Calculus on Manifolds

Jędrzej Śniatycki

The Inverse Function Theorem

Let U be an open subset of $\mathbb{R}^n$ and $F : U \to \mathbb{R}^n$ be smooth. If the derivative DF is invertible at a point $x_0 \in U$, then there exists an open neigbourhood V of x_0 contained in U such that $F(V)$ is open in $\mathbb{R}^n$, the restriction $F_{|V}$ of F to V is a one-to-one map of V onto $F(V)$ with a smooth inverse $(F_{|V})^{-1} : F(V) \to V$.

Reference: Lee

Proof. We begin with a special case when $x_0 = 0 \in \mathbb{R}^n$, and $DF(0)$ is the identity $n \times n$ matrix I.

We first show that the map F is one-to-one in a neighbourhood of $0 \in \mathbb{R}^n$. Let $H(x) = x - F(x)$. Then $H(0) = 0$ and $DH(0) = 0$. Hence, there exists $r > 0$ such that the matrix norm $\|DH(x)\| \leq \frac{1}{2}$ for all x in the closure $\bar{B}_{0,r}$ of the open ball $B_{0,r}$. In other words, for every vector $u \in \mathbb{R}^n$, and every $x \in \bar{B}_{0,r}$, we have $\|DH(x)u\| \leq \frac{1}{2}\|u\|$. For every $x \in \bar{B}_{0,r}$, the linear approximation of $H(x)$ gives $H(x) = H(0) + DH(x')x$ for some x' in the line interval joining x to 0. Since $x' \in \bar{B}_{0,r}$ if $x \in \bar{B}_{0,r}$, and $H(0) = 0$, it follows that

$$\|H(x)\| \leq \frac{1}{2}\|x\| \text{ for all } x \in \bar{B}_{0,r}.$$

Similarly, $H(x) - H(x') = DH(x'')(x - x')$ for some x'' in the line interval joing x and x', so that

$$\|H(x) - H(x')\| \leq \frac{1}{2}\|x - x'\|$$

for every $x, x' \in \bar{B}_{0,r}$. Hence, the definition of H yields

$$\begin{aligned} \|x - x'\| &= \|H(x) + F(x) - H(x') - F(x')\| \leq \|H(x) - H(x')\| + \|F(x) - F(x')\| \\ &\leq \frac{1}{2}\|x - x'\| + \|F(x) - F(x')\|. \end{aligned}$$

This implies that

$$\|F(x) - F(x')\| \geq \frac{1}{2}\|x - x'\| \quad (1)$$

for every $x, x' \in \bar{B}_{0,r}$. Thus, the restriction of F to $\bar{B}_{0,r}$ is one-to-one.

Next, we show that F maps the open ball $B_{0,r}$ onto the open ball $B_{0,r/2}$. For $y \in B_{0,r/2}$, let $G(x) = y + H(x) = y + x - F(x)$. Observe that $G(x) = x$ if and only if $F(x) = y$. Thus, to show that there exists $x \in B_{0,r}$ such that $F(x) = y$, it suffices to show that there exists a point $x \in B_{0,r}$ such that $G(x) = x$. For $x \in \bar{B}_{0,r}$,

$$\|G(x)\| \leq \|y\| + \|H(x)\| \leq \frac{r}{2} + \frac{1}{2}\|x\| \leq r.$$

Hence, G maps $\bar{B}_{0,r}$ to itself. Moreover,

$$\|G(x) - G(x')\| = \|H(x) - H(x')\| \leq \frac{1}{2}\|x - x'\|$$

for every $x, x' \in \bar{B}_{0,r}$. In order to find $x \in \bar{B}_{0,r}$ such that $G(x) = x$, we consider a sequence (x_k) defined inductively by $x_1 = G(0)$ and $x_k = G(x_{k-1})$ for $k \geq 2$. For any $k > 1$ we have $\|x_{k+1} - x_k\| = \|G(x_k) - G(x_{k-1})\| \leq \frac{1}{2}\|x_k - x_{k-1}\|$. By induction, we obtain $\|x_{k+1} - x_k\| \leq \left(\frac{1}{2}\right)^k \|x_1\|$. Therefore, for $k \geq l \geq m$,

$$\begin{aligned}
\|x_{k+1} - x_l\| &= \|(x_{k+1} - x_k) + (x_k - x_{k-1}) + ... + (x_{l+1} - x_l)\| \\
&\leq \|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + ... + \|x_{l+1} - x_l\| \\
&\leq \left(\left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^{k-1} + ... + \left(\frac{1}{2}\right)^l\right)\|x_1\| \\
&= \left(\frac{1}{2}\right)^l \left(\left(\frac{1}{2}\right)^{k-l} + \left(\frac{1}{2}\right)^{k-l-1} + ... + \left(\frac{1}{2}\right)^0\right)\|x_1\| \\
&\leq \left(\frac{1}{2}\right)^l \sum_{i=0}^{\infty}\left(\frac{1}{2}\right)^i \|x_1\| \leq \left(\frac{1}{2}\right)^m \|x_1\|.
\end{aligned}$$

This implies that (x_k) is a Cauchy sequence. Hence, it has the limit $x = \lim_{k\to\infty} x_k$ in $\bar{B}_{0,r}$. Since G is a continuous function, it follows that $G(x) = G(\lim_{k\to\infty} x_k) = \lim_{k\to\infty} G(x_k) = \lim_{k\to\infty}(x_{k+1}) = x$, so that x is a fixed point of G. Hence, $F(x) = y$. Moreover, equation (1) for $x' = 0$ together with the assumption that y is contained in the open ball $\in B_{0,r/2}$ imply that

$$\|x\| \leq 2\|F(x)\| = 2\|y\| < r,$$

so that x is contained in the open ball $B_{0,r}$. Thus, F restricted to $B_{0,r}$ is one-to-one and onto $B_{0,r/2}$, which implies that there exists the inverse function $F^{-1} : B_{0,r/2} \to B_{0,r}$. Equation (1) can be rewritten as

$$\left\|F^{-1}(y) - F^{-1}(y')\right\| \leq 2\|y - y'\| \tag{2}$$

for every $y, y' \in B_{0,r/2}$. Hence, F^{-1} is continuous.

It remains to show that F^{-1} is smooth. If F^{-1} were differentiable at $y = F(x) \in B_{0,r/2}$, it would satisfy the chain rule have

$$DF(x) \circ DF^{-1}(y) = D(F \circ F^{-1})(y) = I,$$

and the derivative if F^{-1} at y would be given by

$$DF^{-1}(y) = (DF(x))^{-1}.$$

We show $(DF(x))^{-1}$ is in fact the derivative of F^{-1} at $y = F(x)$. Consider two functions

$$\begin{aligned} P(\Delta x) &= F(x + \Delta x) - F(x) - DF(x)\Delta x, \\ Q(\Delta y) &= F^{-1}(y + \Delta y) - F^{-1}(y) - (DF(x))^{-1}\Delta y \end{aligned}$$

defined for Δx and Δy in a neigbourhood of $0 \in \mathbb{R}^n$ such that that $x + \Delta x \in B_{0,r}$ and $y + \Delta y \in B_{0,r/2}$. Since F is smooth, it follows that $P(\Delta x)/\|\Delta x\| \to 0$ as $\Delta x \to 0$. In order to prove that F^{-1} is differentiable at y with derivative $(DF(x))^{-1}$ it suffices to show that $Q(\Delta y)/\|\Delta y\| \to 0$ as $\Delta y \to 0$. Let

$$R(\Delta y) = F^{-1}(y + \Delta y) - F^{-1}(y) = F^{-1}(y + \Delta y) - x.$$

Since $F^{-1}(y + \Delta y) = F^{-1}(y) + R(\Delta y) = x + R(\Delta y)$, it follows that $y + \Delta y = F(F^{-1}(y + \Delta y)) = F(x + R(\Delta y))$. The Mean Value Theorem implies that $\Delta y = F(x + R(\Delta y)) - F(x) = DF(x')R(\Delta y)$ for x' in the line segment between x and $x + R(\Delta y)$. By construction, for every $x'' \in B_{0,r}$, the operator norm $\|DH(x'')\| \leq \frac{1}{2}$, where $DH(x'') = DF(x'') - I$. Therefore, $DF(x'') = DH(x'') + I$ and the triangle inequality implies that

$$\|DF(x'')\| \leq 1 + \|DH(x'')\| \leq \frac{3}{2}$$

for all $x'' \in B_{0,r}$. Hence,

$$\|\Delta y\| \leq \|DF(x')\| \, \|R(\Delta y)\| \leq \frac{3}{2} \|R(\Delta y)\| . \tag{3}$$

On the other hand inequality (2) implies that

$$\|R(\Delta y)\| \leq 2 \|\Delta y\| . \tag{4}$$

Combining inequalities (3) and (4) we get an estimate

$$\frac{2}{3} \|\Delta y\| \leq \|R(\Delta y)\| \leq 2 \|\Delta y\| . \tag{5}$$

It implies that $R(\Delta y)$ is a continuous function of Δy and it vanishes only when $\Delta y = 0$. Next, we rewrite the expression for Q as follows:

$$\begin{aligned} Q(\Delta y) &= F^{-1}(y+\Delta y) - F^{-1}(y) - (DF(x))^{-1}\Delta y = R(\Delta y) - (DF(x))^{-1}\Delta y \\ &= (DF(x))^{-1}(DF(x)R(\Delta y) - \Delta y) \\ &= (DF(x))^{-1}\left(DF(x)R(\Delta y) - F(x + R(\Delta y)) + F(x)\right) \\ &= -(DF(x))^{-1}(P(R(\Delta y)). \end{aligned}$$

Hence, for $\Delta y \neq 0$,

$$\begin{aligned} \frac{\|Q(\Delta y)\|}{\|\Delta y\|} &\leq \left\|(DF(x))^{-1}\right\| \frac{\|P(R(\Delta y))\|}{\|\Delta y\|} = \left\|(DF(x))^{-1}\right\| \frac{\|P(R(\Delta y))\|}{\|R(\Delta y)\|}\frac{\|R(\Delta y)\|}{\|\Delta y\|} \\ &\leq 2\left\|(DF(x))^{-1}\right\| \frac{\|P(R(\Delta y))\|}{\|R(\Delta y)\|} \to 0 \text{ as } \|\Delta y\| \to 0 \end{aligned}$$

because $P(\Delta x)/\|\Delta x\| \to 0$ as $\Delta x \to 0$ since F is smooth with derivative $DF(x)$ at x. Therefore $Q(\Delta y)/\|\Delta y\| \to 0$ as $\Delta y \to 0$, which implies that F^{-1} is differentiable at y with derivative $(DF(x))^{-1}$.

Thus $DF^{-1}(y)$ exists for every $y \in B_{0,r/2}$ and is given by the chain rule. Hence, we get a map $DF^{-1} : B_{0,r/2} \to Gl(n, \mathbb{R})$. Let $i : Gl(n, R) \to Gl(n, R)$ be given by $i(A) = A^{-1}$ for each $A \in GL(n, R)$. We have

$$DF^{-1} = i \circ DF \circ F^{-1}.$$

Since F^{-1} is continuous and DF and i are smooth it follows that DF^{-1} is continuous. Thus, F^{-1} has continuous first derivatives. We can proceed by induction. If F^{-1} has continuous derivatives up to an order k, then DF^{-1} has continuous derivatives up to the order k. Therefore, F^{-1} has continuous derivatives up to the order $k+1$. It follows that F^{-1} is smooth (of class C^∞). This completes the proof in the special case in which $F(0) = 0$ and $DF(0) = I$.

In order to complete the proof of the theorem we need to reduce the general case to the special case considered above. As in the statement of the theorem, we assume that $A = DF(x_0)$ is invertible and set $y_0 = F(x_0)$. A function $\tilde{F}$, defined by

$$\tilde{F}(x) = A^{-1}(F(x + x_0) - y_0),$$

is smooth, and satisfies the conditions $\tilde{F}(0) = 0$ and $D\tilde{F}(0) = A^{-1}DF(x_0) = I$. Hence, by the above argument, there exists $r > 0$ such that the restriction of $\tilde{F}$ to the ball $B_{0,r}$ has smooth inverse $\tilde{F}^{-1} : B_{0,r/2} \to B_{0,r}$. Since $F(x) = y_0 + A\tilde{F}(x - x_0)$, it follows that to get $F^{-1}(y)$ we need to solve the equation $y = y_0 + A\tilde{F}(x - x_0)$ for x in terms of y. It is easy to see that the solution is $x = x_0 + \tilde{F}(A^{-1}(y - y_0))$. Hence, the function

$$F^{-1}(y) = x_0 + \tilde{F}(A^{-1}(y - y_0)).$$

for all $y \in W = \{y \in \mathbb{R}^n \mid A^{-1}(y - y_0) \in B_{0,r/2}\}$ is the inverse of F restricted to $U = \{x \in R^n \mid A^{-1}(F(x + x_0) - y_0) \in B_{0,r/2}$. Clearly, F^{-1} is smooth as the composition of smooth functions. □

An important consequence of the Inverse Function Theorem is the Implicit Function Theorem stated below. Here, we use the canonical isomorphism

$$\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n+m} : ((x^1, ..., x^n), (y^1, ..., y^m)) \longmapsto (x^1, ..., x^n, y^1, ..., y^m).$$

The Implicit Function Theorem Let (x_0, y_0) be a point in an open subset U of $\mathbb{R}^{n+m} \equiv \mathbb{R}^n \times \mathbb{R}^m$ and

$$F : U \to R^m : (x^1, ..., x^n, y^1, ..., y^m) \longmapsto (f^1(x, y), ..., f^m(x, y))$$

be a smooth map such that

$$F(x_0, y_0) = 0 \text{ and } \det(\partial f^a/\partial x^i)(x_0, y_0) \neq 0.$$

Then, there exists a open neighbourhood V of x_0 in $\mathbb{R}^n$ an open neighbourhood W of y_0 in $\mathbb{R}^m$ such that $V \times W \subseteq U$ and there exists a smooth map $H : V \to W$ such that, for every $(x, y) \in V \times W$,

$$F(x, y) = 0 \text{ if and only if } y = H(x).$$

Proof. Let $K : U \to \mathbb{R}^n \times \mathbb{R}^m$ be given by $K(x, y) = (x, F(x, y))$. Its derivative at (x_0, y_0) is

$$DK(x_0, y_0) = \begin{pmatrix} I_n & 0 \\ (\partial f^a/\partial x^i)(x_0, y_0) & (\partial f^a/\partial y^b)(x_0, y_0) \end{pmatrix},$$

and $\det DK(x_0, y_0) = \det(\partial f^a/\partial y^b)(x_0, y_0) \neq 0$. By the Inverse Function Theorem, there exists an open neighbourhood $V_0 \times W_0$ of $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ contained in U such that $K(V_0 \times W_0)$ is open in $\mathbb{R}^n \times \mathbb{R}^m$, the restriction $K_{|V_0 \times W_0}$ of K to $V_0 \times W_0$ is a one-to-one map of $V_0 \times W_0$ onto $K(V_0 \times W_0)$ with a smooth inverse $(K_{|V_0 \times W_0})^{-1} : K(V_0 \times W_0) \to V_0 \times W_0$. Let $V_1 \times W_1$ be an pen neighbourhood of $K(x_0, y_0) = (x_0, 0)$ contained in $K(V_0 \times W_0)$. The restriction of $(K_{|V_0 \times W_0})^{-1}$ to $V_1 \times W_1$ can be expressed in the form

$$(K_{|V_0 \times W_0})^{-1}_{|V_1 \times W_1} : (x, y) \mapsto (k(x, y), h(x, y)).$$

Composing this expression with K, we get $K \circ (K_{|V_0 \times W_0})^{-1}_{|V_1 \times W_1} = I_{|V_1 \times W_1}$ which implies that

$$(k(x, y), F(k(x, y), h(x, y)) = (x, y) \text{ for all } (x, y) \in V_1 \times W_1.$$

Hence, $k(x, y) = x$ and $F(x, h(x, y)) = y$. Thus, setting $y = 0$ we get

$$F(x, h(x, 0)) = 0.$$

Taking the composition in the opposite order we get $(K_{|V_0\times W_0})^{-1}_{|V_1\times W_1} \circ K = I$, so that

$$(k(x, F(x,y)), h(x, F(x,y)) = (x,y).$$

Hence,

$$h(x, F(x,y)) = y.$$

Therefore, $y = h(x,0)$ whenever $F(x,y) = 0$, which completes the proof. $\square$