Lecture 3

2 Differentiation

2.1 Differentiation in n dimensions

We are setting out to generalize to n dimensions the notion of differentiation in one-dimensional calculus. We begin with a review of elementary one-dimensional calculus.

Let $I \subseteq \mathbb{R}$ be an open interval, let $f : I \to \mathbb{R}$ be a map, and let $a \in I$.

Definition 2.1. The *derivative of f at a* is

$$f'(a) = \lim_{t \to 0} \frac{f(a+t) - f(a)}{t}, \tag{2.1}$$

provided that the limit exists. If the limit exists, then *f is differentiable at a.*

There are half a dozen or so possible reasonable generalizations of the notion of derivative to higher dimensions. One candidate generalization which you have probably already encountered is the directional derivative.

Definition 2.2. Given an open set U in $\mathbb{R}^n$, a map $f : U \to \mathbb{R}^m$, a point $a \in U$, and a point $u \in \mathbb{R}^n$, the *directional derivative of f in the direction of u at a* is

$$D_u f(a) = \lim_{t \to 0} \frac{f(a+tu) - f(a)}{t}, \tag{2.2}$$

provided that the limit exists.

In particular, we can calculate the directional derivatives in the direction of the standard basis vectors $e_1, \ldots, e_n$ of $\mathbb{R}^n$, where

$$e_1 = (1, 0, \ldots, 0), \tag{2.3}$$
$$e_2 = (0, 1, 0, \ldots, 0), \tag{2.4}$$
$$\vdots \tag{2.5}$$
$$e_n = (0, \ldots, 0, 1). \tag{2.6}$$

Notation. The directional derivative in the direction of a standard basis vector e_i of $\mathbb{R}^n$ is denoted by

$$D_i f(a) = D_{e_i} f(a) = \frac{\partial}{\partial x_i} f(a). \tag{2.7}$$

We now try to answer the following question: What is an adequate definition of differentiability at a point a for a function $f : U \to \mathbb{R}^m$?

- Guess 1: Require that $\frac{\partial f}{\partial x_i}(a)$ exists.

  However, this requirement is inadequate. Consider the function defined by

$$f(x_1, x_2) = \begin{cases} 0, & \text{if } (x_1, x_2) \text{ lies on the } x_1\text{-axis or the } x_2\text{-axis}, \\ 1, & \text{otherwise.} \end{cases} \tag{2.8}$$

  Then, both

$$\frac{\partial f}{\partial x_1}(0) = 0 \text{ and } \frac{\partial f}{\partial x_2}(0) = 0, \tag{2.9}$$

  but the function f is not differentiable at $(0, 0)$ along any other direction.

- Guess 2: Require that all directional derivatives exist at a.

  Unfortunately, this requirement is still inadequate. For example (from Munkres chapter 5), consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} \frac{xy^2}{x^2+y^4}, & (x, y) \neq (0, 0) \\ 0, & x = y = 0. \end{cases} \tag{2.10}$$

  Claim. *The directional derivative $D_u f(0)$ exists for all u.*

  Proof. Let $u = (h, k)$. Then

$$\begin{aligned} \lim_{t\to 0} \frac{f(tu) - f(0)}{t} &= \lim_{t\to 0} \frac{f(tu)}{t} \\ &= \lim_{t\to 0} \left(\frac{t^3 hk^2}{t^2h^2 + t^4k^4} \right) \frac{1}{t} \\ &= \begin{cases} 0, & h = 0 \\ k^2/h, & h \neq 0. \end{cases} \end{aligned} \tag{2.11}$$

  So the limit exists for every u. □

  However, the function is a non-zero constant on a parabola passing through the origin: $f(t^2, t) = \frac{t^4}{2t^4} = \frac{1}{2}$, except at the origin where $f(0, 0) = 0$. The function f is discontinuous at the origin despite the existence of all directional derivatives.

- Guess 3. This guess will turn out to be correct.

  Remember than in one-dimensional calculus we defined

$$f'(a) = \lim_{t\to 0} \frac{f(a + t) - f(a)}{t}, \tag{2.12}$$

  for a function $f : I \to \mathbb{R}$ and a point $a \in I$. Now consider the function $\lambda : \mathbb{R} \to \mathbb{R}$ defined by

$$\lambda(t) = f'(a)t. \tag{2.13}$$

Then,

$$\lim_{t\to 0} \frac{f(a+t)-f(a)-\lambda(t)}{t} = \lim_{t\to 0} \frac{f(a+t)-f(a)}{t} - f'(a) \\ = 0. \tag{2.14}$$

So, $\lambda(t) \approx f(a+t) - f(a)$ when t is small.

Now we generalize to n dimensions.

Definition 2.3. Given an open subset U of $\mathbb{R}^n$, a map $f : U \to \mathbb{R}^m$, and a point $a \in U$, the function f is *differentiable at* a if there exists a linear mapping $B : \mathbb{R}^n \to \mathbb{R}^m$ such that for every $h \in \mathbb{R}^n - \{0\}$,

$$\frac{f(a+h)-f(a)-Bh}{|h|} \to 0 \text{ as } h \to 0. \tag{2.15}$$

That is, $f(a+h) - f(a) \approx Bh$ when h is small.

Theorem 2.4. *If f is differentiable at a, then for every u the directional derivative of f in the direction of u at a exists.*

Proof. The function f is differentiable at a, so

$$\frac{f(a+tu)-f(a)-B(tu)}{|tu|} \to 0 \text{ as } t \to 0. \tag{2.16}$$

Furthermore,

$$\begin{aligned} \frac{f(a+tu)-f(a)-B(tu)}{|tu|} &= \frac{t}{|tu|}\frac{f(a+tu)-f(a)-B(tu)}{t} \\ &= \frac{t}{|t|}\frac{1}{|u|}\left(\frac{f(a+tu)-f(a)}{t} - Bu\right) \\ &\to 0, \end{aligned} \tag{2.17}$$

as $t \to 0$, so

$$\frac{f(a+tu)-f(a)}{t} \to Bu \text{ as } t \to 0. \tag{2.18}$$

□

Furthermore, the linear map B is unique, so the following definition is well-defined.

Definition 2.5. The *derivative of f at a* is $Df(a) = B$, where B is the linear map defined above.

Note that $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map.

Theorem 2.6. *If f is differentiable at a, then f is continuous at a.*

Sketch of Proof. Note that for $h \neq 0$ in $\mathbb{R}^n$,

$$\frac{f(a+h) - f(a) - Bh}{|h|} \to 0 \text{ as } h \to 0 \tag{2.19}$$

implies that

$$f(a+h) - f(a) - Bh \to 0 \text{ as } h \to 0. \tag{2.20}$$

From this you can conclude that f is continuous at a. □

Remark. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map and $a \in \mathbb{R}^n$. The point a can be written as a sum $a = \sum_{j=1}^n a_j e_j = (a_1, \dots, a_n)$. The point La can be written as the sum $La = \sum a_j L e_j$, and L can be written out in components as $L = (L_1, \dots, L_m)$, where each $L_j : \mathbb{R}^n \to \mathbb{R}$ is a linear map. Then $Le_j = (L_1, e_j, \dots, L_m e_j)$, and $L_i e_j = \ell_{i,j}$. The numbers $\ell_{i,j}$ form an $n \times n$ matrix denoted by $[\ell_{i,j}]$.

Remark. Let $U \subseteq \mathbb{R}^n$, and let $f_1 : \mathbb{R}^n \to \mathbb{R}^{m_1}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}^{m_2}$ be differentiable maps. Let $m = m_1 + m_2$, so that $\mathbb{R}^{m_1} \times \mathbb{R}^{m_2} = \mathbb{R}^m$. Now, construct a function $f : \mathbb{R}^n \to \mathbb{R}^m$ defined in component form by $f = (f_1, f_2)$. The derivative of f at a is

$$Df(a) = (Df_1(a), Df_2(a)). \tag{2.21}$$

Remark. Let $f : U \to \mathbb{R}^m$ be a map. The action of f on input x written out in component form is $f(x) = (f_1(x), \dots, f_m(x))$. So, the map can be represented in component form as $f = (f_1, \dots, f_m)$, where each f_i as a map of the form $f_i : U \to \mathbb{R}$. The derivative of f acting on the standard basis vector e_j is

$$\begin{aligned} Df(a)e_j &= (Df_1(a)e_j, \dots, Df_m(a)e_j) \\ &= (\frac{\partial f_1}{\partial x_j}(a), \dots, \frac{\partial f_m}{\partial x_j}(a)). \end{aligned} \tag{2.22}$$

So, the derivative $(Df)(a)$ can be represented by an $m \times n$ matrix

$$(Df)(a) \cong J_f(a) = \left[\frac{\partial f_i}{\partial x_j}(a)\right] \tag{2.23}$$

called the Jacobian matrix of f at a, which you probably recognize.