i f(p)$ exist, then $A := Df(p)$ satisfies the above equation.

Proof. Let $e_1, \ldots, e_n$ be the standard basis for $\mathbf{R}^n$. Then

$$D_i f(p) = \lim_{t \to 0} \frac{f(p + te_i) - f(p)}{t} = \lim_{t \to 0} \frac{A(te_i) + r(p + te_i, p)}{t} = A(e_i).$$

Thus all the partial derivatives of f exist at p, and $D_i f(p)$ coincides with the i^{th} column of (the matrix representation) of A. In particular, $A = Df(p)$ and therefore A is unique.

Conversely, suppose that all the partial derivatives $D_i f(p)$ exist and set

$$r(x,p) := f(x) - f(p) - Df(p)(p-x).$$

By the mean value theorem,

$$r(x,p) = (Df(s) - Df(p))(p-x)$$

for some s on the line segment joining p and s. Thus it follows that

$$\lim_{x\to p} \frac{r(x,p)}{\|x-p\|} = \lim_{x\to p}(Df(s) - Df(p))\left(\frac{p-x}{\|p-x\|}\right) = 0,$$

as desired. □

Now we are finally ready to prove the main result of this section.

Proof of Theorem 1. By 3 we may assume that $M = \mathbf{R}^n = N$. Further, after replacing $f(x)$ with $(Df(p))^{-1}f(x-p) - f(p)$ we may assume, via Lemma 2, that

$$p = o, \quad f(o) = o, \quad \text{and} \quad Df(o) = I,$$

where I denotes the identity matrix. Now define $g\colon \mathbf{R}^n \to \mathbf{R}^n$ by

$$g(x) = x - f(x).$$

Then $g(o) = o$, and $Dg(o) = 0$. Thus, by Proposition 8, there exists $r > 0$ such that for all x_1, $x_2 \in B_r(o)$, the closed ball of radius r centered at o,

$$\|g(x_1) - g(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|.$$

In particular, $\|g(x)\| = \|g(x) - g(o)\| \leq \|x\|/2$. So $g(B_r(o)) \subset B_{r/2}(o)$. Now, for every $y \in B_{r/2}(o)$ and $x \in B_r(o)$ define

$$T_y(x) := y + g(x) = y + x - f(x).$$

Then, by the triangle inequality, $\|T_y(x)\| \leq r$. Thus $T_y\colon B_r(o) \to B_r(o)$. Further note that

$$T_y(x) = x \iff y = f(x).$$

in particular, T_y has a unique fixed point on $B_r(o)$ if and only if f is one-to-one on $B_r(o)$. But

$$\|T_y(x_1) - T_y(x_2)\| = \|g(x_1) - g(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\|.$$

Thus by Lemma 4, T_y does indeed have a unique fixed point, and we conclude that f is one-to-one on $B_r(o)$. In particular, we let U be the interior of $B_r(o)$.

Next we show that $f(U)$ is open. To this end it suffices to prove that $f^{-1}\colon f(B_r(o)) \to B_r(o)$ is continuous. To see this note that, by the definition of g and the triangle inequality,

$$\|g(x_1)-g(x_2)\| = \|(x_1-x_2)-(f(x_1)-f(x_2))\| \geq \|x_1-x_2\|-\|f(x_1)-f(x_2)\|.$$

Thus,

$$\|f(x_1) - f(x_2)\| \geq \|x_1 - x_2\| - \|g(x_1) - g(x_2)\| = \frac{1}{2}\|x_1 - x_2\|,$$

which in turn implies

$$\|y_1 - y_2\| \geq \frac{1}{2}\|f^{-1}(y_1) - f^{-1}(y_2)\|.$$

So f^{-1} is continuous.

It remains to show that f^{-1} is smooth on $f(U)$. To this end, note that by Lemma 9, for every $p \in U$,

$$f(x) - f(p) = Df(p)(x - p) + r(x, p).$$

Now multiply both sides of the above equality by $A := (Df(p))^{-1}$, and set $y := f(x)$, $q := f(p)$. Then

$$A(y - q) = f^{-1}(y) - f^{-1}(q) + Ar(f^{-1}(y), f^{-1}(q)),$$

which we may rewrite as

$$f^{-1}(y) - f^{-1}(q) = A(y - q) + \overline{r}(y, q),$$

where

$$\overline{r}(y, q) := Ar(f^{-1}(y), f^{-1}(q)).$$

Finally note that

$$\lim_{y\to q}\frac{\overline{r}(y,q)}{\|y-q\|} = A\lim_{y\to q}\frac{r(f^{-1}(y),f^{-1}(q))}{\|y-q\|} \leq 2A\lim_{y\to q}\frac{r(f^{-1}(y),f^{-1}(q))}{\|f^{-1}(y)-f^{-1}(q)\|} = 0.$$

Thus, again by Lemma 9, f^{-1} is differentiable at all $p \in U$ and

$$D(f^{-1})(p) = \Big(Df\big(f^{-1}(p)\big)\Big)^{-1}.$$

Since the right hand side of the above equation is a continuous function of p (because f is C^1 and f^{-1} is continuous), it follows that f^{-1} is C^1. But if f is C^r, then the right hand side of the above equation is C^r (since Df is C^∞ everywhere), which in turn yields that f^{-1} is C^{r+1}. So, by induction, f^{-1} is C^∞. □

Exercise 10. Give a simpler proof of the inverse function theorem for the special case of mappings $f\colon \mathbf{R} \to \mathbf{R}$.

2.6 The rank theorem

The inverse function theorem we proved in the last section yields the following more general result:

Theorem 11 (The rank theorem). *Let $f\colon M \to N$ be a smooth map, and suppose that rank$(df_p) = k$ for all $p \in M$, then, for each $p \in M$, there exists local charts (U,ϕ) and (V,ψ) of M and N centered at p and $f(p)$ respectively such that*

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

Exercise 12. Show that to prove the above theorem it suffices to consider the case $M = \mathbf{R}^n$ and $N = \mathbf{R}^m$. Furthermore, show that we may assume that $p = o$, $f(o) = o$, and the $k \times k$ matrix in the upper left corner of the jacobian matrix $Df(o)$ is nonsingular.

Proof. Suppose that the conditions of the previous exercise hold. Define $\phi\colon \mathbf{R}^n \to \mathbf{R}^n$ by

$$\phi(x) := (f^1(x),\dots,f^k(x),x^{k+1},\dots,x^n).$$

Then

$$D\phi(o) = \begin{pmatrix} \frac{\partial(f^1,\ldots,f^k)}{(x^1,\ldots,x^k)}(o) & * \\ 0 & I_{n-k} \end{pmatrix}.$$

Thus $D\phi(o)$ is nonsingular. So, by the inverse function theorem, ϕ is a local diffeomorphism at o. In particular ϕ^{-1} is well defined on some open neighborhood U of o. Let $\pi_i \colon \mathbf{R}^\ell \to \mathbf{R}$ be the projection onto the i^{th} coordinate. Then, for $1 \leq i \leq k$, $\pi_i \circ \phi = f^i$. Consequently, $f^i \circ \phi^{-1} = \pi_i$. Thus, if we set $\tilde{f}^i := f^i \circ \phi^{-1}$, for $k+1 \leq i \leq m$, then

$$f \circ \phi^{-1}(x) = (x^1, \ldots, x^k, \tilde{f}^{k+1}(x), \ldots, \tilde{f}^m(x))$$

for all $x \in U$. Next note that

$$D(f \circ \phi^{-1})(o) = \begin{pmatrix} I_k & 0 \\ * & \frac{\partial(\tilde{f}^{k+1},\ldots,\tilde{f}^m)}{(x^{k+1},\ldots,x^n)}(o) \end{pmatrix}.$$

On the other hand, $D(f \circ \phi^{-1})(o) = D(f)(p) \circ D(\phi^{-1})(o)$. Thus

$$rank(D(f \circ \phi^{-1})(o)) = rank(D(f)(p)) = k,$$

because $D(\phi^{-1}) = D(\phi)^{-1}$ is nonsingular. The last two equalities imply that

$$\frac{\partial(\tilde{f}^{k+1}, \ldots, \tilde{f}^m)}{(x^{k+1}, \ldots, x^n)}(o) = 0,$$

where 0 here denotes the matrix all of whose entries is zero. So we conclude that the functions $\tilde{f}^{k+1}, \ldots, \tilde{f}^m$ do not depend on $x^{k+1}, \ldots, x^n$. In particular, if V is a small neighborhood of o in $\mathbf{R}^m$, then the mapping $T \colon V \to \mathbf{R}^m$ given by

$$T(y) := \big(y^1, \ldots, y^k, y^{k+1} + f^{k+1}(y^1, \ldots, y^k), \ldots, y^m + f^m(y^1, \ldots, y^k)\big)$$

is well defined. Now note that

$$DT(o) = \begin{pmatrix} I_k & * \\ 0 & I_{m-k} \end{pmatrix}.$$

Thus, by the inverse function theorem, $\psi := T^{-1}$ is well defined on an open neighborhood of o in $\mathbf{R}^m$. Finally note that

$$\begin{aligned} \psi \circ f \circ \phi^{-1}(x) &= \psi(x^1, \ldots, x^k, \tilde{f}^{k+1}(x), \ldots, \tilde{f}^m(x)) \\ &= \psi \circ T(x^1, \ldots, x^k, 0, \ldots, 0) \\ &= (x^1, \ldots, x^k, 0, \ldots, 0), \end{aligned}$$

as desired. □

Exercise 13. Show that there exists no C^1 function $f\colon \mathbf{R}^2 \to \mathbf{R}$ which is one-to-one.