Multivariable Calculus Notes

Contents

1 Total Derivatives

Definition 1. Let V and W be finite dimensional normed vector spaces. Let U be an open subset of V. We say that a function $F : U \to W$ is **differentiable at** $a \in U$ if there is a linear map $L : V \to W$ such that

$$\lim_{v \to 0} \frac{|F(a+v) - F(a) - Lv|}{|v|} = 0. \tag{1}$$

Remark 1. Equivalently, $F : U \to W$ is differentiable at $a \in U$ iff there is a linear map $L : V \to W$ such that

$$\lim_{v \to a} \frac{|F(v) - F(a) - L(v-a)|}{|v-a|} = 0.$$

Proposition 1. *If $F : U \to W$ is differentiable at $a \in U$, then the linear map L satisfying equation* (1) *is unique.*

Proof. Let L and L' be two such linear maps. Then $L0 = 0 = L'0$ by linearity. Now let $v \in V - \{0\}$, and let $\epsilon > 0$. Then there is $\delta > 0$ such that when $0 < |u| < \delta$,

$$\frac{|F(a+u) - F(a) - Lu|}{|u|} < \epsilon$$

and

$$\frac{|F(a+u) - F(a) - L'u|}{|u|} < \epsilon.$$

Let $t = \frac{\delta}{2|v|}$ and let $u = tv$. Then

$$0 < |u| = t|v| = \frac{\delta}{2} < \delta.$$

Therefore

$$\begin{aligned} |Lv - L'v| &\leq \frac{|u|}{t}\left(\frac{|Lu - F(a+u) + F(a)|}{|u|} + \frac{|F(a+u) - F(a) - L'u|}{|u|}\right) \\ &< 2\epsilon|v| \end{aligned}$$

for all $\epsilon > 0$. Hence $Lv = L'v$ for all $v \in V$, so L is unique. □

Definition 2. If F is differentiable at a, the linear map L satisfying equation (1) is denoted by $DF(a)$ and is called the **total derivative of F at a**.

Remark 2. Equation (1) can be rewritten as

$$F(a+v) = F(a) + DF(a)v + R_F(v) \tag{2}$$

where $R_F(v) = F(a+v) - F(a) - DF(a)v$ satisfies $|R(v)|/|v| \to 0$ as $v \to 0$.

Proposition 2. *Let V, W, and X be finite dimensional vector spaces. Let $U \subset V$ be open. Let $a \in U$. Let $F, G : U \to W$, and let $f, g : U \to \mathbb{R}$. Then*

1. *If F is differentiable at a, then F is continuous at a.*
2. *If F is constant, then F is differentiable at a and $DF(a) = 0$.*
3. *If F and G are differentiable at a, and if $c \in \mathbb{R}$, then $cF+G$ is differentiable at a and*
$$D(cF + G)(a) = cDF(a) + DG(a).$$
4. *If f and g are differentiable at a, then fg is differentiable at a, and*
$$D(fg)(a) = f(a)Dg(a) + g(a)Df(a).$$
5. *If f and g are differentiable at a, and if $g(a) \neq 0$, then f/g is differentiable at a and*
$$D(f/g)(a) = \frac{g(a)Df(a) - f(a)Dg(a)}{g(a)^2}.$$

6. *If* $T : V \to W$ *is linear, then* T *is differentiable at every* $v \in V$*, with* $DT(v) = T$.

7. *If* $B : V \times W \to X$ *is bilinear, then* B *is differentiable at every* $(v, w) \in V \times W$*, and*
$$DB(v,w)(x,y) = B(v,y) + B(x,w).$$

Proof. 1. Since $a \in U$ is open, there is a neighborhood N of 0 such that $a + v \in U$ for all $v \in N$. Then for any $v \in N - \{0\}$,

$$\begin{aligned}|F(a+v) - F(a)| &= \frac{|F(a+v) - F(a) - DF(a)v|}{|v|}|v| + |DF(a)v| \\ &\le |v|(|R(v)|/|v| + |DF(a)|)\end{aligned}$$

Since $|R(v)|/|v| + |DF(a)| \to |DF(a)|$ and $|v| \to 0$ as $v \to 0$, we conclude that F is continuous at a.

2. Since $a \in U$ is open, there is a neighborhood N of 0 such that $a + v \in U$ for all $v \in N$. Then for any $v \in N - \{0\}$,

$$\frac{|F(a+v) - F(a) - 0v|}{|v|} = 0.$$

Therefore $0 : V \to W$ satisfies the differentiability condition, so F is differentiable at a. By uniqueness, $DF(a) = 0$.

3. Since $a \in U$ is open, there is a neighborhood N of 0 such that $a + v \in U$ for all $v \in N$. Then the conclusion follows from the fact that

$$\frac{|N(v)|}{|v|} \le c\frac{|R_F(v)|}{|v|} + \frac{|R_G(v)|}{|v|}$$

for all $v \in N - \{0\}$, where

$$N(v) = (cF + G)(a+v) - (cF + G)(a) - cDF(a)v - DG(a)v,$$

$R_F(v) = F(a+v) - F(a) - DF(a)v$, and $R_G(v) = G(a+v) - G(a) - DG(a)v$.

4. Since $a \in U$ is open, there is a neighborhood N of 0 such that $a + v \in U$ for all $v \in N$. Let $v \in N - \{0\}$. We have that

$$f(a+v) = f(a) + Df(a)v + R_f(v)$$

where
$$R_f(v) = f(a+v) - f(a) - Df(a)v.$$

Similarly,
$$g(a+v) = g(a) + Dg(a)v + R_g(v)$$

where
$$R_g(v) = g(a+v) - g(a) - Dg(a)v.$$

Hence

$$(fg)(a+v) = (fg)(a) + f(a)Dg(a)v + g(a)Df(a)v + R(v),$$

where

$$\begin{aligned} R(v) = f(a)R_g(v) + Df(a)vDg(a)v + Df(a)vR_g(v) + \\ R_f(v)g(a) + R_f(v)Dg(a)v + R_f(v)R_g(v). \end{aligned}$$

We have that $|R_g(v)|/|v| \to 0$ and $|R_f(v)|/|v| \to 0$ as $v \to 0$. We also have that $|R_g(v)| \to 0$ and

$$|Df(a)vDg(a)v|/|v| \leq |Df(a)||Dg(a)||v| \to 0$$

as $v \to 0$. From this, we see that $|R(v)|/|v| \to 0$ as $v \to 0$. Since we also have that

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

this proves the result.

5. Since $a \in U \subset V$ is open, since $g(a) \neq 0$, and since g is continuous at a, there is an open neighborhood $N \subset V$ of 0 such that $a+v \in U$ for all $v \in N$ and $g(a+v) \neq 0$ for all $v \in N$. Then $a + N = \{a + v : v \in N\} \subset U$ is an open neighborhood of a. Let $h : a + N \to \mathbb{R}$ be defined by $h(u) = 1/g(u)$ for all $u \in a + N$. We will show that h is differentiable at a and

$$Dh(a) = -\frac{1}{g(a)^2} Dg(a).$$

Indeed, we have that for any $v \in N - \{0\}$,

$$\begin{aligned} h(a+v) - h(a) &= 1/g(a+v) - 1/g(a) \\ &= \frac{g(a) - g(a+v)}{g(a)g(a+v)} \\ &= -\frac{1}{g(a)g(a+v)}(Dg(a)v + R_g(v)). \end{aligned}$$

Then

$$h(a+v) - h(a) + \frac{1}{g(a)^2} Dg(a)v = R(v),$$

where

$$R(v) = \left(\frac{1}{g(a)^2} - \frac{1}{g(a)g(a+v)}\right) Dg(a)v - \frac{1}{g(a)g(a+v)} R_g(v).$$

Since

$$|Dg(a)v|/|v| \leq |Dg(a)|$$

and since $|R_g(v)|/|v| \to 0$ and

$$\left| \frac{1}{g(a)^2} - \frac{1}{g(a)g(a+v)} \right| \to 0$$

as $v \to 0$, we conclude that

$$\frac{\left| h(a+v) - h(a) + \frac{1}{g(a)^2} Dg(a)v \right|}{|v|} = \frac{|R(v)|}{|v|} \to 0$$

as $v \to 0$. Hence h is differentiable at a and

$$Dh(a) = -\frac{1}{g(a)^2} Dg(a).$$

Now since $f/g = fh$ on $a + N$, we have that f/g is differentiable at a and

$$D(f/g)(a) = \frac{1}{g(a)} Df(a) - \frac{f(a)}{g(a)^2} Dg(a) = \frac{g(a)Df(a) - f(a)Dg(a)}{g(a)^2}.$$

6. For any $v \in V$, any $v' \in V - \{0\}$,

$$\frac{|T(v - v') - Tv - Tv'|}{|v'|} = 0,$$

so the result follows immediately.

7. For any $v \in V$, $w \in W$, $(x, y) \in V \times W - \{(0, 0)\}$, we have that

$$\frac{|B(v + x, w + y) - B(v, w) - B(v, y) - B(x, w)|}{|(x, y)|} = \frac{|B(x, y)|}{|(x, y)|}.$$

We will show that $|B(x, y)|/|(x, y)| \to 0$ as $(x, y) \to 0$, which will prove the result.

First, suppose that $\{e_1, \ldots, e_n\}$ is a basis for V and $\{f_1, \ldots, f_m\}$ is a basis for W. Suppose that V and W are endowed with the ℓ^∞-norms with respect to these bases, ie

$$|v| = |\sum_i \alpha_i e_i| = \max_i |\alpha_i|$$

for all $v = \sum_i \alpha_i e_i$ in V and

$$|w| = |\sum_j \beta_j f_j| = \max_j |\beta_j|$$

for all $w = \sum_j \beta_j f_j \in W$. Also suppose that we have that ℓ^∞ product norm on $V \times W$, ie $|(v, w)| = \max\{|v|, |w|\}$ for all $v \in V$, $w \in W$. Let $x =$

$\sum_i x_i e_i$ and $y = \sum_j y_j f_j$ be arbitrary such that $(x, y) \in V \times W - \{(0, 0)\}$. Then we have that

$$B(x, y) = \sum_{i,j} x_i y_j B(e_i, f_j),$$

so that

$$\begin{aligned}|B(x, y)|/|(x, y)| &\leq nm \max_{i,j} |B(e_i, f_j)| \frac{|x||y|}{\max\{|x|, |y|\}} \\ &\leq nm \max_{i,j} |B(e_i, f_j)| \min\{|x|, |y|\}\end{aligned}$$

and the last quantity converges to 0 as (x, y) converges to $(0, 0)$. This proves the result when V, W, and $V \times W$ all have the norms that we specified. Now if V, W, and $V \times W$ all have arbitrary norms on them, since all norms on a finite dimensional vector space are equivalent, the general result follows from what we just showed. In other words,

$$|B(x, y)|/|(x, y)| \to 0$$

as $(x, y) \to (0, 0)$ independent of choice of norms for V, W, $V \times W$, and X. This completes the proof.

□

Proposition 3 (Chain Rule for Total Derivatives)**.** *Let V, W, and X be finite dimensional vector spaces. Let $U \subset V$ and $\tilde{U} \subset W$ be open subsets. Let $F : U \to \tilde{U}$ and $G : \tilde{U} \to X$. If F is differentiable at $a \in U$ and G is differentiable at $F(a) \in \tilde{U}$, then $G \circ F$ is differentiable at a with*

$$D(G \circ F)(a) = DG(F(a)) \circ DF(a).$$

Proof. Since $F(a) \in \tilde{U}$ is open, there is an open neighborhood $\tilde{N} \subset W$ of 0 such that $F(a) + w \in \tilde{U}$ for all $w \in \tilde{N}$. Hence $F(a) + \tilde{N} = \{F(a) + w : w \in N\} \subset \tilde{U}$ is an open neighborhood of $F(a)$. Since F is continuous at a and $a \in U$ is open, there is an open neighborhood $N \subset V$ of 0 such that $a + v \in U$ and $F(a + v) \in F(a) + \tilde{N}$ for all $v \in N$. Then for any $v \in N - \{0\}$, we have that

$$w(v) = F(a + v) - F(a) = DF(a)v + R_F(v) \in \tilde{N}.$$

Therefore, for $v \in N - \{0\}$ such that $w(v) \neq 0$,

$$\begin{aligned}G(F(a + v)) - G(F(a)) &= G(F(a) + w(v)) - G(F(a)) \\ &= DG(F(a))w(v) + R_G(w(v)) \\ &= DG(F(a))DF(a)v + DG(F(a))R_F(v) + R_G(w(v)).\end{aligned}$$

Hence

$$G(F(a + v)) - G(F(a)) - DG(F(a))DF(a)v = R(v),$$

where

$$R(v) = DG(F(a))R_F(v) + R_G(w(v)).$$

We have that $|R_F(v)|/|v| \to 0$ as $v \to 0$. We also have that

$$\begin{aligned} \frac{|R_G(w(v))|}{|v|} &= \frac{|R_G(w(v))|}{|v|(|DF(a)| + |R_F(v)|/|v|)} \left(|DF(a)| + \frac{|R_F(v)|}{|v|}\right) \\ &\leq \frac{|R_G(w(v))|}{|DF(a)v + R_F(v)|} \left(|DF(a)| + \frac{|R_F(v)|}{|v|}\right) \\ &= \frac{|R_G(w(v))|}{|w(v)|} \left(|DF(a)| + \frac{|R_F(v)|}{|v|}\right). \end{aligned}$$

Now since $w(v) \to 0$ as $v \to 0$, and since $|R_G(w)|/|w| \to 0$ as $w \to 0$, the inequality above shows that

$$\frac{|G(F(a+v)) - G(F(a)) - DG(F(a))DF(a)v|}{|v|} = \frac{|R(v)|}{|v|} \to 0$$

as $v \to 0$. This completes the proof. □

2 Partial Derivatives

Definition 3. Let $U \subset \mathbb{R}^n$ be open, and let $f : U \to \mathbb{R}$. Let $e_1, \ldots, e_n$ be the standard basis vectors of $\mathbb{R}^n$. For any $a \in U$ and any $j \in \{1, \ldots, n\}$, the **jth partial derivative of f at** a is

$$\frac{\partial f}{\partial x^j}(a) = \lim_{h \to 0} \frac{f(a + he_j) - f(a)}{h}$$

if the limit exists.

Remark 3. We can use any symbol in place of x in the notation above.

Definition 4. Let $U \subset \mathbb{R}^n$ be open, and let $F : U \to \mathbb{R}^m$. The partial derivatives of F are the partial derivatives of the **component functions** $F^i : U \to \mathbb{R}$ where $F(x) = (F^1(x), \ldots, F^m(x))$ for all $x \in U$. The matrix $(\partial F^i/\partial x^j)$ of partial derivatives is called the **Jacobian matrix of** F, and its determinant is the **Jacobian determinant of** F.

Proposition 4. *Let $U \subset \mathbb{R}^n$ be open, and let $F : U \to \mathbb{R}^m$. If F is differentiable, then each of its partial derivatives exist at all points of U, and for each $a \in U$, the matrix representing $DF(a)$ with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is the Jacobian matrix $(\partial F^i/\partial x^j(a))$.*

Proof. Let $a \in U$ and let $j \in \{1, \ldots, n\}$. Since U is open, there is an $\epsilon > 0$ such that when $|t| < \epsilon$, $a + te_j \in U$. Then for all $0 < |t| < \epsilon$,

$$F(a + te_j) - F(a) = tDF(a)e_j + R_F(te_j).$$

Then for each i,

$$\frac{F^i(a+te_j)-F^i(a)}{t} = (DF(a))^i_j + \frac{R_F(te_j)^i}{t}. \tag{3}$$

Observe that for any norm $|\cdot|$ on $\mathbb{R}^m$, there is a constant $C>0$ such that for all $x \in \mathbb{R}^m$, all $i \in \{1,\ldots,m\}$, $|x_i| \le C|x|$. Indeed, this holds for $C=1$ with the ℓ^∞ norm on $\mathbb{R}^m$, and since all norms are equivalent on $\mathbb{R}^m$, the general result follows. In particular, there is a constant $C>0$ such that

$$|R_F(te_j)^i|/|t| \le C|R_F(te_j)|/|te_j|$$

for all $0<|t|<\epsilon$.

Then since

$$\frac{|R_F(te_j)|}{|te_j|} \to 0$$

as $t \to 0$, taking the limit as $t \to 0$ in equation (3) implies that

$$\frac{\partial F^i}{\partial x^j}(a) = (DF(a))^i_j$$

for all i, j, and a as desired. □

Proposition 5. *Let $U \subset \mathbb{R}^n$ be open, and let $F : U \to \mathbb{R}^m$. Then F is differentiable iff each component function $F^i : U \to \mathbb{R}$ is differentiable, where the F^i satisfy $F(x) = (F^1(x),\ldots,F^n(x))$ for all $x \in U$.*

Proof. If F is differentiable, then for each $a \in U$, the linear map $DF(a) : \mathbb{R}^n \to \mathbb{R}^m$ exists, and its standard matrix is given by

$$(DF(a))^i_j = \frac{\partial F^i}{\partial x^j}(a).$$

Then since $a \in U$ is open, there is an $\epsilon > 0$ such that $a+v \in U$ for all $|v|<\epsilon$. Then for each i and each $0<|v|<\epsilon$, we have that

$$F^i(a+v) - F^i(a) = \sum_j \frac{\partial F^i}{\partial x^j}(a)v^j + R_F(v)^i.$$

Then the linear map $v \mapsto \sum_j \partial F^i/\partial x^j(a)v^j$ from $\mathbb{R}^n$ to $\mathbb{R}$ satisfies

$$F^i(a+v) - F^i(a) - \sum_j \frac{\partial F^i}{\partial x^j}(a)v^j = R_F(v)^i \tag{4}$$

for all $0<|v|<\epsilon$. From equivalence of norms on $\mathbb{R}^m$, there is a constant $C>0$ such that

$$|R_F(v)^i|/|v| \le C|R_F(v)|/|v| \to 0$$

as $v \to 0$. Therefore equation (4) implies that each F^i is differentiable at each $a \in U$.

Conversely, if each F^i is differentiable, then for each $a \in U$, the linear map $DF^i(a) : \mathbb{R}^n \to \mathbb{R}$ exists and its standard matrix is given by

$$(DF^i(a))_j = \frac{\partial F^i}{\partial x^j}(a).$$

Then for $a \in U$ and for v sufficiently small where $a + v \in U$,

$$F^i(a+v) - F^i(a) - \sum_j \frac{\partial F^i}{\partial x^j}(a)v^j = R_{F^i}(v)$$

for each i, where $|R_{F^i}(v)|/|v| \to 0$ as $v \to 0$. For each v sufficiently small, let $R(v)$ be the vector in $\mathbb{R}^m$ given by

$$R(v)^i = R_{F^i}(v).$$

Also, for each $a \in U$, let $L(a) : \mathbb{R}^n \to \mathbb{R}^m$ be the linear map given by

$$(L(a)v)^i = \sum_j \frac{\partial F^i}{\partial x^j}(a)v^j$$

for all i. Then we have that for all $a \in U$, for all v sufficiently small,

$$F(a+v) - F(a) - L(a)v = R(v).$$

Observe that $|R(v)|/|v| \to 0$ as $v \to 0$ when $\mathbb{R}^m$ is given the ℓ^1 norm. Since all norms on $\mathbb{R}^m$ are equivalent, we then conclude that $|R(v)|/|v| \to 0$ independent of choice of norms on $\mathbb{R}^m$ and $\mathbb{R}^n$. Hence F is differentiable at all $a \in U$. □

Remark 4. The proof of the previous proposition also shows that

$$(DF(a))^i_j = (DF^i(a))_j,$$

that is, the i-th row of the standard matrix of $DF(a)$ corresponds to the standard matrix of $DF^i(a)$, provided that either F is differentiable at a or all F^i are differentiable at a.

3 Continuously Differentiable Functions

Definition 5. Let $U \subset \mathbb{R}^n$ be open. If $F : U \to \mathbb{R}^m$ is a function where each of its partial derivatives exist at all points of U, and each of the functions $\partial F^i/\partial x^j : U \to \mathbb{R}$ so defined are continuous, then F is said to be of **class** C^1 or **continuously differentiable**.

Remark 5. It follows immediately from the definitions that a function $F : U \to \mathbb{R}^m$ defined on an open subset U of $\mathbb{R}^n$ is C^1 iff each $F^i : U \to \mathbb{R}$ is C^1.

Proposition 6. *Let $U \subset \mathbb{R}^n$ be open. If $F : U \to \mathbb{R}^m$ is C^1, then F is differentiable at each point of U.*

Proof. First suppose that $m = 1$ and $n = 2$. Let $a = (a^1, a^2) \in U$. Since U is open, there is an $\epsilon > 0$ such that when $v \in B(0, \epsilon) - \{0\}$, $a + v \in U$. Given $v = (v_1, v_2)$ such that $0 < |v| < \epsilon$, we have that

$$F(a+v)-F(a) = [F(a^1+v^1, a^2+v^2)-F(a^1, a^2+v^2)]+[F(a^1, a^2+v^2)-F(a^1, a^2)].$$

Since F is C^1, we can apply the mean value theorem twice to conclude that there is $w^1(v)$ between a^1 and $a^1 + v^1$ and $w^2(v)$ between a^2 and $a^2 + v^2$ such that

$$F(a + v) - F(a) = \frac{\partial F}{\partial x^1}(w^1(v), a^2 + v^2)v^1 + \frac{\partial F}{\partial x^2}(a^1, w^2(v))v^2.$$

This defines functions $w^1, w^2 : B(0, \epsilon) - \{0\} \to \mathbb{R}$ such that $w^1(v) \to a^1$ and $w^2(v) \to a^2$ as $v \to 0$. Now let

$$\begin{aligned} R(v) = &\left(\frac{\partial F}{\partial x^1}(w^1(v), a^2 + v^2) - \frac{\partial F}{\partial x^1}(a^1, a^2)\right) v^1 + \\ &\left(\frac{\partial F}{\partial x^2}(a^1, w^2(v)) - \frac{\partial F}{\partial x^2}(a^1, a^2)\right) v^2, \end{aligned}$$

so that

$$F(a + v) - F(a) - \frac{\partial F}{\partial x^1}(a)v^1 - \frac{\partial F}{\partial x^2}(a)v^2 = R(v).$$

From the equivalence of norms on $\mathbb{R}^n$, we have that there is a $C > 0$ such that

$$\begin{aligned} \frac{|R(v)|}{|v|} \leq\ & C\left|\frac{\partial F}{\partial x^1}(w^1(v), a^2 + v^2) - \frac{\partial F}{\partial x^1}(a^1, a^2)\right| + \\ & C\left|\frac{\partial F}{\partial x^2}(a^1, w^2(v)) - \frac{\partial F}{\partial x^2}(a^1, a^2)\right| \end{aligned}$$

and, by continuity of the partial derivatives, both terms on the right converge to 0 as $v \to 0$. Hence $|R(v)|/|v| \to 0$ as $v \to 0$, so this shows that F is differentiable. Therefore the result holds for $m = 1$ and $n = 2$.

The case for $m = 1$ and general n is a straightforward generalization of the argument we just gave, just with more notation: write $F(a + v) - F(a)$ as a telescoping sum and apply the mean value theorem to each of the relevant pieces. The case for arbitrary m and n proceeds as follows: If F is C^1, then each of the component functions $F^i : U \to \mathbb{R}$ are C^1, so we can apply our $m = 1$ case to each component function to conclude that each $F^i : U \to \mathbb{R}$ is differentiable. But then that implies that $F : U \to \mathbb{R}^m$ is differentiable. This completes the proof. □

Remark 6. If U is an open subset of $\mathbb{R}^n$ and $F : U \to \mathbb{R}^m$ is C^1, then since the matrix representing DF has entries given by the partial derivatives of F, we have that $DF : U \to L(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{nm}$ is continuous.

4 Higher Order Derivatives

Definition 6. Let $U \subset \mathbb{R}^n$ be open and $F : U \to \mathbb{R}^m$. If F is of class C^1, then we can differentiate the partial derivatives to obtain **second-order partial derivatives**

$$\frac{\partial^2 F^i}{\partial x^k \partial x^j} = \frac{\partial}{\partial x^k}\left(\frac{\partial F^i}{\partial x^j}\right)$$

whenever they exist. Continuing in this way, the **partial derivatives of F of order k** are the partial derivatives of those of order $k-1$ whenever they exist.

Definition 7. Let $U \subset \mathbb{R}^n$ be open and let $F : U \to \mathbb{R}^m$. We say that F is of **class C^k** or k **times continuously differentiable** if all the partial derivatives of F of order less than or equal to k exist and are continuous functions on U. In particular, C^0 is the class of continuous functions.

Remark 7. Let $U \subset \mathbb{R}^n$ be open and $F : U \to \mathbb{R}^m$. Then F is C^k iff for all $x \in U$, there is an open neighborhood N of x such that $F : N \cap U \to \mathbb{R}^m$ is C^k.

Remark 8. If a function is C^{k+1}, then it is also C^k. Furthermore, a function is C^{k+1} iff its partial derivatives are C^k, and a function is C^k iff all of its component functions are C^k.

Definition 8. A function that is class C^k for all $k \geq 0$ is said to be **class C^∞**, **smooth**, or **infinitely differentiable**.

Remark 9. A function is smooth iff its partial derivatives are smooth iff its partial derivatives of all orders are smooth iff all of its component functions are smooth.

Proposition 7. *Let $U \subset \mathbb{R}^n$ be open, and let $F : U \to \mathbb{R}^m$ be C^2. Then the mixed second-order partial derivatives of F do not depend on the order of differentiation:*

$$\frac{\partial^2 F^i}{\partial x^j \partial x^k} = \frac{\partial^2 F^i}{\partial x^k \partial x^j}$$

for all i, j, and k.

Proof. Let $a \in U$. Since U is open, there is $\epsilon > 0$ such that when $v \in B^n(0, \epsilon)$, $a + v \in U$. Let $\Delta : B^1(0, \epsilon/2) \to \mathbb{R}$ be defined by

$$\Delta(s) = F^i(a + se_j + se_k) - F^i(a + se_j) - F(a + se_k) + F(a).$$

Let $G_s : B^1(0, \epsilon/2) \to \mathbb{R}$ be defined by

$$G_s(t) = F^i(a + se_j + te_k) - F^i(a + te_k)$$

for each $s \in B^1(0, \epsilon/2)$. Then each G_s is C^1, and

$$\Delta(s) = G_s(s) - G_s(0)$$

for all $s \in B^1(0, \epsilon/2)$. By the mean value theorem, there is $\delta : B^1(0, \epsilon/2) \to \mathbb{R}$ such that $0 < |\delta(s)| < |s|$ for all $s \in B^1(0, \epsilon/2)$ and

$$\frac{\Delta(s)}{s} = G_s'(\delta(s)) = \frac{\partial F^i}{\partial x^k}(a + se_j + \delta(s)e_k) - \frac{\partial F}{\partial x^k}(a + \delta(s)e_k) \tag{5}$$

for all $s \in B^1(0, \epsilon/2) - \{0\}$. Since $\partial F^i/\partial x^k$ is C^1, and hence differentiable, we have that

$$\frac{\partial F^i}{\partial x^k}(a+se_j+\delta(s)e_k) = \frac{\partial F^i}{\partial x^k}(a)+\frac{\partial^2 F^i}{\partial x^j \partial x^k}(a)s+\frac{\partial^2 F^i}{\partial x^k \partial x^k}(a)\delta(s)+R(se_j+\delta(s)e_k)$$

and

$$\frac{\partial F^i}{\partial x^k}(a + \delta(s)e_k) = \frac{\partial F^i}{\partial x^k}(a) + \frac{\partial^2 F^i}{\partial x^k \partial x^k}(a)\delta(s) + R(\delta(s)e_k)$$

for all $s \in B^1(0, \epsilon/2)$. Substituting our last two equations into equation (5) implies that

$$\frac{\Delta(s)}{s^2} - \frac{\partial^2 F^i}{\partial x^j \partial x^k}(a) = \frac{R(se_j + \delta(s)e_k)}{s} - \frac{R(\delta(s)e_k)}{s} \tag{6}$$

for all $s \in B^1(0, \epsilon/2) - \{0\}$.

Now since $|\delta(s)| \leq |s|$ for each s, we have that

$$\frac{|R(\delta(s)e_k)|}{|s|} \leq \frac{|R(\delta(s)e_k)|}{|\delta(s)e_k|} \to 0 \tag{7}$$

as $s \to 0$. If we give $\mathbb{R}^n$ the ℓ^∞ norm, we also have that $|se_j + \delta(s)e_k|_\infty \leq |s|$. Therefore, by equivalence of norms, for the given arbitrary norm on $\mathbb{R}^n$ there is a constant $C > 0$ such that

$$|se_j + \delta(s)e_k| \leq C|s|$$

for all s. Therefore

$$\frac{|R(se_j + \delta(s)e_k)|}{|s|} \leq C\frac{|R(se_j + \delta(s)e_k)|}{|se_j + \delta(s)e_k|} \to 0 \tag{8}$$

as $s \to 0$. Equation (6) and inequalities (7) and (8) then imply that

$$\frac{\Delta(s)}{s^2} \to \frac{\partial^2 F^i}{\partial x^j \partial x^k}(a)$$

as $s \to 0$.

Now for each $s \in B^1(0, \epsilon/2)$, let $H_s : B^1(0, \epsilon/2) \to \mathbb{R}$ be defined by

$$H_s(t) = F^i(a + te_j + se_k) - F^i(a + te_j).$$

Then by following a similar argument as before, using H_s in place of G_s and $\partial F^i/\partial x^j$ in place of $\partial F^i/\partial x^k$, we can also show that

$$\frac{\Delta(s)}{s^2} \to \frac{\partial^2 F^i}{\partial x^k \partial x^j}(a)$$

as $s \to 0$. Hence the second order mixed partials agree at all $a \in U$, which is what we wanted to show. $\square$

Corollary 1. *If $U \subset \mathbb{R}^n$ is open and $F : U \to \mathbb{R}^m$ is smooth, then the mixed partials of order $k+2$ do not depend on the order of differentiation for all k:*

$$\frac{\partial^{k+2} F^i}{\partial x^{j_{k+2}} \cdots \partial x^{j_1}} = \frac{\partial^{k+2} F^i}{\partial x^{j_{\sigma(k+2)}} \cdots \partial x^{j_{\sigma(1)}}}$$

for all i, all k, all $(k+2)$-tuples $(j_1, \dots, j_{k+2})$ where each $1 \le j_l \le n$, and all permutations $\sigma : \{1, \dots, k+2\} \to \{1, \dots, k+2\}$.

Proof. We prove this by induction. The base case $k = 0$ was proved by the last proposition. Suppose this holds for some $k \ge 0$. Now let $(j_1, \dots, j_{k+3})$ be a $(k+3)$-tuple where each $1 \le j_l \le k+3$, and let $\sigma : \{1, \dots, k+3\} \to \{1, \dots, k+3\}$ be a permutation. If $\sigma(k+3) = k+3$, then $\sigma : \{1, \dots, k+2\} \to \{1, \dots, k+2\}$ is a permutation. Therefore, for any i, we have that

$$\begin{aligned}
\frac{\partial^{k+3} F^i}{\partial x^{j_{\sigma(k+3)}} \cdots \partial x^{j_{\sigma(1)}}} &= \frac{\partial}{\partial x^{j_{k+3}}} \left(\frac{\partial^{k+2} F^i}{\partial x^{j_{\sigma(k+2)}} \cdots \partial x^{j_{\sigma(1)}}} \right) \\
&= \frac{\partial}{\partial x^{j_{k+3}}} \left(\frac{\partial^{k+2} F^i}{\partial x^{j_{k+2}} \cdots \partial x^{j_1}} \right) \\
&= \frac{\partial^{k+3} F^i}{\partial x^{j_{k+3}} \cdots \partial x^{j_1}}.
\end{aligned}$$

If instead $k+3 \in \sigma(\{1, \dots, k+2\})$, then we also have that $\sigma(k+3) \in \{1, \dots, k+2\}$. Let $l \in \{1, \dots, k+2\}$ be such that $\sigma(l) = k+3$. For convenience, assume that $1 < l < k+2$. Then for all i,

$$\begin{aligned}
\frac{\partial^{k+3} F^i}{\partial x^{j_{\sigma(k+3)}} \cdots \partial x^{j_{\sigma(1)}}} &= \frac{\partial}{\partial x^{j_{\sigma(k+3)}}} \left(\frac{\partial^{k+2} F^i}{\partial x^{j_{\sigma(k+2)}} \cdots \partial x^{j_{\sigma(1)}}} \right) \\
&= \frac{\partial}{\partial x^{j_{\sigma(k+3)}}} \left(\frac{\partial^{k+2} F^i}{\partial x^{j_{\sigma(l)}} \partial x^{j_{\sigma(k+2)}} \cdots \partial x^{j_{\sigma(1)}}} \right) \\
&= \frac{\partial^2}{\partial x^{j_{\sigma(k+3)}} \partial x^{j_{\sigma(l)}}} \left(\frac{\partial^{k+1} F^i}{\partial x^{j_{\sigma(k+2)}} \cdots \partial x^{j_{\sigma(1)}}} \right) \\
&= \frac{\partial^2}{\partial x^{j_{\sigma(l)}} \partial x^{j_{\sigma(k+3)}}} \left(\frac{\partial^{k+1} F^i}{\partial x^{j_{\sigma(k+2)}} \cdots \partial x^{j_{\sigma(1)}}} \right) \\
&= \frac{\partial}{\partial x^{j_{k+3}}} \left(\frac{\partial^{k+2} F^i}{\partial x^{j_{\sigma(k+3)}} \cdots \partial x^{j_{\sigma(1)}}} \right) \\
&= \frac{\partial^{k+3} F^i}{\partial x^{j_{k+3}} \cdots \partial x^{j_1}}.
\end{aligned}$$

The case when $l = 1$ or $l = k+2$ follows almost exactly as above, just with some slight modifications to the notation. Therefore the proof is finished by induction. ☐

5 Diffeomorphisms

Definition 9. If U and V are open subsets of Euclidean space, a function $F : U \to V$ is a **diffeomorphism** if it is smooth, bijective, and its inverse is smooth.

Remark 10. Every diffeomorphism between open subsets of Euclidean space is a homeomorphism.

Proposition 8. *Let $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$, be open, and let $F : U \to V$ be a diffeomorphism. Then $m = n$, and for each $a \in U$, the total derivative $DF(a)$ is invertible with $DF(a)^{-1} = D(F^{-1})(F(a))$.*

Proof. Since F is a diffeomorphism, in particular F and F^{-1} are both C^1 and hence differentiable, so $DF(a)$ exists at each $a \in U$ and $D(F^{-1})(b)$ exists at each $b \in V$. Hence $F^{-1} \circ F = I_U$ is differentiable, and it is easy to verify that

$$DI_U(a) = I_{\mathbb{R}^n},$$

so

$$I_{\mathbb{R}^n} = D(F^{-1} \circ F)(a) = D(F^{-1})(F(a)) \circ DF(a).$$

Similarly, since $F \circ F^{-1} = I_V$, we also have that

$$I_{\mathbb{R}^m} = DF(a) \circ D(F^{-1})(F(a)).$$

Hence $DF(a)$ is an invertible linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ with inverse

$$DF(a)^{-1} = D(F^{-1})(F(a)),$$

and thus $n = m$. □

6 Smooth Real-Valued Functions

Definition 10. If $U \subset \mathbb{R}^n$ is open, we let $C^k(U)$ denote the set of all C^k functions from U to $\mathbb{R}$, and we let $C^\infty(U)$ denote the set of all smooth functions from U to $\mathbb{R}$. Sums, scalar multiples, and products are all defined pointwise: given $f, g : U \to \mathbb{R}$ and $c \in \mathbb{R}$,

$$(f + g)(x) = f(x) + g(x),$$

$$(cf)(x) = c(f(x)),$$

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

Proposition 9. *Let $U \subset \mathbb{R}^n$ be open and let $f, g \in C^\infty(U)$ and $c \in \mathbb{R}$. Then $f + g$, cf, and fg all belong to $C^\infty(U)$. Thus $C^\infty(U)$ is a commutative ring and a commutative and associative algebra over $\mathbb{R}$.*

Proof. From the definitions:

$$\frac{\partial(cf + g)}{\partial x^j}(x) = c\frac{\partial f}{\partial x^j}(x) + \frac{\partial g}{\partial x^j}(x)$$

for all j and all x. Thus $cf + g$ is C^1. In fact, this shows that taking partial derivatives is a linear operation. Now if $cf+g$ is $C^1, C^2, \dots, C^k$, and if an order k partial derivative of $f + g$ is of the form

$$\frac{\partial^k (cf+g)}{\partial x^{j_k} \cdots \partial x^{j_1}}(x) = c\frac{\partial^k f}{\partial x^{j_k} \cdots \partial x^{j_1}}(x) + \frac{\partial^k g}{\partial x^{j_k} \cdots \partial x^{j_1}}(x),$$

then an order $k + 1$ partial derivative of $cf + g$ is of the form

$$\frac{\partial^{k+1} (cf+g)}{\partial x^{j_{k+1}} \cdots \partial x^{j_1}}(x) = c\frac{\partial^{k+1} f}{\partial x^{j_{k+1}} \cdots \partial x^{j_1}}(x) + \frac{\partial^{k+1} g}{\partial x^{j_{k+1}} \cdots \partial x^{j_1}}(x)$$

which is continuous. Hence, by induction, $cf+g$ is smooth. Taking $c = 1$ shows that $f + g$ is smooth for all smooth f and g, and taking $g = 0$ shows that cf is smooth for all c and all smooth f.

Now

$$\begin{aligned}
\frac{\partial(fg)}{\partial x^j}(x) &= \lim_{h\to 0} \frac{f(x+he_j)g(x+he_j) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \left(\frac{f(x+he_j) - f(x)}{h} g(x+he_j) + f(x)\frac{g(x+he_j) - g(x)}{h} \right) \\
&= \frac{\partial f}{\partial x^j}(x)g(x) + f(x)\frac{\partial g}{\partial x^j}(x)
\end{aligned}$$

for all x and all j, so we conclude that fg is C^1, and the partial derivatives of fg of order 1 are sums of products of partial derivatives of f and g of order at most 1.

Now suppose that fg is C^1, C^2, …, C^k and the partial derivatives of fg of order k are sums of products of partial derivatives of f and g of order at most k. A particular term in a kth order partial derivative of fg is of the form

$$\frac{\partial^i f}{\partial x^{j_i} \cdots \partial x^{j_1}}(x)\frac{\partial^l g}{\partial x^{j_l} \cdots \partial x^{j_1}}(x)$$

where $0 \le i, l \le k$ (a partial derivative of order 0 is just $f(x)$ or $g(x)$). Therefore, differentiating one of these terms gives us a term of the form

$$\frac{\partial^{i+1} f}{\partial x^j \partial x^{j_i} \cdots \partial x^{j_1}}(x)\frac{\partial^l g}{\partial x^{j_l} \cdots \partial x^{j_1}}(x) + \frac{\partial^i f}{\partial x^{j_i} \cdots \partial x^{j_1}}(x)\frac{\partial^{l+1} g}{\partial x^j \partial x^{j_l} \cdots \partial x^{j_1}}(x).$$

Since taking partial derivatives is a linear operation, differentiating an order k partial derivative of fg to obtain an order $k+1$ partial derivative of fg will give us some of terms like above, which shows that all order $k+1$ partial derivatives of fg are continuous. Hence, by induction, fg is smooth when f and g are smooth.

It immediately follows from the algebraic properties of $\mathbb{R}$ that $C^\infty(U)$ is a commutative ring and a commutative and associative algebra over $\mathbb{R}$. The additive identity is the 0 function, the multiplicative identity is the constant 1 function, and the additive inverse of f is the function $-f = (-1)f$. □

Proposition 10. *Let $U \subset \mathbb{R}^n$ and $\tilde{U} \subset \mathbb{R}^m$ be open.*

1. *If $F : U \to \tilde{U}$ and $G : \tilde{U} \to \mathbb{R}^p$ are C^1, then $G \circ F : U \to \mathbb{R}^p$ is C^1, and its partial derivatives are given by*

$$\frac{\partial (G^i \circ F)}{\partial x^j}(x) = \sum_{j=1}^{m} \frac{\partial G^i}{\partial y^k}(F(x)) \frac{\partial F^k}{\partial x^j}(x).$$

2. *If F and G are smooth, then $G \circ F$ is smooth.*

Proof. Since F and G are C^1, they are differentiable, so $G \circ F$ is also differentiable, and for each $x \in U$, the matrix of $D(G \circ F)(x)$ is given by

$$\begin{aligned}
\frac{\partial (G^i \circ F)}{\partial x^j}(x) &= [D(G \circ F)(x)]^i_j \\
&= [DG(F(x)) \circ DF(x)]^i_j \\
&= \sum_{k=1}^{m} [DG(F(x))]^i_k [DF(x)]^k_j \\
&= \sum_{k=1}^{m} \frac{\partial G^i}{\partial y^k}(F(x)) \frac{\partial F^k}{\partial x^j}(x).
\end{aligned}$$

This shows that the partial derivatives of $G \circ F$ are sums of products of continuous functions, which is continuous. Hence $G \circ F$ is C^1. Thus the composition of C^1 functions is C^1.

Suppose now that the composition of C^k functions is C^k. If F and G are C^{k+1}, then let

$$H^i_l(y) = \frac{\partial G^i}{\partial y^l}(y)$$

for all i, l, and y. Then our computation above shows that

$$\frac{\partial G^i \circ F}{\partial x^j}(x) = \sum_{l=1}^{n} (H^i_l \circ F(x)) \frac{\partial F^k}{\partial x^j}(x)$$

for all i, j, and x. Since G is C^{k+1}, each H^i_l is C^k. Since F is C^{k+1}, and hence is also C^k, we have that $H^i_l \circ F$ is C^k and $\partial F^k / \partial x^j$ is also C^k. Therefore the partials of $G^i \circ F$ are sums of products of C^k functions, and hence is C^k. Therefore each $G^i \circ F$ is C^{k+1}, so $G \circ F$ is C^{k+1} whenever G and F are C^{k+1}. Hence, by induction, the composition of C^k functions is C^k for all k. From this, it follows that the composition of smooth functions is smooth. □

Corollary 2. *Let $U \subset \mathbb{R}^n$ be open, and let $f, g : U \to \mathbb{R}$. If f and g are smooth, and if g never vanishes on U, then f/g is smooth.*

Proof. Let $h : \mathbb{R} - \{0\} \to \mathbb{R}$ be defined by $h(x) = 1/x$. Then for any $x \neq 0$,

$$h'(x) = -1/x^2$$

which is continuous. Therefore h is C^1, and

$$h'(x) = \frac{(-1)^1 1!}{x^{1+1}}$$

for all $x \in \mathbb{R} - \{0\}$. Now suppose that h is C^k and

$$\frac{d^k h}{dx^k}(x) = \frac{(-1)^k k!}{x^{1+k}} = (-1)^k k! h(p(x))$$

for all $x \in \mathbb{R} - \{0\}$, where $p : \mathbb{R} \to \mathbb{R}$ is defined by $p(x) = x^{1+k}$. Then since $p'(x) = (k+1)x^k$, we have from the chain rule that

$$\frac{d^{k+1} h}{dx^{k+1}}(x) = \frac{(-1)^{k+1}(k+1)!}{x^{1+k+1}}$$

for all $x \in \mathbb{R} - \{0\}$. Hence, by induction, h is smooth. Since $f/g = f \cdot (h \circ g)$ on U, and since the multiplication and composition of smooth functions is smooth, we conclude that f/g is smooth. □

7 Extension to Non-Open Subsets

Definition 11. If $A \subset \mathbb{R}^n$, then $F : A \to \mathbb{R}^m$ is **smooth on** A if for all $x \in A$, there is an open neighborhood $U \subset \mathbb{R}^n$ of x and a smooth function $\tilde{F} : U \to \mathbb{R}^m$ such that $\tilde{F} = F$ on $U \cap A$. We call such an $\tilde{F}$ a **smooth extension of** F **on an open neighborhood of** x.

Remark 11. If $U \subset \mathbb{R}^n$ is open, then $F : U \to \mathbb{R}^m$ is smooth on U as above iff $F : U \to \mathbb{R}^m$ is smooth in the previously defined sense.

Remark 12. Let $A \subset \mathbb{R}^m$. If $F : A \to \mathbb{R}^n$ is smooth, then F is continuous.

Proposition 11. *Let $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$, $F : A \to \mathbb{R}^m$, $G : B \to \mathbb{R}^p$, and $F(A) \subset B$. If F and G are smooth, then $G \circ F : A \to \mathbb{R}^p$ is smooth.*

Proof. Let $x \in A$. Then there is an open neighborhood V of $f(x)$ and a smooth function $\tilde{G} : V \to \mathbb{R}^p$ such that $\tilde{G} = G$ on $V \cap B$, and there is an open neighborhood U of x and a smooth function $\tilde{F} : U \to \mathbb{R}^m$ such that $\tilde{F} = F$ on $U \cap A$. Then $U \cap \tilde{F}^{-1}(V)$ is an open neighborhood of x and $\tilde{G} \circ \tilde{F} : U \cap \tilde{F}^{-1}(V) \to \mathbb{R}^p$ is a smooth function such that $\tilde{G} \circ \tilde{F} = G \circ F$ on $U \cap \tilde{F}^{-1}(V) \cap A$. Hence $G \circ F : A \to \mathbb{R}^p$ is smooth. □

Definition 12. Given $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$, a **diffeomorphism from** A **to** B is a smooth bijection $F : A \to B$ with smooth inverse.

Remark 13. Every diffeomorphism between subsets of Euclidean space is a homeomorphism.

8 Directional Derivatives

Definition 13. Let $f : U \to \mathbb{R}$ be a smooth real-valued function on an open subset U of $\mathbb{R}^n$. For each $v \in \mathbb{R}^n$, each $a \in U$, the **directional derivative of** f **in the direction of** v **at** a is the number

$$D_v f(a) = \left.\frac{d}{dt}\right|_{t=0} f(a + tv).$$

Remark 14. More precisely, given $a \in U$ and $v \in \mathbb{R}^n$, since U is open, there is an $\epsilon > 0$ such that $a + tv \in U$ for all $t \in \mathbb{R}$ such that $|t| < \epsilon$. Let $g : (-\epsilon, \epsilon) \to U$ be defined by

$$g(t) = a + tv.$$

Since the map g is smooth, $f \circ g$ is smooth. Then

$$D_v f(a) = (f \circ g)'(0) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(a) v^i,$$

where the last equality follows from the chain rule.

9 The Inverse Function Theorem and the Implicit Function Theorem

Definition 14. Let (X, d) be a metric space. A map $G : X \to X$ is a **contraction** if there is a constant $\lambda \in (0, 1)$ such that $d(G(x), G(y)) \leq \lambda d(x, y)$ for all $x, y \in X$. We call such a λ a **contraction constant** for G.

Remark 15. Every contraction is continuous.

Definition 15. Let X be a set. A **fixed point** of a map $G : X \to X$ is a point $x \in X$ such that $G(x) = x$.

Lemma 1 (Contraction Lemma)**.** *Let (X, d) be a nonempty complete metric space. Every contraction $G : X \to X$ has a unique fixed point.*

Proof. Let $x_0 \in X$. Let $x_{i+1} = G(x_i)$ for all $i \geq 0$. Let λ be a contraction constant for G. Then the sequence $(x_n) \subset X$ satisfies

$$d(x_i, x_{i+1}) = d(G(x_{i-1}), G(x_i)) \leq \lambda d(x_{i-1}, x_i)$$

for all $i \geq 1$. By induction, we conclude that

$$d(x_i, x_{i+1}) \leq \lambda^i d(x_0, x_1)$$

for all i. Hence for any $i < j$, we have that

$$\begin{aligned}d(x_i, x_j) &\leq d(x_i, x_{i+1}) + d(x_{i+1}, x_{i+2}) + \cdots + d(x_{j-1}, x_j) \\ &\leq \lambda^i(1 + \lambda + \cdots + \lambda^{j-i-1})d(x_0, x_1) \\ &= \lambda^i \frac{1-\lambda^{j-i}}{1-\lambda} d(x_0, x_1) \\ &\leq \lambda^i \frac{d(x_0, x_1)}{1-\lambda}.\end{aligned}$$

Since $0 \leq d(x_i, x_j) = d(x_j, x_i)$ for all i, j, and since $0 = d(x_i, x_i)$ for all i, we then conclude that

$$0 \leq d(x_i, x_j) \leq \lambda^{\min\{i,j\}} \frac{d(x_0, x_1)}{1-\lambda}$$

for all i, j. Now since the last term converges to 0 as $i, j \to \infty$, we conclude that (x_n) is a Cauchy sequence in X. Therefore, since X is complete, there is an $x \in X$ such that $x_n \to x$. Since contractions are continuous, we then have that $G(x_n) \to G(x)$. However, since $G(x_n) = x_{n-1}$ for all $n \geq 1$, we then conclude that $x_n \to G(x)$ as well. In other words, $G(x) = x$, so x is a fixed point of G.

If x' is another fixed point, then

$$d(x, x') = d(G(x), G(x')) \leq \lambda d(x, x').$$

Since $0 < \lambda < 1$, this implies that $d(x, x') = 0$, so that $x = x'$. Therefore G has exactly one fixed point. □

Proposition 12 (Lipschitz Estimate for C^1 Functions)**.** *Let $U \subset \mathbb{R}^n$ be open, and let $F : U \to \mathbb{R}^m$ be C^1. Then F is Lipschitz continuous on every compact convex subset $K \subset U$, with Lipschitz constant* $\sup_{x\in K} |DF(x)|$, *where*

$$|DF(x)| = \sqrt{\sum_{i,j}([DF(x)]^i_j)^2}$$

and $[DF(x)]$ is the standard matrix representation of $DF(x)$.

Proof. Let $a, b \in K$. Then for all $0 \leq t \leq 1$, $a + t(b - a) \in K$. From the fundamental theorem of calculus and the chain rule,

$$\begin{aligned}F(b) - F(a) &= \int_0^1 \frac{d}{dt} F(a + t(b - a))\, dt \\ &= \int_0^1 [DF(a + t(b - a))](b - a)\, dt.\end{aligned}$$

Therefore, by properties of the integral and properties of the given matrix norm,

$$|F(b) - F(a)| \leq \left(\sup_{x\in K} |DF(x)|\right) |b - a|.$$

□

Lemma 2 (The Inverse Function Theorem, Special Case)**.** *Let U and V be open neighborhoods of 0 in $\mathbb{R}^n$. Let $F : U \to V$ be smooth and such that $F(0) = 0$ and $DF(0) = I_n$. Also, suppose that $DF(x)$ is invertible for all $x \in U$. Then there are connected open neighborhoods $U_0 \subset U$ and $V_0 \subset V$ of 0 such that $F : U_0 \to V_0$ is a diffeomorphism.*

Proof. Step 1: Finding a neighborhood of 0 for which F is injective. Let $H(x) = x - F(x)$ for each $x \in U$. Then $DH(0) = I_n - I_n = 0$. Observe that the matrix entries of $[DH] = [I_n - DF]$ are continuous functions on U. Hence $DH : U \to L(\mathbb{R}^n, \mathbb{R}^n) \cong \mathbb{R}^{n^2}$ is continuous at $0 \in U$. Therefore, there is a $\delta > 0$ such that $K := \overline{B_\delta(0)} \subset U$ and for all $x \in K$,

$$|DH(x) - DH(0)| = |DH(x)| \leq 1/2.$$

From the Lipschitz estimate for C^1 functions applied to the function H and the compact set K,

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

for all $x, x' \in K$. Taking $x' = 0$ gives us

$$|H(x)| \leq \frac{1}{2}|x| \tag{9}$$

for all $x \in K$. Since

$$x - x' = F(x) - F(x') + H(x) - H(x')$$

for all $x, x' \in U \supset K$, we also have that

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

for all $x, x' \in K$. This implies that

$$0 \leq |x - x'| \leq 2|F(x) - F(x')| \tag{10}$$

for all $x, x' \in K$, so that F is injective on K.

Step 2: Finding a neighborhood of 0 for which F is bijective. Let $y \in B_{\delta/2}(0) \subset K$. We will show that there is $x \in B_\delta(0) \subset K$ such that $F(x) = y$. For all $x \in K$, let $G(x) = y + H(x) = y + x - F(x)$. Then $G(x) = x$ iff $F(x) = y$. Now for all $x \in K$, equation (9) implies that

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

Then $G : K \to B_\delta(0) \subset K$, and

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

for all $x, x' \in K$. Hence $G : K \to B_\delta(0) \subset K$ is a contraction map, so by the contraction mapping lemma, there is a unique $x \in K$ such that $G(x) = x \in B_\delta(0)$. Therefore there is a unique $x \in B_\delta(0)$ such that $F(x) = y$.

Step 3: Finding U_0, V_0, and $F^{-1} : V_0 \to U_0$. Let $V_0 = B_{\delta/2}(0) \subset K \subset U$ and let $U_0 = B_\delta(0) \cap F^{-1}(V_0) \subset K \subset U$. Then $U_0 \subset U$ and V_0 are open and steps 1 and 2 show that $F : U_0 \to V_0$ is bijective. Hence $F^{-1} : V_0 \to U_0$ exists. Given $y \in V_0$, we have that $F^{-1}(y) \in U_0 \subset U$, so $y = F(F^{-1}(y)) \in F(U) \subset V$. Hence $V_0 \subset V$ as well. Since equation (10) applies to all $x, x' \in K \supset U_0$, for any $y, y' \in V_0$, we have that

$$|F^{-1}(y) - F^{-1}(y')| \leq 2|y - y'|,$$

so that $F^{-1} : V_0 \to U_0$ is continuous (even Lipschitz continuous). Therefore $F : U_0 \to V_0$ is a homeomorphism, so since V_0 is connected, U_0 is also connected.

Step 4: Showing $F^{-1} : V_0 \to U_0$ is differentiable. Let $y \in V_0$, and let $y' \in V_0 - \{y\}$. Let $x = F^{-1}(y) \in U_0$ and let $L = DF(x)$. Since $F^{-1}(V_0) = U_0 \subset K \subset U$, we have that L^{-1} exists by assumption and is linear since L is linear. Let $x' = F^{-1}(y') \in U_0$. Since F^{-1} is injective, $x \neq x'$. We also have that $y = F(x)$ and $y' = F(x')$. Therefore, all of our observations imply that

$$\frac{|F^{-1}(y') - F^{-1}(y) - L^{-1}(y' - y)|}{|y' - y|} = \frac{|x' - x|}{|y' - y|} \frac{|L^{-1}(L(x' - x) - F(x') + F(x))|}{|x' - x|}.$$

Since equation (10) applies to all $x, x' \in K \supset U_0$, we have that

$$\frac{|x' - x|}{|y' - y|} \leq 2.$$

Since L^{-1} is a linear map between finite dimensional vector spaces, there is a constant $C > 0$ such that

$$|L^{-1}(L(x' - x) - F(x') + F(x))| \leq C|F(x') - F(x) - L(x' - x)|.$$

Therefore

$$0 \leq \frac{|F^{-1}(y') - F^{-1}(y) - L^{-1}(y' - y)|}{|y' - y|} \leq 2C \frac{|F(x') - F(x) - L(x' - x)|}{|x' - x|}.$$

Now if $y' \to y$, since F^{-1} is continuous, $x' \to x$. Then since $L = DF(x)$ and since F is differentiable,

$$\frac{|F(x') - F(x) - L(x' - x)|}{|x' - x|} \to 0$$

as $x' \to x$. Hence

$$\frac{|F^{-1}(y') - F^{-1}(y) - L^{-1}(y' - y)|}{|y' - y|} \to 0$$

as $y' \to y$, so F^{-1} is differentiable at each $y \in V_0$ and

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

Step 5: Showing $F^{-1} : V_0 \to U_0$ is C^1. Since F^{-1} is differentiable, the partial derivatives of F^{-1} exist and are the entries of the matrix-valued function $y \mapsto [D(F^{-1})(y)] = [DF(F^{-1}(y))]^{-1}$. This map can be realized as the composition of the maps

$$y \mapsto F^{-1}(y) \mapsto [DF(F^{-1}(y))] \mapsto [DF(F^{-1}(y)]^{-1}. \tag{11}$$

We have that F^{-1} is continuous, $x \mapsto [DF(x)]$ is smooth as a map from U_0 to $\mathbb{R}^{n^2} \cong GL(n, \mathbb{R})$, and, because of Cramer's rule, taking inverses of invertible matrices is smooth when thought of as a map from $GL(n, \mathbb{R}) \cong \mathbb{R}^{n^2} \to GL(n, \mathbb{R}) \cong \mathbb{R}^{n^2}$. Therefore all the intermediate maps in the composition are at least continuous, so the entries of the map $y \mapsto [D(F^{-1})(y)]$ are continuous maps on V_0. In other words, all the partial derivatives of F^{-1} exist and are continuous, so F^{-1} is C^1.

Step 6: Showing F^{-1} is smooth. Suppose that F^{-1} is C^k. Then each of the maps in (11) is C^k, which implies that the entries of $y \mapsto [D(F^{-1})(y)]$ are C^k. In other words, all the partial derivatives of F^{-1} are C^k, so F^{-1} is C^{k+1}. Therefore, by induction, F^{-1} is smooth. □

Theorem 1 (The Inverse Function Theorem, General Case)**.** *Let U and V be open subsets of $\mathbb{R}^n$, and let $F : U \to V$ be smooth. Let $a \in U$, and suppose that $DF(a) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible. Then there exist connected open neighborhoods $U_0 \subset U$ of a and $V_0 \subset V$ of $F(a)$ such that $F : U_0 \to V_0$ is a diffeomorphism.*

Proof. First we reduce to the special case. Let $a \in U$. Then since V is open, there is an $r > 0$ such that $F(a) \in B_r(F(a)) \subset V$, and there is an $s > 0$ such that $a \in B_s(a) \cap F^{-1}(B_r(F(a))) \subset U$. Since $B_s(a) \cap F^{-1}(B_r(F(a)))$ is open, there is an $s' > 0$ such that $a \in B_{s'}(a) \subset B_s(a) \cap F^{-1}(B_r(F(a)))$. Observe that when $x \in U_1 := B_{s'}(0)$, $a + x \in B_{s'}(a)$ and $F_1(x) := F(a + x) - F(a) \in B_r(0) := V_1$. Then $F_1 : U_1 \to V_1$ is a smooth map between connected open neighborhoods U_1 of 0 and V_1 of $F_1(0) = 0$. Also, we have that $D(F_1)(0) = DF(a)$, so $D(F_1)(0)$ is invertible. Now let $F_2(x) = D(F_1)(0)^{-1}(F_1(x))$ for all $x \in U_1$. Then since $F_2 : U_1 \to \mathbb{R}^n$ is the composition of a smooth map and a linear map, we have that $F_2 : U_1 \to \mathbb{R}^n$ is smooth. We also have that $F_2(0) = 0$ since $F_1(0) = 0$ and $D(F_1)(0)^{-1}$ is linear. Furthermore, by the chain rule and linearity of $D(F_1)(0)^{-1}$, we have that

$$D(F_2)(0) = D(D(F_1)(0)^{-1})(F_1(0)) \circ D(F_1)(0) = D(F_1)(0)^{-1} \circ D(F_1)(0) = I_n.$$

Since F_2 is smooth, there is an $s' > s'' > 0$ such that $F_2(B_{s''}(0)) \subset B_r(0) = V_1$. Let $U_2 = B_{s''}(0)$, so that $U_2 \subset U_1$. The map $x \mapsto \det[D(F_2)(x)]$ is smooth on U_2 since it is a polynomial of the partial derivatives of F_2 which are smooth functions. Therefore, there is an $0 < s''' < s''$ such that when $|x| < s'''$, we have that

$$1 - |\det[D(F_2)(x)]| \leq |\det[D(F_2)(0)] - \det[D(F_2)(x)]| < 1/2.$$

Hence when $x \in U_3 := B_{s'''}(0) \subset U_2$, we have that

$$|\det[D(F_2)(x)]| > 1/2,$$

and thus $D(F_2)(x)$ is invertible for all $x \in U_3$. Hence the map $F_2 : U_3 \to V_1$ is a smooth map between connected open neighborhoods U_3 and V_1 of 0 which satisfies $DF_2(0) = I_n$, $F_2(0) = 0$, and $D(F_2)(x)$ is invertible for all $x \in U_3$.

We now apply the special case to $F_2 : U_3 \to V_1$ to conclude that there are connected open neighborhoods $U_4 \subset U_3$ and $V_2 \subset V_1$ of 0 such that $F_2 : U_4 \to V_2$ is a diffeomorphism. Then since $F_1 = D(F_1)(0) \circ F_2 : U_4 \to V_3 := D(F_1)(0)(V_2)$, which is the composition of smooth maps between connected open neighborhoods of 0, we have that $F_1 : U_4 \to V_3$ is smooth. Furthermore, we also have that $F_1^{-1} = F_2^{-1} \circ D(F_1)(0)^{-1} : V_3 \to U_4$ exists and is smooth, so $F_1 : U_4 \to V_3$ is a diffeomorphism between connected open neighborhoods of 0. Now if $x \in U_0 := a + U_4 \subset U$, then $x - a \in U_4$ and

$$F(x) = F_1(x - a) + F(a) \in V_0 := F(a) + V_3.$$

Therefore $F : U_0 \to V_0$ is smooth. Given $y \in F(a) + V_3$, $y = F(a) + z$ for some $z \in V_3 = F_1(U_4)$, so $y = F(a) + F_1(z')$ for some $z' \in U_4$. Then $a + z' \in U_0$ and

$$F(a + z') = F_1(z') + F(a) = y.$$

Hence $F : U_0 \to V_0$ is bijective. Thus, given $y \in V_0$, $F^{-1}(y) \in U_0 \subset U$, and thus $y \in F(U) \subset V$. Thus $U_0 \subset U$ is a connected open neighborhood of a and $V_0 \subset V$ is a connected open neighborhood of $F(a)$. Finally, we also have that

$$F^{-1}(y) = F_1^{-1}(y - F(a)) + a$$

for all $y \in V_0$. Hence $F^{-1} : U_0 \to V_0$ is smooth, so $F : U_0 \to V_0$ is a diffeomorphism between a connected open neighborhood $U_0 \subset U$ of a and $V_0 \subset V$ of $F(a)$. This completes the proof. □

Theorem 2 (The Implicit Function Theorem). *Let $U \subset \mathbb{R}^n \times \mathbb{R}^k$ be open and let $\phi : U \to \mathbb{R}^k$ be smooth. Let $(x, y) = (x^1, \ldots, x^n, y^1, \ldots, y^k)$ denote the standard coordinates on U. Let $(a, b) \in U$ and let $c = \phi(a, b)$. Suppose that the $k \times k$ matrix*

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

is nonsingular. Then there are open neighborhoods $V_0 \subset \mathbb{R}^n$ of a and $W_0 \subset \mathbb{R}^k$ of b and a smooth function $F : V_0 \to W_0$ such that for all $x \in V_0$ and $y \in W_0$, $\phi(x, y) = c$ iff $y = F(x)$.

Proof. First we define a smooth function for which to apply the inverse function to. Let $\psi : U \to \mathbb{R}^n \times \mathbb{R}^k$ be defined by

$$\psi(x, y) = (x, \phi(x, y)).$$

This is smooth, and

$$[D\psi(a,b)] = \begin{bmatrix} I_n & 0 \\ \frac{\partial\phi^i}{\partial x^j}(a,b) & \frac{\partial\phi^i}{\partial y^j}(a,b) \end{bmatrix}$$

is nonsingular by our assumptions. Therefore, by the inverse function theorem, there are connected open neighborhoods $U_0 \subset U$ of (a,b) and $Y_0 \subset \mathbb{R}^n \times \mathbb{R}^k$ of $\psi(a,b) = (a,c)$ such that $\psi : U_0 \to Y_0$ is a diffeomorphism.

Next, we define V_0 and W_0. Since U_0 is an open subset of $\mathbb{R}^n \times \mathbb{R}^k$, there are open sets $V \subset \mathbb{R}^n$ and $W_0 \subset \mathbb{R}^k$ such that $(a,b) \in V \times W_0 \subset U_0$. Then $\psi(a,b) = (a,c) \in \psi(V \times W_0) \subset Y_0$ and $\psi : V \times W_0 \to \psi(V \times W_0)$ is a diffeomorphism. Let $V_0 = \{x \in V : (x,c) \in \psi(V \times W_0)\}$, so that $a \in V_0 \subset V$ is open and $b \in W_0$ is open.

Now, we define $F : V_0 \to W_0$. Since $\psi^{-1} : \psi(V \times W_0) \to V \times W_0$, is smooth, there are smooth functions $A : \psi(V \times W_0) \to V$ and $B : \psi(V \times W_0) \to W_0$ such that $\psi^{-1}(x,y) = (A(x,y), B(x,y))$ for all $(x,y) \in \psi(V \times W_0)$. Let $F : V_0 \to W_0$ be defined by $F(x) = B(x,c)$ for all $x \in V_0$.

Now before we show that F has all of the desired properties, we make one observation. Let $(x,y) \in \psi(V \times W_0)$. Then

$$\begin{aligned}(x,y) &= \psi(\psi^{-1}(x,y)) \\ &= \psi(A(x,y), B(x,y)) \\ &= (A(x,y), \phi(A(x,y), B(x,y)).\end{aligned}$$

Comparing the first coordinates shows us that

$$A(x,y) = x$$

for all $(x,y) \in \psi(V \times W_0)$.

Now we show that F has all of the desired properties. First, since B is smooth, F is smooth. Next, if $x \in V_0$ and $y \in W_0$ is such that $\phi(x,y) = c$, then $\psi(x,y) = (x,c) \in \psi(V \times W_0)$, so that $A(x,c) = x$. Therefore,

$$(x,y) = \psi^{-1}(x,c) = (A(x,c), B(x,c)) = (x, F(x)).$$

Hence $y = F(x)$.

Conversely, if $x \in V_0$ and $y \in W_0$ is such that $y = F(x)$, then $(x,c) \in \psi(V \times W_0)$, so that $A(x,c) = x$. Therefore,

$$\begin{aligned}(x,c) &= \psi(\psi^{-1}(x,c)) \\ &= (A(x,c), \phi(A(x,c), B(x,c)) \\ &= (x, \phi(x, F(x))) \\ &= (x, \phi(x,y)).\end{aligned}$$

Comparing the second coordinates then implies that $\phi(x,y) = c$. This completes the proof. □