Differential Calculus

Jayadev Naram

Contents

1 Total Derivatives in Finite-Dimensional Vector spaces

Definition 1. *Let V, W be finite-dimensional vector spaces, which we may assume to be endowed with norms. If $U \subseteq V$ is an open subset $a \in U$, a map $F : U \to W$ is said to be* ***differentiable at a*** *if there exists a linear map $L : V \to W$ such that*

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

Remark 1. *It can be easily seen that*

$$\underset{v\to 0}{Lt} \frac{\|F(a+v) - F(a) - Lv\|}{\|v\|} = 0 \Leftrightarrow \underset{v\to 0}{Lt} \frac{F(a+v) - F(a) - Lv}{\|v\|} = 0. \tag{2}$$

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

Proof. Let $L_1 : V \to W$ and $L_2 : V \to W$ satisfy (1). Define $A = L_1 - L_2$, then we have

$$\begin{aligned} \|Av\| = &\ \|(F(a+v) - F(a) - L_2v) - (F(a+v) - F(a) - L_1v)\| \\ \leq &\ \|F(a+v) - F(a) - L_2v\| + \|F(a+v) - F(a) - L_1v\| \end{aligned}$$

Dividing by $\|v\|$ and taking limit we get

$$\implies \underset{v\to 0}{Lt} \frac{\|Av\|}{\|v\|} \leq \underset{v\to 0}{Lt} \left(\frac{\|F(a+v) - F(a) - L_1v\|}{\|v\|} + \frac{\|F(a+v) - F(a) - L_2v\|}{\|v\|} \right) = 0$$

Therefore we have $Av = 0 \; \forall \; v \in V$ which implies that $A = 0$, i.e., $L_1 = L_2$. ☐

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

Remark 2. *Condition* (1) *can also be written as*

$$F(a+v) - F(a) = DF(a)v + R(v), \tag{3}$$

where the remainder term $R(v)$ satisfies $\|R(v)\|/\|v\| \to 0$ as $v \to 0$. Thus the total derivative represents the "best linear approximation" to $F(a+v) - F(a)$ near a. Note that $\|R(v)\|/\|v\| \to 0$ implies that eventually $\|R(v)\|/\|v\| \leq 1$, i.e., $\|R(v)\| \leq \|v\|$.

Proposition 2. *Suppose V, W are finite-dimensional vector spaces, $U \subseteq V$ is an open subset, $a \in U$, and $F : U \to W$ is a map. If F is differentiable at a, then it is continuous at a.*

Proof. In (3) take norm and apply limit $v \to 0$ on both sides

$$\begin{aligned} 0 \leq \underset{v\to 0}{Lt}\|F(a+v) - F(a)\| = \underset{v\to 0}{Lt}\|DF(a)v + R(v)\| &\leq \underset{v\to 0}{Lt}\|DF(a)v\| + \|R(v)\| \\ &\leq \underset{v\to 0}{Lt}(\|DF(a)\| + 1)\|v\| = 0, \end{aligned}$$

where $\|DF(a)\|$ is the operator norm. Thus F is continuous at a. □

Proposition 3. *Suppose V, W, X are finite-dimensional vector spaces. Then*

(a) If $T : V \to W$ is a linear map, then T is differentiable at every point $v \in V$, with total derivative equal to T itself: $DT(v) = T$.

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

Proof. ***(a)*** Setting $L = T$ in (1) and using the linearity of T, we see that T is differentiable everywhere with the total derivative equal to T itself.

(b) We use (3) to show that bilinear map is differentiable. Note that

$$B(v + x, w + y) = B(v, w) + B(v, y) + B(x, w) + B(x, y).$$

But $B(x, y) \leq \|B\|\|x\|\|y\|$ where $\|B\|$ is operator norm which is finite by continuity of B. Then comparing with (3)

$$B(v + x, w + y) - B(v, w) = DB(v, w)(x, y) + R(x, y),$$

where $DB(v, w)(x, y) = B(v, y) + B(x, w)$ and $R(x, y) \leq \|B\|\|x\|\|y\| \to 0$ as $(x, y) \to 0$. □

Proposition 4 (***The Chain Rule for Total Derivatives***)**.** *Suppose V, W, X are finite-dimensional vector spaces, $U \subseteq V$ and $\tilde{U} \subseteq W$ are open subsets, and $F : U \to \tilde{U}$ and $G : \tilde{U} \to X$ are maps. 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 and $D(G \circ F)(a) = DG(F(a)) \circ DF(a)$.*

Proof. Let $A = DF(a)$ and $B = DG(F(a))$. We need to show that

$$\underset{v\to 0}{Lt}\frac{\|G(F(a + v)) - G(F(a)) - BAv\|}{\|v\|} = 0. \tag{4}$$

Let us write $b = F(a)$ and $w = F(a + v) - F(a)$. With these substitutions, we can rewrite the quotient in (4) as

$$\begin{aligned} \frac{\|G(b + w) - G(b) - BAv\|}{\|v\|} &= \frac{\|G(b + w) - G(b) - Bw + Bw - BAv\|}{\|v\|} \\ &\leq \frac{\|G(b + w) - G(b) - Bw\|}{\|v\|} + \frac{\|B(w - Av)\|}{\|v\|} \qquad (\dagger) \end{aligned}$$

The differentiability of F at a means that for any $\epsilon > 0$, we can ensure that

$$\|w - Av\| = \|F(a + v) - F(a) - Av\| \leq \epsilon\|v\|$$

as long as v lies in a small enough neighborhood of 0. Moreover, as $v \to 0$, $\|v\| = \|F(a + v) - F(v)\| \to 0$ by continuity of F. Therefore by differentiability of G at b means that by making $\|v\|$ even smaller if necessary, we can also achieve

$$\|G(b + w) - G(b) - Bw\| \leq \epsilon\|w\|.$$

Also note that $\|B(w-Av)\| \leq \|B\|\|w-Av\|$. Putting all of these estimates together, we see that for $\|v\|$ sufficiently small, (†) is bounded by

$$\begin{aligned}\epsilon\frac{\|w\|}{\|v\|}+\|B\|\frac{\|w-Av\|}{\|v\|} =&\epsilon\frac{\|w-Av+Av\|}{\|v\|}+\|B\|\frac{\|w-Av\|}{\|v\|}\\ \leq&\epsilon\frac{\|w-Av\|}{\|v\|}+\frac{\|Av\|}{\|v\|}+\|B\|\frac{\|w-Av\|}{\|v\|}\\ \leq&\epsilon^2+\epsilon\|A\|+\epsilon\|B\|,\end{aligned}$$

which can be made as small as desired. □

Lemma 1. *The addition operation $+ : \mathbb{R}^2 \to \mathbb{R}$ defined as $(+)(x,y) = x+y$ is differentiable and $D(+)(a,b) = +$. The multiplication operation $\times : \mathbb{R}^2 \to \mathbb{R}$ defined as $(\times)(x,y) = xy$ is differentiable and $D(\times)(a,b)(x,y) = bx+ay$. The reciprocal operation $h : \mathbb{R} \to \mathbb{R}$ defined as $h(x) = 1/x$ is differentiable and $Dh(x) = -\frac{1}{x^2}$.*

Proposition 5. *Suppose V, W are finite-dimensional vector spaces, $U \subseteq V$ is an open subset, a is a point in U, and $F, G : U \to W$ and $f, g : U \to \mathbb{R}$ are maps. Then*

(a) If F is a constant map, then F is differentiable at a and $DF(a) = 0$.

(b) If F and G are differentiable at a, then $F+G$ is also, and

$$D(F+G)(a) = DF(a) + DG(a).$$

(c) If f and g are differentiable at a, then fg is also, and

$$D(fg)(a) = f(a)Dg(a) + g(a)Df(a).$$

(d) If f and g are differentiable at a and $g(a) \neq 0$, then f/g is differentiable at a, and

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

Proof. **(a)** Let $F(v) = c \in W$ for all $v \in V$. Then setting $L = 0$ in (1) satifies the equation showing that F is differentiable at a and $DF(a) = 0$.

(b) Note that $F+G = (+) \circ (F,G)$, then using chain rule

$$\begin{aligned}D(F+G)(a) = D((+)\circ(F,G))(a) =&D(+)(F(a),G(a))\circ D(F,G)(a)\\ =&D((+)(F(a),G(a)))\circ(DF(a),DG(a))\\ =&(+)(DF(a),DG(a)) = DF(a)+DG(a).\end{aligned}$$

(c) Note that $fg = (\times) \circ (f,g)$, then using chain rule

$$\begin{aligned}D(fg)(a) = D((\times)\circ(f,g))(a) =&D((\times)(f(a),g(a))\circ D(f,g)(a)\\ =&D(\times)(f(a),g(a))(Df(a),Dg(a))\\ =&g(a)Df(a)+f(a)Dg(a)\end{aligned}$$

(d) Note that $f/g = (\times) \circ (f, 1/g)$ and $1/g = h \circ g$, so by chain rule

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

Then

$$
\begin{aligned}
D(f/g)(a) = D((\times) \circ (f, 1/g))(a) =& D(\times)(f(a), 1/g(a)) \circ D(f, 1/g)(a) \\
=& D(\times)(f(a), 1/g(a))(Df(a), D(1/g)(a)) \\
=& \frac{1}{g(a)} Df(a) - f(a)\frac{1}{g(a)^2} Dg(a) \\
=& \frac{g(a)Df(a) - f(a)Dg(a)}{g(a)^2}.
\end{aligned}
$$

□

2 Total and Partial Derivatives in $\mathbb{R}^n$

Definition 3. *Suppose $U \subseteq \mathbb{R}^n$ is open and $f : U \to \mathbb{R}$ is a real-valued function. For any $a = (a^1, \ldots, a^n) \in U$ and any $j \in \{1, \ldots, n\}$, the* **j-th partial derivative of f at a** *is defined to be the ordinary derivative of f w.r.t. x^j while holding the other variables fixed:*

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

if the limit exists.

Definition 4. *For a vector-valued function $F : U \to \mathbb{R}^m$, we can write the coordinates of $F(x)$ as $F(x) = (F^1(x), \ldots, F^m(x))$. This defines m functions $F^1, \ldots, F^m : U \to \mathbb{R}$ called the* **component functions of F**. *The partial derivatives of F are defined simply to be the partial derivatives $\partial F^i/\partial x^j$ of its component functions. The matrix $(\partial F^i/\partial x^j)$ of partial derivatives is called the* **Jacobian matrix of F**, *and its determinant is called the* **Jacobian determinant of F**.

Definition 5. *If $F : U \to \mathbb{R}^m$ is a function for which each partial derivative exists at each point in U and the functions $\partial F^i/\partial x^j : U \to \mathbb{R}$ so defined are all continuous, then F is said to be of class* **C^1** *or* **continuously differentiable**. *If this is the case, we can differentiate the functions $\partial F^i/\partial x^j$ to obtain* **second-order partial derivatives**

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

if they exist. Continuing this way leads to higher-order partial derivatives: the **partial derivatives of F of order k** *are the (first) partial derivatives of those of order $k - 1$, when they exist.*

Definition 6. *If $U \subseteq \mathbb{R}^n$ is an open subset and $k \geq 0$, a function $F : U \to \mathbb{R}^m$ is said to be 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.*

Remark 3. *Thus a function of class C^0 is just a continuous function. Because existence and continuity of derivatives are local properties, clearly F is C^k iff it has the property in a neighborhood of each point in U.*

Definition 7. *A function that is of class C^k for every $k \geq 0$ is said to be of* **class C^∞**, *or* **smooth**, *or* **infinitely differentiable**. *If U and V are open subsets of Euclidean spaces, a function $F : U \to V$ is called a* **diffeomorphism** *if it is smooth and bijective and its inverse function is also smooth.*

Proposition 6. *Suppose $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ are open subsets and $F : U \to V$ is a diffeomorphism. Then $m = n$, and for each $a \in U$, the total derivative is invertible, with $DF(a)^{-1} = D(F^{-1})(F(a))$.*

Proof. Because $F^{-1} \circ F = Id_U$, the chain rule implies that for each $a \in U$,

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

Similarly, $F \circ F^{-1} = Id_V$ implies that for each $F(a) \in V$, we have

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

This implies that $DF(a)$ is invertible with inverse $D(F^{-1})(F(a))$, and therefore $m = n$. $\square$

Definition 8 (***Smoothness on Arbitrary Domains***)**.** *If $A \subseteq \mathbb{R}^n$ is an arbitrary subset, a function $F : A \to \mathbb{R}^m$ is said to be **smooth on A** if it admits a smooth extension to an open neighborhood of each point, or more precisely, if for every $x \in A$, there exists an open neighborhood $U_x \subseteq \mathbb{R}^n$ and a smooth function $\tilde{F} : U_x \to \mathbb{R}^m$ that agrees with F on $U_x \cap A$. The notion of diffeomorphism extends to arbitrary subsets in the obvious way: given arbitrary subsets $A, B \subseteq \mathbb{R}^n$, a **diffeomorphism from A to B** is a smooth bijective map $f : A \to B$ with smooth inverse.*

Definition 9. *If $U \subseteq \mathbb{R}^n$ is open, the set of all real-valued functions of class C^k on U is denoted by $C^k(U)$, and the set of all smooth real-valued functions by $C^\infty(U)$. Sums, constant multiples, and products of functions are defined pointwise: for $f, g : U \to \mathbb{R}$ and $c \in \mathbb{R}$,*

$$\begin{aligned}(f+g)(x) =& f(x) + g(x),\\ (cf)(x) =& c(f(x)),\\ (fg)(x) =& f(x)g(x).\end{aligned}$$

Proposition 7 (***Equality of Mixed Partial Derivatives***)**.** *If U is an open subset of $\mathbb{R}^n$ and $F : U \to \mathbb{R}^m$ is a function of class C^2, then the mixed second-order partial derivatives of F do not depend on the order of differentiation:*

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

Corollary. *If $F : U \to \mathbb{R}^m$ is smooth, then the mixed partial derivatives of F of any order are independent of the order of differentiation.*

Proposition 8. *Let $U \subseteq \mathbb{R}^n$ be open, and suppose $F : U \to \mathbb{R}^m$ is differentiable at $a \in U$. Then all of the partial derivatives of F at a exist, and $DF(a)$ is the linear map whose matrix is the Jacobian of F at a:*

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

Proof. Let $B = DF(a)$, and for $v \in \mathbb{R}^n$ small enough that $a + v \in U$, let $R(v) = F(a+v) - F(a) - Bv$. The fact that F is differentiable at a implies that each component of the vector-valued function $R(v)/\|v\|$ goes to zero as $v \to 0$. The i-th partial derivative of F^j at a, if it exists, is

$$\frac{\partial F^j}{\partial x^i}(a) = \underset{t\to 0}{Lt} \frac{F^j(a+te_i) - F^j(a)}{t} = \underset{t\to 0}{Lt} \frac{B^j_i t + R^j(te_i)}{t} = B^j_i + \underset{t\to 0}{Lt} \frac{R^j(te_i)}{t}.$$

The norm of the quotient on the right above is $\|R^j(te_i)\|/\|te_i\|$, which approaches zero as $t \to 0$. It follows that $\partial F^j/\partial x^i(a)$ exists and is equal to B^j_i as claimed. $\square$

Proposition 9. *Suppose $U \subseteq \mathbb{R}^n$ is open. Then $F : U \to \mathbb{R}^m$ is differentiable at $a \in U$ iff each of its component functions $F^1, \dots, F^m$ is differentiable at a and*

$$DF(a) = \begin{pmatrix} DF^1(a) \\ \vdots \\ DF^m(a) \end{pmatrix}$$

Proof. Using the fact that $y = (y_1, \dots, y_m) \to 0 \Leftrightarrow y_i \to 0$ for each i, from Remark 1 we see that

$$\begin{aligned} \underset{v\to 0}{Lt} \frac{\|F(a+v) - F(a) - DF(a)v\|}{\|v\|} = 0 &\Leftrightarrow \underset{v\to 0}{Lt} \frac{F(a+v) - F(a) - DF(a)v}{\|v\|} = 0 \\ &\Leftrightarrow \underset{v\to 0}{Lt} \frac{F^i(a+v) - F^i(a) - DF^i(a)v}{\|v\|} = 0, \ \forall i \\ &\Leftrightarrow \underset{v\to 0}{Lt} \frac{\|F^i(a+v) - F^i(a) - DF^i(a)v\|}{\|v\|} = 0, \ \forall i. \end{aligned}$$

□

Proposition 10. *Let $U \subseteq \mathbb{R}^n$ be open. If $F : U \to \mathbb{R}^n$ is of class C^1, then it is differentiable at each point of U.*

Proposition 11. *Let $U \subseteq \mathbb{R}^n$ be an open subset, and suppose $f, g \in C^\infty(U)$ and $c \in \mathbb{R}$.*

(a) Then $f + g$, cf, and fg are smooth.

(b) If g never vanishes on U, then f/g is smooth.

Proof. The result follows immediately by noting that each of the partial derivatives of $f + g$, cf, fg and f/g of any order are continuous as they can be written as sum, product, quotient of partial derivatives of f and g which are assumed to be continuous. □

Proposition 12 (***The Chain Rule for Partial Derivatives***)**.** *Let $U \subseteq \mathbb{R}^n$ and $\tilde{U} \subseteq \mathbb{R}^m$ be open subsets, and let $x = (x^1, \dots, x^n)$ denote the standard coordinates on U and $y = (y^1, \dots, y^m)$ those on $\tilde{U}$.*

(a) A composition of C^1 functions $F : U \to \tilde{U}$ and $G : \tilde{U} \to \mathbb{R}^p$ is again of class C^1, with partial derivatives given by

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

(b) If F and G are smooth, then $G \circ F$ is smooth.

Proof. ***(a)*** From the chain rule of total derivative(Prop. 4) and the Jacobian matrix formulation of total derivative(Prop. 8), the matrix of $D(G \circ F)$ will be the product of the Jacobian matrices of G and F. Since $H = G \circ F : U \to \mathbb{R}^p$, the components of $H = (H^1, \dots, H^p)$ can be written as $H^i = G^i \circ F : U \to \mathbb{R}$. Then we have

$$(\partial H^i / \partial x^j) = (\partial G^i / \partial y^k)(\partial F^k / \partial x^j),$$

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

Then each component $\partial H^i / \partial x^j$ is continuous because it is sum of product of continuous functions. Thus $G \circ F$ is also C^1.

(b) Repeated application of chain rule shows that $G \circ F$ is smooth. □

Proposition 13. *Suppose $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ are arbitrary subsets, and $F : A \to \mathbb{R}^m$ and $G : B \to \mathbb{R}^p$ are smooth maps (according to Def. 8) such that $F(A) \subseteq B$. Then $G \circ F : A \to \mathbb{R}^p$ is smooth.*

Proof. Let $x \in A$, then by smoothness of F, there exists a neighborhood U of x and a smooth map $\tilde{F} : U \to \mathbb{R}^m$ such that $\tilde{F}\Big|_{U \cap A} = F$. But $F(x) \in B$, so by smoothness of G, we find a neighborhood V of $F(x)$ and a smooth map $\tilde{G} : V \to \mathbb{R}^p$ such that $\tilde{G}\Big|_{V \cap B} = G$. Define $\tilde{U} = U \cap A \cap F^{-1}(V \cap B)$. Then $\tilde{U}$ is a neighborhood of x, and $\tilde{G} \circ \tilde{F} : \tilde{U} \to \mathbb{R}^p$ is a smooth map (by Prop. 12) such that $\tilde{G} \circ \tilde{F}\Big|_{\tilde{U}} = G \circ F$. ☐

Definition 10. *Suppose $f : U \to \mathbb{R}$ is a smooth real-valued function on an open subset $U \subseteq \mathbb{R}^n$ and $a \in U$. For each vector $v \in \mathbb{R}^n$, we define the* ***directional derivative of f in the direction v at a*** *to be the number*

$$D_v f(a) = \frac{d}{dt}\Big|_{t=0} f(a + tv). \tag{5}$$

Remark 4. *This definition makes sense for any vector v; we do not require v to be a unit vector as one sometimes does in elementary calculus.*

Remark 5. *Since $D_v f(a)$ is the ordinary derivative of the composite function $t \mapsto a + tv \mapsto f(a + tv)$, by chain rule it can be written more concretely as*

$$D_v f(a) = \sum_{i=1}^{n} v^i \frac{\partial f}{\partial x^i}(a) = Df(a)v.$$

Proposition 14 (***Differentiation Under an Integral Sign***)**.** *Let $U \subseteq \mathbb{R}^n$ be an open subset, let $a, b \in \mathbb{R}$, and let $f : U \times [a, b] \to \mathbb{R}$ be continuous function such that partial derivatives $\partial f / \partial x^i : U \times [a, b] \to \mathbb{R}$ exist and are continuous on $U \times [a, b]$ for $i = 1, \dots, n$. Define $F : U \to \mathbb{R}$ by*

$$F(x) = \int_a^b f(x, t) dt.$$

Then F is of class C^1, and its partial derivatives can be computed by differentiating under the integral sign:

$$\frac{\partial F}{\partial x^i}(x) = \int_a^b \frac{\partial f}{\partial x^i}(x, t) dt.$$

For any m-tuple $I = (i_1, \dots, i_m)$ of indices with $1 \leq i_j \leq n$, we let $|I| = m$ denote the number of indices in I, and

$$\begin{aligned} \partial_I =& \frac{\partial^m}{\partial x^{i_1} \dots \partial x^{i_m}}, \\ (x - a)^I =& (x^{i_1} - a^{i_1}) \dots (x^{i_m} - a^{i_m}). \end{aligned}$$

Proposition 15 (***Taylor's Theorem***)**.** *Let $U \subseteq \mathbb{R}^n$ be an open subset, and let $a \in U$ be fixed. Suppose $f \in C^{k+1}(U)$ for some $k \geq 0$. If W is any convex subset of U containing a, then for all $x \in W$,*

$$f(x) = P_k(x) + R_k(x), \tag{6}$$

where P_k is the ***k-th order Taylor polynomial of f at a****, defined by*

$$P_k(x) = f(a) + \sum_{m=1}^{k} \frac{1}{m!} \sum_{I : |I| = m} \partial_I f(a)(x - a)^I, \tag{7}$$

and R_k is the ***k-th remainder term****, given by*

$$R_k(x) = \frac{1}{k!} \sum_{I\colon |I|=k+1} (x-a)^I \int_0^1 (1-t)^k \partial_I f(a+t(x-a))dt. \tag{8}$$

Proof. For $k=0$ (where we interpret P_0 to mean $f(a)$), this is just the fundamental theorem of calculus (Prop. 14) applied to the function $u(t) = f(a+t(x-a))$, together with the chain rule. Assume the result holds for some k, integration by parts applied to the integral in the remainder term yield

$$\begin{aligned}
&\int_0^1 (1-t)^k \partial_I f(a+t(x-a))dt \\
&\quad = \Big[-\frac{(1-t)^{k+1}}{k+1} \partial_I f(a+t(x-a)) \Big]_{t=0}^{t=1} + \int_0^1 \frac{(1-t)^{k+1}}{k+1} \frac{\partial}{\partial t}(\partial_I f(a+t(x-a)))dt \\
&\quad = \frac{1}{k+1}\partial_I f(a) + \frac{1}{k+1} \sum_{j=1}^n (x^j - a^j) \int_0^1 (1-t)^{k+1} \frac{\partial}{\partial x^j} \partial_I f(a+t(x-a))dt.
\end{aligned}$$

When we insert this into (6), we obtain the analogous formula with k replaced by $k+1$. $\square$

Corollary. *Suppose $U \subseteq \mathbb{R}^n$ is an open subset, $a \in U$, and $f \in C^{k+1}(U)$ for some $k \geq 0$. If W is a convex subset of U containing a on which all of the $(k+1)$-st partial derivatives of f are bounded in absolute value by a constant M, then for all $x \in W$,*

$$|f(x) - P_k(x)| \leq \frac{n^{k+1}M}{(k+1)!} |x-a|^{k+1},$$

where P_k is the k-th Taylor polynomial of f at a, defined by (7).

Proof. There are n^{k+1} terms on the right-hand side of (8), each term is bounded in absolute value by $(1/(k+1)!)|x-a|^{k+1}M$. $\square$

Proposition 16 (***Lipschitz Estimate for C^1 Functions***)**.** *Let $U \subseteq \mathbb{R}^n$ be an open subset, and suppose $F : U \to \mathbb{R}^m$ is of class C^1. Then F is Lipschitz continuous on every compact convex subset $K \subseteq U$. The Lipschitz constant can be taken to be $sup_{x \in K}\|DF(x)\|$.*

Proof. Since $\|DF(x)\|$ is a continuous function of x, it is bounded on the compact set K. Let $M = sup_{x\in K}\|DF(x)\|$. For arbitrary $a, b \in K$, we have $a + t(b-a) \in K$ for all $t \in [0,1]$ because K is convex. By the fundamental theorem of calculus applied to each component of F, together with the chain rule,

$$\begin{aligned}
F(b) - F(a) &= \int_0^1 \frac{d}{dt} F(a+t(b-a))dt \\
&= \int_0^1 DF(a+t(b-a))(b-a)dt. \\
\implies \|F(b) - F(a)\| &\leq \int_0^1 \|DF(a+t(b-a))\| \|b-a\| dt \\
&\leq \int_0^1 M\|b-a\|dt = M\|b-a\|.
\end{aligned}$$

$\square$

Corollary. *If $U \subseteq \mathbb{R}^n$ is an open subset and $F : U \to \mathbb{R}^m$ is of class C^1, then f is locally Lipschitz continuous.*

Proof. Each point of U is contained in a ball whose closure is contained in U, and Prop. 16 shows that the restriction of F to such a ball is Lipschitz continuous. $\square$

3 The Inverse Function Theorem and Related Results

Definition 11. *Let (X, d) be a metric space. A map $G : X \to X$ is said to be a* ***contraction*** *if there is a constant $\lambda \in (0, 1)$ such that $d(G(x), G(y)) \le \lambda d(x, y)$ for all $x, y \in X$. A* ***fixed point*** *of a map $G : X \to X$ is a point $x \in X$ such that $G(x) = x$.*

Remark 6. *Clearly, every contraction is continuous.*

Proposition 17 (***Contraction Lemma***)**.** *Let X be a nonempty complete metric space. Every contraction $G : X \to X$ has a unique fixed point.*

Proof. Uniqueness is immediate, for if x, x' are both fixed points of G, the contraction property implies that $d(x, x') = d(G(x), G(x')) \le \lambda d(x, x')$, which is possible only if $x = x'$.

To prove the existence of a fixed point, let x_0 be an arbitrary point in X, and define a sequence $(x_n)_{n=0}^{\infty}$ inductively by $x_{n+1} = G(x_n)$. For any $i \ge 1$ we have $d(x_i, x_{i+1}) = d(G(x_{i-1}), G(x_i)) \le \lambda d(x_{i-1}, x_i)$, and therefore by induction $d(x_i, x_{i+1}) \le \lambda^i d(x_0, x_1)$. If N is a positive integer and $j \ge i \ge N$,

$$\begin{aligned} d(x_i, x_j) &\le d(x_i, x_{i+1}) + d(x_{i+1}, x_{i+2}) + \ldots + d(x_{j-1}, x_j) \\ &\le (\lambda^i + \ldots + \lambda^{j-1}) d(x_0, x_1) \le \lambda^i \Big(\sum_{n=0}^{\infty} \lambda^n \Big) d(x_0, x_1) \\ &\le \lambda^N \Big(\frac{1}{1 - \lambda} \Big) d(x_0, x_1), \end{aligned}$$

where we have used that $\lambda^N \ge \lambda^i$ for $i \ge N$. Since the last expression can be made as small as desired by choosing N large, the sequence (x_n) is Cauchy and therefore converges to a limit $x \in X$. Because G is continuous,

$$x_n \to x \implies G(x_n) \to G(x), \text{ but } G(x) = \underset{n \to \infty}{Lt}\, G(x_n) = \underset{n \to \infty}{Lt}\, x_{n+1} = x,$$

so x is the desired fixed point. □

Proposition 18 (***Inverse Function Theorem***)**.** *Suppose U, V are open subsets of $\mathbb{R}^n$, and $F : U \to V$ is a smooth function. If $DF(a)$ is invertible, i.e., Jacobian determinant is nonzero, at some point $a \in U$, then there exists connected neighborhoods $U_0 \subseteq U$ of a and $V_0 \subseteq V$ of $F(a)$ such that $F|_{U_0} : U_0 \to V_0$ is a diffeomorphism.*

Proof. We begin by making some simple modifications to the function F to streamline the proof. First, the function F_1 defined by $F_1(x) = F(x + a) - F(a)$ is smooth on a neighborhood of 0 and satisfies $F_1(0) = 0$ and $DF_1(0) = DF(a)$; clearly, F is a diffeomorphism on a connected neighborhood of a iff F_1 is a diffeomorphism on a connected neighborhood of 0. Second, the function $F_2 = DF_1(0)^{-1} \circ F_1$ is smooth on the same neighborhood of 0 and satisfies $F_2(0) = 0$ and $DF_2(0) = I_n$; and F_2 is a diffeomorphism on a connected neighborhood of 0 iff F_1 is a diffeomorphism and therefore also F. Henceforth, replacing F by F_2, we assume that F is defined in a neighborhood U of 0, $F(0) = 0$ and $DF(0) = I_n$. Because the determinant of $DF(x)$ is a continuous function of x, by shrinking U if necessary, we may assume that $DF(x)$ is invertible for each $x \in U$.

Let $H(x) = x - F(x)$ for each $x \in U$. Then $DH(0) = I_n - I_n = 0$. Because the matrix entries of $DH(x)$ are continuous functions of x, there is a number $\delta > 0$ such that $\mathbb{B}_0(\delta) \subseteq U$ and for all $x \in \bar{\mathbb{B}}_0(\delta)$, we have $\|DH(x)\| \le \frac{1}{2}$. If $x, x' \in \bar{\mathbb{B}}_0(\delta)$, the Lipschitz estimate for smooth functions (Prop. 16) implies that

$$\|H(x) - H(x')\| \le \frac{1}{2} \|x - x'\|. \tag{9}$$

In particular, taking $x' = 0$, this implies

$$\|H(x)\| \leq \frac{1}{2}\|x\|. \tag{10}$$

Since $x' - x = F(x') - F(x) + H(x') - H(x)$, it follows that

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

and rearranging gives

$$\|x' - x\| \leq 2\|F(x') - F(x)\| \tag{11}$$

for all $x, x' \in \bar{\mathbb{B}}_0(\delta)$. In particular, this shows that F is injective on $\bar{\mathbb{B}}_0(\delta)$.

Now let $y \in \mathbb{B}_0(\delta/2)$ be arbitrary. We will show that there exists a unique point $x \in \mathbb{B}_0(\delta)$ such that $F(x) = y$. Let $G(x) = y + H(x) = y + x - F(x)$, so that $G(x) = x$ iff $F(x) = y$. If $\|x\| \leq \delta$, (10) implies

$$\|G(x)\| \leq \|y\| + \|H(x)\| < \frac{\delta}{2} + \frac{1}{2}\|x\| \leq \delta, \tag{12}$$

so G maps $\bar{\mathbb{B}}_0(\delta)$ to itself. It follows from (9) that $\|G(x') - G(x)\| = \|H(x) - H(x')\| \leq \frac{1}{2}\|x - x'\|$, so G is a contraction. Since $\bar{\mathbb{B}}_0(\delta)$ is a complete metric space, the contraction lemma implies that G has a unique fixed point $x \in \bar{\mathbb{B}}_0(\delta)$. From (12), $\|x\| = \|G(x)\| < \delta$, so in fact $x \in \mathbb{B}_0(\delta)$, thus proving the claim.

Let $V_0 = \mathbb{B}_0(\delta/2)$ and $U_0 = \mathbb{B}_0(\delta) \cap F^{-1}(V_0)$. Then U_0 is open in $\mathbb{R}^n$, and the argument above shows that $F : U_0 \to V_0$ is bijective, so $F^{-1} : V_0 \to U_0$ exists. Substituting $x = F^{-1}(y)$ and $x' = F^{-1}(y')$ into (11) shows that F^{-1} is continuous. Thus $F : U_0 \to V_0$ is a homeomorphism, and it follows that U_0 is connected because V_0 is.

The only thing that remains to be proved is that F^{-1} is smooth. If we knew it were smooth, Prop. 6 would imply that $D(F^{-1})(y) = DF(x)^{-1}$, where $x = F^{-1}(y)$. We begin by showing that F^{-1} is differentiable to each point of V_0, with total derivative given by this formula.

Let $y \in V_0$ be arbitrary, and set $x = F^{-1}(y)$ and $L = DF(x)$. We need to show that

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

Given $y' \in V_0 - \{y\}$, write $x' = F^{-1}(y') \in U_0 - \{x\}$. Then

$$\begin{aligned}
\frac{F^{-1}(y') - F^{-1}(y) - L^{-1}(y' - y)}{\|y' - y\|} &= L^{-1}\left(\frac{L(x' - x) - (y' - y)}{\|y' - y\|}\right) \\
&= \frac{\|x' - x\|}{\|y' - y\|} L^{-1}\left(-\frac{F(x') - F(x) - L(x' - x)}{\|x' - x\|}\right).
\end{aligned}$$

The factor $\|x' - x\|/\|y' - y\|$ above is bounded due to (11), and because L^{-1} is linear and therefore bounded, $\|L^{-1}\|$ is bounded. As $y' \to y$, it follows that $x' \to x$ by continuity of F^{-1}, and then the term in the bracket of last equation goes to zero because $L = DF(x)$ and F is differentiable. This complete the proof that F^{-1} is differentiable.

By Prop. 8, the partial derivatives of F^{-1} are defined at each point $y \in V_0$. Observe that the formula $D(F^{-1})(y) = DF(F^{-1}(y))^{-1}$ implies that the matrix-valued function $y \mapsto D(F^{-1})(y)$ can be written as the composition

$$y \overset{F^{-1}}{\longmapsto} F^{-1}(y) \overset{DF}{\longmapsto} DF(F^{-1}(y)) \overset{i}{\longmapsto} DF(F^{-1}(y))^{-1}, \tag{13}$$

where i is the matrix inversion. In the composition, F^{-1} is continuous; DF is smooth because its component functions are the partial derivatives of F; and i is smooth because Cramer's rule expresses the entries of an inverse matrix as rational functions of entries of the matrix. Because

$D(F^{-1})$ is composition of continuous functions, it is continuous. Thus, the partial derivatives of F^{-1} are continuous, so F^{-1} is of class C^1.

Now assume by induction that we have shown that F^{-1} is of class C^k. This means that each of the functions in (13) is of class C^k. Because $D(F^{-1})$ is a composition of C^k functions, it is itself C^k; this implies that partial derivatives of F^{-1} are of class C^k, so F^{-1} itself is of class C^{k+1}. Continuing by induction, we conclude that F^{-1} is smooth. □

Corollary. *Suppose $U \subseteq \mathbb{R}^n$ is an open subset, and $F : U \to \mathbb{R}^m$ is a smooth function whose Jacobian determinant is nonzero at every point in U. Then*

(a) F is an open map.

(b) If F is injective, then $F : U \to F(U)$ is a diffeomorphism.

Proof. ***(a)*** For each $a \in U$, the fact that the Jacobian determinant of F is nonzero implies that $DF(a)$ is invertible, so the inverse function theorem implies that there exists open subsets $U_a \subseteq U$ containing a and $V_a \subseteq F(U)$ containing $F(a)$ such that F restricts to a diffeomorphism $F|_{U_a} : U_a \to V_a$. In particular, this means that each point of $F(U)$ has a neighborhood contained in $F(U)$, so $F(U)$ is open. If $U_0 \subseteq U$ is an arbitrary open subset, the same argument with U replaced by U_0 shows that $F(U_0)$ is also open.

(b) If F is injective, then the inverse map $F^{-1} : F(U) \to U$ exists; on a neighborhood of each point $F(a) \in F(U)$ F^{-1} defined above is equal to the inverse of $F|_{U_a}$, so it is smooth. □

Proposition 19 (***Implicit Function Theorem***)**.** *Let $U \in \mathbb{R}^n \times \mathbb{R}^k$ be an open subset, and let $(x, y) = (x^1, \ldots, x^n, y^1, \ldots, y^k)$ denote the standard coordinates on U. Suppose $\Phi : U \to \mathbb{R}^k$ is a smooth function, $(a, b) \in U$, and $c = \Phi(a, b)$. If the $k \times k$ matrix $(\partial\Phi^i(a, b)/\partial y^j)$ is nonsingular, then there exists neighborhoods $V_0 \subseteq \mathbb{R}^n$ of a and $W_0 \in R^k$ of b and a smooth function $F : V_0 \to W_0$ such that $\Phi^{-1}(c) \cap (V_0 \times W_0)$ is the graph of F, i.e., $\Phi(x, y) = c$ for $(x, y) \in V_0 \times W_0$ iff $y = F(x)$.*

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

$$D\Psi(a, b) = \begin{pmatrix} I_n & 0 \\ \dfrac{\partial\Phi^i}{\partial x^j}(a, b) & \dfrac{\partial\Phi^i}{\partial y^j}(a, b) \end{pmatrix},$$

which is nonsingular because it is block lower triangular and the two blocks on the main diagonal are nonsingular. Thus by inverse function theorem there exists connected neighborhood U_0 of (a, b) and Y_0 of (a, c) such that $\Psi : U_0 \to Y_0$ is a diffeomorphism. Since $\Psi : U_0 \to Y_0$ is defined by $\Psi(x, y) = (x, \Phi(x, y))$, the inverse map $\Psi^{-1} : Y_0 \to U_0$ will be of the form $\Psi^{-1}(x, y) = (x, B(x, y))$ for smooth function $B : Y_0 \to \mathbb{R}^k$. Shrinking U_0 and Y_0 if necessary, we may assume that $U_0 = V \times W$ is a product neighborhood.

The two compositions $\Psi \circ \Psi^{-1}$ and $\Psi^{-1} \circ \Psi$ give

$$\begin{aligned} (x, y) &= (\Psi \circ \Psi^{-1})(x, y) = \Psi(x, B(x, y)) = (x, \Phi(x, B(x, y))),\ \forall\ (x, y) \in Y_0 \\ (x, y) &= (\Psi^{-1} \circ \Psi)(x, y) = \Psi^{-1}(x, \Phi(x, y)) = (x, B(x, \Phi(x, y)))\ \forall\ (x, y) \in U_0. \end{aligned} \tag{14}$$

If $\Phi(x, y) = c$, then the second equation of (14) gives $y = B(x, c)$. This suggests that we define $F(x) = B(x, c)$ for all $x \in \mathbb{R}^n$ for which $(x, c) \in Y_0$. Now let $V_0 = \{x \in V : (x, c) \in Y_0\}$ and $W_0 = W$, then $F : V_0 \to W_0$ defined by $F(x) = B(x, c)$.

Let $x \in V_0$. If $\Phi(x, y) = c$ then $y = B(x, c) = F(x)$, so the graph of F is contained in $\Phi^{-1}(c)$. Conversely, suppose $y = F(x)$ and in the first equation of (14) we set $(x, y) = (x, c)$, then $c = \Phi(x, B(x, c)) = \Phi(x, F(x)) = \Phi(x, y)$. This completes the proof. □

Proposition 20. *The implicit function theorem is equivalent to the inverse function theorem.*

Proof. ($\Longrightarrow$)Already shown above.
($\Longleftarrow$)Let $F : U \to V$ be a smooth map defined such that $U, V \subseteq \mathbb{R}^n$ are open subsets such that at some point $p \in U$ the Jacobian determinant is nonzero. Finding a local inverse for $y = F(x)$ near p amounts to solving the equation $G(x, y) = F(x) - y = 0$ for x in terms of y near $(p, F(p))$. Note that $\partial G^i/\partial x^j = \partial F^i/\partial x^j$. Hence,

$$det\left[\frac{\partial G^i}{\partial x^j}(p, F(p))\right] = det\left[\frac{\partial F^i}{\partial x^j}(p, F(p))\right] \neq 0.$$

By the implicit function theorem, x can be expressed in terms of y locally near $(p, F(p))$, i.e., there is a smooth function $x = H(y)$ defined in a neighborhood of $F(p)$ in $\mathbb{R}^n$ such that $G(x, y) = F(x) - y = F(H(y)) - y = 0$. Thus, $y = F(H(y))$. Since $y = F(x)$, $x = H(y) = H(F(x))$. Therefore, F and H are inverse functions defined near p and $F(p)$ respectively and H is smooth by implicit function theorem. $\square$

References

[1] John M. Lee, Introduction to Smooth Manifolds.