Notes on multivariable calculus

Robert

June 2024

Contents

List of Theorems

List of Definitions

Chapter 1

Some prerequisite material

We will quickly remark that the prerequisite chapter is not fully complete.

1.1 Notation

We shall try to be as clear about the type of our variables as much as possible. In particular, vectors will be in bold: $\mathbf{v} \in \mathbb{R}^n$.

1.2 Linear algebra

We shall assume that the reader knows a little bit of linear algebra. Here, we will not bold any of our vectors, since they are all vectors.

Definition 1.2.1 (Inner Product). Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. Then an *inner product on V* is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ such that

1. (**Linear in the 2nd argument**) $\langle u, \lambda v + w \rangle = \lambda \langle u, v \rangle + \langle u, w \rangle$,
2. (**Conjugate symmetric**) $\langle u, v \rangle = \overline{\langle v, u \rangle}$,
3. (**Positive definiteness**) $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0$ if and only if $v = 0$.

A few quick remarks are in order. The third property of an inner product tells us that $\langle v, v \rangle$ must be a real number, so comparison with 0 is a legal move. In addition, if V is a real vector space, conjugate symmetry is simply symmetry, i.e. $\langle u, v \rangle = \langle v, u \rangle$ since $\overline{r} = r$ whenever r is a real number. Thus, any inner product on a real vector space will also be linear in the first argument. In either case, linearity of the second argument allows us to define linear transformations like $\varphi_v(w) = \langle v, w \rangle$ where $v \in V$ is some fixed vector. We shall see that this is used later on in Theorem 1.2.4, and even in the discussion after Equation (2.2).

We shall really only be working with one kind of inner product, the *usual inner product on* $\mathbb{R}^n$. This is defined to be

$$\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \tag{1.1}$$

Given an inner product, we can define a *norm*. Again, we are mostly only interested in the *Euclidean norm on* $\mathbb{R}^n$, which is defined by

$$\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{\sum_{i=1}^{n} v_i v_i}. \tag{1.2}$$

Here, $\langle v, v \rangle$ is taking the usual inner product of v with itself.

This is a useful proposition which is used in a lot of bounding arguments. See [Spi18, Problem 1-1 on p. 4] for the problem.

Proposition 1.2.2. *Let $\|\cdot\|$ be the usual Euclidean norm. Then for any $x \in \mathbb{R}^n$ we have*

$$\|x\| \leq \sum_{i=1}^{n} |x_i| .$$

This proposition will come in handy quite often when proving results later on, so we state it here. See [Spi18, Problem 1-10 on p. 5] for the original exercise.

Proposition 1.2.3. *Suppose $T : \mathbb{R}^m \to \mathbb{R}^n$ is linear. Then, there exists $M \in \mathbb{R}$ such that $\|T(h)\| \leq M \|h\|$ for all $h \in \mathbb{R}^m$.*

Proof. Let us consider the matrix of T,

$$\mathcal{M}(T) = \begin{bmatrix} A_{1,1} & \dots & A_{1,m} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \dots & A_{n,m} \end{bmatrix},$$

where each column of coefficients is the scalars that make the vector that $T(e_i)$ is sent to a linear combination of the standard basis of $\mathbb{R}^n$. □

We may also use the Riesz Representation Theorem. For now, the weaker version will suffice. You can see [Spi18, Problem 1-12 on p. 5] for the original problem.

Theorem 1.2.4 (Weak Riesz Representation Theorem)**.** Let V be a real inner product space of dimension n and let V^* denote the dual space of V. This is the set of linear functionals $\varphi : V \to \mathbb{R}$. Now, for each $v \in V$, define $\varphi_v(w) = \langle v, w \rangle$, and define $F : V \to V^*$ by $F(v) = \varphi_v$. Then F is an isomorphism.

Proof. Since all vector spaces in question are finite dimensional, we shall simply show that F is injective. Let us suppose that $F(v) = F(v')$. Then this means that the functions φ_v and $\varphi_{v'}$ agree on all $w \in V$.

□

1.3 Topology

We shall only briefly touch on topology. For now, the definitions will be introduced in a less general manner. Readers who do know topology can gloss over this section.

Definition 1.3.1 (Ball)**.** Let $\mathbf{x} \in \mathbb{R}^n$ and $\varepsilon > 0$. Then a ***ball** around $\mathbf{x}$ of radius ε*[a] is the set

$$B(\mathbf{x}, \varepsilon) := \{ \mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \|x\|\| < \varepsilon \} . \tag{1.3}$$

[a] Sometimes you will see this being called an *open ball*.

When $n = 1$, balls are simply open intervals. When $n = 2$ they can be seen as circles and when $n = 3$ they can be seen as spheres.

Definition 1.3.2 (Open set)**.** Let $U \subseteq \mathbb{R}^n$. Then the set U is called ***open*** if for every point $\mathbf{x} \in U$, there exists some $\varepsilon > 0$ such that $B(\mathbf{x}, \varepsilon) \subseteq U$.

It is immediate from the definition that every ball is an open set. Additionally, with a little set theory one can prove that every open set is the union of balls.

Definition 1.3.3 (Neighborhoods). Let $\mathbf{x} \in \mathbb{R}^n$. A ***neighborhood** of* $\mathbf{x}$ is an open set $U \subseteq \mathbb{R}^n$ such that $\mathbf{x} \in U$.

Definition 1.3.4 (Open covers). Let $S \subseteq \mathbb{R}^n$. An *open cover of S* is a collection of open sets $\mathcal{O} = \{ O_\alpha : \alpha \in \Lambda \}$ (here Λ is some indexing set, so we can refer to each element in $\mathcal{O}$ by writing O_α) such that the union $\bigcup \mathcal{O} = \bigcup_{\alpha \in \Lambda} O_\alpha$ contains S, i.e.

$$\bigcup_{\alpha \in \Lambda} O_\alpha \supseteq S.$$

If there are finitely many indices $\alpha_1, \ldots, \alpha_n$ such that $\bigcup_{i=1}^n O_{\alpha_i} \supseteq S$, then $\mathcal{O}$ is said to have a *finite subcover*.

This definition is a little hard to get your head around so let us see a few examples.
Example 1.3.5. We can open cover $\mathbb{R}$ with open intervals of the form $(n, n+1)$, for each $n \in \mathbb{Z}$. //

Definition 1.3.6 (Compactness - General definition). Let $S \subseteq \mathbb{R}^n$. Then S is said to be *compact* if given *any* collection of open sets $\mathcal{O} = \{ O_\alpha : \alpha \in \Lambda \}$, there is a finite subcover of S.

Definition 1.3.7 (Compactness - $\mathbb{R}^n$ definition). Let $S \subseteq \mathbb{R}^n$. Then S is said to be *compact* if it is closed and bounded.

Note that the equivalence of those two definitions in $\mathbb{R}^n$ is justified by the following theorem.

Theorem 1.3.8 (Heine-Borel). Let $S \subseteq \mathbb{R}^n$. Then S is compact (in the sense of Definition 1.3.6) if and only if it is closed and bounded.

Proof. Much shorter with topology, and thus omitted. □

You are free to use any definition of compactness that is convenient. The following property of compactness is extremely useful.
Proposition 1.3.9. *Let $S \subseteq \mathbb{R}^n$ be compact. Let $f : S \to \mathbb{R}^n$ be continuous. Then, the image of S under f, $f[S]$, is also compact.*

And as a corollary, we get for free, the
Corollary 1.3.10 (Extreme Value Theorem)**.** *Let $S \subseteq \mathbb{R}^n$ be compact and $f : S \to \mathbb{R}$ be continuous. Then, f attains a maximum and minimum, i.e. there exists $m, M \in \mathbb{R}$ such that for all $s \in S$, $m \leq f(s) \leq M$.*

Chapter 2

Differentiation

2.1 Derivatives

We immediately begin with the most general definition of the derivative. The derivative, in essence, is trying to capture the idea of a linear approximation to a function at a point.

Definition 2.1.1 (Derivative). Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{a} \in \mathbb{R}^n$ Then f is said to be ***differentiable** at* $\mathbf{a}$ if there exists a linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})\|}{\|\mathbf{h}\|} = 0. \tag{2.1}$$

The linear transformation is often denoted $Df(\mathbf{a})$.

Warning: This notation can get a little confusing! If we do $Df(\mathbf{a})(\mathbf{v})$, we are saying evaluate the linear transformation $Df(\mathbf{a})$ at $\mathbf{v}$. For less confusion, the reader can insert brackets like $(Df(\mathbf{a}))(\mathbf{v})$. However, this notation is worth the initial confusion because we will soon see that Df itself can be a function.

When we refer to the matrix of the derivative of f at $\mathbf{a}$, i.e. $\mathcal{M}(Df(\mathbf{a}))$, it is called the *Jacobian of f at* $\mathbf{a}$. This is denoted $f'(\mathbf{a})$ in [Spi18].

Of course, the codomain of f need not be all of $\mathbb{R}^n$, just some open subset of it. When we have $f : \mathbb{R}^n \to \mathbb{R}$, we define

Definition 2.1.2 (Gradient). Let $f : \mathbb{R}^n \to \mathbb{R}$. Then the ***gradient** of f at* $\mathbf{a}$, denoted $\nabla f(\mathbf{a})$ is a vector in $\mathbb{R}^n$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{h}\|}{\|\mathbf{h}\|} = 0. \tag{2.2}$$

It is not too hard to see that the gradient is really just a special case of the derivative. In this case, the linear transformation λ is actually given by $\lambda(\mathbf{h}) = \nabla f(\mathbf{a}) \cdot \mathbf{h}$. We can appeal to Theorem 1.2.4.

Of course, the derivative is unique.

Theorem 2.1.3 (Uniqueness of derivative). Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable. Then, there is a unique linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ that makes Equation (2.1) hold.

We shall refer the reader to [Spi18] for a proof for now, until I update these notes again.

2.2 Consequences of differentiability

Recall that in single variable calculus, if $f : \mathbb{R} \to \mathbb{R}$ is differentiable at c, then it is also continuous at c. The same is true in in multivariable calculus

Proposition 2.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$. Then f is continuous at $\mathbf{a}$.*

Proof. Suppose f is differentiable at $\mathbf{a}$. Let $\varepsilon > 0$, and we would like to find a $\delta > 0$ such that if $0 < \|\mathbf{x} - \mathbf{a}\| < \delta$, we have $\|f(\mathbf{x}) - f(\mathbf{a})\| < \varepsilon$. Since f is differentiable at $\mathbf{a}$, there exists a linear transformation λ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

By definition of the limit, we have some $\delta > 0$ such that when $0 < \|\mathbf{h}\| < \delta$,

$$\frac{\|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})\|}{\|\mathbf{h}\|} < \varepsilon.$$ [1]

Now, if we take $\mathbf{x}$ such that $0 < \|\mathbf{x} - \mathbf{a}\| < \delta$, we will have

$$\frac{\|f(\mathbf{a}+(\mathbf{x}-\mathbf{a})) - f(\mathbf{a}) - \lambda(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} < \varepsilon.$$

□

2.3 The chain rule

Theorem 2.3.1 (Chain rule). Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ be differentiable at $f(\mathbf{a})$. Then $g \circ f : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $\mathbf{a}$ and we have

$$D(g \circ f)(\mathbf{a}) = Dg(f(\mathbf{a})) \circ Df(\mathbf{a}). \tag{2.3}$$

Proof. **To be done** □

In [Spi18, Theorem 2-2 in p. 19], Equation (2.3) can seem a little confusing. Let us try to explain what is going on here. First of all, we use composition because $Dg(f(\mathbf{a}))$ and $Df(\mathbf{a})$ are linear transformations. Next, what is the equation saying in English? Well, what we are really saying here is that the derivative of the composition $g \circ f$ at $\mathbf{a}$ is nothing but the derivative of g at the point $f(\mathbf{a})$ composed with the derivative of f at $\mathbf{a}$. To make this less confusing, let $\mu = Dg(f(\mathbf{a}))$ and $\lambda = Df(\mathbf{a})$. Then, if we evaluate $D(g \circ f)(\mathbf{a})(\mathbf{v})$, we get $\mu(\lambda(\mathbf{v}))$.

Some other useful propositions regarding differentiability are listed below. Note that if we do not quantify over $\mathbf{a}$, it is any vector in $\mathbb{R}^n$.

Theorem 2.3.2 (Useful properties of differentiation).

1. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is a constant function (that is, there is some $\mathbf{y} \in \mathbb{R}^m$ such that $f(\mathbf{x}) = \mathbf{y}$ for all $x \in \mathbb{R}^n$). Then,

$$Df(\mathbf{a}) = 0$$

(Note here that 0 refers to the linear transformation which is always 0.)

2. Suppose $f : \mathbb{R}^m \to \mathbb{R}^m$ is a linear transformation. Then,

$$Df(\mathbf{a}) = f.$$

This is to say, the best linear approximation of a linear transformation is itself (unsurprisingly).

3. Let $f : \mathbb{R}^m \to \mathbb{R}^m$. Then f is differentiable at $\mathbf{a} \in \mathbb{R}^n$ if and only if each component function of f, $f_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, m)$ is differentiable at $\mathbf{a}$. Additionally, we have

$$Df(\mathbf{a}) = (Df_1(\mathbf{a}), \ldots, Df_m(\mathbf{a})).$$

[1] We remark that no absolute value signs are needed since the norm is always positive.

4. If $s : \mathbb{R}^2 \to \mathbb{R}$ is defined by $s(x, y) = x + y$ then $Ds((a, b)) = s$
5. If $p : \mathbb{R}^2 \to \mathbb{R}$ is defined by $p(x, y) = xy$ then

$$Dp((a, b))(x, y) = bx + ay.$$

This means that $\nabla p((a, b)) = (b, a)$.

Proof. (1) and (2) follow quickly from the definition of the derivative.

(3) First suppose that f is differentiable at $\mathbf{a} \in \mathbb{R}^n$. Recall that the projection functions π_i are linear and thus are differentiable everywhere. By (2) we have $D\pi_i(f(\mathbf{a})) = \pi_i$. By Theorem 2.3.1 we have $f_i = \pi_i \circ f$ being differentiable at $f(\mathbf{a})$. Now let us suppose that each component function of f is differentiable at $\mathbf{a}$. Let $\lambda = (Df_1(\mathbf{a}), \ldots, Df_n(\mathbf{a}))$. To make this easier to think about, λ has a matrix given by

$$\mathcal{M}(\lambda) = \begin{bmatrix} \nabla f_1(\mathbf{a}) \\ \nabla f_2(\mathbf{a}) \\ \vdots \\ \nabla f_n(\mathbf{a}) \end{bmatrix},$$

where each row of $\mathcal{M}(\lambda)$ is given by the vector $\nabla f_1(\mathbf{a})$. Now, we can write

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h}) = (f_1(\mathbf{a} + \mathbf{h}) - f_i(\mathbf{a}) - Df_1(\mathbf{a})(\mathbf{h}), \ldots, f_n(\mathbf{a} + \mathbf{h}) - f_n(\mathbf{a}) - Df_n(\mathbf{a})(\mathbf{h})).$$

Now, let's take a look at $\|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})\|$. By Proposition 1.2.2 we have

$$\|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})\| \leq \sum_{i=1}^{n} |f_i(\mathbf{a} + h) - f_i(\mathbf{a}) - Df_i(\mathbf{a})(\mathbf{h})|.$$

This implies that

$$\frac{\|f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \lambda(\mathbf{h})\|}{\|\mathbf{h}\|} \leq \sum_{i=1}^{n} \frac{|f_i(\mathbf{a} + \mathbf{h}) - f_i(\mathbf{a}) - Df_i(\mathbf{a})(\mathbf{h})|}{\|\mathbf{h}\|}.$$

Taking the limit $\mathbf{h} \to \mathbf{0}$ on both sides, the right side is 0 and thus the left side must be 0 too.

(4) follows from (2) as s is easily seen to be linear.

(5) **To be done**. See [Spi18, Theorem 2-3 on p. 21]. □

Corollary 2.3.3. *Suppose $f, g : \mathbb{R}^n \to \mathbb{R}$ are differentiable at $\mathbf{a}$. Then,*

$$D(f + g)(\mathbf{a}) = Df(\mathbf{a}) + Dg(\mathbf{a}),$$
$$D(f \cdot g)(\mathbf{a}) = g(\mathbf{a})Df(\mathbf{a}) + f(\mathbf{a})Dg(\mathbf{a}).$$

If $g(\mathbf{a}) \neq 0$ then we also have

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

Proof. **To be done** □

Note here $f + g$ is a function defined by $(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$. The sum on the right side is actually a sum of vectors. Additionally, $f \cdot g$ is a function defined by $(f \cdot g)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$. Keep in mind that $f(\mathbf{a}), g(\mathbf{a})$ are scalars.

A very useful thing to note is that the determinant function is actually differentiable. The following propositions from [Spi18, Problem 2-14, 2-15 on pp. 23, 24] can be used to deduce this fact.

Proposition 2.3.4. *Let E_i, $i = 1, \ldots, k$ be Euclidean spaces of various dimensions. That is to say, $E_i = \mathbb{R}^{n_i}$ for some n_i. A function $f : E_1 \times \cdots \times E_k \to \mathbb{R}^p$ is* multilinear *if for each $\mathbf{x}_j \in E_j$, where $j \neq i$, the function $g : E_i \to \mathbb{R}^p$ defined by $g(\mathbf{v}) = f(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{v}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_k)$ is linear.*

Suppose f is multilinear, and $i \neq j$. Then if $\mathbf{h} = (\mathbf{h}_1, \ldots, \mathbf{h}_k)$ where $\mathbf{h}_l \in E_l$,

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|f(\mathbf{a}_1, \ldots, \mathbf{h}_i, \ldots, \mathbf{h}_j, \ldots, \mathbf{a}_k)\|}{\|\mathbf{h}\|} = 0.$$

Additionally, we have

$$Df(\mathbf{a}_1, \ldots, \mathbf{a}_k)(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sum_{i=1}^{k} f(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{x}_i, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_k).$$

Proof. **To be done** □

Proposition 2.3.5. *Let M be an $n \times n$ matrix. Treat M as an element of $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ by considering each row of M as an element of $\mathbb{R}^n$. Then, the determinant function $\det : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ is differentiable and we have*

$$D(\det)(\mathbf{r}_1, \ldots, \mathbf{r}_n)(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{i=1}^{n} \det \begin{pmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{x}_i \\ \vdots \\ \mathbf{r}_n \end{pmatrix}.$$

Here, we have $\mathbf{x}_i$ in the ith row and everywhere else remains r_j when $j \neq i$.

Proof. The determinant function is multilinear, now apply the previous proposition. □

2.4 Multivariable MVT

Let us first begin by recalling the mean value theorem from single variable calculus.

> **Theorem 2.4.1** (Mean Value Theorem). Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then, there exists $c \in (a, b)$ such that
> $$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

We omit the proof as this is a single variable calculus result.

> **Theorem 2.4.2** (Multivar MVT). Let $U \subseteq \mathbb{R}^n$ and let $\mathbf{a}, \mathbf{b} \in U$ be such that the graph of the function $\gamma : [0, 1] \to \mathbb{R}^n$ defined by $\gamma(t) = (1 - t)\mathbf{a} + t\mathbf{b}$ has $\gamma(t) \in U$ for all $t \in [0, 1]$. If $f : U \to \mathbb{R}$ is differentiable on U and $f \circ \gamma$ satisifes the hypothesis of Theorem 2.4.1, then there exists $t_0 \in (0, 1)$ such that $\mathbf{c} = \gamma(t_0)$ and we have
> $$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \circ (\mathbf{b} - \mathbf{a}).$$

Proof. Not too hard, just apply Theorem 2.3.1 and Theorem 2.4.1. □

2.5 Partial derivatives

Definition 2.5.1 (Partial derivative)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{a} \in \mathbb{R}^n$. The ***i-th partial derivative*** *of f at* $\mathbf{a}$ is the limit

$$\lim_{h \to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}, \tag{2.4}$$

if it exists. We will denote this limit with $\partial_i f(\mathbf{a})$, or $D_i f(\mathbf{a})$.

The partial derivative is actually the usual single-variable calculus derivative of a certain function. The reader has probably observed that if we define a function $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = f(a_1, \ldots, a_{i-1}, x, a_{i+1}, a_n)$ then $D_i f(\mathbf{a}) = g'(a_i)$.

Partial derivatives are usually quite easy to calculuate. For example, if we let $f(x, y) = x^2 y + 4y$ then we have $D_1 f(x_0, y_0) = 2x_0 y_0$ and $D_2 f(x_0, y_0) = x_0^2 + 4$.

The reader will wonder how does the derivative interact with partial derivatives.

Theorem 2.5.2 (Components of gradient)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ and suppose f is differentiable at $\mathbf{a}$. Then, we have

$$\nabla f(\mathbf{a}) = (\partial_1 f(\mathbf{a}), \partial_2 f(\mathbf{a}), \ldots \partial_n f(\mathbf{a})).$$

Proof. **To be done** □

If we combine Theorem 2.5.2 and part (3) of Theorem 2.3.2, we can obtain the following

Corollary 2.5.3. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and suppose f is differentiable at* $\mathbf{a}$. *Then, we have*

$$\mathcal{M}(Df(\mathbf{a})) = \begin{bmatrix} \nabla f_1(\mathbf{a}) \\ \nabla f_2(\mathbf{a}) \\ \vdots \\ \nabla f_m(\mathbf{a}) \end{bmatrix} = \begin{bmatrix} \partial_1 f_1(\mathbf{a}) & \cdots & \partial_n f_1(\mathbf{a}) \\ \partial_1 f_2(\mathbf{a}) & \cdots & \partial_n f_2(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(\mathbf{a}) & \cdots & \partial_n f_m(\mathbf{a}) \end{bmatrix}$$

Additionally, we may wonder whether the existence of partial derivatives implies the existence of the derivative. This is not true.

Example 2.5.4. Let

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

f is certainly not continuous at $(0, 0)$ and thus is not differentiable at $(0, 0)$. However, it does have partial derivatives at $(0, 0)$. //

Theorem 2.5.5 (Continuous partials implies differentiable)**.** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and we have an open set U that contains $\mathbf{a}$, and for all $\mathbf{x} \in U$, for all i, $D_i f(\mathbf{x})$ exists, and the function $D_i f$ which is $\mathbf{x} \mapsto D_i f(\mathbf{x})$ is continuous at $\mathbf{a}$. Then $Df(\mathbf{a})$ exists.

If f satisifes the hypothesis of theorem above, it is called ***continuously differentiable*** at $\mathbf{a}$. Such a function is also called a C^1 function.

Proof. **To be done** □

This theorem can be easily generalized for $f : \mathbb{R}^n \to \mathbb{R}^m$.

Definition 2.5.6 (Class C^1 function)**.** Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$, and let $\mathbf{a}$ be a point in the interior of A. Then f *is* ***continuously differentiable*** *at a* or $f \in C^1$ at $\mathbf{a}$ if for all $i \in \{1, \ldots, n\}$, $\partial_i f$ is defined on some neighborhood of $\mathbf{a}$ and is continuous at $\mathbf{a}$.

2.5.1 Higher-order partial derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$ and fix some $i \in \{1, \ldots, n\}$. If the partial derivative of f exists everywhere (i.e. $D_i f(\mathbf{x})$ exists for all $\mathbf{x} \in \mathbb{R}^n$), then we have a function $D_i f : \mathbb{R}^n \to \mathbb{R}$ which maps a vector to the i-th partial derivative at that vector, i.e.

$$\begin{aligned} D_i f &: \mathbb{R}^n \to \mathbb{R} \\ \mathbf{x} &\mapsto D_i f(\mathbf{x}). \end{aligned}$$

One might want to take the partial derivatives of the function $D_i f$. For example, the j-th partial derivative of $D_i f$ at $\mathbf{x}$ would be $D_j(D_i f)(\mathbf{x})$. It is very possible that $D_j(D_i f)(\mathbf{x})$ exists for all $\mathbf{x} \in \mathbb{R}^n$ too, in this case we obtain a function

$$\begin{aligned} D_j(D_i f) &: \mathbb{R}^n \to \mathbb{R} \\ \mathbf{x} &\mapsto D_j(D_i f)(\mathbf{x}). \end{aligned}$$

This function is called a ***second-order (mixed) partial derivative*** *of* f. It's not too hard to define higher order partial derivatives. In [Spi18] this function is denoted $D_{i,j} f(\mathbf{x})$. This notation does reverse the order of i and j, but it turns out that for most functions, this does not matter. See Theorem 2.5.7.

Theorem 2.5.7 (Clairut's Theorem). Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and let $i, j \in \{1, \ldots, n\}$. Suppose $Df_j(Df_i)$ and $Df_i(Df_j)$ are continuous in some open set that contains $\mathbf{a}$. Then, $Df_j(Df_i)(\mathbf{a}) = Df_i(Df_j)(\mathbf{a})$.

Proof. **To be done**. □

Note that the converse of this false. The following example from [Spi18, Problem 2-24 on p. 29] will provide a nice counterexample to the converse of Theorem 2.5.7.

Example 2.5.8. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

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

We can show that $D_2 f(x, 0) = x$ for all x, and $D_1 f(0, y) = -y$ for all y. Then $D_{1,2} f(0, 0) \neq D_{2,1} f(0, 0)$. //

Let $f : \mathbb{R}^n \to \mathbb{R}^m$. If f is differentiable everywhere and has continuous partial derivatives, we say that f is C^1. Such a function f satisifes the hypothesis of Theorem 2.5.7.

Theorem 2.5.9. Let $A \subseteq \mathbb{R}^n$. If the maximum or minimum of $f : A \to \mathbb{R}$ occurs at a point $\mathbf{a}$ in the interior of A and $D_i f(\mathbf{a})$ exists, then $D_i f(\mathbf{a}) = 0$.

Proof. Define $g_i : \mathbb{R} \to \mathbb{R}$ by $g_i(x) = f(a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_n)$. Then g_i has a maximum or minimum at a_i as f has a maximum or minimum at $\mathbf{a}$. Now, since $\mathbf{a}$ is in the interior of A, there is some open ball that contains $\mathbf{a}$ and thus there is some open interval of a_i which g_i is defined on. g_i is also differentiable because $g_i'(a_i) = D_i f(\mathbf{a})$[2]. By single variable calculus, we know that $g_i'(a_i) = 0$ as $g(a_i)$ is the maximum or minimum. Now, $D_i f(\mathbf{a}) = g_i'(a_i) = 0$. □

As an immediate corollary, we get the following:

Corollary 2.5.10. *If $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$ and $\mathbf{a}$ is in the interior of A, then $\nabla f(\mathbf{a}) = \mathbf{0}$.*

[2]Technically this justification is very handwavey. However it's not too hard to repair this, so we'll do it at a later time.

2.6 Directional derivatives

Definition 2.6.1 (Directional derivative)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$. The ***directional derivative*** *of f at $\mathbf{a}$ in the direction of $\mathbf{u}$* is defined to be

$$\lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{x}) - f(\mathbf{a})}{t},$$

if the limit exists. We denote this limit as $D_{\mathbf{u}}f(\mathbf{a})$.

While this definition does not impose any conditions on the direction vector $\mathbf{u}$, we usually ask for it to be a unit vector to make life easier. An immediate consequence of this definition is that the i-th partial derivative of f at $\mathbf{a}$ can be thought of as the directional derivative of f at $\mathbf{a}$ in the direction $\mathbf{e}_i$, where $\mathbf{e}_i$ is the i-th standard basis vector of $\mathbb{R}^n$.

Proposition 2.6.2. *If $\partial_i f(\mathbf{a})$ exists then $\partial_i f(\mathbf{a}) = D_{\mathbf{e}_i} f(\mathbf{a})$.*

Proof. Apply the definition. □

Again, the reader may wonder whether the existence of directional derivatives tells us anything about the derivative. It turns out that this is not true, even if every single directional derivative exists. The following example comes from [Spi18, Problem 1-26 on p. 13, Problem 2-31 on p. 33].

Example 2.6.3. Let $A = \{\,(x, y) \in \mathbb{R}^2 : x > 0 \text{ and } 0 < y < x^2\,\}$. If we take any straight line that passes through $(0, 0)$ (which you can think about as the graph of a function $f(x) = mx$ for some $m \in \mathbb{R}$), it contains some interval around $(0, 0)$ which is in $\mathbb{R}^2 \setminus A$.

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x, y) = \begin{cases} 0 \text{ if } x \notin A \\ 1 \text{ if } x \in A \end{cases}.$$

For any $\mathbf{h} \in \mathbb{R}^2$ we define $g_{\mathbf{h}} : \mathbb{R} \to \mathbb{R}$ by $g_{\mathbf{h}}(t) = f(t\mathbf{h})$. Then $g_{\mathbf{h}}$ is continuous at 0, but f is not continuous at $(0, 0)$.
To be completed //

2.7 Optimization

Let us first extend the definition of maximum and minimum points to $\mathbb{R}^n$.

Definition 2.7.1 (Local extrema)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$. Then if there is a neighborhood U of $\mathbf{a}$ such that for all $\mathbf{x} \in U$, we have $f(\mathbf{x}) \le f(\mathbf{a})$, then $\mathbf{a}$ is called a ***local maximum***. If it is the case that for all $\mathbf{x} \in U$ we have $f(\mathbf{x}) \ge f(\mathbf{a})$, then $\mathbf{a}$ is called a ***local minimum***.

Points which are local maximum or minimum points are called *local extrema*.

Definition 2.7.2 (Critical point)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then $\mathbf{c} \in \mathbb{R}^n$ is a ***critical point*** if $\nabla f(\mathbf{c}) = 0$. Additionally, the value $f(\mathbf{c})$ is called a *critical value*.

Theorem 2.7.3 (Lagrange Multipliers)**.** Let $f, G : \mathbb{R}^n \to \mathbb{R}$ be C^1 functions. Let $S = G^{-1}(()\,0)$. If the restriction of f to S, $f : S \to \mathbb{R}$ has a maximum or minimum at $\mathbf{c} \in S$ and $\nabla G(\mathbf{c}) \neq \mathbf{0}$, then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(\mathbf{c}) = \lambda \nabla G(\mathbf{c}).$$

2.8 Implicit function theorem

Theorem 2.8.1 (Implicit function theorem)**.** Let $F : \mathbb{R}^n \times \mathbb{R}$, $F(\mathbf{x}, y)$ be C^1 on some neighborhood $U \subseteq \mathbb{R}^{n+1}$ of the point $(\mathbf{a}, b) \in \mathbb{R}^{n+1}$ and suppose that $F(\mathbf{a}, b) = 0$ and $\partial F_y(\mathbf{a}, b) \neq 0$. Then there is an open ball centered at $\mathbf{a}$ of radius r and a *unique* C^1 function $f : B(\mathbf{a}, r) \to \mathbb{R}$ such that $F(\mathbf{x}, f(\mathbf{x})) = 0$ for all $x \in B(\mathbf{a}, r)$.

2.9 Inverse function theorem

Theorem 2.9.1 (Inverse function theorem). Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in an open set containing $\mathbf{a}$ and suppose $Df(\mathbf{a})$ is invertible. Then, there is some open set $V \subseteq \mathbb{R}^n$ that contains $\mathbf{a}$ and an open set $W \subseteq \mathbb{R}^n$ that contains $f(\mathbf{a})$ such that $f : V \to W$ has a continuous inverse $f^{-1} : W \to V$, which is differentiable and for all $\mathbf{y} \in W$, we have

$$D(f^{-1})(\mathbf{y}) = [Df(f^{-1}(\mathbf{y}))]^{-1}.$$

Bibliography

[Spi18] Michael Spivak. *Calculus on manifolds: a modern approach to classical theorems of advanced calculus.* eng. Mathematics monograph series. Boca Raton London New York: CRC Press, Taylor & Francis Group, 2018. ISBN: 9780805390216.