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 1.2.5, 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.

Proposition 1.2.2 (Cauchy-Schwarz Inequality)**.** *Let V be an inner product space with induced norm. Then, for any vectors $u, v \in V$,*

$$|\langle u, v\rangle| \leq \|u\|\|v\|,$$

with equality if and only if u is a scalar multiple of v (or vice versa).

Proof. See linear algebra done right □

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

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

where x_i is the ith coordinate of x.

Proof. Let $\{ e_1, \ldots, e_n \}$ be the standard basis of $\mathbb{R}^n$. So if $x \in \mathbb{R}^n$, we have $x = \sum_1^n c_i e_i$. Observe that c_i is precisely the ith coordinate of x. Thus,

$$\begin{aligned} \|x\| &= \left\| \sum_1^n c_i e_i \right\| \\ &\leq \sum_1^n |c_i| \|e_i\| \\ &= \sum_1^n |c_i|. \end{aligned}$$

□

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 1.2.4. *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 $\{ v_1, \ldots, v_m \}$ be a basis of $\mathbb{R}^m$, so if $v \in \mathbb{R}^m$, then $v = \sum_1^m a_i v_i$. Then,

$$\begin{aligned} \|T(v)\| &= \left\| T\left(\sum_1^m a_i v_i \right) \right\| \\ &\leq \sum_1^m \|T(a_i v_i)\| \\ &= \sum_1^m |a_i| \|T(v_i)\|. \end{aligned}$$

Notice that the last line looks suspiciously like a dot product. Indeed, it is. Let $x = (|a_1|, \ldots, |a_m|)$ and $y = (\|T(v_1)\|, \ldots, \|T(v_m)\|)$. Then, by Cauchy-Schwarz, we have

$$\langle x, y\rangle = |\langle x, y\rangle| \leq \|x\|\|y\|.$$

So choose $M = \|y\|$. Then the result follows. □

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 1.2.5 (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.

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 1.2.5.

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 $\mathbb{R}^n$.

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

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
$$Df(\mathbf{a}) = (Df_1(\mathbf{a}), \ldots, Df_m(\mathbf{a})).$$
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 1.2.3 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, $i = 1, \dots, k$ 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 $g(\mathbf{v}) = f(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{v}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_k)$ is linear.*

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

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

Additionally, we have

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

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*

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

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. □

Note that we do we need that the codomain of f to be $\mathbb{R}$. The next example shows this. See [Zam24, Example 4.2.3, p. 226]

Example 2.4.3. Let $f : \mathbb{R} \to \mathbb{R}^2$ be defined by $f(t) = (\cos t, \sin t)$. Then $f(0) = (1,0)$ and $f(\pi) = (-1,0)$ but there is no $c \in (0,\pi)$ such that $f(\pi) - f(0) = Df(c) \cdot (\pi - 0)$. This is because the equation means that $(-2,0) = \pi(-\sin c, \cos c)$. But if $\cos c = 0$ then $-\pi \sin c$ is ± 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 next example taken from [Spi18, Problem 2-24 on p. 29] is a 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 leave it to the reader show that $D_2 f(x, 0) = x$ for all x, and $D_1 f(0, y) = -y$ for all y. (Apply the definition of partial derivative). 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 direction $\mathbf{u}$ 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, Prob 1-26 on p. 13, Prob 2-31 on p. 33].

Example 2.6.3. Let $A = \{ (x, y) \in \mathbb{R}^2 : x > 0 \text{ and } 0 < y < x^2 \}$. To see this set better, view the plot of this set.

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 see this, notice that $g_{\mathbf{h}}$ is a function that defines a line passing through the origin. Thus, there is some open interval around 0, call it I, such that if $t \in I$, we have $t\mathbf{h} \notin A$. //

Example 2.6.4. Let $n \geq 1$ and let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x, y) = \begin{cases} \frac{x^n y}{x^{n+1}+y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}.$$

Then f is not continuous at 0. To see this, approach along $y = x^n$. Approaching along this path leads to the limit at 0 being 1.

We do have every directional derivative though. //

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 $f(\mathbf{x}) \leq f(\mathbf{a})$, then $\mathbf{a}$ is called a ***local maximum***. If it is the case that for all $\mathbf{x} \in U$ we have $f(\mathbf{x}) \geq f(\mathbf{a})$, 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 $\nabla f(\mathbf{c}) = 0$ or $\nabla f(\mathbf{c})$ does not exist. Additionally, the value $f(\mathbf{c})$ is called a *critical value*.

Note that if the domain of f is changed to be some subset of $\mathbb{R}^n$, we will require that $\mathbf{c}$ be an interior point (otherwise we cannot discuss $\nabla f(\mathbf{c})$).

Theorem 2.7.3 (Local EVT). Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$. If $\mathbf{a}$ is in the interior of A and $\mathbf{a}$ is a local extremum then $\nabla f(\mathbf{a}) = 0$ or it does not exist.

Proof. This actually follows as a corollary to Theorem 2.5.9. If $\mathbf{a}$ is a local extremum then it is either a maximum or minimum, so all $\partial_i f(\mathbf{a})$ are zero. Thus if $\nabla f(\mathbf{a})$ exists then we must have $\nabla f(\mathbf{a}) = 0$. ☐

As a consequence we note that if $\mathbf{a}$ is a local extremum then it is either a boundary point or a critical point.

As an example of ∇f being undefined at a critical point, we may consider the following [Zam24, Example 4.3.13, p. 234].

Example 2.7.4. Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(\mathbf{x}) = \|\mathbf{x}\|$. Then, $\nabla f(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$. So ∇f is not defined at the origin, but the origin of f is a critical point of f. Observe that $\nabla f(\mathbf{x})$ is not 0 if $\mathbf{x}$ is not zero, so f has no local extremum on $\mathbb{R}^n \setminus \{ \mathbf{0} \}$. Therefore the only possible local minimum of f is at $\mathbf{0}$. It is also the local minimum of f. //

Definition 2.7.5 (Saddle point). Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$. Then, $\mathbf{a}$ is a **saddle point** if it is an interior point of A, $\nabla f(\mathbf{a}) = 0$ and $\mathbf{a}$ is not a local extremum.

Theorem 2.7.6 (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

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

1. Find out whether global extrema exist. Usually this involves checking if the image of the function is compact.
2. Finding the critical points on the interior. This involves calculating the gradient of the function.
3. Checking the boundary for extrema. Usually you can parameterize the boundary. Compose the parameterization with the original function to obtain a single variable function which you can optimize using single variable calculus techniques.

2.8 Tangent vectors

Definition 2.8.1 (Tangent Vector). Let $S \subseteq \mathbb{R}^n$ and $\mathbf{p} \in S$. Then $\mathbf{v}$ *is a* ***tangent vector*** *of* S *at* $\mathbf{p}$ if there is an open interval $I \subseteq \mathbb{R}$ containing 0 and a function $\gamma : I \to S$ such that $\gamma(0) = \mathbf{p}$ and $\gamma'(0) = \mathbf{v}$.

Note that we technically do not need that the open interval contains 0, as any open interval whatsoever is diffeomorphic to an open interval containing 0. Additionally, the condition on the codomain of γ can be changed to $\gamma : I \to \mathbb{R}^n$ and $\gamma(I) \subseteq S$.

Definition 2.8.2 (Tangent space). The set of all tangent vectors of S at $\mathbf{p}$ is called ***tangent space*** *of* S *at* $\mathbf{p}$. It is denoted $T_{\mathbf{p}}S$.

Example 2.8.3. Let $f(x, y) = 9 - x^2 - y^2$ and let $S = \{ (x, y, f(x, y)) : x, y \in \mathbb{R} \} = \Gamma(f)$. //

Definition 2.8.4 (Tangent plane). The tangent plane of S at $\mathbf{p}$ is denoted $\mathbf{p} + T_{\mathbf{p}}S$.

$$\mathbf{p} + T_{\mathbf{p}}S = \{ \mathbf{p} + \mathbf{v} : \mathbf{v} \in T_{\mathbf{p}}S \}.$$

If S is a graph of some C^1 function the tangent space is a vector space and is easy to calculate.

Theorem 2.8.5. Let $V \subseteq \mathbb{R}^k$ be an open set and $F : V \to \mathbb{R}^{n-k}$ be a C^1 function. Let $S = \Gamma(F)$, that is

$$S = \{ (\mathbf{v}, F(\mathbf{v})) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : \mathbf{v} \in V \} .$$

Then for any $\mathbf{a} \in V, \mathbf{p} = (\mathbf{a}, F(\mathbf{a})) \in S$, the tangent space of S at $\mathbf{p}$ satisifies

$$T_{\mathbf{p}}S = \{ (\mathbf{w}, D_{\mathbf{a}}F(\mathbf{w})) \in \mathbb{R}^k \times \mathbb{R}^{n-k} : \mathbf{w} \in \mathbb{R}^k \} .$$

Thus $T_{\mathbf{p}}S$ is a k-dimensional vector space spanned by $\{ (\mathbf{e}_i, \partial_i F(\mathbf{a})) \}$ where $i \in \{ 1, \ldots, k \}$ and $\mathbf{e}_i$ are the standard basis vectors of $\mathbb{R}^k$.

2.9 Smooth Manifolds

Definition 2.9.1 (Smooth Manifold at a point p)**.** Let $S \subseteq \mathbb{R}^n$ and $\mathbf{p} \in \mathbb{R}^n$. Then S is a *k-dimensional **smooth manifold** at* $\mathbf{p}$ if there is an open neighborhood of $\mathbf{p}$ and a C^1 function $f : V \to \mathbb{R}^n$ where V is open such that $S \cap U$ is the graph of f.

Example 2.9.2 (The circle)**.** Let $S^1 = \{ \mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1 \}$ be the unit circle. Then S^1 is a smooth manifold at any point. We leave the reader to verify this fact. Notice that points (x, y) such that $x^2 + y^2 = 1$ lie in S^1, and these are the only points. So write y in terms of x. This actually hints to Theorem 2.10.1. //

Example 2.9.3 (The real numbers)**.** $\mathbb{R}$ is easily seen to be a n-dimensional smooth manifold as it is the graph of the identity function from $\mathbb{R}$ to $\mathbb{R}$. //

Definition 2.9.4 (Smooth Manifold)**.** Let $S \subseteq \mathbb{R}^n$. Then S is a k-dimensional smooth manifold if it is a k-dimensional smooth manifold at every point $\mathbf{p} \in S$.

Note that in both Definition 2.9.4 and Definition 2.9.1, we require that the dimension k of the manifold be lower than the dimension of the space it sits in, i.e. $k < n$.

Now for some non-examples of smooth manifolds.

Example 2.9.5 (A cusp)**.** This comes from [Zam24, Example 4.6.10, p. 268]. Consider the curve determined by $x^2 = y^3$ (see Figure 2.1).

Figure 2.1: Graph of the cusp $x^2 = y^3$

This is not a smooth manifold because it is not a smooth manifold at the origin. Notice that any ball around the origin is not the graph of a C^1 function, as the function $f(x) = x^{2/3}$ is not differentiable at the origin. A more careful proof of this fact can be found in [Zam24, Example 4.6.11, p. 268]. //

Example 2.9.6 (Figure 8)**.** This example comes from [Zam24, Example 4.6.10, p. 268] as well. Consider the figure 8 shape determined by the equation $x^4 = x^2 - y^2$ (see Figure 2.2). Any ball at the origin intersected with this shape

Figure 2.2: Graph of the figure 8 shape $x^4 = x^2 - y^2$

cannot produce a graph, since for each value of x, you need to have 2 different y values. If x is treated as a function of y instead, then 2 different x values need to be provided for a single y value. Consequently S cannot be a graph at the origin so it cannot be a smooth manifold at the origin. //

Here are some non-examples of smooth manifolds in 2 dimensions. **TODO: add these**

2.9.1 Diffeomorphisms

Recall from linear algebra that a vector space isomorphism is a bijective linear transformation. Vector space isomorphisms preserve all the structure we care about in a vector space (algebraically speaking). A diffeomorphism is an isomorphism, but for smooth manifolds instead. It preserves the properties we care about for smooth manifolds.

Definition 2.9.7 (Diffeomorphism)**.** Let $U, V \subseteq \mathbb{R}^n$ be open sets. Then $f : U \to V$ is a **diffeomorphism** if f is bijective, C^1 and $f^{-1} : V \to U$ is also a C^1 function.

Note that the condition of f^{-1} being C^1 is absolutely necessary since f being C^1 does not guarantee that f^{-1} is C^1.

Example 2.9.8. Let $f(x) = x^3$. Then $f^{-1}(y) = \sqrt[3]{y}$. f^{-1} is differentiable, but its derivative is discontinuous at 0. For more details, see [Zam24, Example 5.5.5, p. 326]. //

Proposition 2.9.9. *Let $U, V \subseteq \mathbb{R}^n$ be open and $f : U \to V$ be a diffeomorphism. Then, the following are true:*

1. *$O \subseteq U$ is open (closed) iff $f(O)$ is open (closed).*
2. *$K \subseteq U$ is compact iff $f(K)$ is compact.*
3. *$S \subseteq U$ is (path)-connected iff $f(S)$ is (path)-connected.*

Proof. A diffeomorphism is a homeomorphism. □

Proposition 2.9.10. *Let $f : U \to V$ be a diffeomorphism and $S \subseteq U$, $\mathbf{p} \in S$. Then $\mathbf{v} \in T_{\mathbf{p}}S$ if and only if $\mathbf{v} \in T_{f(\mathbf{p})}f(S)$.*

Proof. Apply the definition. You will need to use the chian rule (Theorem 2.3.1) in the proof. □

Definition 2.9.11 (Local Diffeomorphism)**.** Let $A, B \subseteq \mathbb{R}^n$ be open and $\mathbf{a} \in A$. Let $f : A \to B$. Then f is a **local diffeomorphism** at $\mathbf{a}$ if there is an open set $U \subseteq A$ containing $\mathbf{a}$ such that $f|_U : U \to f(U)$ is a diffeomorphism. The inverse $f|_U^{-1}$ is called a *local inverse for f at* $\mathbf{a}$.

Clearly every global diffeomorphism is a local diffeomorphism, however the converse is untrue. Even if a function is a local diffeomorphism at every point, it may still not be a global diffeomorphism.

Example 2.9.12. Let $f(x, y) = (e^x \cos y, e^x \sin y)$. Then $Df(x, y)$ is invertible for all (x, y) so f is a local diffeomorphism at every point in $\mathbb{R}^2$. However notice that $f(x, y) = f(x, y + 2\pi)$ so f is not injective and thus is not a diffeomorphism. //

Proposition 2.9.13. *If f is a local diffeomorphism at $\mathbf{a}$, then $Df(\mathbf{a})$ is invertible and satisfies*

$$Df^{-1}(f(x)) = (Df(x))^{-1}.$$

Proof. Apply the chain rule. □

The converse of this proposition is the inverse function theorem (Theorem 2.10.2). Thus we summarize that f is a local diffeomorphism at $\mathbf{a}$ if and only if $Df(\mathbf{a})$ is invertible.

2.10 Implicit and inverse functions

Theorem 2.10.1 (Implicit function theorem)**.** Let $F : \mathbb{R}^{n+k} \times \mathbb{R}^k$, $F(\mathbf{x}, \mathbf{y})$ be C^1 on some neighborhood $U \subseteq \mathbb{R}^{n+k}$ of the point $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{n+k}$ and suppose that $F(\mathbf{a}, \mathbf{b}) = 0$ and $\partial F_{\mathbf{y}}(\mathbf{a}, \mathbf{b})$ is invertible. Then there is an open ball centered at $\mathbf{a}$ of radius r and a *unique* C^1 function $f : B(\mathbf{a}, r) \to \mathbb{R}^k$ such that $F(\mathbf{x}, f(\mathbf{x})) = 0$ for all $x \in B(\mathbf{a}, r)$.

Note that

$$\partial F_{\mathbf{y}} = \begin{bmatrix} \dfrac{\partial F_1}{\partial y_1} & \cdots & \dfrac{\partial F_1}{\partial y_k} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_k}{\partial y_1} & \cdots & \dfrac{\partial F_k}{\partial y_k} \end{bmatrix},$$

since we have

$$Df = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} & \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_k} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_k}{\partial x_1} & \cdots & \frac{\partial F_k}{\partial x_n} & \frac{\partial F_k}{\partial y_1} & \cdots & \frac{\partial F_k}{\partial y_k} \end{bmatrix}$$

Proof. Proof from scratch To be done. ☐

Proof using Theorem 2.10.2. To be done. ☐

Theorem 2.10.2 (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

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

Proof from scratch. Very long. See [Spi18, Thm 2-11]. ☐

Proof using Theorem 2.10.1. Define $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by $F(\mathbf{x}, \mathbf{y}) = \mathbf{y} - f(\mathbf{x})$. **To be completed.** ☐

Chapter 3

Integration

3.1 Riemann integration

Recall that a *partition* of an interval $[a, b]$ is a set of points $\{ t_0, \dots, t_n \}$ where $a = t_0 < t_1 < \cdots < t_n = b$. Given a rectangle $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$, a partition P of R is a collection of partitions $P_1, \dots, P_n$, where P_i is a partition of $[a_i, b_i]$. Now suppose we partition $[a_1, b_1] \times [a_2, b_2]$ with $P_1 = \{ t_i \}_0^n$ and $P_2 = \{ x_i \}_0^m$. Given t_{k-1}, t_k and x_{j-1}, x_j we can form a **subrectangle** $[t_{k-1}, t_k] \times [x_{j-1}, x_j]$. Notice here that valid subrectangles would be a product of subintervals of the partition. This notion generalizes easily to higher dimensions.

Let $R \subset \mathbb{R}^n$ be a rectangle, let $f : R \to \mathbb{R}$ be a bounded function. Let P be a partition of R, let S be a subrectangle of P, and define

$$\begin{aligned} m_S(f) &= \inf_{x \in S} f(x) \\ M_S(f) &= \sup_{x \in S} f(x) \end{aligned}$$

The **upper sum** of f on P is denoted $U(f, P)$, and the **lower sum** of f on P is denoted $L(f, P)$. They are defined by

$$\begin{aligned} L(f, P) &= \sum_{S \in R_P} m_S(f) v(S) \\ U(f, P) &= \sum_{S \in R_P} M_S(f) v(S), \end{aligned}$$

where $v(S)$ is the volume of a subrectangle and R_P is the subrectangles of a partition P. The **upper integral** is denoted $U(f)$ and the **lower integral** is denoted $L(f)$. They are defined by

$$\begin{aligned} U(f) &= \inf \{ U(f, P) : P \in \mathcal{P} \} \\ L(f) &= \sup \{ L(f, P) : P \in \mathcal{P} \} \end{aligned}$$

where $\mathcal{P}$ is the set of all partitions of R. Note that $U(f)$ and $L(f)$ always exist whenever f is bounded.

A **refinement** of a partition $P = (P_1, \dots, P_n)$ is a partition $P' = (P_1', \dots, P_n')$ such that $P_i \subseteq P_i'$. It thus follows that if P' refines P, every subrectangle of P' is contained in some subrectangle of P.

Lemma 3.1.1 (Upper lower sum inequalities)**.** *Suppose P' refines P. Then, $L(f, P) \leq L(f, P')$ and $U(f, P) \geq U(f, P')$.*

Proof. Apply the definitions. □

Corollary 3.1.2. *Given any 2 partitions P, P', $L(f, P) \leq U(f, P')$.*

Corollary 3.1.3. *The lower integral is always smaller than the upper integral, i.e. $L(f) \leq U(f)$.*

Definition 3.1.4 (Integrability). If $f : R \to \mathbb{R}$ is a bounded function on a rectangle $R \subset \mathbb{R}^n$, then f is said to be **integrable** if $U(f) = L(f)$.

Proposition 3.1.5 (Integrability criterion)**.** *Let $f : R \to \mathbb{R}$ be a bounded function on a rectangle $R \subseteq \mathbb{R}^n$. Then, f is integrable if and only if for every $\varepsilon > 0$, there is a partition P such that $U(f, P) - L(f, P) < \varepsilon$.*

Proof. For the forward direction, take partitions P, P' such that $U(f, P) - L(f, P') < \varepsilon$ and let P'' refine both P and P'. The converse is immediate. □

Theorem 3.1.6 (Properties of Integrals). Let $f, g : R \to \mathbb{R}^n$ be integrable functions defined on a rectangle $R \subset \mathbb{R}^n$. Then, the following are true:

1. **(Linearity)** Let $c \in \mathbb{R}$. Then, $f + cg$ is integrable, and $\int_R f + cg = \int_R f + c \int_R g$.
2. **(Monotonicity)** Suppose $f \leq g$[a]. Then, $\int_R f \leq \int_R g$.
3. **(Product)** The product function fg is integrable (defined as $fg(x) = f(x)g(x)$).
4. $|f|$ is integrable, and $\int_R |f| \geq \left|\int_R f\right|$.

[a]This is pointwise, i.e. for all x, $f(x) \leq g(x)$

Proof. (1), (2) follow from the definition. For (3), prove that the function f^2 defined by $f^2(x) = f(x)f(x)$ is integrable first, then notice $(f + g)^2 = f^2 + 2fg + g^2$. For (4), just play around with it. □

Definition 3.1.7 (Measure Zero). Let $S \subseteq \mathbb{R}^n$. Then S is said to have **measure 0** if given $\varepsilon > 0$, there is a *countable* collection of rectangles $(R_n)_{n\in\mathbb{N}}$ that cover S (i.e. $\bigcup_{n\in\mathbb{N}} R_n \supseteq S$) and have total volume less than ε, i.e. $\sum_1^\infty v(R_n) < \varepsilon$, where $v(R_n)$ denotes the volume of the rectangle R_n.

Note that this agrees with the definition of a set having 0 Lebesgue measure.

The following theorem tells us that every Riemann integrable function is continuous almost everywhere.

Theorem 3.1.8 (Lebesgue's Criterion). Let R be a rectangle in $\mathbb{R}^n$ and let $f : R \to \mathbb{R}$. Then, f is integrable on R if and only if the set of discontinuities of f has measure 0.

If a function integrates to 0, is it the zero function? Almost everywhere.

Lemma 3.1.9. *Suppose $f : R \to \mathbb{R}$. Then $\{ x : f(x) \neq 0 \}$ if and only if f is integrable and $\int_R f = 0$.*

Proposition 3.1.10. *Let $f, g : R \to \mathbb{R}$ and suppose f is integrable. If the set $\{ x : f(x) \neq g(x) \}$ has measure 0, then g is integrable and $\int_R g = \int_R f$.*

3.2 Partitions of unity

Definition 3.2.1 (Support of a function). Let $f : X \to \mathbb{R}$ be a real valued function. The **support of** f is defined to be the set of points in X where f is not zero, i.e.

$$\operatorname{supp} f = \{ x \in X : f(x) \neq 0 \}.$$

We can additionally talk about the *closed support* of f, which if X is a topological space, will simply involve taking the closure of $\operatorname{supp} f$. If the closed support of f is compact, then f is said to have *compact support.*

Let $A \subseteq \mathbb{R}^n$. Then a collection of sets $(S_\alpha)_{\alpha\in J}$ is said to be *locally finite for* A if for all $x \in A$, there is a neighborhood U of x such that U intersects only finitely many S_α (i.e. there are finitely many indices α such that $U \cap S_\alpha$ is nonempty)[1].

[1]We introduce this definition to make defining

Definition 3.2.2 (Smooth partition of unity). Let $A \subseteq \mathbb{R}^n$ and $\mathcal{U} = \{ U_\alpha : \alpha \in J \}$ be an open cover of A. A C^∞ **partition of unity for A subordinate to $\mathcal{U}$** is a *countable* collection of C^∞ functions $\{ \varphi_i : U \to \mathbb{R}, i \in \mathbb{N} \}$ where U is open and $A \subseteq U$ satisfying the following properties:

1. $0 \leq \varphi_i(x) \leq 1$ for all $x \in A$;
2. The collection of supports is locally finite for A;
3. For all $x \in A$, $\sum_{i=1}^{\infty} \varphi_i(x) = 1$;
4. Each φ_i has compact support;
5. For every φ_i, there is some U_α such that $\operatorname{supp} \varphi_i \subseteq U_\alpha$

A collection of functions that satisfy only 1-3 is simply called a partition of unity for A. Condition (5) is needed for the collection of functions to be subordinate to $\mathcal{U}$. Due to (2), the sum in (3) will always converge.

Note that we are following the definition of [Spi18, Thm 3.11, p. 63].

Chapter 4

Differential Forms

4.1 Tensors and Alternating Tensors

The language of differential forms is of tensors. To understand differential forms, we must first understand its language.

Definition 4.1.1 (Multilinear Function/Tensor). A function $T : V^k \to \mathbb{R}$ is **multilinear** if for every $i \in \{1, \ldots, k\}$, the following are true:

$$\begin{aligned} T(v_1, \ldots, v_i + v_i', \ldots, v_k) &= T(v_1, \ldots, v_i, \ldots, v_k) + T(v_1, \ldots, v_i', \ldots, v_k), \\ T(v_1, \ldots, av_i, \ldots, v_k) &= aT(v_1, \ldots, v_i, \ldots, v_k). \end{aligned}$$

So T is multilinear if it linear in each argument separately.

A multilinear function $T : V^k \to \mathbb{R}$ is called a **k-tensor on V**.

What we have technically defined above is a multilinear *functional*, since it is a linear map into the underlying field. However, we will really only care about such objects since we are going to investigate tensors.

The set of all k-tensors on V will be denoted $\mathfrak{J}^k(V)$. There is a natural vector space structure on $\mathfrak{J}^k(V)$ obtained by passing to V. Namely, we define $S + T$ and aS to be the tensors

$$\begin{aligned} (S + T)(v_1, \ldots, v_k) &= S(v_1, \ldots, v_k) + T(v_1, \ldots, v_k), \\ (aS)(v_1, \ldots, v_k) &= a \cdot S(v_1, \ldots, v_k). \end{aligned}$$

Now, $\mathfrak{J}^k(V)$ is special because not only are the tensors in $\mathfrak{J}^k(V)$ vectors, tensors in $\mathfrak{J}^k(V)$ can also interact with tensors in $\mathfrak{J}^l(V)$. This operation is called a tensor product.

Definition 4.1.2 (Tensor product). Let $S \in \mathfrak{J}^k(V)$ be a k-tensor and $T \in \mathfrak{J}^l(V)$ be a l-tensor. The **tensor product** of S and T, denoted $S \otimes T \in \mathfrak{J}^{k+l}(V)$, is a $k + l$ tensor given by

$$(S \otimes T)(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}) = S(v_1, \ldots, v_k) \cdot T(v_{k+1}, \ldots, v_{k+l}).$$

When defining a new operation, we need to discuss its algebraic properties.
Proposition 4.1.3. *The tensor product satisifies the following:*

1. $(S_1 + S_2) \otimes T = S_1 \otimes T + S_2 \otimes T$.
2. $S \otimes (T_1 + T_2) = S \otimes T_1 + S \otimes T_2$.
3. $(aS) \otimes T = S \otimes (aT) = a(S \otimes T)$.

4. $(S \otimes T) \otimes U = S \otimes (T \otimes U)$.

So it is associative, distributive over addition of tensors, and plays nicely[1] *with scalar multiplication of tensors.*

Proof. Apply the definitions. □

In light of Proposition 4.1.3, we will drop the parentheses when discussing $(S \otimes T) \otimes U$, and simply write $S \otimes T \otimes U$.
Warning 4.1.4. The tensor product is not commutative in general. So, $S \otimes T$ is usually not equal to $T \otimes S$.

We bring our attention to 1-tensors. A 1-tensor is simply a linear functional and so $\mathfrak{J}^1(V)$ is just V^*, the dual space of V. Is such an easy description coincidental? Is there any relation of the space of k-tensors with V^*? Luckily, the answer is yes.

Suppose $v_1, \ldots, v_n$ is a basis of V. Then, the *dual basis*, denoted $\varphi_1, \ldots, \varphi_n$ are linear functionals with the property[2] that $\varphi_i(v_j) = 1$ if $i = j$, and 0 if $i \neq j$.
Proposition 4.1.5. *Let $v_1, \ldots, v_n$ be a basis of V, and let $\varphi_1, \ldots, \varphi_n$ be the dual basis. The set of all possible k-fold tensor products*

$$\varphi_{j_1} \otimes \cdots \otimes \varphi_{j_k}$$

where $j_i \in \{1, \ldots, n\}$ is a basis of $\mathfrak{J}^k(V)$.

It thus follows that $\mathfrak{J}^k(V)$ has dimension n^k.

Proof. Let us first make an important observation. Fix $j_1, \ldots, j_k \in \{1, \ldots, n\}$. For any $i_1, \ldots, i_k \in \{1, \ldots, n\}$, notice that

$$\varphi_{j_1} \otimes \cdots \varphi_{j_k}(v_{i_1}, \ldots, v_{i_k}) = \begin{cases} 1 & \text{if } j_1 = i_1, \ldots, j_k = i_k, \\ 0 & \text{otherwise.} \end{cases}$$

Let us check that these guys span $\mathfrak{J}^k(V)$. Let $T \in \mathfrak{J}^k(V)$. □

Sending a vector space to the set of k tensors on it is a functorial construction. What that means is that any linear transformation between vector spaces also induces a transformation of their corresponding tensor spaces.

Definition 4.1.6 (Tensoring stuff is a contravariant functor). Suppose $f : V \to W$ is a linear transformation. Then, there is a linear transformation $f^* : \mathfrak{J}^k(W) \to \mathfrak{J}^k(V)$ defined by

$$f^*T(v_1, \ldots, v_k) = T(f(v_1), \ldots, f(v_k)).$$

Thus it assigns a k tensor on W, $T \in \mathfrak{J}^k(W)$ to a k tensor on V, f^*.

In this case the transformation is "backward". So, this means $(g \circ f)^* = f^* \circ g^*$, which we leave the reader to verify.

At this stage, we haven't given many examples of tensors. So let us now give some examples.
Example 4.1.7. Of course any element of V^* is a 1-tensor. //
Example 4.1.8. Recall that an inner product is a conjugate symmetric, positive definite and bilinear map. Since we are working over $\mathbb{R}$, our inner products are symmetric. So any inner product on a vector space V qualifies as a 2-tensor.

To be more precise, let V be a real vector space. Then an *inner product on V* is a 2-tensor T which is symmetric, i.e. $T(v, w) = T(w, v)$, and positive definite, i.e. $T(v, v) > 0$ and $T(v, v) = 0$ if and only if $v = 0$. //
Proposition 4.1.9. *Suppose T is an inner product on V. Then, there is an orthonormal basis of V with respect to T. Moreover, there is an isomorphism f from $\mathbb{R}^n$ to V such that $T(f(u), f(v)) = \langle u, v \rangle$.*

Proof. Consequence of linear algebra. In particular, apply Grahm-Schmidt on some basis of V to obtain an orthonormal basis of V. The isomorphism can be defined by sending each standard basis vector of $\mathbb{R}^n$ to a corresponding basis vector in the orthonormal basis we just found. □

[1] I need a better name for this
[2] Note here that specifying its behavior on the basis elements is sufficient, you would simply linearly extend these maps.

Another tensor one might be familiar with is the determinant. The determinant is a n tensor, so $\det \in \mathfrak{J}^n(\mathbb{R}^n)$, with the special property that is is *alternating*. Switching any 2 rows of a matrix changes the sign of the determinant. We now generalize this notion to other tensors.[3]

Definition 4.1.10 (Alternating tensor). A k-tensor $\omega \in \mathfrak{J}^k(V)$ is **alternating** if for all $v_1, \dots, v_k \in V$,

$$\omega(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = -\omega(v_1, \dots, v_j, \dots, v_i, \dots, v_k).$$

So switching any 2 arguments changes its sign.

Let us call the set of alternating k tensors on V by $\Omega^k(V)$. This is not only a set, it is a subspace of $\mathfrak{J}^k(V)$. Now, what lives in $\Omega^k(V)$? We know one element if $k = \dim V$, the determinant. But can we answer this in general?

Let us first begin by recalling that S_k is the symmetric group on the set $\{1, \dots, k\}$, and the sign of a permutation $\operatorname{sgn}\sigma$ is 1 if σ is even and -1 if σ is odd. We define a function $\operatorname{Alt}(T)$ that sends a k tensor $T \in \mathfrak{J}^k(V)$ to an alternating tensor, by the following:

$$\operatorname{Alt(T)}(v_1, \dots, v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}\sigma \cdot T(v_{\sigma(1)}, \dots, v_{\sigma(k)}).$$

The next proposition shows that the function we defined does actually take a tensor to an alternating tensor, it is a linear function and it does not disturb any alternating tensors.

Proposition 4.1.11.

1. *The map* Alt *is linear.*
2. *If T is a k tensor, then $\operatorname{Alt}(T)$ is alternating, so $\operatorname{Alt}(T) \in \Omega^k(V)$.*
3. *The map* Alt *is the identity on $\Omega^k(V)$, i.e. if $\theta \in \Omega^k(V), \operatorname{Alt}(\theta) = \theta$.*
4. *The map* Alt *is idempotent, so that $\operatorname{Alt}(\operatorname{Alt}(T)) = \operatorname{Alt}(T)$.*

Before we see the proof, let us first discuss how we shall approach it. For part 2, we would like to show that $\operatorname{Alt}(T)(v_1, \dots, v_j, \dots, v_i, \dots, v_k) = -\operatorname{Alt}(T)(v_1, \dots, v_i, \dots, v_j, \dots, v_k)$. Looking at the definition of Alt, and armed with some knowledge of group theory, we might have an idea of how to proceed. Since any coset of S_k is simply S_k, if we consider the transposition (i, j) which swaps around the i and j arguments, we can say say that every $\sigma' \in S_k$ can be written as $(i, j) \circ \sigma$ where $\sigma \in S_k$. Now we can basically do the proof

Proof. (1) is easy.

For (2), let (i, j) be the transposition, so (i, j) fixes everything except i and j. For $\sigma \in S_k$, we let $\sigma' = \sigma \circ (i, j)$ Thus,

$$\begin{aligned}
&\operatorname{Alt}(T)(v_1, \dots, v_j, \dots, v_i, \dots, v_k) \\
&= \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}\sigma \cdot T(v_{\sigma(1)}, \dots, v_{\sigma(j)}, \dots, v_{\sigma(i)}, \dots, v_{\sigma(k)}) \\
&= \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn}\sigma \cdot T(v_{\sigma'(1)}, \dots, v_{\sigma'(i)}, \dots, v_{\sigma'(j)}, \dots, v_{\sigma(k)})
\end{aligned}$$

Left multiplying every $\sigma \in S_k$ by (i, j) leaves the sum unchanged since any coset of S_k is S_k itself anyway. Additionally, notice that $\sigma'(i) = \sigma(j)$ and $\sigma'(j) = \sigma(i)$.

Continuing, we see that the above equation is equal to

$$\frac{1}{k!} \sum_{\sigma' \in S_k} -\operatorname{sgn}\sigma' \cdot T(v_{\sigma'(1)}, \dots, T_{v_{\sigma'(k)}}).$$

If σ was even, then σ' is odd. The same is true when swapping the words even and odd. This is why the negative sign appears. The result therefore follows. ☐

[3] Explain why we would want to do such a thing.

4.2 Vector fields and forms

Warning 4.2.1. The content ahead is extremely abstract!

> **Definition 4.2.2** (Tangent Space). For each point $p \in \mathbb{R}^n$, the **tangent space of** $\mathbb{R}^n$ **at** p is the set $\{p\} \times \mathbb{R}^n$, and it is denoted $T_p(\mathbb{R}^n)$ or $\mathbb{R}^n_p$.

Warning 4.2.3. This definition can be confusing, since $\mathbb{R}^n$ has a vector space structure on it. If instead we let $M = \mathbb{R}^n$, then we can write $T_p(M)$. We reserve the notation $\mathbb{R}^n_p$ specifically for when we are discussing the vector space structure of $\mathbb{R}^n_p$ itself. Additionally, we will make sure to keep vectors bold. The reason why we let $M = \mathbb{R}^n$ is because $\mathbb{R}^n$ is a smooth manifold, and it is customary to denote manifolds with M. The concepts here generalize quite easily to arbitrary smooth manifolds.

We make this $T_p(\mathbb{R}^n)$ into a vector space in an obvious way, namely by defining $(p, \mathbf{v}) + (p, \mathbf{w}) = (p, \mathbf{v} + \mathbf{w})$ and $a \cdot (p, \mathbf{v}) = (p, a\mathbf{v})$. Thus, $T_p(\mathbb{R}^n)$ is isomorphic to $\mathbb{R}^n$, when $\mathbb{R}^n$ is regarded as a vector space. Additionally, we notate a vector in $T_p(\mathbb{R}^n)$ by $\mathbf{v}_p$ or $\mathbf{v}|_p$ (we shall prefer the latter notation since it is less confusing). We shall also define the usual orientation of $T_p(\mathbb{R}^n)$ to be $[\mathbf{e}_1|_p, \ldots, \mathbf{e}_n|_p]$.

Since this assignment can be made for each point p, we can intuitively think about each *point* p as having the vector space $T_p(\mathbb{R}^n)$ attached to it. Additionally, since $T_p(\mathbb{R}^n)$ is isomorphic to $\mathbb{R}^n$ as a vector space, the set $\mathfrak{I}^k(T_p(\mathbb{R}^n))$, which is the space of k-tensors on $T_p(\mathbb{R}^n)$ is well defined and it is isomorphic to $\mathfrak{I}^k(\mathbb{R}^n)$.

> **Definition 4.2.4** (Vector field). Let $M = \mathbb{R}^n$. Let TM denote the disjoint union of all the tangent spaces at every point $p \in M$, i.e. $TM = \bigsqcup_{p \in M} T_p(M)$. Then a **vector field** is a function $F : M \to TM$ such that $F(p) \in T_p(M)$.

The reader is probably aware of some vector fields. If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then we can assign to each $p \in \mathbb{R}^n$, its derivative $Df(p)$, thought of as a vector. It would also be useful to be able to attach to each point $p \in M$, a tensor.

> **Definition 4.2.5** (Tensor field). Let $M = \mathbb{R}^n$. Let $\mathfrak{I}^k M$ denote the disjoint union of all k-tensor spaces of $\mathbb{R}^n_p$ at each point p of M, i.e. $\mathfrak{I}^k M = \bigsqcup_{p \in M} \mathfrak{I}^k(T_p(M))$. A **$k$-tensor field** is a function $F : M \to \mathfrak{I}^k M$ such that $F(p) \in \mathfrak{I}^k(T_p(M))$.

So, a tensor field is a function that assigns to each point p, a k-tensor. If a tensor field assigns every point to an alternating tensor, it is special and gets its own name, a differential form. We shall say that F *is a k-tensor field on* $\mathbb{R}^n$.

> **Definition 4.2.6** (Differential form). A k tensor field F for which $F(p) \in \Omega^k(T_p(\mathbb{R}^n))$ is a **differential form** or **k-form**. So a k tensor field assigns an alternating k-tensor to each point p.

Again, this is still extremely abstract, so we shall provide some examples. We shall assume all forms and vector fields are C^∞ if we say that they are differentiable.

Example 4.2.7 (The gradient vector field). Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Now, this means that for each point $p \in \mathbb{R}^n$, $Df(p)$ exists, and it is a linear functional. As such $Df(p) \in \Omega^1(\mathbb{R}^n)$. So we can kind of see how Df which assigns to each point $p \in \mathbb{R}^n$ the 1-tensor $Df(p)$ should be a differential form. However, $Df(p)$ puts out things in $\Omega^1(\mathbb{R}^n)$, and we need it to be a tensor field. Thus, let us define df, which assigns to each $p \in \mathbb{R}^n$, the 1-tensor $Df(p)(-)|_p$. So, we can write

$$df(p)(\mathbf{v}|_p) = Df(p)(\mathbf{v})|_p.$$

//

Example 4.2.8 (The derivative tensor field). Now let us consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$. To reduce confusion further down the line, let $M = \mathbb{R}^n$ and $N = \mathbb{R}^m$. Since f is differentiable, each point $p \in M$ comes with the linear transformation $Df(p) : \mathbb{R}^n \to \mathbb{R}^m$. Recalling from the previous example that Df is basically a vector field, we would now like to do the same here, making Df into a tensor field. As such, we need to make $Df(p)$ into a linear transformation of tangent spaces. In particular we would like it to go from $T_p(M)$ to $T_{f(p)}(N)$. So we shall *define* the

map $f_* : T_p(M) \to T_{f(p)}(N)$ by
$$f(\mathbf{v}|_p) = (Df(p)(\mathbf{v}))|_{f(p)}.$$

//

In the previous example, it happens to be the case that f_* is a linear map. As such, it induces a linear map of tensor spaces. In this case since $f_* : T_p(M) \to T_{f(p)}(N)$, the linear map it induces goes $\mathfrak{J}^k(T_{f(p)}(N)) \to \mathfrak{J}^k(T_p(M))$. We will only care about the linear map it induces between $\Omega^k(T_{f(p)}(N))$ and $\Omega^k(T_p(M))$. Thus we will focus on $(f_*)^* : \Omega^k(T_{f(p)}(N)) \to \Omega^k(T_p(M))$. We can now use $(f_*)^*$ to turn k forms on $T_{f(p)}(N)$ into k forms on $T_p(M)$. The reader may have already guessed how to define the next part. If η is an alternating k tensor on $T_{f(p)}(N)$, and we had k vectors $\mathbf{v}_1|_p, \ldots, \mathbf{v}_k|_p \in T_p(M)$, then by Definition 4.1.6, $(f_*)^*(\eta)(\mathbf{v}_1|_p, \ldots, \mathbf{v}_k|_p) = \eta(f_*(\mathbf{v}_1|_p), \ldots, f_*(\mathbf{v}_k|_p))$.

Let ω be a k-form on $T_{f(p)}(N)$. As a slight abuse of notation, we shall identify ω as a k-form on N instead. We then *define*[4] the map f^* from the k forms on N to the k forms on M. In particular, $f^*\omega$ is a k form on M, and it is defined by
$$(f^*\omega)(p) = (f_*)^*(\omega(f(p))).$$

[4]It is to be noted that the last line of [Spi18, p. 89] has a typo in the definition. The definition as given in these notes is the correct one.

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.

[Zam24] Asif Zaman. *Multivariable Calculus with Proofs*. 2024.