Proposition 8.34 (Chain Rule)

Let D be open in $\mathbb{R}^n$ and E be open in $\mathbb{R}^m$. Let $f: D \to E$ and $g: E \to \mathbb{R}^p$ be two functions, $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$, $g = \begin{pmatrix} g_1 \\ \vdots \\ g_p \end{pmatrix}$. Assume that f is differentiable on D and g is differentiable on E. Then the function $h := g \circ f : D \to \mathbb{R}^p$, $h = \begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix}$ is differentiable on D and for all

$x = (x_1 \dots x_n) \in D$,

$$J_h(x) = J_g(f(x)) \cdot J_f(x).$$

Furthermore, for all $i = 1, \dots, n$ and $k = 1, \dots, p$

$$\frac{\partial h_k}{\partial x_i}(x) = \sum_{j=1}^{m} \frac{\partial g_k}{\partial y_j}(f(x)) \cdot \frac{\partial f_j}{\partial x_i}(x)$$
$$= \sum_{j=1}^{m} \frac{\partial g_k}{\partial y_j}(f_1(x), \dots, f_m(x)) \cdot \frac{\partial f_j}{\partial x_i}(x).$$

Proof

The fact that for all $x \in D$, $J_h(x) = J_g(f(x)) \cdot J_f(x)$ follows immediately from Proposition 8.10 and the fact that the differential of a function at a point is represented by the Jacobi matrix of the function at that point. The equality $J_h(x) = J_g(f(x)) \cdot J_f(x)$ is equivalent with

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_1}(x) & \dots & \frac{\partial h_1}{\partial x_n}(x) \\ \frac{\partial h_2}{\partial x_1}(x) & \dots & \frac{\partial h_2}{\partial x_n}(x) \\ \dots & \dots & \dots \\ \frac{\partial h_p}{\partial x_1}(x) & \dots & \frac{\partial h_p}{\partial x_n}(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(f(x)) & \dots & \frac{\partial g_1}{\partial y_m}(f(x)) \\ \frac{\partial g_2}{\partial y_1}(f(x)) & \dots & \frac{\partial g_2}{\partial y_m}(f(x)) \\ \dots & \dots & \dots \\ \frac{\partial g_p}{\partial y_1}(f(x)) & \dots & \frac{\partial g_p}{\partial y_m}(f(x)) \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \dots & \frac{\partial f_2}{\partial x_n}(x) \\ \dots & \dots & \dots \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}$$

Then for all $i=1,\dots,n$ and $k=1,\dots,p$

$\frac{\partial h_k}{\partial x_i}(x)$ is obtained by multiplying the k-th row of $J_g(f(x))$ with the i-th column of $Jf(x)$. Thus

$$\frac{\partial h_k}{\partial x_i}(x) = \sum_{j=1}^{m} \frac{\partial g_k}{\partial y_j}(f(x)) \cdot \frac{\partial f_j}{\partial x_i}(x).$$

□

Corollary 8.35

Let D be open in $\mathbb{R}^n$ and E be open in $\mathbb{R}^m$. Let $f: D \to E$ and $g: E \to \mathbb{R}$ be two functions s.t. f is differentiable on D and g is differentiable on E. Then the function $h := g \circ f : D \to \mathbb{R}$ is differentiable on D and for all $i=1,\dots,n$ and all $x \in D$

$$\frac{\partial h}{\partial x_i}(x) = \sum_{j=1}^{m} \frac{\partial g}{\partial y_j}(f(x)) \cdot \frac{\partial f_j}{\partial x_i}(x)$$

Proof

Apply Proposition 8.34 with $p=1$. □

Corollary 8.36

Let D be open in $\mathbb{R}$ and E be open in $\mathbb{R}^m$. Let $f: D \to E$ and $g: E \to \mathbb{R}$ be two functions s.t. f is differentiable on D and g is differentiable on E. Then the real function $h := g \circ f : D \to \mathbb{R}$ is differentiable on D and for all $x \in D$

$$h'(x) = \sum_{j=1}^{m} \frac{\partial g}{\partial y_j}(f(x)) \cdot f_j'(x).$$

Proof Apply Corollary 8.35 with $n=1$. □

Example 8.37

(i) Let $f: \mathbb{R}^2 \to \mathbb{R}$, $f(x) = \begin{pmatrix} x \\ x \end{pmatrix}$ and $g: \mathbb{R}^2 \to \mathbb{R}$, $g(y_1, y_2) = y_1^{y_2}$.

Define $h = g \circ f : \mathbb{R} \to \mathbb{R}$, $h(x) = g(f(x)) = g(x,x) = x^x$ for all $x \in \mathbb{R}$.

We have that $f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$, where $f_1(x) = f_2(x) = x$ and $f_1'(x) = f_2'(x) = 1$ for all $x \in D$. Furthermore, $\frac{\partial g}{\partial y_1}(y_1, y_2) = y_2 \cdot y_1^{y_2 - 1}$ and $\frac{\partial g}{\partial y_2}(y_1, y_2) = y_1^{y_2} \cdot \log y_1$. It follows that for all $x \in \mathbb{R}$,

$$h'(x) = \sum_{j=1}^{2} \frac{\partial g}{\partial y_j}(f(x)) \cdot f_j'(x) = \frac{\partial g}{\partial y_1}(x,x) \cdot f_1'(x) + \frac{\partial g}{\partial y_2}(x,x) \cdot f_2'(x) = x \cdot x^{x-1} + x^x \cdot \log x = x^x(\log x + 1).$$

(ii) Let $f: \mathbb{R}^2 \to \mathbb{R}^2$, $f(x_1, x_2) = \begin{pmatrix} x_1^2 + x_2^2 \\ x_1 x_2 \end{pmatrix}$, $g: \mathbb{R}^2 \to \mathbb{R}$,

$g(y_1, y_2) = y_1 y_2$.

Define $h = g \circ f : \mathbb{R}^2 \to \mathbb{R}$, $h(x_1, x_2) = g(f(x_1, x_2)) = g(x_1^2 + x_2^2, x_1 x_2) =$
$= (x_1^2 + x_2^2) \cdot x_1 x_2 = x_1^3 x_2 + x_2^3 x_1$.

We have that $f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$, where $f_1(x_1, x_2) = x_1^2 + x_2^2$, $f_2(x_1, x_2) = x_1 x_2$.

Then for all $(a_1, a_2) \in \mathbb{R}^2$,

$$\frac{\partial h}{\partial x_1}(a_1, a_2) = \frac{\partial g}{\partial y_1}(f(a_1, a_2)) \cdot \frac{\partial f_1}{\partial x_1}(a_1, a_2) + \frac{\partial g}{\partial y_2}(f(a_1, a_2)) \cdot \frac{\partial f_2}{\partial x_1}(a_1, a_2)$$

$$= f_2(a_1, a_2) \cdot 2a_1 + f_1(a_1, a_2) \cdot a_2 =$$

$$= 2a_1^2 a_2 + a_2(a_1^2 + a_2^2) = 3a_1^2 a_2 + a_2^3.$$

$$\frac{\partial h}{\partial x_2}(a_1,a_2) = \frac{\partial g}{\partial y_1}(f(a_1,a_2)) \cdot \frac{\partial f_1}{\partial x_2}(a_1,a_2) + \frac{\partial g}{\partial y_2}(f(a_1,a_2)) \cdot \frac{\partial f_2}{\partial x_2}(a_1,a_2)$$

$$= f_2(a_1,a_2) \cdot 2a_2 + f_1(a_1,a_2) \cdot a_1 =$$

$$= 2a_1 a_2^2 + a_1(a_1^2+a_2^2) = 3a_1 a_2^2 + a_1^3.$$

(iii) Let $f: \mathbb{R}^3 \to \mathbb{R}^3$, $f(x,y,z) = \begin{pmatrix} x^2 \\ x^2 y \\ e^z \end{pmatrix}$ and $g: \mathbb{R}^3 \to \mathbb{R}$,

$g(u,v,w) = u^2+v^2-w^2$.

Define $h = g \circ f: \mathbb{R}^3 \to \mathbb{R}$, $h(x,y,z) = x^4(1+y^2) - e^{2z}$.

We have that $f = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}$, where $f_1(x,y,z) = x^2$, $f_2(x,y,z) = x^2 y$, $f_3(x,y,z) = e^z$.

Then for all $(x,y,z) \in \mathbb{R}^3$,

$$\frac{\partial h}{\partial x}(x,y,z) = \frac{\partial g}{\partial u}(f(x,y,z)) \frac{\partial f_1}{\partial x}(x,y,z) + \frac{\partial g}{\partial v}(f(x,y,z)) \cdot \frac{\partial f_2}{\partial x}(x,y,z)$$

$$+ \frac{\partial g}{\partial w}(f(x,y,z)) \cdot \frac{\partial f_3}{\partial z}(x,y,z)$$

$$= 2f_1(x,y,z) \cdot 2x + 2f_2(x,y,z) \cdot 2xy + (-2f_3(x,y,z)) \cdot 0$$

$$= 4x^3 + 4x^3y^2 = 4x^3(1+y^2).$$

$$\frac{\partial h}{\partial y}(x,y,z) = \frac{\partial g}{\partial u}(f(x,y,z)) \cdot \frac{\partial f_1}{\partial y}(x,y,z) + \frac{\partial g}{\partial v}(f(x,y,z)) \cdot \frac{\partial f_2}{\partial y}(x,y,z)$$

$$+ \frac{\partial g}{\partial w}(f(x,y,z)) \cdot \frac{\partial f_3}{\partial y}(x,y,z) = 2f_1(x,y,z) \cdot 0 +$$

$$+ 2f_2(x,y,z) \cdot x^2 + (-2f_3(x,y,z)) \cdot 0 = 2x^4 y.$$

$$\frac{\partial h}{\partial z}(x,y,z) = \frac{\partial g}{\partial u}(f(x,y,z)) \cdot \frac{\partial f_1}{\partial z}(x,y,z) + \frac{\partial g}{\partial v}(f(x,y,z)) \cdot \frac{\partial f_2}{\partial z}(x,y,z)$$

$$+ \frac{\partial g}{\partial w_3}(f(x,y,z)) \cdot \frac{\partial f_3}{\partial z}(x,y,z) = 2f_1(x,y,z) \cdot 0 + 2f_2(x,y,z) \cdot 0 +$$

$$+ (-2f_3(x,y,z)) \cdot e^{z} = -2e^{2z}.$$

Hence, $J_h(x,y,z) = \begin{pmatrix} 4x^3(1+y^2) & 2x^4y & -2e^{2z} \end{pmatrix}$.

Let us compute the derivative of h at $(1,2,0)$. For all $v \in \mathbb{R}^3$ we have that

$$h'(1,2,0)(v) = J_h(1,2,0) \cdot v = \begin{pmatrix} 20 & 4 & -2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 20v_1 + 4v_2 - 2v_3.$$

$\square$

As an application of the Chain Rule, we shall prove the Mean Value Theorem for real-valued functions.

Definition 9.38

For any $a, b \in \mathbb{R}^n$, the line segment joining a and b, denoted by $[a,b]$, is defined as

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

Thus, $[a,b]$ is the image of the affine path joining a and b (see Definition 8.9).

Theorem 9.39 (Mean Value Theorem)

Let D be open in $\mathbb{R}^n$, $f : D \to \mathbb{R}$ be differentiable and $a, b \in D$ be s.t. $[a,b] \subseteq D$. Then there exists $c \in [a,b]$ s.t.

$$f(b) - f(a) = f'(c)(b-a).$$

Proof

Let us consider the path joining a and b: $\gamma : [0,1] \to \mathbb{R}^n$, $\gamma(t) = (1-t)a + tb = t(b-a) + a$. Then γ is an affine function, so γ is differentiable on $[0,1]$ and for all $t \in [0,1]$,

$$\gamma'(t)(v) = (b-a)v \quad \text{for all } v \in \mathbb{R}.$$

Define $F := f \circ \gamma : [0,1] \to \mathbb{R}$. Since f and γ are differentiable, by the Chain Rule (Proposition 9.10) we get that F is differentiable and for all $t \in [0,1]$,

$$F'(t) = f'(\gamma(t)) \circ \gamma'(t) = f'(\gamma(t))(b-a).$$

Since F is differentiable on $[0,1]$ we can apply the Mean Value Theorem from Analysis I to get a $\xi \in]0,1[$ s.t. $F(1)-F(0)=F'(\xi)$.

Since $F(1)=f(\gamma(1))=f(b)$ and $F(0)=f(\gamma(0))=f(a)$, we get that

$$f(b)-f(a)=F'(\xi)=f'(\gamma(\xi))(b-a)=f'(c)(b-a), \text{ where } c=\gamma(\xi)\in[a,b].$$

<u>Remark 3.40</u>

If D is a (closed) ball, then for all $a,b\in D$ we have that $[a,b]\subseteq D$.

<u>Proof</u>

Let $D=U_r(x_0)$ be an open ball and $a,b\in D$. For all $t\in[0,1]$,

$$\|(1-t)a+tb-x_0\| = \|(1-t)a+tb-(1-t)x_0-tx_0\| = \|(1-t)(a-x_0)+t(b-x_0)\|$$
$$\le \|(1-t)(a-x_0)\|+\|t(b-x_0)\| = (1-t)\|a-x_0\|+t\|b-x_0\| < (1-t)r+tr=r,$$

so $(1-t)a+tb\in U_r(x_0)$.

Similarly for $D=\overline{U_r(x_0)}$.

□

<u>Proposition 3.41</u>

Let $D\subseteq\mathbb{R}^n$ be open and let $f:D\to\mathbb{R}^m$. Then the following statements are equivalent:

(i) f is continuously differentiable on D;

(ii) f is partial differentiable on D and all the partial derivatives

$$\frac{\partial f}{\partial x_i}: D\to\mathbb{R}^m, \quad i=1,\dots,n$$

are continuous on D.

<u>Proof</u>

(i) $\Rightarrow$ (ii) Assume that f is continuously differentiable on D.

That is, f is differentiable on D and the map

$$f' : D \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), \quad a \mapsto f'(a)$$

is continuous.

By Proposition 1.28, we have that f is partial differentiable on D.

Let us prove that for all $i = 1, \ldots, n$ the i-th partial derivative

$\frac{\partial f}{\partial x_i} : D \to \mathbb{R}^m$ is continuous.

Let $x, a \in D$. Then

$$\left\| \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) \right\| = \| D_{e_i} f(x) - D_{e_i} f(a) \| \overset{P\,1.21}{=\!=\!=} \| f'(x)(e_i) - f'(a)(e_i) \| \le \| f'(x) - f'(a) \| \cdot \| e_i \| =$$

$$= \| f'(x) - f'(a) \|.$$

Since f' is continuous, $\lim_{x \to a} f'(x) = f'(a)$, so $\lim_{x \to a} \| f'(x) - f'(a) \| = 0$.

It follows that $\lim_{x \to a} \left\| \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) \right\| = 0$, that is $\lim_{x \to a} \frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial x_i}(a)$.

That is, $\frac{\partial f}{\partial x_i}$ is continuous at a. Since $a \in D$ was arbitrary, we get that $\frac{\partial f}{\partial x_i}$ is continuous on D.

(ii) $\Rightarrow$ (i) Assume that f is partial differentiable and all the partial derivatives $\frac{\partial f}{\partial x_i}$, $i = 1..n$ are continuous. By Proposition 1.32, it follows that f is differentiable on D and for all $a \in D$, $f'(a)$ has as associated matrix the Jacobi matrix $Jf(a)$ of f at a.

It remains to prove that $f' : D \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is continuous.

Let $x, a \in D$. Then

$$\|f'(x) - f'(a)\| = \sup_{\|h\| \leq 1} \|f'(x)(h) - f'(a)(h)\| =$$

$$= \sup_{\|h\|\leq 1} \|\partial f(x) \cdot h - \partial f(a) \cdot h\|$$

$$= \sup_{\|h\|\leq 1} \left\| \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) h_i - \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) \cdot h_i \right\|$$

$$= \sup_{\|h\|\leq 1} \cdot |h_i| \cdot \left\| \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) \right) \right\|$$

$$\leq \left\| \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) \right) \right\|$$, if we consider the $\|\cdot\|_\infty$ on $\mathbb{R}^n$, so $|h_i| \leq \|h\| \leq 1$.

$$\leq \sum_{i=1}^{n} \left\| \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) \right\| .$$

Since $\lim_{x \to a} \left\| \frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a) \right\| = 0$ for all $i = 1 \ldots n$, we get that $\lim_{x \to a} \|f'(x) - f'(a)\| = 0$, that is $\lim_{x \to a} f'(x) = f'(a)$.

Thus, f' is continuous on D. $\square$

The Contraction Mapping Principle

Let (X,d) be a metric space. A function $T: X \to X$ is called Lipschitz or lipschitzian if there is a constant $K \geq 0$ s.t. for all $x, y \in X$

(1) $$d(T(x), T(y)) \leq K \cdot d(x,y).$$

The smallest number K for which (1) holds is called the Lipschitz constant of T.

Definition 8.42

A lipschitzian function $T: X \to X$ with Lipschitz constant $K < 1$ is said to be a contraction.

Theorem 8.43 (Banach's Contraction Mapping Principle)

Let (X,d) be a complete metric space and let $T: X \to X$ be a contraction. Then T has a unique fixed point x_0. Moreover, for each $x \in X$,

$$\lim_{n\to\infty} T^n(x) = x_0$$

and in fact for each $x \in X$

$$d(T^n(x), x_0) \leq \frac{K^n}{1-K} d(x, T(x)), \qquad n = 1, 2, \ldots .$$

Proof

Since T is a contraction, we know that for each $x \in X$,

$$d(T(x), T^2(x)) \leq K \cdot d(x, T(x)).$$

Adding $d(x, T(x))$ to both sides of the above yields

$$d(x, T(x)) + d(T(x), T^2(x)) \leq d(x, T(x)) + K \cdot d(x, T(x)),$$

which can be rewritten

$$d(x, T(x)) - K\, d(x, T(x)) \leq d(x, T(x)) - d(T(x), T^2(x)).$$

This in turn is equivalent to

$$d(x, T(x)) \leq \frac{1}{1-K}\left(d(x, T(x)) - d(T(x), T^2(x))\right).$$

Now define the functions $\varphi : X \to \mathbb{R}^+$ by setting $\varphi(x) = \frac{1}{1-K} d(x, T(x))$, for $x \in X$. This gives us the basic inequality

$$d(x, T(x)) \leq \varphi(x) - \varphi(T(x)), \qquad x \in M.$$

Therefore if we fix $x \in X$ and take $m, n \in \mathbb{N}$ with $n < m$, we obtain

$$d(T^n(x), T^{m+1}(x)) \leq \sum_{i=n}^{m} d(T^i(x), T^{i+1}(x)) \leq \varphi(T^n(x)) - \varphi(T^{m+1}(x)).$$

(Notice that the last inequality comes from cancelation in the telescoping sum.) In particular by taking $n=1$ and letting $m \to \infty$ we conclude that

$$\sum_{i=1}^{\infty} d(T^i(x), T^{i+1}(x)) \leq \varphi(T(x)) < \infty.$$

This implies that $(T^n(x))_{n \in \mathbb{N}}$ is a Cauchy sequence. Since X is complete there exists $x_0 \in X$ s.t.

$$\lim_{n \to \infty} T^n(x) = x_0$$

and since T is continuous

$$x_0 = \lim_{n\to\infty} T^n(x) = \lim_{n\to\infty} T^{n+1}(x) = T(x_0).$$

Thus x_0 is a fixed point of T.

In order to see that x_0 is the only fixed point of T, suppose $T(y)=y$.

Then by what we have just shown

$$x_0 = \lim_{n\to\infty} T^n(y) = y.$$

Returning to the inequality

$$d(T^n(x), T^{m+1}(x)) \leq \varphi(T^n(x)) - \varphi(T^{m+1}(x)),$$

upon letting $m\to\infty$ we see that

$$d(T^n(x), x_0) \leq \varphi(T^n(x)) = \frac{1}{1-K} d(T^n(x), T^{n+1}(x)).$$

Since $\frac{1}{1-K} d(T^n(x), T^{n+1}(x)) \leq \frac{K^n}{1-K} d(x, T(x))$ we obtain

$$d(T^n(x), x_0) \leq \frac{K^n}{1-K} d(x, T(x)).$$

This provides an estimate on the rate of convergence for the sequence $(T^n(x))$ which depends only on $d(x, T(x))$. $\square$

Inverse Function Theorem

Motivation

1. Let $f: D \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be differentiable. Note that the dimension of the domain and the range are the same. Suppose $f(a) = b$. Then an approximation to $f(x)$ for x near a is given by the map

$$(*) \qquad x \mapsto f(a) + f'(a)(x-a).$$

We expect that if $f'(a)$ is an invertible linear function, which implies the map in $(*)$ is bijective, then f should be also bijective near a. We shall see from the Inverse Function Theorem that this is true if f is continuously differentiable.

2. Consider the system of equations $f(x) = y$, where $f: D \subseteq \mathbb{R}^n \to \mathbb{R}^n$, $f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$, $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. Thus,

$$\begin{aligned} f_1(x_1, \dots, x_n) &= y_1 \\ f_2(x_1, \dots, x_n) &= y_2 \\ &\vdots \\ f_n(x_1, \dots, x_n) &= y_n. \end{aligned}$$

Suppose that these equations are satisfied if $x = a = (a_1, \dots, a_n)$ and $y = b = (b_1, \dots, b_n)$ and that f is of class C^1 with $f'(a)$ invertible. Then it follows from the Inverse Function Theorem that for all y in some neighborhood of b the equations have a unique solution

x in some neighborhood of a.

Before we proceed with the Inverse Function Theorem, we give without proof the following result.

Proposition 1.44

Let V be a finite dimensional normed space and consider the set

$$M = \{ T \in \mathcal{L}(V,V) : T \text{ invertible} \}.$$

Then

(i) M is open in $\mathcal{L}(V,V)$. More precisely, if $T_0 \in M$ and $\|T - T_0\| < \frac{1}{2\|T_0^{-1}\|}$ then $T \in M$ and $\|T^{-1}\| \le 2\|T_0^{-1}\|$.

(ii) the function $\varphi : M \to M$, $\varphi(T) = T^{-1}$ is continuous.

Theorem 1.45 (Inverse Function Theorem)

Let D be open in $\mathbb{R}^n$ and $f : D \to \mathbb{R}^n$ be continuously differentiable. Assume that $f'(a)$ is invertible for some $a \in D$.

Then there exist an open neighborhood U of a and an open neighborhood V of $b = f(a)$ s.t.

(i) $f'(x)$ is invertible at every $x \in U$

(ii) $f|_U : U \to V$ is bijective, and hence has an inverse $g : V \to U$

(iii) g is continuously differentiable and

$$g'(f(x)) = (f'(x))^{-1} \quad \text{for every } x \in U.$$

Proof

We consider $\|\cdot\|_\infty$ on $\mathbb{R}^n$.

Step 1

We prove that there exists an open neighborhood E of a s.t. $f'(x)$ is invertible for all $x \in E$.

Since $f'(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ and $f'(a)$ is represented by the Jacobi matrix $\mathcal{J}f(a)$, we have that $f'(a)$ is invertible iff $\det \mathcal{J}f(a) \neq 0$. Let us consider the function

$$G : D \to \mathbb{R}, \quad G(x) = \det \mathcal{J}f(x) = \begin{vmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \\ \frac{\partial f_n}{\partial x_1}(x) & \cdots & \frac{\partial f_n}{\partial x_n}(x) \end{vmatrix}.$$

Since f is continuously differentiable, we get from Proposition 8.41 that all the partial derivatives $\frac{\partial f_j}{\partial x_i} : D \to \mathbb{R}$, $i, j = 1, \dots, n$ are continuous, hence all the coefficients of the matrix $\mathcal{J}f(x)$ are continuous. Since $G(x) = \det \mathcal{J}f(x)$ consists of sums and products of these coefficients, it follows that G is a continuous function. We have that $G(a) = \det \mathcal{J}f(a) \neq 0$, so by continuity, there exists an open neighborhood E of a s.t. $G(x) \neq 0$ for all $x \in E$. That is, $f'(x)$ is invertible at every $x \in E$.

Step 2

Since we can consider in the sequel the restriction $f|_E : E \to \mathbb{R}^n$ of f the open neighborhood E of a, we can assume without loss of generality that $D = E$. Thus, $f'(x)$ is invertible at every $x \in D$.

Step 3

Let
$$y \in B_\delta(f(a)),$$
where $\delta > 0$ will be chosen later. (We will take the set V in the theorem to be $B_\delta(f(a))$.)

For each such y, we want to prove the existence of $x \in D$ s.t.

(1) $$f(x) = y.$$

Since f is differentiable at a with derivative $f'(a)$, there exists $r: D \to \mathbb{R}^n$ s.t.
$$f(x) = f(a) + f'(a)(x-a) + r(x), \quad \text{and} \quad \lim_{x \to a} \frac{r(x)}{\|x-a\|} = 0.$$

Then we obtain the following chain of equivalences:

$$f(x) = y$$
$$\Updownarrow$$
$$f(x) + f'(a)(x-a) + r(x) = y$$
$$\Updownarrow$$
$$f'(a)(x-a) = y - f(a) - r(x)$$
$$\Updownarrow$$
$$x - a = (f'(a))^{-1}(f'(a)(x-a)) = (f'(a))^{-1}(y - f(a) - r(x)) =$$
$$= (f'(a))^{-1}(y - f(a)) - (f'(a))^{-1}(r(x))$$
$$\Updownarrow$$
$$x = a + (f'(a))^{-1}(y - f(a)) - (f'(a))^{-1}(r(x)).$$

For brevity, let us denote $T := (f'(a))^{-1}$. Then (1) $\Longleftrightarrow$

(2) $$x = a + T(y - f(a)) - T(r(x)).$$

Thus, we want to find $x \in D$ s.t. (2) is satisfied.

Step 4

We define $\phi_y : D \to \mathbb{R}^n$,

$$\phi_y(x) = a + T(y - f(a)) - T(r(x)).$$

Note the following:

x is a fixed point of $\phi_y \iff \phi_y(x) = x \iff$ (2) holds $\iff$

$\iff$ x solves the equation (1).

We claim that

$$\phi_y\left(\overline{U_\varepsilon(a)}\right) \subseteq U_\varepsilon(a)$$

and that ϕ_y is a contraction, provided $\varepsilon > 0$ is sufficiently small (ε will depend only on f and a) and provided $y \in B_\delta(f(a))$, where $\delta > 0$ also depends only on f and a.

To prove the claim, let $\varepsilon > 0$ and $x_1, x_2 \in \overline{U_\varepsilon(a)}$. Then

$$\phi_y(x_1) - \phi_y(x_2) = (a + T(y - f(a)) - T(r(x_1))) - (a + T(y - f(a)) - T(r(x_2)))$$
$$= T(r(x_2)) - T(r(x_1)) = T(r(x_2) - r(x_1)).$$

and so

(3) $$\| \phi_y(x_1) - \phi_y(x_2) \| = \| T(r(x_2) - r(x_1)) \| \le \|T\| \cdot \| r(x_2) - r(x_1) \|$$
$$= K \cdot \| r(x_2) - r(x_1) \|, \text{ where } K := \|T\|.$$

We have that

$$r(x_2)-r(x_1)=\Big(f(x_2)-f(a)-f'(a)(x_2-a)\Big)-\Big(f(x_1)-f(a)-f'(a)(x_1-a)\Big)$$
$$=f(x_2)-f(x_1)-f'(a)(x_2-x_1).$$

Since $r, f: D \to \mathbb{R}^n$, $r=\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}$, $f=\begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$, where $r_i, f_i: D\to \mathbb{K}$ for all $i=1,\dots,n$. Moreover, by Proposition 8.15, $f'(a)=\begin{pmatrix} f_1'(a) \\ \vdots \\ f_n'(a) \end{pmatrix}$.

Then we get that for all $i=1,\dots,n$

$$r_i(x_2)-r_i(x_1)=f_i(x_2)-f_i(x_1)-f_i'(a)(x_2-x_1).$$

Since f_i is differentiable and $x_1,x_2\in\overline{U_\varepsilon(a)}$ implies $[x_1,x_2]\in\overline{U_\varepsilon(a)}$ (by Remark 9.40), we can apply the Mean Value Theorem 9.39 to get $c_i\in[x_1,x_2]$ s.t.

$$f_i(x_2)-f_i(x_1)=f_i'(c_i)(x_2-x_1).$$

It follows that

$$r_i(x_2)-r_i(x_1)=f_i'(c_i)(x_2-x_1)-f_i'(a)(x_2-x_1)=$$
$$=\big(f_i'(c_i)-f_i'(a)\big)(x_2-x_1).$$

Hence,

(4) $$|r_i(x_2)-r_i(x_1)|=\big|\big(f_i'(c_i)-f_i'(a)\big)(x_2-x_1)\big|\le\|f_i'(c_i)-f_i'(a)\|\cdot\|x_2-x_1\|$$

Since f' is continuous by the hypothesis, we get that f_i' is also continuous at a, so there exists $\varepsilon_i>0$ (depending only on f_i and a) s.t.

$$(\forall x\in D)\left(x\in\overline{U_{\varepsilon_i}(a)}\Rightarrow\|f_i'(x)-f_i'(a)\|\le\frac{1}{2K}\right).$$

Let $\varepsilon := \min_{i=1,\dots,n} \varepsilon_i$. Then $\overline{U_\varepsilon(a)} \subseteq \overline{U_{\varepsilon_i}(a)}$, so for any $i=1,\dots,n$

(5) $(\forall x \in D)\left(x \in \overline{U_\varepsilon(a)} \Rightarrow \|f_i'(x) - f_i'(a)\| \leq \frac{1}{2K}\right)$.

Using (4) and (5) we get that for all $x_1, x_2 \in \overline{U_\varepsilon(a)}$ and for all $i=1,\dots,n$,

$$|r_i(x_2) - r_i(x_1)| \leq \|f_i'(c_i) - f_i'(a)\| \cdot \|x_1 - x_2\| \leq \frac{1}{2K} \cdot \|x_1 - x_2\|,$$

since $c_i \in [x_1, x_2] \subseteq \overline{U_\varepsilon(a)}$.

Thus, $\|r(x_2) - r(x_1)\| \leq \frac{1}{2K} \cdot \|x_1 - x_2\|$

By (3), it follows that

$$\|\Phi_y(x_1) - \Phi_y(x_2)\| \leq K \cdot \|r(x_2) - r(x_1)\| \leq K \cdot \frac{1}{2K} \cdot \|x_1 - x_2\|$$

$$= \frac{1}{2}\|x_1 - x_2\|.$$

This proves that $\Phi_y|_{\overline{U_\varepsilon(a)}} : \overline{U_\varepsilon(a)} \to \mathbb{R}^n$ is a contraction.

It remains to prove that $\Phi_y(\overline{U_\varepsilon(a)}) \subseteq U_\varepsilon(a)$.

For this, let $x \in \overline{U_\varepsilon(a)}$. Then

$$\|\Phi_y(x) - a\| = \|a + T(y - f(a)) - T(r(x)) - a\| = \|T(y - f(a)) - T(r(x))\|$$
$$\leq \|T(y - f(a))\| + \|T(r(x))\| \leq \|T\| \cdot \|y - f(a)\| +$$
$$+ \|T\| \cdot \|r(x)\| = K \cdot \|y - f(a)\| + K \cdot \|r(x) - r(a)\|,$$

since $r(a) = 0$.

$$\leq K \cdot \|y - f(a)\| + K \cdot \frac{1}{2K}\|\cancel{x} - a\| \leq K \cdot \delta + \frac{\varepsilon}{2}.$$

Let us choose $\delta > 0$ s.t. $K\delta < \frac{\varepsilon}{2}$, so $\delta > \frac{\varepsilon}{2K}$. Then for all $y \in B_\delta(f(a))$, we have that $\|\Phi_y(x) - a\| < \varepsilon$ for all $x \in \overline{U_\varepsilon(a)}$, that is $\Phi_y(x) \in U_\varepsilon(a)$ for all $x \in \overline{U_\varepsilon(a)}$.

Step 5

Thus, we have found $\varepsilon > 0$ and $\delta > 0$ depending only on f_0 and a s.t. for all $y \in B_\delta(f(a))$

$$\phi_y : \overline{U_\varepsilon(a)} \to U_\varepsilon(a), \quad \phi_y(x) = a + T(y - f(a)) - T(r(x))$$

is a contraction. Since $\overline{U_\varepsilon(a)}$ is closed in the complete metric space $\mathbb{R}^n$ with the distance induced by $\|\cdot\|_\infty$, it follows by Proposition 6.15 that $\overline{U_\varepsilon(a)}$ is complete. So, we can apply Banach's Contraction Mapping Principle (Theorem 8.43) to get that there exists a unique $x \in \overline{U_\varepsilon(a)}$ s.t. $\phi_y(x) = x$. Moreover, since $\phi_y(x) \in U_\varepsilon(a)$ we get that $x \in U_\varepsilon(a)$. Thus,

for any $y \in B_\delta(f(a))$ there is a unique $x \in B_\varepsilon(a)$ s.t. $f(x) = y$.

Step 6

Let $U := B_\varepsilon(a) \cap f^{-1}(V)$ and $V := B_\delta(f(a))$. Then U, V are open, $b \in V$ and $a \in U$, since $a \in B_\varepsilon(a)$ and $f(a) = b \in V$, so $a \in f^{-1}(V)$.

From the definition of U, we have that for all $x \in U$, $f(x) \in V$, so $f(U) \subseteq V$. Thus, $f|_U : U \to V$ is well-defined. Since we have proved in Step 5 that for all $y \in V$ there exists a unique $x \in U$ s.t. $f(x) = y$, it follows that $f|_U : U \to V$ is bijective.

Step 7

Let $g : V \to U$ be the inverse of f. Thus for all $y \in V$, we have that $g(y) = x$, where x is unique s.t. $f(x) = y$ or, equivalently, $\phi_y(x) = x$.

We prove that g is Lipschitz, hence continuous. For all $y_1, y_2 \in V$, if we denote $x_1 = g(y_1)$ and $x_2 = g(y_2)$, we get that

$$\begin{aligned}
\|g(y_1) - g(y_2)\| &= \|x_1 - x_2\| = \|\Phi_{y_1}(x_1) - \Phi_{y_2}(x_2)\| = \\
&= \|(a + T(y_1 - f(a)) - T(r(x_1))) - (a + T(y_2 - f(a)) - T(r(x_2)))\| \\
&= \|T(y_1 - y_2) - T(r(x_1) - r(x_2))\| \le \|T\| \cdot \|y_1 - y_2\| + \\
&\quad + \|T\| \cdot \|r(x_1) - r(x_2)\| \le K \cdot \|y_1 - y_2\| + K \cdot \frac{1}{2K} \|x_1 - x_2\| \\
&= K \cdot \|y_1 - y_2\| + \frac{1}{2} \|g(y_1) - g(y_2)\|.
\end{aligned}$$

Thus,

$$\frac{1}{2} \|g(y_1) - g(y_2)\| \le K \cdot \|y_1 - y_2\|$$

and so

$$\|g(y_1) - g(y_2)\| \le 2K \cdot \|y_1 - y_2\|.$$

Step 8

Since g is continuous on V and $f'(x)$ is invertible for all $x \in V$, we can apply Proposition 9.41 to get that g is differentiable on V with derivative

$$g'(y) = [f'(g(y))]^{-1} \quad \text{for all } x \in V,$$

so $g'(f(x)) = (f'(x))^{-1}$ for all $x \in U$.

Step 9.

It remains to prove that g' is continuous.

This follows immediately by using Proposition 9.44 and remarking that $g' = \varphi \circ f' \circ g$, where φ is the continuous function defined in Proposition 9.44. Since φ is continuous, g is continuous as we proved in Step 7 and f' is continuous by the hypothesis, it follows that g' is continuous.

□

Remark 9.46

Local invertibility does not imply global invertibility.

Proof

Let us consider the following example.

$$f : \underbrace{\{(x,y) \in \mathbb{R}^2 : y > 0\}}_{= D} \to \mathbb{R}^2, \quad f(x,y) = \begin{pmatrix} y \cos x \\ y \sin x \end{pmatrix}.$$

Then f is continuously differentiable with $Jf(x,y) = \begin{pmatrix} -y\sin x & \cos x \\ y\cos x & \sin x \end{pmatrix}$

for all $(x,y) \in D$. Then

$$\det Jf(x,y) = -y\sin^2 x - y\cos^2 x = -y \neq 0$$

for all $(x,y) \in D$. Consequently, f is locally invertible at every point in D.

Yet, f is not globally invertible, since f is 2π-periodic with respect to x, so f is not injective.

□