8: DIFFERENTIATION IN SEVERAL VARIABLES

STEVEN HEILMAN

Contents

1. Review

Definition 1.1 (**Derivative on the real line**). Let E be a subset of $\mathbb{R}$, and let x_0 be a limit point of E, and let $f\colon E\to\mathbb{R}$. If the limit

$$\lim_{x\to x_0;x\in E\setminus\{x_0\}}\frac{f(x)-f(x_0)}{x-x_0}.$$

exists and converges to a real number $L\in\mathbb{R}$, then we write $f'(x_0)=L$ and we say that f is **differentiable** at x_0. If this limit does not exist, then we say that f is **not differentiable** at x_0.

Lemma 1.2. *Let E be a subset of $\mathbb{R}$, let $f\colon E\to\mathbb{R}$, let $x_0\in E$, and let $L\in\mathbb{R}$. Then the following two statements are equivalent.*

- *f is differentiable at x_0 and $f'(x_0)=L$.*
- *We have $\lim_{x\to x_0;x\in E\setminus\{x_0\}}\frac{|f(x)-(f(x_0)+L(x-x_0))|}{|x-x_0|}=0$.*

Definition 1.3. Let n be a positive integer. Let $x=(x_1,\ldots,x_n)\in\mathbb{R}^n$. We define the ℓ_2 norm $\|x\|$ of x by

$$\|x\|=\|(x_1,\ldots,x_n)\|:=\left(\sum_{i=1}^n x_i^2\right)^{1/2}.$$

Let $y=(y_1,\ldots,y_n)\in\mathbb{R}^n$. We define the standard inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^n$ by

$$\langle x,y\rangle:=\sum_{i=1}^n x_iy_i.$$

Date: December 22, 2014.

So, $\|x\| = \sqrt{\langle x, x\rangle}$. We also denote the standard basis vectors $e_1, \ldots e_n$ so that

$$e_1 = (1, 0, \ldots, 0), \quad e_2 = (0, 1, 0, \ldots, 0), \quad \cdots \quad e_n = (0, \ldots, 0, 1).$$

Definition 1.4. Let n, m be positive integers. A **linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$** is a function $L\colon \mathbb{R}^n \to \mathbb{R}^m$ which satisfies the following properties.

- For all $x, y \in \mathbb{R}^n$, we have $L(x + y) = L(x) + L(y)$.
- For all $x \in \mathbb{R}^n$ and for all $\alpha \in \mathbb{R}$, we have $L(\alpha x) = \alpha L(x)$.

Remark 1.5. Given a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$, there exists an $m \times n$ matrix A (that is, a matrix A with m rows and n columns) such that

$$L(x) = Ax, \quad \forall\, x \in \mathbb{R}^n.$$

Conversely, given a matrix A, the function $L\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by $L(x) := Ax$ for all $x \in \mathbb{R}^n$, is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. So, on Euclidean spaces, the notions of matrices and linear transformations are interchangeable.

2. Introduction

Our final topic in this course will be differentiation in several variables. Here the theory somewhat resembles the theory of differentiation in one variable, however there are many key differences. The first obstacle we need to overcome is to simply define the derivative in the higher dimensional setting. We therefore begin with this task.

3. Differentiation in multiple variables

Let n, m be positive integers. Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$. In order to define the derivative of f, we cannot simply copy and paste Definition 1.1, since we would need to let $x \in \mathbb{R}^n$ and then divide by x, which is meaningless unless $n = 1$. We instead use the equivalent definition within Lemma 1.2. In this case, we can successfully define differentiation by replacing the absolute values by the appropriate norm, and by replacing L by a linear map.

Definition 3.1 (**Derivatives in multiple variables**)**.** Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, let $x_0 \in E$, and let $L\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We say that f is **differentiable at x_0 with derivative** L if and only if we have

$$\lim_{x \to x_0; x \in E} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0.$$

Example 3.2. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(x_1, x_2) = (x_1^2, x_2^2)$. Define the linear transformation $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ by $L(x_1, x_2) := (2x_1, 4x_2)$. We will show that L is the derivative of f at the point $x_0 = (1, 2)$. We want to show that

$$\lim_{x \to (1,2); x \neq (1,2)} \frac{\|f(x) - (f(1,2) + L(x - (1,2)))\|}{\|x - (1,2)\|} = 0.$$

Now, note that

$$\begin{aligned} f(x) - (f(1,2) + L(x - (1,2))) &= (x_1^2, x_2^2) - ((1,4) + (2x_1, 4x_2) - (2,8)) \\ &= (x_1^2, x_2^2) - (2x_1 - 1, 4x_2 - 4) \\ &= ((x_1 - 1)^2, (x_2 - 2)^2). \end{aligned}$$

So, using the triangle inequality,

$\|f(x)-(f(1,2)+L(x-(1,2)))\| \leq \|((x_1-1)^2,0)\|+\|(0,(x_2-2)^2)\| = (x_1-1)^2+(x_2-2)^2$.

In conclusion,

$$0 \leq \lim_{x\to(1,2);x\neq(1,2)} \frac{(x_1-1)^2+(x_2-2)^2}{\sqrt{(x_1-1)^2+(x_2-2)^2}} = \lim_{x\to(1,2);x\neq(1,2)} \sqrt{(x_1-1)^2+(x_2-2)^2} = 0.$$

So, we have proven our desired statement.

The following lemma shows that a function can have at most one derivative at an interior point of E.

Lemma 3.3. *Let E be a subset of $\mathbb{R}^n$, let $f\colon E\to\mathbb{R}^m$ be a function, and let x_0 be an interior point of E. Let $L_a\colon \mathbb{R}^n\to\mathbb{R}^m$ and let $L_b\colon \mathbb{R}^n\to\mathbb{R}^m$ be linear transformations. Suppose f is differentiable at x_0 with derivative L_a, and f is differentiable at x_0 with derivative L_b. Then $L_a = L_b$.*

Exercise 3.4. Prove Lemma 3.3. (Hint: argue by contradiction. Assume that $L_a \neq L_b$. Then there exists a nonzero vector $v\in\mathbb{R}^n$ such that $L_a v \neq L_b v$. Then, apply the definition of the derivative, and try to specialize to the case where $x = x_0 + tv$ for some scalar t, in order to obtain a contradiction.)

Using Lemma 3.3, we can now talk about the derivative of f at interior points x_0, and we will label this derivative as $f'(x_0)$. That is, if x_0 is an interior point of E, then $f'(x_0)$ is the unique linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ such that

$$\lim_{x\to x_0;x\in E} \frac{\|f(x)-(f(x_0)+f'(x_0)(x-x_0))\|}{\|x-x_0\|} = 0.$$

Informally, we therefore have **Newton's approximation**:

$$f(x) \approx f(x_0)+f'(x_0)(x-x_0).$$

Remark 3.5. We sometimes refer to $f'(x_0)$ as the **total derivative** of f, to distinguish $f'(x_0)$ from the related directional and partial derivatives.

4. Partial and Directional Derivatives

We now relate the total derivative to the partial and directional derivatives. Let n,m be positive integers.

Definition 4.1. Let E be a subset of $\mathbb{R}^n$, let $f\colon E\to\mathbb{R}^m$ be a function, let x_0 be an interior point of E, let $v\in\mathbb{R}^n$, and let t be a real number. If the limit

$$\lim_{t\to0;t\neq0,x_0+tv\in E} \frac{f(x_0+tv)-f(x_0)}{t}.$$

exists, we say that f is **differentiable in the direction** v **at** x_0, and we denote this limit by $D_v f(x_0)$.

$$D_v f(x_0) := \lim_{t\to0;t\neq0,x_0+tv\in E} \frac{f(x_0+tv)-f(x_0)}{t}.$$

Equivalently, we have

$$D_v f(x_0) := \frac{d}{dt} f(x_0+tv)|_{t=0}.$$

Note that in this definition we are dividing by the scalar t, so this division is okay, and $D_v f(x_0) \in \mathbb{R}^m$.

Example 4.2. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(x_1, x_2) = (x_1^2, x_2^2)$. Let $x_0 := (1, 2)$ and let $v := (3, 4)$. We then compute

$$\frac{((1+3t)^2, (2+4t)^2) - (1,4)}{t} = \frac{(1+6t+9t^2, 4+16t+16t^2) - (1,4)}{t} = (6+9t, 16+16t).$$

Therefore,

$$D_v f(x_0) = \lim_{t \to 0; t \neq 0} (6+9t, 16+16t) = (6, 16).$$

If v is a standard basis vector, then we write $\frac{\partial f}{\partial x_j}(x_0)$ or $\frac{\partial}{\partial x_j} f(x_0)$ for $D_{e_j} f(x_0)$. We refer to $\frac{\partial f}{\partial x_j}(x_0)$ as the **partial derivative of f with respect to** x_j. So,

$$\frac{\partial f}{\partial x_j}(x_0) := \lim_{t \to 0; t \neq 0, x_0 + te_j \in E} \frac{f(x_0 + te_j) - f(x_0)}{t} = \frac{d}{dt} f(x_0 + te_j)|_{t=0}.$$

Note that if $f\colon E \to \mathbb{R}^m$, then $\frac{\partial f}{\partial x_j} \in \mathbb{R}^m$. And if we write f in its components as $f = (f_1, \ldots, f_m)$, then

$$\frac{\partial f}{\partial x_j}(x_0) = \left(\frac{\partial f_1}{\partial x_j}(x_0), \ldots, \frac{\partial f_1}{\partial x_j}(x_0) \right).$$

The total derivative and directional derivative are related in the following way.

Lemma 4.3. *Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, let x_0 be an interior point of E, and let $v \in \mathbb{R}^n$. If f is differentiable at x_0, then f is also differentiable in the direction v at x_0, and*

$$D_v f(x_0) = f'(x_0) v.$$

Exercise 4.4. Prove Lemma 4.3.

From Lemma 4.3, total differentiability implies directional differentiability. Unfortunately, the converse is false.

Exercise 4.5. Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) := x^3/(x^2 + y^2)$ when $(x, y) \neq (0, 0)$, and $f(0, 0) := 0$. Show that for any $v \in \mathbb{R}^2$, f is differentiable at $(0, 0)$ in the direction v. However, show that f is not differentiable at $(0, 0)$.

Remark 4.6. From Lemma 4.3, if $E \subseteq \mathbb{R}^n$ and if $f\colon E \to \mathbb{R}^m$ is differentiable at $x_0 \in E$, then all partial derivatives $\frac{\partial f}{\partial x_j}$ exist at x_0, for all $j \in \{1, \ldots, n\}$, and

$$\frac{\partial f}{\partial x_j} = f'(x_0) e_j, \quad \forall j \in \{1, \ldots, n\}.$$

Also, given $v = (v_1, \ldots, v_n) = \sum_{j=1}^n v_j e_j \in \mathbb{R}^n$, we have

$$D_v f(x_0) = f'(x_0) \sum_{j=1}^n v_j e_j = \sum_{j=1}^n v_j f'(x_0) e_j = \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(x_0). \qquad (*)$$

From Exercise 4.5, partial differentiability does not imply differentiability. However, if the partial derivatives of a function are continuous, then partial differentiability does imply differentiability. We will use equation $(*)$ to prove this assertion.

Theorem 4.7. *Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, let F be a subset of E, and let x_0 be an interior point of F. If the partial derivatives $\frac{\partial f}{\partial x_j}$ exist on F and are continuous at x_0 for all $j \in \{1, \ldots, n\}$, then f is differentiable at x_0. Moreover, $f'(x_0)\colon \mathbb{R}^n \to \mathbb{R}^m$ is defined by*

$$f'(x_0)(v_1, \ldots, v_n) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

Proof. Define a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ by

$$L(v_1, \ldots, v_n) := \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

We need to show that

$$\lim_{x \to x_0; x \in E \setminus \{x_0\}} \frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} = 0.$$

Let $\varepsilon > 0$. We will find $\delta > 0$ such that, if x satisfies $0 < \|x - x_0\| < \delta$, then

$$\frac{\|f(x) - (f(x_0) + L(x - x_0))\|}{\|x - x_0\|} < \varepsilon.$$

That is, we will show, if x satisfies $0 < \|x - x_0\| < \delta$, then

$$\|f(x) - (f(x_0) + L(x - x_0))\| < \varepsilon \|x - x_0\|.$$

Since x_0 is an interior point of F, there exists $r > 0$ such that $B(x_0, r) \subseteq F$. Since the partial derivative $\frac{\partial f}{\partial x_j}$ is continuous on F for each $j \in \{1, \ldots, n\}$, there exists $0 < \delta_j < r$ such that $\|\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0)\| < \varepsilon/(nm)$, for every $x \in B(x_0, \delta_j)$, for every $j \in \{1, \ldots, n\}$. Define $\delta := \min_{j=1,\ldots,n} \delta_j$. Then $\|\frac{\partial f}{\partial x_j}(x) - \frac{\partial f}{\partial x_j}(x_0)\| < \varepsilon/(nm)$, for every $x \in B(x_0, \delta)$, for every $j \in \{1, \ldots, n\}$.

Let $x \in B(x_0, \delta)$, and write $x = x_0 + v_1 e_1 + \cdots + v_n e_n$ for some scalars $v_1, \ldots, v_n$. Note that

$$\|x - x_0\| = \sqrt{v_1^2 + \cdots + v_n^2}.$$

In particular, we have $|v_j| \leq \|x - x_0\|$ for all $j \in \{1, \ldots, n\}$. Recall that we need to show

$$\|f(x_0 + v_1 e_1 + \cdots + v_n e_n) - f(x_0) - \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0)\| < \varepsilon \|x - x_0\|.$$

Write f in its components as $f = (f_1, \ldots, f_m)$, so that $f_i\colon E \to \mathbb{R}$ for all $i \in \{1, \ldots, m\}$. Applying the Mean Value Theorem in the first variable, there exists a real number t_i between 0 and v_1 such that

$$f_i(x_0 + v_1 e_1) - f_i(x_0) = \frac{\partial f_i}{\partial x_1}(x_0 + t_i e_1) v_1.$$

Note that, for all $i \in \{1, \ldots, m\}$, for all $j \in \{1, \ldots, n\}$, we have

$$|\frac{\partial f_i}{\partial x_1}(x_0 + t_i e_1) - \frac{\partial f_i}{\partial x_1}(x_0)| \leq \|\frac{\partial f}{\partial x_1}(x_0 + t_i e_1) - \frac{\partial f}{\partial x_1}(x_0)\| \leq \varepsilon/(nm).$$

Therefore,

$$|f_i(x_0 + v_1e_1) - f_i(x_0) - \frac{\partial f_i}{\partial x_1}(x_0)v_1| \le \varepsilon |v_1| /(nm).$$

Summing this inequality over $i \in \{1, \dots, m\}$ and using $\|(y_1, \dots, y_m)\| \le |y_1| + \cdots + |y_m|$, we have

$$\|f(x_0 + v_1e_1) - f(x_0) - \frac{\partial f}{\partial x_1}(x_0)v_1\| \le \varepsilon |v_1| /n \le \varepsilon \|x - x_0\| /n.$$

In the last inequality, we used $|v_1| \le \|x - x_0\|$.

Using a similar argument, we conclude that

$$\|f(x_0 + v_1e_1 + v_2e_2) - f(x_0 + v_1e_1) - \frac{\partial f}{\partial x_2}(x_0)v_2\| \le \varepsilon \|x - x_0\| /n.$$

And so on, until we get

$$\|f(x_0 + v_1e_1 + \cdots + v_ne_n) - f(x_0 + v_1e_1 + \cdots + v_{n-1}e_{n-1}) - \frac{\partial f}{\partial x_n}(x_0)v_n\| \le \varepsilon \|x - x_0\| /n.$$

Summing these n inequalities and using the triangle inequality $\|x + y\| \le \|x\| + \|y\|$, we get a telescoping sum which finally gives

$$\|f(x_0 + v_1e_1 + \cdots + v_ne_n) - f(x_0) - \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0)\| < \varepsilon \|x - x_0\| .$$

□

From Theorem 4.7 and Lemma 4.3, if the partial derivatives of a function $f \colon E \to \mathbb{R}^m$ exist and are continuous on a set F, then all directional derivatives of f exist at every interior point x_0 of F, and

$$D_{(v_1,\dots,v_n)} f(x_0) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

In particular, if $f \colon E \to \mathbb{R}$ is a real-valued function, and if we define the **gradient** $\nabla f(x_0)$ of f at x_0 to be the n-dimensional row vector

$$\nabla f(x_0) := (\frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_n}(x_0)),$$

then we have the formula

$$D_v f(x_0) = \langle \nabla f(x_0), v \rangle.$$

More generally, if $f \colon E \to \mathbb{R}^m$ is a function with $f = (f_1, \dots, f_m)$, and x_0 is in the interior of the region where the partial derivatives of f exist and are continuous, then Theorem 4.7 says

$$f'(x_0)(v_1, \dots, v_n) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0) = \left(\sum_{j=1}^{n} v_j \frac{\partial f_i}{\partial x_j}(x_0) \right)_{i=1}^{m}.$$

So, if we define the matrix

$$Df(x_0) = \left(\frac{\partial f_i}{\partial x_j}(x_0) \right)_{\substack{1 \le i \le m \\ 1 \le j \le n}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \frac{\partial f_2}{\partial x_1}(x_0) & \frac{\partial f_2}{\partial x_2}(x_0) & \cdots & \frac{\partial f_2}{\partial x_n}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix},$$

then we have

$$D_v f(x_0) = f'(x_0)v = Df(x_0)v.$$

The matrix $Df(x_0)$ is sometimes called the **derivative** or the **differential** of f at x_0. We still wish to distinguish the matrix $Df(x_0)$ from the linear transformation $f'(x_0)$, since the latter is defined in a way which does not depend on the chosen basis of Euclidean space.

5. The Chain Rule in Several Variables

Let n, m, p be positive integers. Recall that if $f\colon X \to Y$ and $g\colon Y \to Z$ are functions, then the composition $g \circ f\colon X \to Z$ is defined by $g \circ f(x) := g(f(x))$, for all $x \in X$.

Theorem 5.1 (**The Chain Rule in Multiple Variables**). *Let E be a subset of $\mathbb{R}^n$, let F be a subset of $\mathbb{R}^m$, let $f\colon E \to F$ be a function, and let $g\colon F \to \mathbb{R}^p$. Let x_0 be a point in the interior of E. Assume that f is differentiable at x_0 and that $f(x_0)$ is in the interior of F. Assume also that g is differentiable at $f(x_0)$. Then $g \circ f\colon E \to \mathbb{R}^p$ is also differentiable at x_0, and*

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

Remark 5.2. We can intuitively think of the chain rule as follows. From Newton's approximation, we have

$$f(x) - f(x_0) \approx f'(x_0)(x - x_0).$$

Also, using Newton's approximation again,

$$g(f(x)) - g(f(x_0)) \approx g'(f(x_0))(f(x) - f(x_0)).$$

So, combining these two approximations, we have

$$g(f(x)) - g(f(x_0)) \approx g'(f(x_0))f'(x_0)(x - x_0).$$

That is, $(g \circ f)'(x_0) = g'(f(x_0))f'(x_0)$. The rigorous version of this proof irons out the details inherent in Newton's approximation.

Exercise 5.3.

- Let $L\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that there exists a real number $M > 0$ such that $\|Lx\| \leq M\|x\|$, for all $x \in \mathbb{R}^n$. (Hint: first, using Remark 1.5, write L in terms of a matrix A. Then, set M to be equal to the sum of the absolute values of the entries of A. Use the triangle inequality a lot. There are many different ways to do this exercise, some of which use a different value of M. For example, you could try using the Cauchy-Schwarz inequality.) In particular, conclude that any linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is continuous.
- Let E be a subset of $\mathbb{R}^n$. Assume that $f\colon E \to \mathbb{R}^m$ is differentiable at an interior point x_0 of E. Then f is also continuous at x_0.
- Prove Theorem 5.1. (Hint: it may be helpful to review the proof of the single variable chain rule. It is probably easiest to use the sequence definition of a limit.)

Example 5.4. Suppose $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a differentiable function, and $x_j\colon \mathbb{R} \to \mathbb{R}$ are differentiable functions for all $j \in \{1, \ldots, n\}$. Then

$$\frac{d}{dt} f(x_1(t), \ldots, x_n(t)) = \sum_{j=1}^{n} x_j'(t) \frac{\partial f}{\partial x_j}(x_1(t), \ldots, x_n(t)).$$

This follows from the chain rule.

6. Iterated Derivatives and Clairaut's Theorem

We now investigate what happens when we differentiate a function twice, in two different directions.

Definition 6.1. Let E be a subset of $\mathbb{R}^n$, and let $f\colon E \to \mathbb{R}^m$ be a function. We say that f is **continuously differentiable** if and only if the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ exist and are continuous on E. We say that f is **twice continuously differentiable** if and only if it is continuously differentiable, and the partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ are themselves continuously differentiable.

Continuously differentiable functions are sometimes called C^1 functions. Twice continuously differentiable functions are sometimes called C^2 functions. One can also define C^3 functions, C^4 functions, etc., but we will not do so here.

Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. As you may have learned, it is often true that $\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2} f = \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f$. Unfortunately, this equality does not always hold.

Exercise 6.2. Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) := (x^3 y)/(x^2+y^2)$ when $(x,y) \neq (0,0)$, and $f(0,0) := 0$. Show that f is continuously differentiable, and the double derivatives $\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2} f$ and $\frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1}$ exist, but these derivatives are not equal at $(0,0)$.

Thankfully, if f is twice continuously differentiable, then the order of differentiation does not matter.

Theorem 6.3 (**Clairaut's Theorem**). *Let E be an open subset of $\mathbb{R}^n$, and let $f\colon E \to \mathbb{R}^m$ be a twice continuously differentiable function. Then, for all $1 \leq i, j \leq n$ and for all interior points x_0 of E, we have $\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f(x_0) = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f(x_0)$.*

Proof. The claim is certainly true for $i = j$ so assume that $i \neq j$. By replacing f by $f(x - x_0)$ as necessary, we may assume that $x_0 = 0$.

Define $a := \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f(x_0)$ and define $a' := \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f(x_0)$. We need to show that $a = a'$.

Let $\varepsilon > 0$. Since f is twice continuously differentiable, there exists $\delta > 0$ such that, for all x with $\|x\| < 2\delta$, we have

$$\left|\frac{\partial}{x_i}\frac{\partial}{x_j} f(x) - a\right| < \varepsilon, \qquad \left|\frac{\partial}{x_j}\frac{\partial}{x_i} f(x) - a'\right| < \varepsilon.$$

Define

$$M := f(\delta e_i + \delta e_j) - f(\delta e_i) - f(\delta e_j) + f(0).$$

Applying the Fundamental Theorem of Calculus to the e_i variable, we have

$$f(\delta e_i + \delta e_j) - f(\delta e_j) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) dx_i.$$

And

$$f(\delta e_i) - f(0) = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i) dx_i.$$

Therefore,

$$M = \int_0^\delta \frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) dx_i$$

For each $x_i \in (0, \delta)$, there exists $x_j \in [0, \delta]$ such that, by the Mean Value Theorem, we have

$$\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) = \delta \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x_i e_i + x_j e_j).$$

By our choice of δ (noting that $\|x_i e_i + x_j e_j\| < 2\delta$), we therefore have

$$\left|\frac{\partial f}{\partial x_i}(x_i e_i + \delta e_j) - \frac{\partial f}{\partial x_i}(x_i e_i) - \delta a'\right| < \varepsilon\delta.$$

So, integrating this inequality over $x_i \in [0, \delta]$, we get

$$\left|M - \delta^2 a'\right| < \varepsilon\delta^2.$$

We can run this same argument with the roles of i and j reversed (noting that M is symmetric in i, j) to get

$$\left|M - \delta^2 a\right| < \varepsilon\delta^2.$$

So, from the triangle inequality, we conclude that

$$|a - a'| < 2\varepsilon.$$

Since this inequality holds for all $\varepsilon > 0$, we conclude that $a = a'$, as desired. ☐

7. Appendix: Notation

Let A, B be sets in a space X. Let m, n be a nonnegative integers.

$$
\begin{aligned}
\mathbb{Z} &:= \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}, \text{ the integers} \\
\mathbb{N} &:= \{0, 1, 2, 3, 4, 5, \ldots\}, \text{ the natural numbers} \\
\mathbb{Z}_+ &:= \{1, 2, 3, 4, \ldots\}, \text{ the positive integers} \\
\mathbb{Q} &:= \{m/n \colon m, n \in \mathbb{Z}, n \neq 0\}, \text{ the rationals} \\
\mathbb{R} &\text{ denotes the set of real numbers} \\
\mathbb{R}^* &= \mathbb{R} \cup \{-\infty\} \cup \{+\infty\} \text{ denotes the set of extended real numbers} \\
\mathbb{C} &:= \{x + y\sqrt{-1} : x, y \in \mathbb{R}\}, \text{ the complex numbers} \\
\emptyset &\text{ denotes the empty set, the set consisting of zero elements} \\
\in &\text{ means “is an element of.” For example, } 2 \in \mathbb{Z} \text{ is read as “2 is an element of } \mathbb{Z}.\text{”} \\
\forall &\text{ means “for all”} \\
\exists &\text{ means “there exists”} \\
\mathbb{R}^n &:= \{(x_1, \ldots, x_n) \colon x_i \in \mathbb{R}, \, \forall \, i \in \{1, \ldots, n\}\} \\
A \subseteq B &\text{ means } \forall \, a \in A, \text{ we have } a \in B, \text{ so } A \text{ is contained in } B \\
A \smallsetminus B &:= \{x \in A \colon x \notin B\} \\
A^c &:= X \smallsetminus A, \text{ the complement of } A \\
A \cap B &\text{ denotes the intersection of } A \text{ and } B \\
A \cup B &\text{ denotes the union of } A \text{ and } B
\end{aligned}
$$

Let (X, d) be a metric space, let $x_0 \in X$, let $r > 0$ be a real number, and let E be a subset of X. Let $(x_1, \ldots, x_n)$ be an element of $\mathbb{R}^n$, and let $p \geq 1$ be a real number.

$$
\begin{aligned}
B_{(X,d)}(x_0, r) = B(x_0, r) &:= \{x \in X \colon d(x, x_0) < r\}. \\
\overline{E} &\text{ denotes the closure of } E \\
\mathrm{int}(E) &\text{ denotes the interior of } E \\
\partial E &\text{ denotes the boundary of } E
\end{aligned}
$$

$$
\begin{aligned}
\|(x_1, \ldots, x_n)\|_{\ell_p} &:= (\sum_{i=1}^{n} |x_i|^p)^{1/p} \\
\|(x_1, \ldots, x_n)\|_{\ell_\infty} &:= \max_{i=1,\ldots,n} |x_i|
\end{aligned}
$$

Let $f, g\colon (X, d_X) \to (Y, d_Y)$ be maps between metric spaces. Let $V \subseteq X$, and let $W \subseteq Y$.

$$f(V) := \{f(v) \in Y\colon v \in V\}.$$
$$f^{-1}(W) := \{x \in X\colon f(x) \in W\}.$$
$$d_\infty(f, g) := \sup_{x \in X} d_Y(f(x), g(x)).$$

$B(X;Y)$ denotes the set of functions $f\colon X \to Y$ that are bounded.

$$C(X;Y) := \{f \in B(X;Y)\colon f \text{ is continuous}\}.$$

Let $f, g\colon \mathbb{R} \to \mathbb{C}$ be $\mathbb{Z}$-periodic functions.

$$\|f\|_\infty := \sup_{x \in [0,1]} |f(x)|.$$
$$\langle f, g\rangle := (\int_0^1 f(x)\overline{g(x)}dx)^{1/2}.$$
$$\|f\|_2 := \sqrt{\langle f, f\rangle} = (\int_0^1 |f(x)|^2\, dx)^{1/2}$$
$$d_{L_2}(f, g) := \|f - g\|_2 = (\int_0^1 |f(x) - g(x)|^2\, dx)^{1/2}.$$

Let n, m be positive integers, let $(e_1, \ldots, e_n)$ denote the standard basis of $\mathbb{R}^n$, let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}^m$ be a function, let $x_0 \in E$ be an interior point of E, let $v \in \mathbb{R}^n$, and let $j \in \{1, \ldots, n\}$.

$f'(x_0)$ denotes the total derivative of f.

$D_v f(x_0)$ denotes the derivative of f in the direction v.

$$\frac{\partial f}{\partial x_j}(x_0) = \frac{\partial}{\partial x_j} f(x_0) = D_{e_j} f(x_0).$$

Let E be a subset of $\mathbb{R}^n$, let $f\colon E \to \mathbb{R}$ be a function, and let x_0 be an interior point of E.

$$\nabla f(x_0) = (\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0)).$$

7.1. **Set Theory.** Let X, Y be sets, and let $f\colon X \to Y$ be a function. The function $f\colon X \to Y$ is said to be **injective** (or **one-to-one**) if and only if: for every $x, x' \in V$, if $f(x) = f(x')$, then $x = x'$.

The function $f\colon X \to Y$ is said to be **surjective** (or **onto**) if and only if: for every $y \in Y$, there exists $x \in X$ such that $f(x) = y$.

The function $f\colon X \to Y$ is said to be **bijective** (or a **one-to-one correspondence**) if and only if: for every $y \in Y$, there exists exactly one $x \in X$ such that $f(x) = y$. A function $f\colon X \to Y$ is bijective if and only if it is both injective and surjective.

Two sets X, Y are said to have the same **cardinality** if and only if there exists a bijection from X onto Y.

UCLA Department of Mathematics, Los Angeles, CA 90095-1555
E-mail address: heilman@math.ucla.edu