Lecture 5

2.3 Chain Rule

Let U and v be open sets in $\mathbb{R}^n$. Consider maps $f : U \to V$ and $g : V \to \mathbb{R}^k$. Choose $a \in U$, and let $b = f(a)$. The composition $g \circ f : U \to \mathbb{R}^k$ is defined by $(g \circ f)(x) = g(f(x))$.

Theorem 2.9. *If f is differentiable at a and g is differentiable at b, then $g \circ f$ is differentiable at a, and the derivative is*

$$(Dg \circ f)(a) = (Dg)(b) \circ Df(a). \tag{2.43}$$

Proof. This proof follows the proof in Munkres by breaking the proof into steps.

- Step 1: Let $h \in \mathbb{R}^n - \{0\}$ and $h \doteq 0$, by which we mean that h is very close to zero. Consider $\Delta(h) = f(a+h) - f(a)$, which is continuous, and define

  $$F(h) = \frac{f(a+h) - f(a) - Df(a)h}{|a|}. \tag{2.44}$$

  Then f is differentiable at a if and only if $F(h) \to 0$ as $h \to 0$.

  $$F(h) = \frac{\Delta(h) - Df(a)h}{|h|}, \tag{2.45}$$

  so

  $$\Delta(h) = Df(a)h + |h|F(h). \tag{2.46}$$

  Lemma 2.10.

  $$\frac{\Delta(h)}{|h|} \textit{ is bounded.} \tag{2.47}$$

  Proof. Define

  $$|Df(a)| = \sup_i \left| \frac{\partial f}{\partial x_i}(a) \right|, \tag{2.48}$$

  and note that

  $$\frac{\partial f}{\partial x_i}(a) = Df(a)e_i, \tag{2.49}$$

  where the e_i are the standard basis vectors of $\mathbb{R}^n$. If $h = (h_1, \ldots, h_n)$, then $h = \sum h_i e_i$. So, we can write

  $$Df(a)h = \sum h_i Df(a)e_i = \sum h_i \frac{\partial f}{\partial x_i}(a). \tag{2.50}$$

It follows that

$$\begin{aligned}|Df(a)h| &\le \sum_{i=1}^{m} h_i \left|\frac{\partial f}{\partial x_i}(a)\right| \\ &\le m|h||Df(a)|.\end{aligned} \tag{2.51}$$

By Equation 2.46,

$$|\Delta(h)| \le m|h||Df(a)| + |h|F(h), \tag{2.52}$$

so

$$\frac{|\Delta(h)|}{|h|} \le m|Df(a)| + F(h). \tag{2.53}$$

□

- Step 2: Remember that $b = f(a)$, $g : V \to \mathbb{R}^k$, and $b \in V$. Let $k \doteq 0$. This means that $k \in \mathbb{R}^n - \{0\}$ and that k is close to zero. Define

$$G(k) = \frac{g(b+k) - g(b) - (Dg)(b)k}{|k|}, \tag{2.54}$$

so that

$$g(b+k) - g(b) = Dg(b)k + |k|G(k). \tag{2.55}$$

We proceed to show that $g \circ f$ is differentiable at a.

$$\begin{aligned}g \circ f(a+h) - g \circ f(a) &= g(f(a+h)) - g(f(a)) \\ &= g(b + \Delta(h)) - g(b),\end{aligned} \tag{2.56}$$

where $f(a) = b$ and $f(a+h) = f(a) + \Delta(h) = b + \Delta(h)$. Using Equation 2.55 we see that the above expression equals

$$Dg(b)\Delta(h) + |\Delta(h)|G(\Delta(h)). \tag{2.57}$$

Substituting in from Equation 2.46, we obtain

$$\begin{aligned}g \circ f(a+h) - g \circ f(a) &= \dots \\ &= Dg(b)(Df(a)h + |h|F(h)) + \dots \\ &= Dg(b) \circ Df(a)h + |h|Dg(b)F(h) + |\Delta(h)|G(\Delta(h))\end{aligned} \tag{2.58}$$

This shows that

$$\frac{g \circ f(a+h) - g \circ f(a) - Dg(b) \circ Df(a)h}{|h|} = Dg(b)F(h) + \frac{\Delta(h)}{|h|}G(\Delta(h)). \tag{2.59}$$

We see in the above equation that $g \circ f$ is differentiable at a if and only if the l.h.s. goes to zero as $h \to 0$. It suffices to show that the r.h.s. goes to zero as $h \to 0$, which it does: $F(h) \to 0$ as $h \to 0$ because f is differentiable at a; $G(\Delta(h)) \to 0$ because g is differentiable at b; and $\Delta(h)/|h|$ is bounded.

□

We consider the same maps g and f as above, and we write out f in component form as $f = (f_1, \ldots, f_n)$ where each $f_i : U \to \mathbb{R}$. We say that f is a $\mathcal{C}^r$ map if each $f_i \in \mathcal{C}^r(U)$. We associate $Df(x)$ with the matrix

$$Df(x) \sim \left[\frac{\partial f_i}{\partial x_j}(x)\right]. \tag{2.60}$$

By definition, f is $\mathcal{C}^r$ (that is to say $f \in \mathcal{C}^r(U)$) if and only if Df is $\mathcal{C}^{r-1}$.

Theorem 2.11. *If $f : U \to V \subseteq \mathbb{R}^n$ is a $\mathcal{C}^r$ map and $g : V \to \mathbb{R}^p$ is a $\mathcal{C}^r$ map, then $g \circ f : U \to \mathbb{R}^p$ is a $\mathcal{C}^r$ map.*

Proof. We only prove the case $r = 1$ and leave the general case, which is inductive, to the student.

- Case $r = 1$:

$$Dg \circ f(x) = Dg(f(x)) \circ Df(x) \sim \left[\frac{\partial g_i}{\partial x_j} f(x)\right]. \tag{2.61}$$

  The map g is $\mathcal{C}^1$, which implies that $\partial g_i / \partial x_j$ is continuous. Also,

$$Df(x) \sim \left[\frac{\partial f_i}{\partial x_j}\right] \tag{2.62}$$

  is continuous. It follows that $Dg \circ f(x)$ is continuous. Hence, $g \circ f$ is $\mathcal{C}^1$.

□

2.4 The Mean-value Theorem in n Dimensions

Theorem 2.12. *Let U be an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$ a $\mathcal{C}^1$ map. For $a \in U$, $h \in \mathbb{R}^n$, and $h \doteq 0$,*

$$f(a + h) - f(a) = Df(c)h, \tag{2.63}$$

where c is a point on the line segment $a + th, 0 \leq t \leq 1$, joining a to $a + h$.

Proof. Define a map $\phi : [0, 1] \to \mathbb{R}$ by $\phi(t) = f(a + th)$. The Mean Value Theorem implies that $\phi(1) - \phi(0) = \phi'(c) = (Df)(c)h$, where $0 < c < 1$. In the last step we used the chain rule. □

2.5 Inverse Function Theorem

Let U and V be open sets in $\mathbb{R}^n$, and let $f : U \to V$ be a $\mathcal{C}^1$ map. Suppose there exists a map $g : V \to U$ that is the inverse map of f (which is also $\mathcal{C}^1$). That is, $g(f(x)) = x$, or equivalently $g \circ f$ equals the identity map.

Using the chain rule, if $a \in U$ and $b = f(a)$, then

$$(Dg)(b) = \text{ the inverse of } Df(a). \tag{2.64}$$

That is, $Dg(b) \circ Df(a)$ equals the identity map. So,

$$Dg(b) = (Df(a))^{-1} \tag{2.65}$$

However, this is not a trivial matter, since we do not know if the inverse exists. That is what the inverse function theorem is for: if $Df(a)$ is invertible, then g exists for some neighborhood of a in U and some neighborhood of $f(a)$ in V. We state this more precisely in the following lecture.