Math 302: Vector Analysis and Integration on Manifolds

Haim Grebnev

1 Introduction

- My name is Haim Grebnev, I'm a postdoctoral scholar. My job is to conduct research and to teach. My field of research is inverse problems with a focus on geometric analysis. Inverse problems is a field that studies math arising from various imaging techniques such as CT scans, sonar sensing, electric impedance tomography, etc. Currently I work on a generalization of the equations that arise in X-ray imaging that is used in polarimetric neutron tomography called the non-Abelian X-ray transform.

- In this course we will push the ideas of differential and integral calculus to multivariable (and sometimes multivalued) functions by proving all results from multivariable calculus. Even though we require single variable analysis as a prerequisite, we will need to redo the theory of differentiation and integration because surprisingly calculus in higher dimensions requires new ideas due to the appearance of obstacles not present in the single-variable theory. Nevertheless, we will be citing results and building off ideas from single-variable theory.

- Precisely, we expect you to have studied limits and Cauchy sequences in $\mathbb{R}^n$, continuity of functions of the form $f : \mathbb{R}^m \to \mathbb{R}^n$, and differential and integral calculus of single-variable single-valued functions $f : \mathbb{R} \to \mathbb{R}$. We do not expect this course to be heavy on metric space topology since we will mostly be working with the basic topology of $\mathbb{R}^n$, but we do expect you to have seen metric space topology at some point. We do expect you to have studied linear algebra because in several places we will make use of determinants, matrix multiplication, and eigenvalues/eigenvectors of symmetric matrices. It will be very helpful if you have taken multivariable calculus before so that this will not be your first time seeing results covered in this course.

- In the second part of the course, we will generalize even further by studying differentiation and integration on smooth submanifolds of Euclidean space (i.e. $\mathbb{R}^n$). In particular, we will end with the generalized Stoke's Theorem. To give you a preview, "smooth submanifolds" of Euclidean space are generalizations of surfaces to any dimensions. This is an important subject: in particular it provides a rigorous foundation for surface and curve integrals.

- Homework will be due every week (with some exceptions), most likely on Fridays at 11:59 p.m. You will submit homework via gradescope: you will have two penalty-free 24-hour extensions. Please rotate your homework properly in Gradescope and label the pages correctly to avoid losing points. Unless stated otherwise, everything in the homework must be proven rigorously. In the homework and exams, you can cite results from class or which were proven in the prerequisite courses. The ULA will have walk-in sessions, and I will have office hours. I'm open to suggestions, and the ULA is a great way to pass anonymous feedback to me.

- For most of the course we will be following the excellent textbook *Advanced Calculus 2^{nd} Ed* by Gerald Folland. You can get a free legal copy from the author's website at:

https://sites.math.washington.edu//~folland/AdvCalc24.pdf

You will be responsible for everything I cover in lecture. If at some point I see that we're running out of time, I may assign readings from the above book (I don't expect this to happen).

2 Differentiation in Several Variables

- We begin by defining differentiation in several variables. The following is *not* the definition of the derivative of a multivariable function, but it's an important and natural place to start.

- **Definition 2.1:** Suppose $U \subseteq \mathbb{R}^m$ is an open set and that $f : U \to \mathbb{R}$ is a function. Explicitly, f is of the form

$$f(x_1, \dots, x_m)$$

Take any $a \in U$, which we can explicitly write as $a = (a_1, \dots, a_m)$. We define the **partials (or partial derivatives) of f** as follows. For any $i = 1, \dots, m$, the i^{th} partial of f at a is

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a_1, \dots, a_i + h, \dots, a_m) - f(a_1, \dots, a_i, \dots, a_m)}{h}$$

$$= \lim_{x_i \to a_i} \frac{f(a_1, \dots, x_i, \dots, a_m) - f(a_1, \dots, a_i, \dots, a_m)}{x_i - a_i}$$

if the limits exist (they either both exist or both don't exist). In other words, set all of the variables except x_i to be equal the components of a and take the ordinary single-variable derivative of f in x_i at a_i if it exists. Obviously at every $a \in U$ there are m possible partials:

$$\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_m}(a)$$

if they exist.

- **Example 2.2:** Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x_1, x_2) = (x_1)^2 \sin(x_2)$ (in this case $U = \mathbb{R}^2$). Let us compute $\frac{\partial f}{\partial x_2}(2, 3)$ (in this case $a = (2, 3)$). We have that

$$\frac{\partial f}{\partial x_2}(2, 3) = \lim_{h \to 0} \frac{2^2 \sin(3 + h) - 2^2 \sin(3)}{h}.$$

But this limit is easy to compute because we observe that it is simply the single-variable derivative of the function $2^2 \sin(x_2)$ at $x_2 = 3$, which we know is $2^2 \cos(3)$ (alternatively, use L'Hôpital's rule). So

$$\frac{\partial f}{\partial x_2}(2, 3) = 2^2 \cos(3).$$

In reality we don't compute partials like this, but rather we compute the partials of f and then plug in $(2, 3)$ into the partials – see Example 2.5 below.

- **Note 2.3:** When working in $\mathbb{R}^2$ and $\mathbb{R}^3$, we will often not write (x_1, x_2) and (x_1, x_2, x_3) but rather (x, y) and (x, y, z) respectively. In that case, had we wrote the f in the previous example as $f(x, y) = x^2 \sin(y)$, then the answer would have been

$$\frac{\partial f}{\partial y}(2, 3) = 2^2 \cos(3).$$

- In single variable theory, when we had a function $f(x)$ we didn't only work with derivatives at preset points, such as $f'(2)$, $f'(3.6)$, etc. If the function was differentiable everywhere, we defined a new function $f'(x)$. We will do the same for partials, which we state next. Please note that the following is *not* yet the definition of a multivariable function – we'll get to that soon.

- **Definition 2.4:** Suppose that $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ is such that its partials exist everywhere (i.e. exist at every $a \in U$) where U is open. Then for every $i = 1, \dots, m$ we can form the function

$$\frac{\partial f}{\partial x_i} : U \to \mathbb{R}$$

where at every $x = (x_1, \dots, x_m) \in U$, $\frac{\partial f}{\partial x_i}(x_1, \dots, x_n)$ is the i^{th} partial of f at x.

- **Example 2.5:** Take our function $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 \sin(y)$. Then

$$\frac{\partial f}{\partial x}(x) = 2x \sin(y) \quad \text{and} \quad \frac{\partial f}{\partial y}(x) = x^2 \cos(y).$$

Hence again $\frac{\partial f}{\partial y}(2,3) = 2^2 \cos(3)$.

- **Remark 2.6:** You should be aware that there are other notations for the i^{th} partial of f, including:

$$\frac{\partial f}{\partial x_j}, \qquad \partial_{x_j} f, \qquad \partial_j f, \qquad f_{x_j}, \qquad f_j.$$

We will mostly use the first, and perhaps the next two as well. The last one is often used in differential geometry where calculations can get extremely long.

- **Notation 2.7:** For any vector $x \in \mathbb{R}^m$, we let $|x| = \sqrt{(x_1)^2 + \dots + (x_m)^2}$ where $x = (x_1, \dots, x_m)$.

- Next we discuss differentiability of multivariable functions. In single variable theory, the derivative represented the slope of the line that best approximates the behavior of the function (i.e. to "first order") which turned out to be a tangent line. In multiple variables we want to do the same thing. In this case, the graph of the function will be a surface and hence the best linear approximation will be a tangent plane! Unfortunately, simply the existence of partials is not enough for a good tangent plane to exist, which is illustrated by the following example. Let

$$f(x,y) = \frac{xy}{x^2+y^2}.$$

As an exercise, try to show that both $\frac{\partial f}{\partial x}(0,0) = 0$ and $\frac{\partial f}{\partial y}(0,0) = 0$. This would indicate that the tangent plane should be flat, however if you plot the function, you'll see that this is a terrible approximation to the function.

- So let's try to guess what the correct definition of differentiability of a multivariable function should be. We could try

$$\lim_{h \to 0} \frac{f(a+h) - (a)}{h}$$

where now a and h are vectors. But this is ridiculous since we're trying to divide a vector by a vector, which is not a well-defined operation. We could try

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

Unfortunately, this limit won't exist for most functions. One way to see this is that the limit will be different as h approaches zero from different directions. In particular, if you let h approach 0 along the x-axis from the right, you will get $\frac{\partial f}{\partial x}(a)$. On the other hand, if you let h approach 0 along the x-axis from the left, you will get $-\frac{\partial f}{\partial x}(a)$.

So what is the derivative of f? When f was a function of a single variable: $f(x)$, we defined its derivative at a as the (unique) number m given by

$$m = \lim_{x \to a} \frac{f(x) - f(a)}{x-a}.$$

This statement is equivalent to

$$0 = \lim_{x \to a} \left(\frac{f(x) - f(a)}{x-a} - m \right) = \lim_{x \to a} \frac{f(x) - f(a) - m(x-a)}{x-a}$$
$$= \lim_{x \to a} \frac{f(x) - [m(x-a) + f(a)]}{x-a}.$$

Notice that $m(x-a) + f(a)$ is the equation for the tangent line to $f(x)$ at $\big(a, f(a)\big)$. Let's do the same thing for a multivariable function!

For simplicity consider a two-variable function $f(x_1, x_2)$ whose graph is a 2D surface in $\mathbb{R}^3$. An equation for a plane in $\mathbb{R}^3$ that passes through (a_1, a_2, b) is $x_3 = c_1(x_1 - a_1) + c_2(x_2 - a_2) + b$. Hence the analog of the above limit statement is

$$0 = \lim_{x \to a} \frac{f(x_1, x_2) - [c_1(x_1 - a_1) + c_2(x_2 - a_2) + f(a_1, a_2)]}{|x-a|}.$$

Notice that a shorter way to write this is

$$0 = \lim_{x \to a} \frac{f(x) - [c \cdot (x-a) + f(a)]}{|x-a|}$$

where " $\cdot$ " here is the dot product and $c = (c_1, c_2)$, $x = (x_1, x_2)$, and $a = (a_1, a_2)$. This way $x_3 = c \cdot (x-a) + f(a)$ should be the equation for the tangent plane to the graph of $f(x)$ at $\big(a_1, a_2, f(a_1, a_2)\big)$ in $\mathbb{R}^3$. This is precisely how differentiability is defined:

- **Definition 2.8:** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ where U is open. Fix a point $a \in U$. We say that $\boldsymbol{f}$ **is differentiable at** $\boldsymbol{a}$ if there exists a vector $c \in \mathbb{R}^m$ such that

$$(2.9) \qquad 0 = \lim_{x \to a} \frac{f(x) - [c \cdot (x-a) + f(a)]}{|x-a|} = \lim_{h \to 0} \frac{f(a+h) - [c \cdot h + f(a)]}{|h|}.$$

If such a $c \in \mathbb{R}^m$ exists, then it is unique (proved in the next theorem). In this case c is called the **gradient of** $\boldsymbol{f}$ **at** $\boldsymbol{a}$ and is denote by $\nabla f(a) \in \mathbb{R}^m$.

If the gradient exists for all $a \in U$, then we say that f is differentiable everywhere. In this case, we get a function $\nabla f : U \to \mathbb{R}^m$ (i.e. the function $\nabla f(x)$).

Remark: In the case when $m = 1$ (i.e. f is a function of one variable), $f'(x) = \nabla f(x)$. In this case we typically use $f'(x)$ instead of $\nabla f(x)$. Also, we'll explain later where the term "gradient" comes from.

- **Theorem 2.10:** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$, where U is open, and that f is differentiable at $a \in U$. Then the partials of f exist at a and the gradient of f at a is given by

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_m}(a) \right).$$

In particular, the gradient is unique. If f is differentiable everywhere, then clearly

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_m} \right)$$

everywhere in U.

Proof: Let c be as in (2.9). We simply need to show that the i^{th} component of c is $\frac{\partial f}{\partial x_i}(a)$. Fix an index $i = 1, \dots, m$. The limit in (2.9) does not depend on the direction from which x approach a. So let x approach a along the i^{th} axis: let $x = (a_1, \dots, x_i, \dots, a_m)$ where $x_i \to a_i$. In that case, in (2.9)

$$c \cdot (x-a) = (c_1, \dots, c_m) \underbrace{\big((a_1, \dots, x_i, \dots, a_m) - (a_1, \dots, a_i, \dots, a_m)\big)}_{(0,\dots,0,x_i - a_i,0,\dots,0)} = c_i(x_i - a_i).$$

and $\quad |x-a| = \sqrt{0^2 + \cdots + 0^2 + (x_i - a_i)^2 + 0^2 + \cdots + 0^2} = |x_i - a_i| = \pm 1 \cdot (x_i - a_i).$

Hence (2.9) becomes

$$0 = \pm 1 \cdot \lim_{x_i \to a_i} \frac{f(a_1, \dots, x_i, \dots, a_m) - [c_i(x_i - a_i) + f(a_1, \dots, a_i, \dots, a_m)]}{x_i - a_i}$$

$$\text{Algebraic manipulation} \quad \Longrightarrow \quad c_i = \lim_{x_i \to a_i} \frac{f(a_1, \dots, x_i, \dots, a_m) - f(a_1, \dots, a_i, \dots, a_m)}{x_i - a_i}.$$

Hence we get two things: the last limit implies that the partial $\frac{\partial f}{\partial x_i}(a)$ indeed exists and that it is equal to c_i.

■

- It follows from the above theorem and our previous discussion that the equation for the tangent plane to the graph of a differentiable function f at $\big(a, f(a)\big)$ is

$$x_{m+1} = \nabla f(a) \cdot (x - a) + f(a).$$

For a function of two variables $f(x, y)$, this reduces to

$$z = \frac{\partial f}{\partial x}(a)(x_1 - a_1) + \frac{\partial f}{\partial y}(a)(x_1 - a_2) + f(a).$$

- **Note 2.11:** Suppose $f : U \to \mathbb{R}$ is differentiable at a as above. By plugging $c = \nabla f(a)$ into (2.9) we get that

$$(2.12) \quad 0 = \lim_{x \to a} \frac{f(x) - [\nabla f(a) \cdot (x - a) + f(a)]}{|x - a|} = \lim_{h \to 0} \frac{f(a + h) - [\nabla f(a) \cdot h + f(a)]}{|h|}$$

By letting $E_a(h)$ denote the numerator in the second limit, it directly follows that

$$(2.13) \qquad f(a + h) = f(a) + \nabla f(a) \cdot h + E_a(h)$$

where E_a satisfies

$$(2.14) \qquad \lim_{h \to 0} \frac{E_a(h)}{|h|} = 0$$

Plugging in $h = x - a$, this takes the equivalent form

$$f(x) = f(a) + \nabla f(a) \cdot (x - a) + E_a(x - a).$$

In other words E_a is an "error function" that measures how well the tangent plane $f(a) + \nabla f(a) \cdot (x - a)$ approximates $f(x)$ near a. It's such a good approximation that the error decays faster than $|h|$ by (2.14). As we'll see later, this is a special case of a Taylor's expansion.

- **Corollary 2.15:** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$, where U is open, and that f is differentiable at $a \in U$. Then f is continuous at a.

Clearly it follows then that if f is differentiable everywhere, then it is continuous everywhere.

Proof: We only need to prove the first statement. We have that

$$\lim_{x\to a}\big(f(x)-f(a)\big)=\lim_{x\to a}(f(x)-[\nabla f(a)\cdot(x-a)+f(a)])$$
$$=\lim_{x\to a}|x-a|\lim_{x\to a}\frac{f(x)-[\nabla f(a)\cdot(x-a)+f(a)]}{|x-a|}=0.$$

Hence indeed $f(a)=\lim\limits_{x\to a}f(x)$ (i.e. f is continuous at a).

■

- We've defined differentiability for multivariable functions, but we have no way of demonstrating that any function is differentiable. The following theorem is a popular way to do this:
- **Theorem 2.16:** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ where U is open. Suppose all partials of f exist on some ball $B \subseteq U$ centered at $a \in U$. Suppose also that all partials $\frac{\partial f}{\partial x_i}$ are continuous at a. Then f is differentiable at a.

  Clearly if all partials of f exist and are continuous everywhere in U, then f is differentiable everywhere in U.

  Proof: Let us suppose $m=2$ for simplicity of the notation: we'll come back to the general case. We want to show that

$$\lim_{h\to 0}\frac{f(a+h)-\left(\frac{\partial f}{\partial x_1}(a),\frac{\partial f}{\partial x_2}(a)\right)\cdot h-f(a)}{|h|}=0 \tag{2.17}$$

(we removed the unnecessary square brackets "[...]" in (2.12)). Let us take a look at $f(a+h)-f(a)$. We make sure that h is small enough so that $a+h$ is still inside B. Writing $a=(a_1,a_2)$ and $h=(h_1,h_2)$, this is given by

$$f(a_1+h_1,a_2+h_2)-f(a_1,a_2)$$
$$=f(a_1+h_1,a_2+h_2)\underbrace{-f(a_1,a_2+h_2)+f(a_1,a_2+h_2)}_{\text{added zero}}-f(a_1,a_2).$$

By the mean value theorem

$$\frac{f(a_1+h_1,a_2+h_2)-f(a_1,a_2+h_2)}{h_1}=\frac{\partial f}{\partial x_1}\Big(\underbrace{c_{1,h_1,h_2}},a_2+h_2\Big).$$
$$\frac{f(a_1,a_2+h_2)-f(a_1,a_2)}{h_2}=\frac{\partial f}{\partial x_2}\Big(a_1,\underbrace{c_{2,h_2}}\Big).$$

for some c_{1,h_1,h_2} between a_1 and a_1+h_1, and some c_{2,h_2} between a_2 and a_2+h_2. Thus the quantity inside the limit in (2.17) is given by

$$\frac{\frac{\partial f}{\partial x_1}(c_{1,h_1,h_2}, a_2+h_2)h_1 + \frac{\partial f}{\partial x_2}(a_1, a_2+c_{2,h_2})h_2 - \frac{\partial f}{\partial x_1}(a_1,a_2)h_1 - \frac{\partial f}{\partial x_2}(a_1,a_2)h_2}{|h|}$$

$$= \left(\frac{\partial f}{\partial x_1}(c_{1,h_1,h_2}, a_2+h_2) - \frac{\partial f}{\partial x_1}(a_1,a_2)\right)\frac{h_1}{|h|} + \left(\frac{\partial f}{\partial x_2}(a_1, a_2+c_{2,h_2}) - \frac{\partial f}{\partial x_2}(a_1,a_2)\right)\frac{h_2}{|h|}.$$

Since $\frac{\partial f}{\partial x_1}$ and $\frac{\partial f}{\partial x_2}$ are continuous at a, the last two "$(\dots)$" quantities go to zero since h_1, h_2, c_{1,h_1,h_2}, and c_{2,h_2} all go to zero as $h \to 0$. And the quantities $\frac{h_1}{|h|}$ and $\frac{h_2}{|h|}$ are bounded in size by 1. So by the squeeze theorem, the above quantity goes to zero. Hence we proved (2.17).

For functions of more variables $f(x_1, \dots, x_m)$, this is done similarly but at the beginning of the proof you instead do

$$f(a_1+h_1, a_2+h_2, \dots, a_m+h_m) \underbrace{-f(a_1, a_2+h_2, \dots, a_m+h_m) + f(a_1, a_2+h_2, \dots, a_m+h_m)}_{\text{added zero}}$$

$$\underbrace{-f(a_1, a_2, a_3+h_3, \dots, a_m+h_m) + f(a_1, a_2, a_3+h_3, \dots, a_m+h_m)}_{\text{added zero}} - \cdots - f(a_1, a_2, \dots, a_m).$$

■

- **Definition 2.18:** For any open set $U \subseteq \mathbb{R}^n$ we let $C^1(U)$ denote the set of functions f such that all of their first partial derivatives exists and are continuous.

- **Remark 2.19:** A good visualization/summary of Theorem 2.10 and Theorem 2.16 is the following diagram:

$$f \in C^1(U) \quad \Longrightarrow \quad f \text{ is differentiable} \quad \Longrightarrow \quad \text{each partial } \frac{\partial f}{\partial x_i} \text{ exists.}$$

  We point out that when working with usual functions, it's almost always obvious that a function is in C^1.

- **Note 2.20:** Derivatives describe rates of change of a function and hence describe well the changes of a function on small scales. This can be encoded in the concept of the **differential**. At the moment the following discussion is not rigorous, but later in the course we'll study how to turn this into a rigorous concept once we get to rank-1 tensors.

  Let's look at differentiable functions $f : \mathbb{R}^2 \to \mathbb{R}$ for simplicity. From (2.13) we have that

$$f(a+h) - f(a) = \nabla f(a) \cdot h + E_a(h)$$

  where $E_a(h)$ is an "error term" that goes to zero really fast: faster than $|h|$ - see (2.14). The vector h denotes the "step away" from a, and so let us write this as $h = (dx, dy)$ where "dx" and "dy" denote small changes in x and y respectively. Let us also denote the small change in f on the left-hand side of the above equation as df. Since $E_a(h)$ becomes negligible in size to dx

and dy on small scales, on the "differential level" the above equation gives the equality (the partials of f are being evaluated at a).

$$df = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \cdot (dx, dy)$$

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy.$$

Thus this equation gives a good way to measure changes in f for small changes in x and y. For functions of more variables $f(x_1, \dots, x_m)$ this takes the form

$$(2.21) \qquad df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_m} dx_n.$$

- A step towards making the concept of differentials rigorous is the notion of directional derivatives which asks the following natural question. The partial derivatives $\frac{\partial f}{\partial x_i}$ gives us information on the rate of change of the function along each x_i-axis. But how does one compute the rate of change of a function in a diagonal direction that doesn't necessarily lie on any axis?

- **Definition 2.22:** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ where U is open. Take any point $a \in U$ and any unit vector $u \in \mathbb{R}^m$ (recall "unit vector" means "vector of length one"). Consider the line $l(t) = a + tu$ which goes through a at $t = 0$ with velocity/direction u. The **directional derivative** of f at a in the direction u, denote by $\partial_u f(a)$ or $\frac{\partial f}{\partial u}(a)$, is defined as

$$(2.23) \qquad \partial_u f(a) = \frac{d}{dt}\big(f \circ l(t)\big)\Big|_{t=0} = \frac{d}{dt}\big(f(a + tu)\big)\Big|_{t=0} = \lim_{s \to 0} \frac{f(a + su) - f(a)}{s}$$

if the limit exists.

- It turns out that there is a very simple equation for the directional derivative when the function is differentiable:

- **Theorem 2.24:** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$, where U is open, and that f is differentiable at a. Then all directional derivatives of f exist at a and for any unit vector $u \in \mathbb{R}^m$ it is given by

$$\partial_u f(a) = \nabla f(a) \cdot u.$$

Proof: Fix a unit vector $u \in \mathbb{R}^m$. Since f is differentiable at a, recall from (2.13) that

$$f(a + h) = f(a) + \nabla f(a) \cdot h + E_a(h) \quad \text{where} \quad \lim_{h \to 0} \frac{E_a(h)}{|h|} = 0$$

(keep in mind that h is a vector). Plugging this into the definition of directional derivative (2.23) gives that

$$\partial_u f(a) = \lim_{s \to 0} \frac{f(a + su) - f(a)}{s} = \lim_{s \to 0} \frac{\nabla f(a) \cdot (su) + E_a(su)}{s} = \lim_{s \to 0} \left(\frac{s \nabla f(a) \cdot u}{s} + \frac{E_a(su)}{s}\right)$$

$$= \nabla f(a) \cdot u + \lim_{s \to 0} \frac{E_a(su)}{|su|}.$$

The last limit is the limit $\lim_{h \to 0} E_a(h)/|h|$ but along the direction u and hence is also zero. Thus we indeed get that $\partial_u f(a) = \nabla f(a) \cdot u$.

■

- Next we discuss the chain rule, which takes on a slightly more complicated form in higher dimensions compared to its single variable version. We begin with the easier case when the inside function only depends on one variable.

- **Theorem 2.25:** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ where U is open and $g : I \subseteq \mathbb{R} \to U \subseteq \mathbb{R}^m$ where I is open. As usual we explicitly write these as $f(x_1, \dots, x_m)$ and $g(t) = \big(g_1(t), \dots, g_m(t)\big)$. Consider the function $\varphi = f \circ g : I \to \mathbb{R}$.

  Suppose that g is differentiable at $a \in I$ (i.e. every component g_k is differentiable at a) and that f is differentiable at $g(a)$. Then φ is differentiable at a and

$$\frac{d\varphi}{dt}(a) = \nabla f\big(g(a)\big) \cdot g'(a) = \frac{\partial f}{\partial x_1}\big(g(a)\big)\frac{dg_1}{dt}(a) + \cdots + \frac{\partial f}{\partial x_m}\big(g(a)\big)\frac{dg_m}{dt}(a).$$

  Clearly if f and g are differentiable everywhere, then φ is differentiable everywhere over I and

$$\frac{d\varphi}{dt} = \nabla f \cdot g' = \frac{\partial f}{\partial x_1}\frac{dg_1}{dt} + \cdots + \frac{\partial f}{\partial x_m}\frac{dg_m}{dt} \tag{2.26}$$

  (important: each $\frac{\partial f}{\partial x_i}$ is evaluated at $g(t)$).

  Remark: Sometimes people instead write $f(t) = f \circ g(t)$ and $g(t) = \big(x_1(t), \dots, x_m(t)\big)$ and so the above becomes

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_m}\frac{dx_m}{dt}.$$

  Notice that if you <u>not rigorously</u> multiply through by dt you recover the equation satisfied by differentials: $df = \frac{\partial f}{\partial x_1}dx_1 + \cdots + \frac{\partial f}{\partial x_m}dx_n$.

  Proof: For shorthand, let $b = g(a)$. We have that

$$\frac{d\varphi}{dt}(a) = \lim_{s \to 0} \frac{\varphi(a+s) - \varphi(a)}{s} = \lim_{s \to 0} \frac{f\big(g(a+s)\big) - f\left(\overbrace{g(a)}^{b}\right)}{s}. \tag{2.27}$$

  We have to show that this is equal to $\nabla f\big(g(a)\big) \cdot g'(a)$. Because f and g are differentiable at b and a respectively

$$(2.28)\qquad f(x) - f(b) = \nabla f(b) \cdot (x-b) + E_{b,f}(x-b) \text{ with } \lim_{h\to 0} \frac{E_{b,f}(h)}{|h|},$$

$$(2.29)\quad \underbrace{\begin{bmatrix} g_1(a+s) \\ \vdots \\ g_m(a+s) \end{bmatrix}}_{g(a+s)} - \underbrace{\begin{bmatrix} g_1(a) \\ \vdots \\ g_m(a) \end{bmatrix}}_{g(a)} = \underbrace{\begin{bmatrix} g_1'(a)s \\ \vdots \\ g_m'(a)s \end{bmatrix}}_{sg'(a)} + \underbrace{\begin{bmatrix} E_{a,g,1}(s) \\ \vdots \\ E_{a,g,m}(s) \end{bmatrix}}_{E_{a,g}(s)} \text{ with each } \lim_{s\to 0} \frac{E_{a,g,k}(s)}{s} = 0$$

$$\Longrightarrow \quad \lim_{s\to 0} \frac{E_{a,g}(s)}{s}.$$

Set both $E_{b,f}(0) = 0$ and $E_{a,g}(0) = 0$. Plugging $x = g(a+s)$ in (2.28) (and recalling $b = g(a)$) and then plugging that into the last quantity in (2.27) gives (we drop writing "$\lim_{s\to 0}$" for now)

$$\frac{\nabla f(b) \cdot \big(g(a+s) - g(a)\big) + E_{b,f}\big(g(a+s) - g(a)\big)}{s}.$$

Plugging in (2.29) into this gives

$$\frac{\nabla f(b) \cdot \left(sg'(a) + E_{a,g}(s)\right) + E_{b,f}\left(sg'(a) + E_{a,g}(s)\right)}{s}$$

$$(2.30)\qquad = \nabla f(b) \cdot g'(a) + \nabla f(b) \cdot \frac{E_{a,g}(s)}{s} + \frac{E_{b,f}\left(sg'(a) + E_{a,g}(s)\right)}{s}.$$

The second term in (2.30) goes to zero as $s \to 0$ by (2.29). The third term is bounded in absolute value by

$$(2.31)\quad \begin{cases} \dfrac{E_{b,f}\left(sg'(a) + E_{a,g}(s)\right)}{\left|sg'(a) + E_{a,g}(s)\right|} \underbrace{\dfrac{\overbrace{\left|sg'(a) + E_{a,g}(s)\right|}^{\text{numerator} \le s|g'(a)| + |E_{a,g}(s)|}}{s}}_{Q} & \text{if } sg'(a) + E_{a,g}(s) \neq 0 \\ 0 & \text{if } sg'(a) + E_{a,g}(s) = 0 \end{cases}$$

We have that $sg'(a) + E_{a,g}(s) \to 0$ as $s \to 0$ and so the quantity $E_{b,f}(\ldots)/|\ldots|$ in the first case goes to zero. The fraction labeled "Q" is bounded in size since $s|g'(a)|/s = |g'(a)|$ and $\left|E_{a,g}(s)\right|/s$ goes to zero as $s \to 0$. So by the squeeze theorem (2.31) goes to zero as $s \to 0$. Tracing the above logic back, this shows that indeed the limit in (2.27) is equal to $\nabla f\big(g(a)\big) \cdot g'(a)$.

∎

- **Corollary 2.32:** Suppose the same situation as the previous theorem, but instead suppose that $f \in C^1(U)$ and $g \in C^1(I)$. Then $\varphi = f \circ g$ is also in $C^1(I)$.

Proof: This follows immediately from (2.26) since all $\frac{\partial f}{\partial x^i}$ and $\frac{dg_i}{dt}$ are continuous. ■

- The above theorem and its proof generalizes directly to the case when g depends on multiple variables:

- **Theorem 2.33:** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ where U is open and $g : V \subseteq \mathbb{R}^k \to U \subseteq \mathbb{R}^m$ where V is open. As usual we explicitly write these as $f(x_1, \dots, x_m)$ and $g(t_1, \dots, t_k) = \big(g_1(t_1, \dots, t_k), \dots, g_m(t_1, \dots, t_k)\big)$. Consider the function $\varphi = f \circ g : V \to \mathbb{R}$.

  Suppose that g is differentiable at $a \in V$ and that f is differentiable at $g(a)$. Then φ is differentiable at a and each partial

$$\frac{d\varphi}{dt_i}(a) = \frac{\partial f}{\partial x_1}\big(g(a)\big)\frac{dg_1}{dt_i}(a) + \cdots + \frac{\partial f}{\partial x_m}\big(g(a)\big)\frac{dg_m}{dt_i}(a).$$

  Clearly if f and g are differentiable everywhere, then φ is differentiable everywhere over I and

$$\frac{d\varphi}{dt_i} = \frac{\partial f}{\partial x_1}\frac{dg_1}{dt_i} + \cdots + \frac{\partial f}{\partial x_m}\frac{dg_m}{dt_i}.$$

  Remark: Sometimes people instead write $f(t_1, \dots, t_k) = f \circ g(t_1, \dots, t_k)$ and $g(t_1, \dots, t_k) = \big(x_1(t_1, \dots, t_k), \dots, x_m(t_1, \dots, t_k)\big)$ and so the above becomes

$$\frac{df}{dt_i} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt_i} + \cdots + \frac{\partial f}{\partial x_m}\frac{dx_m}{dt_i}.$$

  Proof: The proof is very similar to the proof of the previous theorem: we leave it to the interested reader to figure out what changes are needed in the proof. ■

- **Corollary 2.34:** Suppose the same situation as the previous theorem, but instead suppose that $f \in C^1(U)$ and $g \in C^1(V)$. Then $\varphi = f \circ g$ is also in $C^1(V)$.

- We mention an interpretation of gradients in the context of topographic maps. Suppose the surface of a mountain is represented by the graph of a function $z = f(x, y)$. In the x, y plane we can draw contours representing sets on which f has the same value: they are called "isolines." Suppose such a contour is parametrized by a curve $g : I \subseteq \mathbb{R} \to \mathbb{R}^2$. Then $\varphi = f \circ g$ is constant and so for all $a \in I$,

$$\frac{d\varphi}{dt}(a) = 0 \quad \text{and so} \quad \nabla f\big(g(a)\big) \cdot g'(a) = 0.$$

  We have that $g'(a)$ is tangent to the contour, and hence the second equation says that ∇f is always perpendicular to the isolines.

- The next important theorem that we generalize from single variable theory is the mean value theorem which, as we'll see right after, often plays the role of giving quantitative estimates on how fast functions can grow. First we need the notion of a line segment:

- **Definition 2.35:** For any points $a, b \in \mathbb{R}^m$, the **line segment** from a to b is the curve $l : [0, 1] \to \mathbb{R}^m$ given by

$$l(t) = a + t(b - a).$$

The image $L = \operatorname{Im} l$ is also referred to as the "line segment." Sometimes we will also consider other parametrizations of the same curve $l : [\alpha, \beta] \to \mathbb{R}^m$ (i.e. obtained by a change of variables).

- **Theorem 2.36 (Mean Value Theorem):** Suppose we have $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ where U is open. Let $a, b \in U$ be such that the line segment L between them lies in U and such that f is continuous on L and differentiable at every point of L except possibly at the endpoints a and b. Then there exists a $c \in L$ not equal to the endpoints a nor b such that

$$f(b) - f(a) = \nabla f(c) \cdot (b - a). \tag{2.37}$$

Proof: Take the line segment $l : [0, 1] \to \mathbb{R}^m$ going from a to b:

$$l(t) = a + t(b - a)$$

and form the function $\varphi(t) = f \circ l(t)$ (note that $\varphi : [0, 1] \to \mathbb{R}$). By the single variable mean value theorem there exists some $\tilde{c} \in (0, 1)$ such that

$$\frac{\varphi(1) - \varphi(0)}{1 - 0} = \varphi'(c) = \nabla f\big(l(\tilde{c})\big) \cdot l'(\tilde{c}). \tag{2.38}$$

Notice that

$$\varphi(1) = f(b), \quad \varphi(0) = f(a), \quad l'(c) = b - a, \tag{2.39}$$

and so setting $c = l(\tilde{c})$ and plugging (2.39) into (2.38) gives (2.37).

$\blacksquare$

- The following definition is useful in many fields of math, in particular in optimization.

- **Definition 2.40:** A set $S \subseteq \mathbb{R}^m$ is called **convex** if for any two points $a, b \in S$ the line segment L between them is contained in S: $L \subseteq S$.

- **Example 2.41:** As you will prove in the homework, any ball $B \subseteq \mathbb{R}^m$ is convex.

- **Corollary 2.42:** Suppose that $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ is differentiable, where U is open and convex, and that $|\nabla f| \leq M$ everywhere. Then for any $a, b \in U$,

$$|f(a) - f(b)| \leq M|b - a|.$$

Remark: The assumption "convex" is necessary since otherwise a slowly rising spiraling staircase gives a counterexample.

Proof: Take any $a, b \in U$ and the line segment L between them. By Theorem 2.36 there exists a $c \in L$ such that $f(b) - f(a) = \nabla f(c) \cdot (b - a)$. Then (the first "≤" below uses the Cauchy-Schwarz inequality)

$$|f(b) - f(a)| = |\nabla f(c) \cdot (b-a)| \leq |\nabla f(c)||(b-a)| \leq M|(b-a)|.$$

■

- **Corollary 2.43:** Suppose that $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ is differentiable, where U is open and convex, and that $\nabla f = 0$ everywhere. Then f is constant on U.

  Proof: The condition $\nabla f = 0$ is trivially equivalent to $|\nabla f| \leq 0$. So the previous corollary implies that for any $a, b \in U$

  $$|f(b) - f(a)| \leq 0|b-a| = 0 \quad \Longrightarrow \quad f(b) = f(a).$$

  In other words, the value of f is the same at any two points and hence must be constant.

  ■

- We now improve the above corollary to the case when U is connected rather than simply convex. Thus the following theorem will make the above corollary obsolete (though we will use the above corollary to prove the following theorem). First we need a topological lemma, which is essentially trivial.

- **Lemma 2.44:** Suppose that $U \subseteq \mathbb{R}^m$ is open and that $\widetilde{U} \subseteq U$ is open in the metric topology of U. Then $\widetilde{U}$ is open in $\mathbb{R}^m$.

  Proof: Take any point $a \in \widetilde{U}$. Since $\widetilde{U}$ is open in U, there is a ball $B \subseteq \widetilde{U}$ centered at a. Since B is also a ball in $\mathbb{R}^m$ this shows that a is an interior point of $\widetilde{U}$ where $\widetilde{U}$ is thought of as a subset of $\mathbb{R}^m$. Hence $\widetilde{U}$ is open in $\mathbb{R}^m$ as well.

  ■

- **Theorem 2.45:** Suppose that $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ is differentiable, where U is open and connected, and that $\nabla f = 0$ everywhere. Then f is constant on U.

  Proof: Take any $a \in U$ and consider the sets

  $$U_1 = \{x \in U : f(x) = f(a)\} \neq \emptyset \quad \text{(nonempty since } a \in U_1\text{)}$$
  $$\text{and} \quad U_2 = \{x \in U : f(x) \neq f(a)\} = f^{-1}[(-\infty, b) \cup (b, \infty)].$$

  Note that $U = U_1 \cup U_2$. Note also that U_2 is open in U, and hence in $\mathbb{R}^m$ by Lemma 2.44, since it is the preimage of an open set by a continuous function. The set U_1 is also open since if you take any $b \in U_1$ (i.e. $f(b) = f(a)$) you can draw a ball $B \subseteq U$ centered at b on which f is constant by Corollary 2.43 (i.e. f is equal to $f(a)$ in B) and thus $B \subseteq U_1$. Since both U_1 and U_2 are open, we have that

  $$\overline{U_1} \cap U_2 = \emptyset \quad \text{and} \quad U_1 \cap \overline{U_2} = \emptyset$$

  (exercise!). Thus U_1 and U_2 will form a disconnection of U unless one of them is empty. Since U is connected and $U_1 \neq \emptyset$, we must have that $U_2 = \emptyset$. Hence $U_1 = U$ and so f is constantly equal to $f(a)$ everywhere.

■

- So far we've been taking only one derivative of a multivariable function, which describes how the graph of the function moves through space in the "directional sense" accurately described by the tangent plane. If we take two derivatives, then we get information on how the graph of the function "curves" through space.

- **Note 2.46:** Suppose we have a multivariable function $f(x_1, \dots, x_m)$. If the i^{th} partial of f exists everywhere (i.e. $\frac{\partial f}{\partial x_i}$), then $\frac{\partial f}{\partial x_i}(x_1, \dots, x_m)$ defines a new multivariable function. Thus, if possible, we may take another partial derivative of this new function, say the j^{th} partial:
$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) \quad \text{which is also denoted by} \quad \frac{\partial^2 f}{\partial x_j \partial x_i}, \;\; \partial_{x_j}\partial_{x_i} f, \;\; \partial_j \partial_i f, \;\; f_{x_j x_i}, \;\; f_{ji}.$$
These are called **second order partial derivatives**. Higher order partials:
$$\frac{\partial^k f}{\partial x_{i_1} \dots x_{i_k}} \quad \left(\text{a } k^{\text{th}} \text{ order partial derivative of } f\right)$$
are defined similarly. Note that we always read the order of differentiation right to left. When you take the partials only in one variable, this is often called **pure partials**. When you take partials in various variables, this is called **mixed partials**:
$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial^2 f}{\partial x_i \partial x_i} \text{ is a pure partial} \quad \text{and} \quad \frac{\partial^2 f}{\partial x_i \partial x_j} \text{ is a mixed partials when } i \neq j.$$

- A natural question is whether mixed partials depend on the order of differentiation:
$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}?$$
A very surprising theorem of calculus says that for most "nice" functions the order of differentiation does not matter: we will prove this in Theorem 2.57 below. However there are exceptional function when this is not the case. The book gives the example of the function
$$f(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2}$$
where we define $f(0,0) = 0$. A routine calculation (which is a good exercise) shows that for this function
$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1 \neq -1 = \frac{\partial^2 f}{\partial y \partial x}(0,0).$$

- **Definition 2.47:** For any open set $U \subseteq \mathbb{R}^m$, $C^k(U)$ denotes the set of all functions of the form $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ such that all partials of f of order less than or equal to k exist and are continuous. By convention $C^0(U)$ simply denotes the set of all continuous functions of the form $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ (i.e. no need to take partial derivatives).

We let $C^\infty(U)$ denote the set of all functions of the form $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ such that partials of all orders of f exist and are continuous (a short and interesting exercise is to show that $C^\infty(U) = \bigcap_{k=0}^{\infty} C^k(U)$).

- **Remark 2.48:** Notice that $C^k(U) \subseteq C^j(U)$ if $k \geq j$ since if $f \in C^k(U)$ is k times differentiable and all of its partials up to order k are continuous, then this definitely holds if you only consider all partials of order only up to j (i.e. thus $f \in C^j(U)$).

- **Theorem 2.49:** Suppose we have a function $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$, where U is open, which we write out explicitly as

$$f(x_1, \dots, x_m).$$

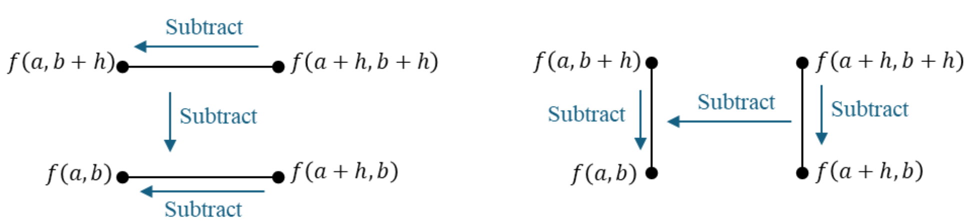
Let $i, j \in \{1, \dots, m\}$ be indices such that the partials $\partial_i f$, $\partial_j f$, $\partial_i \partial_j f$, and $\partial_j \partial_i f$ all exist (on U). Suppose also that $\partial_i \partial_j f$ and $\partial_j \partial_i f$ are continuous at a point $a \in U$. Then

$$\partial_i \partial_j f(a) = \partial_j \partial_i f(a). \tag{2.50}$$

Proof: If $i = j$ then (2.50) is obvious. So suppose that $i \neq j$. Since we're only studying the partials of f with respect to two variables, we can suppose that f is of the form $f(x, y)$ and that we want to prove that

$$\frac{\partial^2 f}{\partial x \partial y}(a, b) = \frac{\partial^2 f}{\partial y \partial x}(a, b) \tag{2.51}$$

at a point $(a, b) \in U$ (we changed the meaning of a). Intuitively $\partial_x \partial_y f$ studies how much the changes of f in the y direction change in the x direction. Similarly $\partial_y \partial_x f$ studies how much the changes of f in the x direction change in the y direction. If we change both variables by some nonzero h, these described changes can be read off of the following diagrams:



The first diagram (left) and the second diagram (right) are respectively equal to

$$[f(a+h, b+h) - f(a, b+h)] - [f(a+h, b) - f(a, b)], \tag{2.52}$$

$$[f(a+h, b+h) - f(a+h, b)] - [f(a, b+h) - f(a, b)]. \tag{2.53}$$

Notice that (2.52) and (2.53) are equal! Let's study them by applying the mean value theorem in each direction separately. To make the notation easier, let

$$\varphi(t) = [f(a+h,t) - f(a,t)],$$

$$\psi(t) = [f(t,b+h) - f(t,b)].$$

in which case the equality (2.52) = (2.53) is given by

$$(2.54) \qquad \varphi(b+h) - \varphi(b) = \psi(b+h) - \psi(b).$$

By the single-variable mean value theorem

$$\varphi(b+h) - \varphi(b) = \varphi'(c_h)h = \left[\partial_y f(a+h,c_h) - \partial_y f(a,c_h)\right]h$$

$$= \left[\partial_x \partial_y f(\tilde{c}_h, c_h)\right]h^2.$$

for some c_h between b and $b+h$ and $\tilde{c}_h$ between a and $a+h$. A similar calculation gives that

$$\psi(b+h) - \psi(b) = \left[\partial_y \partial_x f(e_h, \tilde{e}_h)\right]h^2$$

for some e_h between a and $a+h$ and $\tilde{e}_h$ between b and $b+h$. Plugging this into (2.54) and canceling the h^2 gives

$$(2.55) \qquad \partial_x \partial_y f(\tilde{c}_h, c_h) = \partial_y \partial_x f(e_h, \tilde{e}_h).$$

Since both partials $\partial_x \partial_y f$ and $\partial_y \partial_x f$ are continuous at a and $c_h, e_h \to a$ and $e_h, \tilde{e}_h \to b$ as $h \to 0$, taking the limit of (2.55) as $h \to 0$ gives us (2.51).

■

- **Corollary 2.56:** Suppose that $U \subseteq \mathbb{R}^m$ is open and that $f \in C^2(U)$. Then for any $i, j \in \{1, \dots, m\}$ we have that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \quad \text{everywhere in } U.$$

  Proof: This follows immediately from the previous theorem since here all second order partials are continuous everywhere. ■

- **Theorem 2.57:** Suppose that $U \subseteq \mathbb{R}^m$ is open and that $f \in C^k(U)$. For any collection of indices $i_1, \dots, i_r \in \{1, \dots, m\}$ where $r \leq k$ and any reordering/permutation $j_1, \dots, j_r$ of $i_1, \dots, i_r$,

$$\frac{\partial^r f}{\partial x^{i_1} \dots \partial x^{i_r}} = \frac{\partial^r f}{\partial x^{j_1} \dots \partial x^{j_r}} \quad \text{everywhere in } U.$$

  Proof: This is proven by applying Corollary 2.56 inductively onto the partials of f. ■

- We next introduce a powerful notation, one of whose many purposes is to give concise notation for higher order partials of a multivariable function.

- **Note 2.58:** Suppose we're working over $\mathbb{R}^m$. A **multi-index** $\alpha = (\alpha_1, \dots, \alpha_m)$ is an m-tuple of nonnegative integers (i.e. each α_i is a nonnegative integer). We define its size by

$$|\alpha| = \alpha_1 + \cdots + \alpha_m.$$

Note that this is different from our definition of length of vectors since we don't think of these as vectors but rather as collections of indices (see below): admittingly we use the same notation $|\cdot|$ for the two. For any such multi-index α and any function $f(x_1, \dots, x_m)$ we define

$$\partial^\alpha f = \partial_1^{\alpha_1}\partial_2^{\alpha_2} \dots \partial_m^{\alpha_m} f = \frac{\partial^{\overbrace{\alpha_1+\alpha_2 \dots + \alpha_m}^{|\alpha|}} f}{(\partial x_1)^{\alpha_1}(\partial x_2)^{\alpha_2} \dots (\partial x_m)^{\alpha_m}}.$$

From here we see the origin of the name "multi-index" because α is a collection of information on how many times to differentiate f in every index $i \in \{1, \dots, m\}$. A few more notations that will be useful later are

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_m^{\alpha_m} \quad \forall x \in \mathbb{R}^m \quad \text{(use the convention } 0^0 = 1\text{)}$$

$$\alpha! = \alpha_1!\,\alpha_2! \dots \alpha_m!$$

- **Example 2.59:** Suppose we take a three-variable function $f(x_1, x_2, x_3)$ and the multi-index $\alpha = (2,0,3)$. Then

$$\partial^\alpha f = \partial_1^2 \partial_3^3 f = \frac{\partial^5 f}{(\partial x_1)^2 (\partial x_3)^3}, \qquad \alpha! = 2!\,0!\,3! = 12,$$

$$x^\alpha = x_1^2 x_3^3 \quad \forall x \in \mathbb{R}^3.$$

- An application of multi-indices is to generalize the binomial theorem to higher dimensions as in the following theorem, which is called the "multinomial theorem."

- **Theorem 2.60:** For any $x \in \mathbb{R}^m$ and any integer $j \geq 1$,

$$(x_1 + \cdots + x_m)^j = \sum_{|\alpha|=j} \frac{j!}{\alpha!} x^\alpha = \sum_{|\alpha|=j} \frac{j!}{(\alpha_1)! \dots (\alpha_m)!} x_1^{\alpha_1} \dots x_m^{\alpha_m}.$$

Proof: It's proved by induction on either m or j (your choice). The book proves it by induction on m (Theorem 2.52 there), and I'm assigning it as homework to read. ■

- **Note 2.61:** We now embark on deriving Taylor series for multivariable functions. Surprisingly, this is not difficult since it turns out to follow from single variable Taylor series and the multivariable chain rule. Suppose $U \subseteq \mathbb{R}^m$ is an open set and that we have a function $f \in C^{k+1}(U)$. Fix a point $a \in U$ and consider an open ball $B \subseteq U$ contained in U centered at a.

  Take any point $x \in B$ and consider the line segment l from a to x: $l(s) = a + s(x - a)$ (recall $s \in [0, 1]$). For notational ease, let $h = x - a$ denote our "step." Now we expand f in a Taylor polynomial along this line segment. Precisely, applying the single variable Taylor polynomial to $f \circ l(s)$ centered at $s = 0$ gives

$$(2.62) \qquad f \circ l(s) = \sum_{j=0}^{k} \frac{(f \circ l)^{(j)}(0)}{j!} s^j + E_{k,h}(s)$$

($E_{k,h}(s)$ is not in the sum Σ ...) where the "error" function

$$E_{k,h}(s) = \left[\frac{1}{k!}\int_0^1 (1-t)^k (f \circ l)^{(k+1)}(ts)dt\right] s^{k+1}$$

To get a result explicitly in terms of the partials of f, we need to compute each derivative $(f \circ l)^{(j)}(s)$ explicitly. By the chain rule we have that

$$(2.63) \qquad (f \circ l)^{(j)}(s) = \underbrace{\frac{d}{ds} \dots \frac{d}{ds}}_{j} f(a+th) = \underbrace{\frac{d}{ds} \dots \frac{d}{ds}}_{j-1} \left(\underbrace{\frac{\partial f}{\partial x_1}}_{\substack{\text{evaluated} \\ \text{at } (a+sh)}} h_1 + \dots + \frac{\partial f}{\partial x_m} h_m \right)$$

$$= \underbrace{\frac{d}{ds} \dots \frac{d}{ds}}_{j-1} (h_1 \partial_1 + \dots + h_m \partial_m) f = \text{repeat } j-1 \text{ more times ...}$$

$$= \underbrace{(h_1 \partial_1 + \dots + h_m \partial_m) \dots (h_1 \partial_1 + \dots + h_m \partial_m)}_{j} f = (h_1 \partial_1 + \dots + h_m \partial_m)^j f.$$

Now a proof just like the one for the multinomial theorem (Theorem 2.60 above) gives that

$$(2.64) \qquad (h_1 \partial_1 + \dots + h_m \partial_m)^j f = \sum_{|\alpha|=j} \frac{j!}{\alpha!} h^\alpha \partial^\alpha f.$$

Since we're interested in the value of f at x, we now plug in $s = 1$ into (2.62) to get (here we change $E_{k,h}(s) \to E_k(h)$ since by setting $s = 1$ there is no more dependence on s)

$$f \circ l(1) = f(a+h) = \sum_{j=0}^{k} \frac{\sum_{|\alpha|=j} \frac{j!}{\alpha!} h^\alpha \partial^\alpha f(a)}{j!} + E_k(h) = \sum_{j=0}^{k} \sum_{|\alpha|=j} \frac{\partial^\alpha f(a)}{\alpha!} h^\alpha + E_k(h)$$

$$= \sum_{|\alpha| \le k} \frac{\partial^\alpha f(a)}{\alpha!} h^\alpha + E_k(h).$$

where

$$E_k(h) = \frac{1}{k!}\int_0^1 (1-t)^k \sum_{|\alpha|=k+1} \frac{(k+1)!}{\alpha!} h^\alpha \partial^\alpha f(a+th)\, dt$$

Simplifying, rearranging, and finally substituting $h = x - a$ gives Taylor's theorem in multiple variables:

- **Theorem 2.65:** Suppose that $U \subseteq \mathbb{R}^m$ is an open set, $f \in C^{k+1}(U)$, $a \in U$ is a point, and $B \subseteq U$ is a ball contained in U centered at a such that $\bar{B} \subseteq U$. Then for any $x \in B$ (the following two equations are equivalent)

$$(2.66) \qquad f(x) = \sum_{j=0}^{k} \underbrace{\sum_{|\alpha|=j} \frac{\partial^\alpha f(a)}{\alpha!}(x-a)^\alpha}_{\text{called "}j^{\text{th}}\text{ order term"}} + E_k(x-a),$$

$$f(x) = \sum_{|\alpha| \le k} \frac{\partial^\alpha f(a)}{\alpha!}(x-a)^\alpha + E_k(x-a).$$

where

$$E_k(x-a) = (k+1) \sum_{|\alpha|=k+1} \frac{(x-a)^\alpha}{\alpha!} \int_0^1 (1-t)^k \partial^\alpha f\big(a + t(x-a)\big) dt$$

which implies that

$$(2.67) \qquad |E_k(x-a)| \le \sqrt{m}^{k+1} \frac{\max\limits_{|\alpha|=k+1} \sup\limits_{y \in \bar{B}} |\partial^\alpha f(y)|}{(k+1)!} |x-a|^{k+1}$$

$$\Longrightarrow \quad \frac{|E_k(x-a)|}{|x-a|^k} \to 0 \quad \text{as} \quad x \to a.$$

Remark: The very last implication "$\Longrightarrow$" is very important: it says that the error term $E_k(h)$ vanishes faster than $|h|^k$ at $h = 0$.

Proof: The proof is contained in our discussion Note 2.61 above except for the last implication (2.67), which you'll prove on the homework. ■

- We record the following important equation which gives a useful interpretation of every term in the Taylor series (2.66). Its proof follows immediately from our calculation (2.63) and (2.64).

- **Theorem 2.68:** Suppose that $U \subseteq \mathbb{R}^m$ is open and that $f \in C^k(U)$. Fix any $a \in U$, a vector $h \in \mathbb{R}^m$, and consider the line $l(t) = a + th$. Then for any integer $0 \le j \le k$

$$(2.69) \qquad (f \circ l)^{(j)}(0) = \sum_{|\alpha|=j} \frac{j!}{\alpha!} h^\alpha \partial^\alpha f(a).$$

- *Remark:* In other words, the j^{th} derivative of f along the line l with velocity h at a (i.e. the left-hand side of (2.69)) is equal to the j^{th} order term in the Taylor series of f centered at a evaluated at h (i.e. the right-hand side of (2.69)).

Another thing that we remark is that if h is a unit vector, then it's not hard to see that the left-hand side of (2.69) is equal to the j^{th} directional derivative $\partial_h^j f(a)$.

- **Note 2.70:** Let's see what the first few terms of the Taylor series of f look like. For simplicity, let's first look at the case $m = 2$. Setting $k = 2$ in (2.66) and $h = x - a$ to make the notation simpler gives

$$f(x,y) = f(a_1,a_2) + \left(\frac{\partial_1 f(a)}{1!}h_1 + \frac{\partial_2 f(a)}{1!}h_2\right)$$
$$+\left(\frac{\partial_1\partial_1 f(a)}{2!}h_1^2 + \frac{\partial_1\partial_2 f(a)}{1!\,1!}h_1h_2 + \frac{\partial_2\partial_2 f(a)}{2!}h_2^2\right) + E_2(h)$$

and so

$$f(x,y) = f(a_1,a_2) + \nabla f(a)\cdot h + \underbrace{\frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(a)h_1^2 + 2\frac{\partial^2 f}{\partial x \partial y}(a)h_1h_2 + \frac{\partial^2 f}{\partial y^2}(a)h_2^2\right)}_{\text{second order term in Taylor polynomial}} + E_2(h).$$

Notice that the second order term in the Taylor polynomial without the $1/2$ factor can be written as (here "$\cdot$" is the dot product; we omit writing "(a)")

$$= \left[\begin{pmatrix} \partial_1^2 f & \partial_1\partial_2 f \\ \partial_2\partial_1 f & \partial_2^2 f\end{pmatrix}\begin{pmatrix} h_1 \\ h_2\end{pmatrix}\right]\cdot\begin{pmatrix} h_1 \\ h_2\end{pmatrix} = \underbrace{\left\langle\begin{pmatrix} \partial_1\partial_1 f & \partial_1\partial_2 f \\ \partial_2\partial_1 f & \partial_2\partial_2 f\end{pmatrix}\begin{pmatrix} h_1 \\ h_2\end{pmatrix}, \begin{pmatrix} h_1 \\ h_2\end{pmatrix}\right\rangle}_{\text{Another notation for dot product}}$$

Note that this matrix is symmetric because $\partial_1\partial_2 f = \partial_2\partial_1 f$ since we assumed that $f \in C^k(U)$ where $k = 2$: a very special property of matrices!

For function of more variables $f(x_1, \dots, x_m)$ this will take the form

$$(2.71) \quad f(x) = f(a) + \nabla f(a)\cdot h + \underbrace{\frac{1}{2}\left(\sum_{i=1}^{m}\partial_i^2 f(a)h_i^2 + 2\sum_{1\le i<j\le m}\partial_i\partial_j f(a)h_ih_j\right)}_{\text{second order term in Taylor polynomial}} + E_2(x-a)$$

where the second order order term in the Taylor polynomial without the $1/2$ factor can be written as

$$(2.72) \qquad \left\langle\begin{pmatrix} \partial_1\partial_1 f & \cdots & \partial_1\partial_m f \\ \vdots & \ddots & \vdots \\ \partial_m\partial_1 f & \cdots & \partial_m\partial_m f\end{pmatrix}\begin{pmatrix} h_1 \\ \vdots \\ h_m\end{pmatrix}, \begin{pmatrix} h_1 \\ \vdots \\ h_m\end{pmatrix}\right\rangle.$$

As before this matrix is also symmetric since the entry in the i^{th} column and j^{th} row (i.e. $\partial_i\partial_j f$) is equal to the entry in the j^{th} column and i^{th} row (i.e. $\partial_j\partial_i f$). This matrix has a special name: it is called the **Hessian** of f at a and is denote by "$H_f(a)$" (in more advanced courses they denote the above as $\nabla^2 f(h,h)$, but we avoid this notation for now). The concept of the Hessian is powerful because it allows to use the power of linear algebra to study the second derivatives of f (or how its graph "curves through space").

- **Definition 2.73:** Suppose we have a function $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ where U is open.
  - A point $a \in U$ is called a **critical point** if either f is not differentiable at a or $\nabla f(a) = 0$.
  - A point $a \in U$ is called a **local minimum** of f if
  $$\exists \delta > 0 \quad \forall x \in U : |x - a| < \delta, \;\; f(x) \geq f(a).$$
  - A point $a \in U$ is called a **local maximum** of f if
  $$\exists \delta > 0 \quad \forall x \in U : |x - a| < \delta, \;\; f(x) \leq f(a).$$
  - A point $a \in U$ is called a **global minimum** of f if
  $$\forall x \in U, \;\; f(x) \geq f(a)$$
  - A point $a \in U$ is called a **global maximum** of f if
  $$\forall x \in U, \;\; f(x) \leq f(a)$$

  Observe that a global maximum or global minimum is automatically a local maximum or local minimum respectively.

- The following theorem is a direct generalization of the result in single variable calculus about the relation between critical points and first derivatives.

- **Theorem 2.74:** Suppose we have a function $f : U \subseteq \mathbb{R}^m \to \mathbb{R}$ where U is open. If $a \in U$ is a local minimum or local maximum, then it is a critical point.

  Proof: Take such a point $a \in U$. If f is not differentiable at a, then it's a critical point and so we're done. So suppose that f is differentiable at a. We will show that each $\frac{\partial f}{\partial x_i}(a) = 0$ and hence $\nabla f(a) = 0$. Let's start with showing that $\frac{\partial f}{\partial x_1}(a) = 0$. Take the line $l(t) = a + t(1,0,\dots,0) = (a_1 + t, a_2, \dots, a_m)$. Since f has a local minimum at a, $f \circ l$ will have a local minimum at $t = 0$ and hence by single variable calculus its derivative at $t = 0$ is zero. Thus, by the chain rule (or (2.69))
  $$0 = \frac{d}{dt}\big(f \circ l(t)\big)\Big|_{t=0} = \frac{d}{dt}\big(f(a + t(1,0,\dots,0)\big)\Big|_{t=0} = \frac{\partial f}{\partial x_1}(a) \cdot 1.$$
  which is what we wanted. Showing that $\frac{\partial f}{\partial x_i}(a) = 0$ for $i = 2, \dots, m$ is handled similarly.

  ■

- In single variable calculus we used the second derivative to test if something is a local minimum or local maximum. In higher dimensions, this is much more complicated since we have many directions (including diagonal ones) to take into account. First we recall three results from linear algebra, starting with the following called the **spectral theorem**.

- **Theorem 2.75:** Suppose that A is a symmetric $m \times m$ matrix. Then there exists a unitary $m \times m$ matrix U such that

$$(2.76) \qquad A = U^{\top} \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_m \end{pmatrix} U$$

where $\lambda_1, \dots, \lambda_m$ are eigenvalues of A. The form of the matrix in between $U^{\top}$ and U on the right-hand is called a "diagonal matrix" because it only has nonzero entries on the diagonal. Recall that $U^{\top}$ is the matrix obtained by flipping U across its diagonal running from the top-left corner to the bottom-right corner. Recall also that U being unitary means that $|Uh| = |h|$.

- **Lemma 2.77:** For any square matrix B and any vectors x, y,

$$\langle Bx, y \rangle = \langle x, B^{\top} y \rangle \quad \text{and} \quad \langle x, By \rangle = \langle B^{\top} x, y \rangle.$$

- **Theorem 2.78:** Suppose that $U \subseteq \mathbb{R}^m$ is an open set, $f \in C^2(U)$, and that $\nabla f(a) = 0$ at some point $a \in U$.

  a) If $a \in U$ is a local minimum, then all of the eigenvalues of $H_f(a)$ are nonnegative.

  b) If all of the eigenvalues of $H_f(a)$ are positive, then $a \in U$ is a local minimum.

  c) If $a \in U$ is a local maximum, then all of the eigenvalues of $H_f(a)$ are nonpositive.

  d) If all of the eigenvalues of $H_f(a)$ are negative, then $a \in U$ is a local maximum.

Proof: We will only prove a) and b) since c) and d) are proved similarly. Let's start with a). Choose an eigenvalue λ of $H_f(a)$: we want to show that $\lambda \geq 0$. Let $v \in \mathbb{R}^m$ be an eigenvector of λ, meaning $v \neq 0$ and $H_f(a)v = \lambda v$. Consider the line $l(t) = a + tv$. As argued before, since f has a local minimum at a, $f \circ l$ has a local minimum at $t = 0$. Thus by single variable calculus, the second derivative of $f \circ l$ is nonnegative at $t = 0$. Thus (2.69) and the comment leading to (2.72) give

$$(2.79) \qquad 0 \leq (f \circ l)''(0) = \frac{1}{2}\langle H_f(a)v, v \rangle = \frac{1}{2}\langle \lambda v, v \rangle = \frac{1}{2}\lambda\langle v, v \rangle = \frac{1}{2}\lambda|v|^2.$$

Since $|v|^2 > 0$, this shows that indeed $\lambda \geq 0$.

Now suppose that all of the eigenvalues of $H_f(a)$ are positive. We want to show that $a \in U$ is a local minimum, which note is equivalent to showing that for some small $\delta > 0$, $f(x) - f(a) \geq 0$ for $x \in U : |x - a| < \delta$ (recall that $\nabla f(a) = 0$ by assumption). Let U be a unitary matrix such that $H_f(a) = U^{\top} D U$ where D is a diagonal matrix with the eigenvalues $\lambda_1, \dots, \lambda_m$ of $H_f(a)$

running down the diagonal as in (2.76) (this is possible since $H_f(a)$ is symmetric which recall follows from $f \in C^2(U)$). Letting $h = x - a$ for shorthand, by (2.71) and (2.72) we have that

$$f(x) - f(a) = \frac{1}{2}\langle H_f(a)h, h\rangle + E_2(h) = \frac{1}{2}\langle U^\top DUh, h\rangle + E_2(h) = \frac{1}{2}\langle D\underbrace{Uh}_{\tilde{h}}, \underbrace{Uh}_{\tilde{h}}\rangle + E_2(h)$$

$$= \frac{1}{2}\langle D\tilde{h}, \tilde{h}\rangle + E_2(h) = \frac{1}{2}\begin{pmatrix} \lambda_1\tilde{h}_1 \\ \lambda_2\tilde{h}_2 \\ \vdots \\ \lambda_m\tilde{h}_m \end{pmatrix} \cdot \begin{pmatrix} \tilde{h}_1 \\ \tilde{h}_2 \\ \vdots \\ \tilde{h}_m \end{pmatrix} + E_2(h) = \frac{1}{2}\left(\lambda_1\tilde{h}_1^2 + \cdots + \lambda_m\tilde{h}_m^2\right) + E_2(h).$$

Let $\lambda_{\min}$ be the smallest of the eigenvalues of $\lambda_1, \ldots, \lambda_m$ which observe is positive by assumption of b) (i.e. $\lambda_{\min} > 0$). Thus the above is bigger than or equal to

$$\geq \frac{1}{2}\lambda_{\min}\left(\tilde{h}_1^2 + \cdots + \tilde{h}_m^2\right) + E_2(h) = \frac{1}{2}\left|\tilde{h}\right|^2 + E_2(h) = \frac{1}{2}\lambda_{\min}|h|^2 + E_2(h)$$

$$(2.80) \qquad = \left(\frac{1}{2}\lambda_{\min} + \frac{E_2(h)}{|h|^2}\right)|h|^2.$$

By (2.67) we have that $E_2(h)/|h|^2 \to 0$ as $h \to 0$. Hence there is a $\delta > 0$ such that

$$\left|\frac{E_2(h)}{|h|^2}\right| < \frac{1}{2}\lambda_{\min} \quad \text{for} \quad |h| < \delta.$$

Thus for $|h| < \delta$, or equivalently $|x - a| < \delta$, (2.80) is bigger than or equal to

$$\geq \left(\frac{1}{2}\lambda_{\min} - \frac{1}{2}\lambda_{\min}\right)|h|^2 = \frac{1}{4}\lambda_{\min}|h|^2 \geq 0$$

which is what we wanted.

■

- We record an important technique that appears in the above proof as a separate lemma:

- **Lemma 2.81:** Suppose that $U \subseteq \mathbb{R}^m$ is open, $f \in C^2(U)$, $a \in U$, and v is an eigenvector of $H_f(a)$ with eigenvalue of λ. Consider the line $l(t) = a + tv$. Then

  $$(f \circ l)''(0) = \frac{1}{2}\langle H_f(a)v, v\rangle = \frac{1}{2}\lambda|v|^2.$$

  Remark: It's not hard to see that if v is a unit vector then $(f \circ l)''(0)$ is in fact the second order directional derivative $\partial_v^2 f(a)$.

  Proof: This follows from the calculation (2.79) after the "$0 \leq$". ■

- In other words, the above lemma gives an important geometric interpretation of the eigenvalues of $H_f(a)$ in terms of the graph of f. It says that an eigenvector of $H_f(a)$ whose eigenvalue is positive indicates a direction in which the graph of f will "curve up" and an eigenvector of

$H_f(a)$ whose eigenvalue is negative indicates a direction in which the graph of f will "curve down."

- **Definition 2.82:** Suppose that $U \subseteq \mathbb{R}^m$ is open, $f \in C^2(U)$, and $a \in U$ is such that $\nabla f(a) = 0$. If some eigenvalues of $H_f(a)$ are positive and some are negative and <u>none</u> are zero, then a is called a **saddle point** of f.
- **Remark 2.83:** If $\nabla f(a) = 0$ and one of the eigenvalues of $H_f(a)$ is zero, then you can't conclude if it's a local minimum, local maximum, or a saddle point. You would have to analyze higher order derivatives. A simple demonstration of this is given by $f(x, y) = x^3 - y^2$ (graph it!).
- For two variable functions, there is actually an easy test for local minimums, local maximums, and saddle points. First recall a lemma from linear algebra:
- **Lemma 2.84:** Suppose A is an $m \times m$ matrix and let $\lambda_1, \dots, \lambda_m$ denote its eigenvalues counting algebraic multiplicity. Then

$$\det A = \lambda_1 \cdot \ldots \cdot \lambda_m \quad \text{and} \quad \operatorname{trace}(A) = \lambda_1 + \cdots + \lambda_m.$$

- **Theorem 2.85:** Suppose that $U \subseteq \mathbb{R}^2$ is an open set, $f \in C^2(U)$, and that $\nabla f(a) = 0$ at some point $a \in U$.
  a. If $\det H_f(a) < 0$ then a is a saddle point.
  b. If $\det H_f(a) > 0$ and $\frac{\partial^2 f}{\partial x^2}(a) > 0$, then a is a local minimum.
  c. If $\det H_f(a) > 0$ and $\frac{\partial^2 f}{\partial x^2}(a) < 0$, then a is a local maximum.
  d. If $\det H_f(a) = 0$, then at least one eigenvalue of $H_f(a)$ is zero and so no conclusion can be drawn without looking at higher order partials.

Remark: Parts b) and c) work if you replace "$\frac{\partial^2 f}{\partial x^2}(a)$" with "$\frac{\partial^2 f}{\partial y^2}(a)$" – the proof is essentially identical. Parts a) and d) also hold in higher dimensions (i.e. $\mathbb{R}^m$) – the proof is essentially identical.

Proof: Let λ_1 and λ_2 be eigenvalues of $H_f(a)$.

Proof of a): If $\det H_f(a) = \lambda_1\lambda_2 < 0$, then one of λ_1 and λ_2 is negative and the other is positive and hence a is a saddle point.

Proof of b) and c): If $\det H_f(a) = \lambda_1\lambda_2 > 0$, then either both λ_1 and λ_2 are positive or both are negative. If $\frac{\partial^2 f}{\partial x^2}(a) > 0$, then the second derivative of f along the x axis is positive and hence f must have a local minimum at a. If $\frac{\partial^2 f}{\partial x^2}(a) < 0$, then similar reasoning gives that f must instead have a local maximum at a.

Part d): If $\det H_f(a) = \lambda_1\lambda_2 = 0$, then either $\lambda_1 = 0$ or $\lambda_2 = 0$.

■

3 Implicit Function Theorem and Submanifolds of Euclidean Space

- We've developed differential calculus on flat Euclidean space. Soon we will do the same for integral calculus. Humans have discovered that both theories can be lifted to more general domains, in particular "curved spaces" sitting in Euclidean spaces. To help visualize these, think of smooth curves and 2-dimensional surfaces sitting in $\mathbb{R}^3$ – though keep in mind that we will be working in much higher dimensions as well. This generalization is important as it arises in the theory of optimization of functions with systems of constraints, in the study of the integral form of Maxwell's equations, etc. The essential tool to construct such a theory will be the implicit function theorem, which we now discuss. First let's recall a concept from topology:

- **Definition 3.1:** For any point $p \in \mathbb{R}^m$, an **open neighborhood** U of p is an open set $U \subseteq \mathbb{R}^m$ such that $p \in U$.

- **Note 3.2:** Typically we think of a neighborhood U of p as a "small" open set containing p. When we say that something occurs locally near p, we mean that it occurs in a neighborhood U of p.

- **Note 3.3:** Consider the equation for a sphere S of radius one in $\mathbb{R}^3$:

$$x^2 + y^2 + z^2 - 1 = 0.$$

Suppose you have a function $h : S \to \mathbb{R}$ and you wanted to find its maximum. At the moment it's difficult to perform such 2-dimensional calculus on this 2-dimensional curved object since we've only done calculus on Euclidean space. Suppose for simplicity that we know that the maximum of h occurs on the upper hemisphere. Notice that for points on the upper hemisphere, we can solve for z in the above equation:

$$z = \sqrt{1 - x^2 - y^2}$$

to describe the upper hemisphere of S as the graph of the function $f : B(0; 1) \subseteq \mathbb{R}^2 \to \mathbb{R}$ (here $B(0; 1)$ denotes the ball of radius 1 centered at 0) given by:

$$z = f(x, y) = \sqrt{1 - x^2 - y^2}.$$

This way we can find the maximum of h over the upper hemisphere instead by finding the maximum of $h \circ f : B(0; 1) \to \mathbb{R}$ using the differential calculus that we already devloped.

Notice what we've done: to analyze a surface we've reduced it to the graph of a function. This is a powerful technique, but it must be applied with care. Notice that if on the other hand we knew that h had a maximum on the right hemisphere of S (i.e. the one intersecting the positive x-axis), then we could not have solved for z in terms of x and y to describe that portion of S. However, in that case we can solve for x in terms of y and z and use the fact that the right hemisphere of S is given by the graph of the function

$$x = f(y, z) = \sqrt{1 - y^2 - z^2}.$$

This is a general trick that can be done for all surfaces including (plot these!)

- Hyperboloid of one sheet: $x^2 + y^2 - z^2 - \text{postive constant} = 0$
- Paraboloid: $x^2 + y^2 - z - \text{constant} = 0$
- Corner of a room: $x + e^{-y} + e^z - \text{constant} = 0$.

- **Theorem 3.4: (Implicit Function Theorem I)** Suppose that $F : U \subseteq \mathbb{R}^m \to \mathbb{R}$ is a C^k function where $k \geq 1$ and U is open (k can be infinity). Let $S \subseteq U$ be the set

$$S = \{(x_1, \dots, x_m) \in \mathbb{R}^m : F(x_1, \dots, x_m) = 0\}. \tag{3.5}$$

Fix a point $p = (p_1, \dots, p_m) \in S$ and suppose that $\frac{\partial F}{\partial x_m}(p) \neq 0$. Then there exist r_0 and r_1 so that if you consider the cylinder

$$R = B\big((p_1, \dots, p_{m-1}); r_0\big) \times (p_m - r_1, p_m + r_1), \tag{3.6}$$

then

$$R \cap S = \underbrace{\{(x_1, \dots, x_m) : x_m = f(x_1, \dots, x_{m-1})\}}_{\text{i.e. the graph of } f} \tag{3.7}$$

for a unique function $f \in C^k\big[B\big((p_1, \dots, p_{m-1}); r_0\big)\big]$. Furthermore, for any $i \in \{1, \dots, m-1\}$,

$$\partial_i f(x_1, \dots, x_{m-1}) = -\frac{\partial_i F\big(x_1, \dots, x_{m-1}, f(x_1, \dots, x_{m-1})\big)}{\partial_m F\big(x_1, \dots, x_{m-1}, f(x_1, \dots, x_{m-1})\big)}. \tag{3.8}$$

Remark 1: The above theorem is formulated as taking the equation $F(x_1, \dots, x_m) = 0$ in (3.5) and "solving" for x_m to get (3.7). The above theorem and its proof work equally well if you want to solve for a different variable x_j instead. In that case you need to assume that $\frac{\partial F}{\partial x_j}(p) \neq 0$ in which case (3.6), (3.7), and (3.8) will instead take the form

$$R = B\Big(\big(p_1, \dots, p_{j-1}, p_{j+1}, , \dots, p_m\big), r_0\Big) \times \big(p_j - r_1, p_j + r_1\big),$$

$$E \cap S = \Big\{(x_1, \dots, x_m) : x_j = f\big(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m\big)\Big\}$$

$$\partial_i f\big(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m\big) = -\frac{\partial_i F\big(x_1, \dots, x_{j-1}, f\big(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m\big), x_{j+1}, \dots, x_m\big)}{\partial_j F\big(x_1, \dots, x_{j-1}, f\big(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m\big), x_{j+1}, \dots, x_m\big)}.$$

Remark 2: Intuitively, equation (3.8) indicates why we need the condition $\frac{\partial F}{\partial x_m}(p) \neq 0$ in the statement of the above theorem since if we set $x = p$ in that equation, then $\frac{\partial F}{\partial x_m}(p)$ is in the denominator and we don't want to divide by zero.

Remark 3: Sometimes people aren't interested in interpreting the above theorem in terms of sets S, but rather as a theorem of when you can solve for one variable in terms of other variables.

Proof: We omit the proof of (3.7) – you may find it on pages 114 – 115 in the book. Let us prove (3.8). By the chain rule we have that (we omit writing the arguments of f here)

$$0 = \frac{\partial}{\partial x_i}[0] = \frac{\partial}{\partial x_i}[F(x_1, \dots, x_{m-1}, f)] = \frac{\partial F}{\partial x_i}(x_1, \dots, x_{m-1}, f) \cdot 1 + \frac{\partial F}{\partial x_m}(x_1, \dots, x_{m-1}, f) \cdot \frac{\partial f}{\partial x_i}.$$

Rearranging gives (3.8).

■

- **Note 3.9:** At the beginning of Note 3.3, the sphere S is equal to $\{(x, y, z) \in \mathbb{R}^3 : F(x, y, z) = 0\}$ where $F(x, y, z) = x^2 + y^2 + z^2 - 1$. For a point $p = (p_1, p_2, p_3)$ on the upper hemisphere, $\frac{\partial F}{\partial z}(p) = 2p_3 \neq 0$ and so the above theorem says that we can solve for z in terms of x and y near p. Indeed we saw this in Note 3.3! However, if p is on the equator in the right hemisphere and hence $p_3 = 0$ but $p_2 \neq 0$, we have that $\frac{\partial F}{\partial z}(p) = 2p_3 = 0$ and so we can't solve for z in terms of x and y. But notice that $\frac{\partial F}{\partial y}(p) = 2p_2 \neq 0$ and so we can solve for x in terms of y and z.

- **Note 3.10:** The above can be extended for functions of the form $F : U \subseteq \mathbb{R}^m \to \mathbb{R}^n$. This is important for describing "curved spaces" when the difference in their dimension and the dimension of the Euclidean space that they sit in is bigger. To illustrate this, consider S that is the set of points $(x, y, z) \in \mathbb{R}^3$ satisfying

$$F(x, y, z) = \begin{pmatrix} x^2 + y^2 + z^2 - 1 \\ y - z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

In other words, this is a function $F : \mathbb{R}^3 \to \mathbb{R}^2$ set to zero. The set of points (x, y, z) is the circle in $\mathbb{R}^3$ centered at zero making a 45-degree angle with the y-axis and a 0-degree angle with the x-axis. Let us solve for y and z in the above equation in terms of x in the region $y > 0$. The above equation is equivalent to the system of equations

$$x^2 + y^2 + z^2 - 1 = 0,$$

$$y - z = 0.$$

The second equation gives us that $z = y$. Plugging this into the first equation gives $x^2 + 2z^2 - 1 = 0$ which gives that $z = \sqrt{(1/2)(1 - x^2)}$. Plugging this back into $z = y$ gives $y = \sqrt{(1/2)(1 - x^2)}$. Hence S in the region $y > 0$ is the graph of the function $f : (-1, 1) \to \mathbb{R}^2$ given by

$$(y, z) = f(x) = \left(\sqrt{(1/2)(1 - x^2)}, \sqrt{(1/2)(1 - x^2)}\right).$$

Notice that we solved for 2 variables in terms of $3 - 2 = 1$ varables since we had a system of 2 equations above with three unknowns. For a general $F : \mathbb{R}^m \to \mathbb{R}^n$ we would expect to be able to solve for n variables in terms of $m - n$ variables because we'll have a system of n equations for m unknowns. This hints that the general theory should be developed for $m \geq n$. To develop this theory, we need the notion of the Jacobian:

- **Definition 3.11:** Suppose that we have $F : U \subseteq \mathbb{R}^m \to \mathbb{R}^n$ where U is open. Explicitly

$$F(x_1, \dots, x_m) = \begin{pmatrix} F_1(x_1, \dots, x_m) \\ \vdots \\ F_n(x_1, \dots, x_m) \end{pmatrix}.$$

Suppose that F is differentiable (i.e. each F_j is differentiable). Thus for each $a \in U$, setting $h = x - a$ for shorthand,

$$\begin{pmatrix} F_1(x) \\ \vdots \\ F_n(x) \end{pmatrix} = \begin{pmatrix} F_1(a) \\ \vdots \\ F_n(a) \end{pmatrix} + \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(a)h_1 + \cdots + \frac{\partial F_1}{\partial x_m}(a)h_m \\ \vdots \\ \frac{\partial F_n}{\partial x_1}(a)h_1 + \cdots + \frac{\partial F_n}{\partial x_m}(a)h_m \end{pmatrix} + \underbrace{\begin{pmatrix} E_{1,a}(h) \\ \vdots \\ E_{n,a}(h) \end{pmatrix}}_{E(a)}$$

where each $\lim_{h \to 0} E_{j,a}(h)/|h| = 0$. This can be rewritten as

$$F(x) = F(a) + \underbrace{\begin{pmatrix} \frac{\partial F_1}{\partial x_1}(a) & \cdots & \frac{\partial F_1}{\partial x_m}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial x_1}(a) & \cdots & \frac{\partial F_n}{\partial x_m}(a) \end{pmatrix}}_{\text{"}DF(a)\text{" or "}D_xF(a)\text{"}} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} + E_a(h).$$

It's easy to check that $\lim_{h \to 0} |E_a(h)|/|h| = 0$. The matrix $DF(a)$ in the middle is called the **Jacobian matrix** of F at a. The determinant $J_F(a) = \det DF(a)$ is called the **Jacobian determinant**. Sometimes both are simply referred to as the **Jacobian** (you have to rely on context to tell which is being discussed).

- **Note 3.12:** Look at the above definition, intuitively speaking, since $E_a(h)$ is negligible in size to the rest of the terms, this says that $F(x) - F(a)$ is a linear map on the differential scale (not a rigorous statement at the moment), a central theme in analysis!

- **Theorem 3.13: (Chain Rule with Jacobians)** Suppose that $G : V \subseteq \mathbb{R}^k \to U \subseteq \mathbb{R}^m$ and $F : U \subseteq \mathbb{R}^m \to \mathbb{R}^n$ are differentiable where U and V are open. Then (here DF is being evaluated at $G(x)$)

$$D(F \circ G)(x) = \big[(DF)\big(G(x)\big)\big]\big(DG(x)\big) = [(DF) \circ G](DG)$$

$$\Longrightarrow \qquad J_{F \circ G} = \big(J_F(G(x)\big)J_G(x) = (J_F \circ G)J_G.$$

where in the last quantity on each line we omitted writing the arguments.

Proof: Will be assigned as homework. ■

- **Note 3.14:** We aim to extend the technique explored in Note 3.10 to general F. As a toy model, consider the situation when $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear map with $m \geq n$:

$$F(x_1,\dots,x_m)=\begin{pmatrix}F_1(x_1,\dots,x_m)\\ \vdots\\ F_n(x_1,\dots,x_m)\end{pmatrix}=\begin{pmatrix}a_{1,1}&\cdots&a_{1,m}\\ \vdots&\ddots&\vdots\\ a_{n,1}&\cdots&a_{n,m}\end{pmatrix}\begin{pmatrix}x_1\\ \vdots\\ x_m\end{pmatrix}=\begin{pmatrix}a_{1,1}x_1+\cdots+a_{1,m}x_m\\ \vdots\\ a_{n,1}x_1+\cdots+a_{n,m}x_m\end{pmatrix}$$
$$=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}.$$

For later use, observe that

$$(3.15)\qquad \frac{\partial F_i}{\partial x_j}=a_{i,j}.$$

Now we ask, when can we solve for n of these variables in terms of the other $m-n$ variables? Answer: we need an invertible $n\times n$ submatrix sitting in the above system. Let's illustrate this. For simplicity, suppose that the submatrix consisting of the last n columns of the above matrix is invertible. To make the notation easier, let us instead write points in $\mathbb{R}^m$ as

$$(x_1,\dots,x_m)=(y_1,\dots,y_{m-n},z_1,\dots,z_n)$$

and write

$$(3.16)\quad F(y_1,\dots,y_{m-n-1},z_1,\dots,z_n)=\begin{pmatrix}b_{1,1}y_1+\cdots+b_{1,m-n}y_{m-n-1}+c_{1,1}z_1+\cdots+c_{1,n}z_n\\ \vdots\\ b_{n,1}y_1+\cdots+b_{n,m-n}y_{m-n-1}+x_{n,1}z_1+\cdots+c_{n,n}z_n\end{pmatrix}$$
$$=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}$$

This system can be rewritten as

$$\underbrace{\begin{pmatrix}b_{1,1}&\cdots&b_{1,m-n}\\ \vdots&\ddots&\vdots\\ b_{n,1}&\cdots&b_{n,m-n}\end{pmatrix}}_{B}\underbrace{\begin{pmatrix}y_1\\ \vdots\\ y_{m-n}\end{pmatrix}}_{y}+\underbrace{\begin{pmatrix}c_{1,1}&\cdots&c_{1,n}\\ \vdots&\ddots&\vdots\\ c_{n,1}&\cdots&c_{n,n}\end{pmatrix}}_{C\text{ is invertible}}\underbrace{\begin{pmatrix}z_1\\ \vdots\\ z_n\end{pmatrix}}_{z}=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix},$$
$$\implies\qquad z=-C^{-1}By$$

Just as we got (3.15), it's not hard to see from (3.16) that this last equation can be rewritten as

$$(3.17)\qquad z=-\underbrace{\begin{pmatrix}\dfrac{\partial F_1}{\partial z_1}&\cdots&\dfrac{\partial F_1}{\partial z_n}\\ \vdots&\ddots&\vdots\\ \dfrac{\partial F_n}{\partial z_1}&\cdots&\dfrac{\partial F_n}{\partial z_n}\end{pmatrix}^{-1}}_{(D_zF)^{-1}}\underbrace{\begin{pmatrix}\dfrac{\partial F_1}{\partial y_1}&\cdots&\dfrac{\partial F_1}{\partial y_{m-n}}\\ \vdots&\ddots&\vdots\\ \dfrac{\partial F_n}{\partial y_1}&\cdots&\dfrac{\partial F_n}{\partial y_{m-n}}\end{pmatrix}}_{D_yF}y\underbrace{=f(y)}_{\substack{\text{define }f\\ \text{this way}}}.$$

It's not hard to see (I'll most likely assign it as homework) that this implies that

$$(3.18) \qquad Df = -(D_zF)^{-1}(D_yF).$$

The solution (3.17) won't hold for a general function F since F may not be linear. However, we observed in Note 3.12 that on a differential scale functions, such as F, look linear. Thus (3.18) on the other hand will hold for general functions F (with assumptions). Here is the precise statement:

- **Theorem 3.19: (Implicit Function Theorem II)** Suppose that $F : U \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is a C^k function where $k \geq 1$ and U is open (k can be infinity). Let $S \subseteq U$ be the set

$$S = \{(y_1, \ldots, y_{m-n}, z_1, \ldots, z_n) \in \mathbb{R}^m : F(y_1, \ldots, y_{m-n}, z_1, \ldots, z_n) = 0\}.$$

Fix a point $p = (p_1, \ldots, p_{m-n}, \tilde{p}_1, \ldots, \tilde{p}_n) \in S$ and suppose that $\det D_zF(p) \neq 0$. Then there exist r_0 and r_1 so that if you consider the cylinder

$$R = B\big((p_1, \ldots, p_{m-n}), r_0\big) \times B\big((\tilde{p}_1, \ldots, \tilde{p}_n), r_1\big),$$

then

$$(3.20) \qquad R \cap S = \underbrace{\{(y_1, \ldots, y_{m-n}, z_1, \ldots, z_n) : (z_1, \ldots, z_n) = f(y_1, \ldots, y_{m-n})\}}_{\text{i.e. the graph of } f}$$

for a unique function $f \in C^k\big[B\big((p_1, \ldots, p_{m-n}); r_0\big)\big]$. Furthermore, writing $= (y_1, \ldots, y_{m-n})$ and $z = (z_1, \ldots, z_n)$,

$$(3.21) \qquad D_yf(y) = -(D_zF)^{-1}(D_yF).$$

where the Jacobian matrices of F on the right-hand side are being evaluated at $\big(y, f(y)\big)$.

Remark: The Implicit Function Theorem I is a special case of this theorem with $n = 1$. In addition, similar to Remark 1 made after Theorem 3.4, if you want to solve for a different set of variables $z_1', \ldots, z_n'$ in terms of $y_1', \ldots, y_{m-n-1}'$ when writing points in $\mathbb{R}^m$ as some other permutation of z's and y's: $(y_1', y_2', z_1', y_3', \ldots)$, then you'll need to make obvious modifications to the above theorem including that $\det D_{z'}F(p) \neq 0$ and $D_{y'}f = -(D_{z'}F)^{-1}\big(D_{y'}F\big)$. We leave the details to the reader.

Proof: We omit the proof of (3.20) – it can be found on page 420 – 422 in the book. Just like in the proof of Theorem 3.4, (3.21) follows from the chain rule

$$0 = D(0) = D\Big(F\big(y, f(y)\big)\Big) = D\left(F \circ \begin{pmatrix} y_1 \\ \vdots \\ y_{m-n} \\ f_1(y) \\ \vdots \\ f_n(y) \end{pmatrix}\right)$$

$$= \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_{m-n}} & \frac{\partial F_1}{\partial z_1} & \cdots & \frac{\partial F_1}{\partial z_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_n}{\partial y_1} & \cdots & \frac{\partial F_n}{\partial y_{m-n}} & \frac{\partial F_n}{\partial z_1} & \cdots & \frac{\partial F_n}{\partial z_n} \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \cdots & \frac{\partial f_1}{\partial y_{m-n}} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \cdots & \frac{\partial f_2}{\partial y_{m-n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_1} & \frac{\partial f_n}{\partial y_2} & \cdots & \frac{\partial f_n}{\partial y_{m-n}} \end{pmatrix}$$

$$= (D_y F)(\mathrm{id}) + (D_z F)(D_y f).$$

Rearranging gives $D_y f = -(D_z F)^{-1}(D_y F)$ (we omitted writing arguments).

■

- We record an important corollary for later use:
- **Corollary 3.22: (Inverse Mapping/Function Theorem)** Suppose that $F : U \subseteq \mathbb{R}^m \to V \subseteq \mathbb{R}^m$ is C^k where $k \geq 1$ and U and V are open (k can be infinity). Suppose also that $a \in U$ is such that $DF(a)$ is invertible. Then there exist open neighborhoods $\widetilde{U}$ of a and $\widetilde{V}$ of $F(a)$ such that the restriction $F : \widetilde{U} \to \widetilde{V}$ is bijective and has a C^k inverse $F^{-1} : \widetilde{V} \to \widetilde{U}$. Furthermore, over $\widetilde{U}$

$$D(F^{-1})(y) = (DF)^{-1}\big(F^{-1}(y)\big)$$

(the left-hand side is being evaluated at y and the right-hand side is being evaluated at $F^{-1}(y)$).

Proof: Consider the function $G : U \times V \to \mathbb{R}^m$ given by $G(x, y) = F(x) - y$ and conider the set

$$S = \{(x, y) \in \mathbb{R}^{2m} : G(x, y) = F(x) - y = 0\}$$

It's easy to check that DG is given by attaching the negative $m \times m$ identity matrix to the right of DF (write it out!). Hence by the implicit function theorem there exists $r_0, r_1 > 0$ such that

$$\underbrace{[B(a; r_0) \times B(f(a); r_1)]}_{\text{small cylinder}} \cap S = \{(x, y) \in \mathbb{R}^{2m} : x = f(y)\}$$

for some unique $f \in C^k[B(f(a); r_1)]$. Setting

$$\widetilde{V} = B(f(a); r_1) \quad \text{and} \quad \widetilde{U} = B(a; r_0) \cap F^{-1}[\widetilde{V}]$$

and $F^{-1} = f$ satisfies the conclusion of the theorem. Finally, letting I_m denote the $m \times m$ identity matrix, by (3.21) we have that (we omit writing arguments)

$$D(F^{-1}) = -(D_x G)^{-1}(D_y G) = -(D_x F)^{-1}(-I_m) = (D_x F)^{-1}.$$

■

- Next we address the question of how does one define "curved spaces" sitting inside $\mathbb{R}^m$, more precisely called "manifolds." We're already set up to do this using the implicit function theorem or representing them as graphs of functions. However, we'll take a more powerful approach which the former will fit into nicely as a special case. In addition to representing curved spaces as solutions to systems of equations or graphs of functions, one can obtain them by parametrizing them as in the following examples:

- **Example 3.23:** Angle parametrization of a circle S of radius r centered at zero in $\mathbb{R}^2$ minus the point $(r, 0)$ (here "Im" stands for "image")

$$S = \{(x, y) = f(\theta) : \theta \in (0, 2\pi)\} = \operatorname{Im} f$$

$$\text{where} \quad f : (0, 2\pi) \subseteq \mathbb{R}^1 \to \mathbb{R}^3 \quad \text{given by} \quad f(\theta) = (r\cos(\theta), r\sin(\theta)).$$

Notice that $\dim \operatorname{dom} f = 1$ and $\dim S = 1$. Interesting!

- **Example 3.24:** Spherical coordinate parametrization of a sphere of radius r centered at zero in $\mathbb{R}^3$ minus an arc pointing in the x-axis direction:

$$S = \{(x, y, z) = f(\theta, \varphi) : \theta \in (0, 2\pi) \text{ and } \varphi \in (0, \pi)\} = \operatorname{Im} f$$

$$\text{where} \quad f : (0, 2\pi) \times (0, \pi) \subseteq \mathbb{R}^2 \to \mathbb{R}^3$$

$$\text{given by} \quad f(\theta, \varphi) = (r\sin\varphi\cos\theta, r\sin\varphi\sin\theta, r\cos\varphi).$$

Notice that $\dim \operatorname{dom} f = 2$ and $\dim S = 2$. Seeing a pattern?!

Observe one more thing: suppose we restrict our function f to $[\pi/2, 3\pi/2] \times (0, \pi)$. Then the surface S that it parametrizes would only be half the sphere and notice that a boundary was included along the image of $\theta = \pi/2$ and $\theta = 3\pi/2$ under f.

- **Definition 3.25:** We define the **upper-half space** as

$$\mathbb{H}^j = \{(u_1, \dots, u_j) : u_j \geq 0\}$$

For any subset $V \subseteq \mathbb{H}^j$ that is open in $\mathbb{H}^j$ we say that a function $f : V \to \mathbb{R}^m$ is C^k on V if there exists a C^k function $\tilde{f} : \tilde{V} \subseteq \mathbb{R}^j \to \mathbb{R}^m$ where $\tilde{V} \supseteq V$ is open in $\mathbb{R}^j$ and $\tilde{f}(u) = f(u)$ for $u \in V$ (i.e. $\tilde{f}$ extends f). For future references observe that

$$\partial\mathbb{H}^j = \{(u_1, \dots, u_j) : u_j = 0\}.$$

- **Definition 3.26:** If $j \leq m$ and $k \geq 1$ (k can be infinity), a subset $S \subseteq \mathbb{R}^m$ is called a **j-dimensional embedded C^k submanifold possibly with boundary of $\mathbb{R}^m$** if the following holds. For any point $p \in S$ there exists an open neighborhood $U \subseteq \mathbb{R}^m$ of p and a C^k bijective function $f : V \to U \cap S$ such that

  1. $\operatorname{rank} Df = j$ (i.e. Df is of maximal rank)
  2. Either

a. V is an open subset of $\mathbb{R}^j$

b. or V is an open subset of $\mathbb{H}^j = \{(u_1, \dots, u_j) : j \geq 0\}$ and $V \cap \partial\mathbb{H}^j \neq \emptyset$.

3. f is an embedding (see Note 3.27 below).

Such an f is called a C^k **parametrization** of S. For any such f,

$$\varphi = f^{-1} : U \cap S \to V$$

is called a C^k **chart** of S. If 2.a) above holds for V, then φ is called an **interior chart**. If 2.b) above holds for V, then φ is called a **boundary chart**. People also refer to S as a C^k **manifold possibly with boundary** embedded in $\mathbb{R}^m$. The variables $(u_1, \dots, u_j)$ in the domain of f (or range of φ) are referred to as **coordinates** of S associated to f or φ.

- **Note 3.27:** For property 3 above, recall from topology that an injective map is called an embedding if it's a homeomorphism onto its image (i.e. it is continuous and its inverse is continuous). The reason for requiring this is to prevent the weird figure 8 from being a manifold. *Important:* Recall from topology that if $S \subseteq \mathbb{R}^m$, then $W \subseteq S$ is open in S if and only if it is of the form $W = U \cap S$.

  The idea behind property 1 is that if we look at Df explicitly

$$Df = \begin{bmatrix} \dfrac{\partial f_1}{\partial u_1} & \cdots & \dfrac{\partial f_1}{\partial u_j} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial u_1} & \cdots & \dfrac{\partial f_m}{\partial u_j} \end{bmatrix}$$

  (note that this is a tall matrix since $m \geq j$), then $\operatorname{rank} Df = j$ implies that the columns here are linearly independent. As we'll discuss later, the columns represent velocity of curves along the manifold in different directions and hence for f to parametrize a j^{th}-dimensional manifold, intuitively speaking, you need these velocities to define a j-dimensional plane tangent to the surface (or else the surface could suffer a "collapsing effect").

  The reason for considering the two types of parametrization/charts in property 2 is so that we can define boundaries for manifolds (different from the notion of boundary in topology). We'll discuss this in detail soon.

- The reason for using charts to define and study manifolds is that they provide natural coordinates that we can use to perform calculations on manifolds. Passing one's attention to only using coordinates however has certain dangers since you need to make sure that when you define new concepts for manifolds in coordinates (e.g. connections, curvatures tensors, etc.), you need to make sure that your definitions do not depend on what charts you use. For this reason, it's important to understand how different coordinates are related to each other, which is the content of the following lemma. Before we can state it, we need a definition:

- **Definition 3.28:** If $k \geq 1$ (k can be infinity), a bijective C^k map $F : U \subseteq \mathbb{R}^m \to V \subseteq \mathbb{R}^m$, where U and V are open, that has a C^k inverse $F^{-1} : V \to U$ is called a C^k **diffeomorphism**.

- **Note 3.29:** In the context of the above definition, observe that since

$$F \circ F^{-1} = \mathrm{id} \quad \text{and} \quad F^{-1} \circ F = \mathrm{id},$$

by the chain rule we have that (here we omit arguments)

$$(DF)[D(F^{-1})] = \mathrm{id} \quad \text{and} \quad [D(F^{-1})](DF) = \mathrm{id},$$

and so both DF and $D(F^{-1})$ are invertible everywhere.

Observe also that we can reformulate the inverse function theorem (see Corollary 3.22) as the existence of diffeomorphisms obtained by restricting to smaller open neighborhoods $F : \widetilde{U} \to \widetilde{V}$.

- **Note 3.30:** "Diffeomorphisms" are the differential calculus version of "homeomorphisms." In particular, notice that since differentiability implies continuity, all diffeomorphisms are automatically homeomorphisms.

- **Lemma 3.31:** Suppose that S is a C^k j-dimensional manifold possibly with boundary embedded in $\mathbb{R}^m$. Suppose that $\varphi : U \cap S \to V$ and $\hat{\varphi} : \widehat{U} \cap S \to \widehat{V}$ are C^k charts of S and that $U \cap \widehat{U} \neq \emptyset$. Then the functions

$$\varphi \circ \hat{\varphi}^{-1} : \hat{\varphi}\big[U \cap \widehat{U} \cap S\big] \to \varphi\big[U \cap \widehat{U} \cap S\big]$$
$$\hat{\varphi} \circ \varphi^{-1} : \varphi\big[U \cap \widehat{U} \cap S\big] \to \hat{\varphi}\big[U \cap \widehat{U} \cap S\big]$$

are diffeomorphisms.

Proof: It's difficult to follow this proof without drawing a diagram, so please draw one as you read along! Let $f = \varphi^{-1}$ and $\hat{f} = \hat{\varphi}^{-1}$ be the parametrizations associated with these charts. We start by proving that $\varphi \circ \hat{\varphi}^{-1}$ is a C^k map, which note is equal to $\varphi \circ \hat{f}$. Fix any point $\hat{a} \in \hat{\varphi}\big[U \cap \widehat{U} \cap S\big]$, we will show that $\varphi \circ \hat{f}$ is C^k in a neighborhood of $\hat{a}$. Let $p = \hat{f}(a)$ and let $a = \varphi(p)$. Let's take a look at

$$Df = \begin{bmatrix} \dfrac{\partial f_1}{\partial u_1} & \cdots & \dfrac{\partial f_1}{\partial u_j} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial u_1} & \cdots & \dfrac{\partial f_m}{\partial u_j} \end{bmatrix}$$

(note that this is f, not $\hat{f}$). Since this matrix is of maximum rank j, some square $j \times j$ submatrix here is invertible. Let's suppose that it is the topmost $j \times j$ submatrix since the proofs in the other cases are similar. Consider the projection map $\pi : \mathbb{R}^m \to \mathbb{R}^j$ given by

$$\pi\big(x_1, \dots, x_j, x_{j+1}, \dots, x_m\big) = \big(x_1, \dots, x_j\big).$$

and consider the composition $\pi \circ f$. By the chain rule, the Jacobian matrix of $\pi \circ f$ is given by

$$D(\pi\circ f)=(D\pi)(Df)=\begin{bmatrix}\underbrace{\begin{matrix}1&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&1\end{matrix}}_{j\times j\text{ identity}}&\begin{matrix}0&\cdots&0\\ \vdots&\ddots&\vdots\\ 0&\cdots&0\end{matrix}\end{bmatrix}\begin{bmatrix}\frac{\partial f_1}{\partial u_1}&\cdots&\frac{\partial f_1}{\partial u_j}\\ \vdots&\ddots&\vdots\\ \frac{\partial f_m}{\partial u_1}&\cdots&\frac{\partial f_m}{\partial u_j}\end{bmatrix}=\begin{bmatrix}\frac{\partial f_1}{\partial u_1}&\cdots&\frac{\partial f_1}{\partial u_j}\\ \vdots&\ddots&\vdots\\ \frac{\partial f_j}{\partial u_1}&\cdots&\frac{\partial f_j}{\partial u_j}\end{bmatrix},$$

which we assumed is invertible. Hence by the inverse function theorem there exist neighborhoods $\mathcal{V}\subseteq V$ (open either in $\mathbb{R}^j$ or $\mathbb{H}^j$) of a and W of $\pi(p)$ such that the restriction $\pi\circ f:\mathcal{V}\to W$ is a diffeomorphism. Now consider the map

$$(3.32)\qquad \varphi\circ\hat{f}=(\pi\circ f)^{-1}\circ\pi\circ\hat{f}:\hat{f}^{-1}\big[\pi^{-1}[W]\big]\to V.$$

where observe that $\hat{f}^{-1}\big[\pi^{-1}[W]\big]$ is an open neighborhood of $\hat{a}$ since it's the preimage of an open set by continuous functions. Notice that (3.38) is a composition of C^k maps and hence is also C^k. As discussed above, this proves that $\varphi\circ\hat{\varphi}^{-1}$ is C^k.

Proving that $\hat{\varphi}\circ\varphi^{-1}$ is C^k is done similarly. Since $\varphi\circ\hat{\varphi}^{-1}$ and $\hat{\varphi}\circ\varphi^{-1}$ are inverses of each other, they are diffeomorphisms.

■

- Next we define tangent planes/spaces which, as we'll learn later, are used to define vector and tensor fields on manifolds that play a foundational role in differential geometry.

- **Definition 3.33:** Suppose that S is a C^k j-dimensional manifold possibly with boundary embedded in $\mathbb{R}^m$. Take any point $p\in S$. We define the **tangent plane at** p (or **tangent space at** p), denoted by T_pS, as follows. Take any parametrization $f:V\to S$ of S such that $p\in\operatorname{Im}f$. Then we define

$$(3.34)\qquad T_pS=\operatorname{span}\left\{\begin{pmatrix}\frac{\partial f_1}{\partial u_1}\\ \vdots\\ \frac{\partial f_m}{\partial u_1}\end{pmatrix},\dots,\begin{pmatrix}\frac{\partial f_1}{\partial u_j}\\ \vdots\\ \frac{\partial f_m}{\partial u_j}\end{pmatrix}\right\}.$$

Any vector $v\in T_pS$ is called a **tangent vector**. Note that since by definition Df is of maximal rank j, we have that T_pS is a j-dimensional plane (sitting in $\mathbb{R}^m$). The set of all tangent planes together is called the **tangent space of** S:

$$TS=\bigcup_{p\in S}T_pS.$$

- **Note 3.35:** We need to prove that the above definition is well defined, in particular that T_pS does not depend on the parametrization that you chose. Suppose that $\hat{f}:V\to S$ is another parametrization such that $p\in\operatorname{Im}f$. We need to prove that

$$(3.36) \qquad \operatorname{span}\left\{\begin{pmatrix}\frac{\partial f_1}{\partial u_1}\\ \vdots \\ \frac{\partial f_m}{\partial u_1}\end{pmatrix}, \dots, \begin{pmatrix}\frac{\partial f_1}{\partial u_j}\\ \vdots \\ \frac{\partial f_m}{\partial u_j}\end{pmatrix}\right\} = \operatorname{span}\left\{\begin{pmatrix}\frac{\partial \hat{f}_1}{\partial \hat{u}_1}\\ \vdots \\ \frac{\partial \hat{f}_m}{\partial \hat{u}_1}\end{pmatrix}, \dots, \begin{pmatrix}\frac{\partial \hat{f}_1}{\partial \hat{u}_j}\\ \vdots \\ \frac{\partial \hat{f}_m}{\partial \hat{u}_j}\end{pmatrix}\right\}.$$

First let's prove that the inclusion "⊆" here holds. This will follow if we show that each vector $(\partial f_1/\partial u_k, \dots, \partial f_m/\partial u_k)$ inside the span on the left-hand side is contained in the "span" on the right-hand side (here $1 \le k \le j$). Let φ and $\hat{\varphi}$ be the charts associated to f and $\hat{f}$ respectively. Observe that

$$\begin{pmatrix}\frac{\partial f_1}{\partial u_k}\\ \vdots \\ \frac{\partial f_m}{\partial u_k}\end{pmatrix} = \underbrace{\begin{pmatrix}\frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_j}\\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial u_1} & \cdots & \frac{\partial f_m}{\partial u_j}\end{pmatrix}}_{Df} \underbrace{\begin{pmatrix}\left.\begin{matrix}0\\ \vdots \\ 1\end{matrix}\right\} k\\ 0 \\ \vdots \\ 0\end{pmatrix}}_{e_k} = (Df)e_k = D\left(\hat{f} \circ (\hat{\varphi} \circ \varphi^{-1})\right) e_k$$

$$= \left(D\hat{f}\right) \underbrace{(D[\hat{\varphi} \circ \varphi])e_k}_{\text{call this } v} = \begin{pmatrix}\frac{\partial \hat{f}_1}{\partial \hat{u}_1} & \cdots & \frac{\partial \hat{f}_1}{\partial \hat{u}_j}\\ \vdots & \ddots & \vdots \\ \frac{\partial \hat{f}_m}{\partial \hat{u}_1} & \cdots & \frac{\partial \hat{f}_m}{\partial \hat{u}_j}\end{pmatrix}\begin{pmatrix}v_1\\ \vdots \\ v_j\end{pmatrix} = v_1\begin{pmatrix}\frac{\partial \hat{f}_1}{\partial \hat{u}_1}\\ \vdots \\ \frac{\partial \hat{f}_m}{\partial \hat{u}_1}\end{pmatrix} + \dots + v_j\begin{pmatrix}\frac{\partial \hat{f}_1}{\partial \hat{u}_j}\\ \vdots \\ \frac{\partial \hat{f}_m}{\partial \hat{u}_j}\end{pmatrix}$$

The right-hand side is in the "span" on the right-hand side of (3.36), and hence we've proved the inclusion "⊆" in (3.36). The inclusion "⊇" there is proved similarly, and so indeed (3.36) holds.

- To prove Stokes's Theorem later, we will need the notion of a boundary of a manifold:
- **Definition 3.37:** Suppose that S is C^k j-dimensional manifold possibly with boundary embedded in $\mathbb{R}^m$. A point $p \in S$ is called a **boundary point** if there exists a boundary chart φ that contains p in its domain and

$$(3.38) \qquad \varphi(p) \in \partial \mathbb{H}^j = \{(u_1, \dots, u_j) : u_j = 0\}.$$

A point $p \in S$ is called an **interior point** if it is not a boundary point. The set of all boundary points of S is called the **boundary of** S:

$$\partial S = \{p \in S : p \text{ is a boundary point of } S\}$$

Warning: The notion of boundary point and interior point are not the same thing as those concepts defined in topology with the same name. There is a way to connect these two concepts by placing S inside a bigger manifold of the same dimension, though we will not pursue this question in this course.

- **Lemma 3.39:** Suppose that S is C^k j-dimensional manifold possibly with boundary embedded in $\mathbb{R}^m$. If $p \in S$ is a boundary point, then any chart φ that contains p in its domain is a boundary chart that satisfies (3.38).

  Proof: You will prove this on the HW, or I will come back and prove it. You can use result on the homework and exams. ■

- **Theorem 3.40:** Suppose that S is C^k j-dimensional manifold with boundary embedded in $\mathbb{R}^m$. Then the boundary ∂S is a C^k $(j-1)$-dimensional manifold without boundary embedded in $\mathbb{R}^m$.

  Proof: We leave this as an exercise while listing out the main arguments. This is proved by taking a parametrization $f : U \subseteq \mathbb{H}^j \to S$ such that $\varphi = f^{-1}$ is a boundary chart and showing that its restriction $\tilde{f} : U \cap \partial\mathbb{H}^j \to \partial S$ is a parametrization of ∂S. The fact that $\tilde{f}$ maps into ∂S is given by Lemma 3.39. To show that $\tilde{f}$ is of maximum rank, simply observe that $D\tilde{f}$ consists of the first $j-1$ columns of Df and hence is of max rank. The reason ∂S does not have boundary is that $U \cap \partial\mathbb{H}^j$ (i.e. the domain of $\tilde{f}$) can be viewed as a subset of $\mathbb{R}^{j-1}$. We leave the details to the reader.

  ■

4 Integration in Several Variables

- We digress from the theory of manifolds for a while to study integration in several variables. As the next definition shows, the construction of integration in higher dimensions is very similar to its single variable cousin. In several cases we will formulate the definitions/theorems/proofs in $\mathbb{R}^2$ and then note that the generalization to higher dimensions $\mathbb{R}^m$ is trivial.

- **Definition 4.1:** Suppose that $R = [a,b] \times [c,d] \subseteq \mathbb{R}^2$ is a closed rectangle (or "closed box") in $\mathbb{R}^2$ (here a, b, c, d are finite) and that $f : R \to \mathbb{R}$ is a <u>bounded</u> function. Consider a partition $a = x_0 < x_1 < \cdots < x_J \leq b$ of $[a,b]$ and a partition $c = y_0 < y_1 < \cdots < y_K = d$ of $[c,d]$. Together:

$$P = \{x_0, x_1, \dots, x_j; y_0, y_1, \dots, y_K\}$$

is called a **partition of** R. This partition generates sub-rectangles of R:

$$R_{jk} = [x_j, x_{j+1}] \times [y_k, y_{k+1}]$$

whose area we denote and define as $\Delta A_{jk} = (x_{j+1} - x_j)(y_{k+1} - y_k)$.

The **lower and upper Riemann sums of** f **over** R corresponding to the partition P are respectively defined as:

$$s_P f = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \inf_{x \in R_{jk}} \{f(x)\} \Delta A_{jk} \quad \text{and} \quad S_P f = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \sup_{x \in R_{jk}} \{f(x)\} \Delta A_{jk}.$$

We define the **lower and upper Riemann integrals of** f **over** R respectively as

$$\underline{I}_R f = \sup_P s_P f \quad \text{and} \quad \overline{I}_R f = \inf_P S_P f.$$

It's easy to show that $s_P f \leq S_Q f$ for any two partitions P and Q of R – the proof is essentially identical as in the single variable theory. Hence $\underline{I}_R f \leq \overline{I}_R f$. If the lower and upper Riemann integrals of f are equal: $\underline{I}_R f = \overline{I}_R f$, then we say that f is **(Riemann) integrable on** R and we define the **(Riemann) integral of** f **over** R as

$$\iint_R f\, dA = \iint_R f(x,y) dx dy = \underline{I}_R f = \overline{I}_R f.$$

- **Note 4.2:** The generalization of the above definition to $\mathbb{R}^m$ is trivial. Simply use closed boxes of the form

$$R = [a_1, b_1] \times \dots \times [a_m, b_m] \subseteq \mathbb{R}^m$$

If one takes a partition $a_i = (x_i)_0 < (x_i)_1 < \cdots < (x_i)_{J_i} = b_i$ of each $[a_i, b_i]$, then

$$P = \{(x_1)_0, \dots, (x_1)_{J_1}; \dots; (x_m)_0, \dots, (x_m)_{J_i}\}$$

forms a partition of R. In this case, the definition defines integrability of a bounded function of the form $f : R \to \mathbb{R}$. In the case of $\mathbb{R}^3$, we write $\iiint_R f\, dV$ for the integral of f. In dimensions 4 and higher, we typically simply write $\int_R f\, dV$. Other notation for the integral include

$$\int_R f, \quad \int_R f(x)\, dx_1 \dots dx_m, \quad \int_R f(x) d^m x.$$

(warning: the second and third integrals are not iterated – we'll discuss iterated integrals soon).

- In practice, integration in multiple variables needs to be done over more complicated regions rather than simply boxes. The simplest way to perform this generalization is through the notion of characteristic/indicator functions.

- **Definition 4.3:** Suppose that $S \subseteq \mathbb{R}^m$ is a set. The **characteristic/indicator function** of S is the function $\chi_S : \mathbb{R}^m \to \mathbb{R}$ given by

$$\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases}$$

An alternative notation for χ_S is $\mathbb{1}_S$.

- **Definition 4.4:** Suppose that $S \subseteq \mathbb{R}^m$ is a bounded set and that $f : S \to \mathbb{R}$ is a bounded function. Let R be a closed box that contains S (i.e. $S \subseteq R$). We say that f is **(Riemann) integrable on** S if $(f \cdot \chi_S) : R \to \mathbb{R}$ is integrable (on R), in which case we define the **Riemann integral of f over S** as

$$\int_S f\, dV = \int_R f \chi_S dV.$$

- **Note 4.5:** One has to check that the above definition is well defined, in particular that it does not depend on the closed box that you choose. Precisely one has to show that if $\hat{R}$ is another closed box that contains S, then either both $f\chi_S : R \to \mathbb{R}$ and $f\chi_S : \hat{R} \to \mathbb{R}$ are integrable or both are not integrable, and if both are integrable then

$$\int_R f\chi_S dV = \int_{\hat{R}} f\chi_S dV.$$

This is not hard to do; we leave it to the reader to work out the details if they're interested. In particular, this follows from the fact that $f\chi_S : R \to \mathbb{R}$ and $f\chi_S : \hat{R} \to \mathbb{R}$ are equal on S and are zero everywhere else.

- **Theorem 4.6:** The following are true:

    a) (Linearity) If $S \subseteq \mathbb{R}^m$ is a bounded set, $f_1, f_2 : S \to \mathbb{R}$ are integrable (and hence bounded), and $c_1, c_2 \in \mathbb{R}$, then

$$\int_S (c_1 f_1 + c_2 f_2) = c_1 \int_S f_1 + c_2 \int_S f_2.$$

b) If $S_1, S_2 \subseteq \mathbb{R}^m$ are bounded sets with no points in common (i.e. $S_1 \cap S_2 = \emptyset$) and f is an integrable function on S_1 and S_2, then f is integrable on $S_1 \cup S_2$ and

$$\int_{S_1 \cup S_2} f = \int_{S_1} f + \int_{S_2} f.$$

c) If $S \subseteq \mathbb{R}^m$ is a bounded set, $f, g : S \to \mathbb{R}$ are integrable, and $f(x) \leq g(x)$ everywhere, then

$$\int_S f \leq \int_S g$$

d) If $S \subseteq \mathbb{R}^m$ is a bounded set and $f : S \to \mathbb{R}$ is integrable, then

$$\left| \int_S f \right| \leq \int_S |f|.$$

- We have the definition of integrability, however we have no examples of integrable functions at the moment. We will prove soon that continuous functions over closed boxes are integrable, but what about continuous functions over bounded sets $S \subseteq \mathbb{R}^m$? We defined such integrals as $\int_R f\chi_S$ where R is a closed box that contain S. The issue is that even if $f : S \to \mathbb{R}$ is continuous, $f\chi_S$ is typically discontinuous on the boundary ∂S and hence it's not guaranteed that $f\chi_S$ will be integrable over R. The typical way to fix this issue is to assume that ∂S (i.e. the set of discontinuity of $f\chi_S$) is negligibly small with respect to integration. A precise way to do this is using the concept of zero content:

- **Definition 4.7:** A set $Z \subseteq \mathbb{R}^m$ is said to have **zero content** if for any $\varepsilon > 0$ there exist a <u>finite</u> collection of boxes $R_1, \dots, R_k \subseteq \mathbb{R}^m$ such that $Z \subseteq \bigcup_{i=1}^k R_i$ (i.e. the R_i's cover Z) and the sum of the areas/volumes of the boxes $R_1, \dots, R_k$ is less than ε (we define area/volume of a box as the product of the lengths of all of its sides). You can use open or closed boxes here, the definition is equivalent.

- In application, common examples of sets with zero content are manifolds in Euclidean spaces (such as curves and surfaces) since they will be forming boundaries of regions of integration such as S above.

- **Theorem 4.8:** Suppose that $R \subseteq \mathbb{R}^m$ is a closed box and that $f : R \to \mathbb{R}$ is a bounded function that is continuous everywhere in R except on a set $Z \subseteq R$ of Jordan content zero. Then f is integrable.

  Proof: We will do the proof in $\mathbb{R}^2$, the proof in $\mathbb{R}^m$ is similar. It will suffice to show that for any $\varepsilon > 0$ there exists a partition P of R such that

$$(4.9) \qquad S_P f - s_P f \le \varepsilon.$$

Let $R_1 = (a_1, b_1) \times (c_1, d_1), \dots, R_k = (a_k, b_k) \times (c_k, d_k)$ be boxes that cover S such that the sum of their areas is less than ε. Consider the restriction $f : R \setminus \bigcup_{i=1}^k R_i \to \mathbb{R}$, which is continuous and hence uniformly continuous since its domain is compact. Thus there exists a $\delta > 0$ such that if $x, y \in R \setminus \bigcup_{i=1}^k R_i$ and $|x - y| < \delta$, then $|f(x) - f(y)| < \varepsilon$. Writing $R = [a, b] \times [c, d]$, take a partition

$$(4.10) \qquad P' = \{a = x_0 < x_1 < \dots < x_{J'} = b; c = y_0 < y_1 < \dots < y_{K'} = d\}$$

of R such that the diameter of each sub-rectangle R_{jk} (i.e. longest length inside) is less than δ. Throw in $a_1, \dots, a_k, b_1, \dots, b_k$ and $c_1, \dots, c_k, d_1, \dots, d_k$into the above partitions of $[a, b]$ and $[c, d]$ respectively (i.e. that are sitting in the "{...}" in (4.10)) to get a new partition P of R.

Now, let $B > 0$ be a constant such that $|f(x)| \le B$ everywhere. Then

$$S_P f - s_P f = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \left(\sup_{x \in R_{jk}} \{f(x)\} - \inf_{x \in R_{jk}} \{f(x)\} \right) \Delta A_{jk}$$

$$= \underbrace{\sum_{R_{jk} \in \bigcup_{j=1}^k R_j} \underbrace{\left(\sup_{x \in R_{jk}} \{f(x)\} - \inf_{x \in R_{jk}} \{f(x)\} \right)}_{\le 2B} \Delta A_{jk}}_{2B \sum_{R_{jk} \in \bigcup_{j=1}^k R_j} \Delta A_{jk} = 2B \cdot \left(\text{Area of } \bigcup_{j=1}^k R_j\right)} + \underbrace{\sum_{R_{jk} \notin \bigcup_{j=1}^k R_j} \underbrace{\left(\sup_{x \in R_{jk}} \{f(x)\} - \inf_{x \in R_{jk}} \{f(x)\} \right)}_{\le \varepsilon} \Delta A_{jk}}_{\le \varepsilon \sum_{R_{jk} \notin \bigcup_{j=1}^k R_j} \Delta A_{jk} \le \varepsilon \cdot (\text{Area of } R)}$$

$$\le 2B\varepsilon + \varepsilon(b - a)(d - c) = \big(2B + (b - a)(d - c)\big)\varepsilon.$$

Oops… we didn't get the last quantity to be $\varepsilon > 0$ as desired on the right-hand side of (4.9), but that ok: simply go back and divide every instance of ε appropriately by B or $(b - a)(d - c)/2$ to make this happen.

■

- We now seek to extend Theorem 4.8 to more general bounded sets $S \subseteq \mathbb{R}^m$.
- **Lemma 4.11:** Suppose that $S \subseteq \mathbb{R}^m$. The characteristic function $\chi_S : \mathbb{R}^m \to \mathbb{R}$ is continuous everywhere except on ∂S.

  Proof: This is a basis exercise in topology and continuity and is left to the reader. You can find a proof written out on page 162 of the book. ■
- **Definition 4.12:** A set $S \subseteq \mathbb{R}^m$ is called **Jordan measurable** if it is bounded and its boundary ∂S has Jordan content zero.
- **Theorem 4.13:** Suppose that $S \subseteq \mathbb{R}^m$ is Jordan measurable and that $f : S \to \mathbb{R}$ is continuous and bounded. Then f is integrable on S.

Proof: Take any closed rectangle $R \subseteq \mathbb{R}^m$ that contains S (i.e. $S \subseteq R$). By Lemma 4.11 $f\chi_S : R \to \mathbb{R}$ is discontinuous only on ∂S, and observe that ∂S has Jordan content zero since we're assuming that S is Jordan measurable. Hence by Theorem 4.8, $f\chi_S : R \to \mathbb{R}$ is integrable on R. Thus f is integrable on S.

■

- **Note 4.14:** Over the break, please read the following (it's not a long reading).
  - The statement of Proposition 4.22 in the book.
  - The statement and proof of Corollary 4.23 in the book.
  - The statement and proof of Theorem 4.24 in the book.
  - The statement of Corollary 4.25 in the book (volume of S is defined as the quantity $\int_S 1$. Perhaps you can see why now, or simply take it as a definition and we'll talk about it after the break).
- **Definition 4.15:** Suppose that $S \subseteq \mathbb{R}^m$ is a Jordan measurable set. Then we define

$$\text{Area}(S) = \int_S 1\, dA \quad \text{if} \;\; m = 2,$$

$$\text{Volume}(S) = \int_S 1\, dV \quad \text{if} \;\; m > 2.$$

- Using the upper and lower Riemann sums of $\int_S 1$, one can in addition define outer and inner areas/volumes respectively, however we don't pursue this topic. The interested reader can find a description on page 164 in the textbook.

4.1 Fubini's Theorem and Iterated Integrals

- **Note 4.16:** We've defined integrals of functions of several variables, but at the moment we have no convenient way of actually computing them. This is done by "iterating" integrals, for which the idea is the following. Suppose we want to integrate a continuous two-variable function f over a closed rectangle $R = [a, b] \times [c, d] \subseteq \mathbb{R}^2$. The integral $\int_R f\, dA$ represents the volume underneath the graph of f. Intuitively speaking, this can be obtained by taking the volumes of thin slices along the y-axis of width Δx, which are given by $\int_c^d f(x, y) dy\, \Delta x$, and then adding up the slices to get the full volume. Making the steps Δx smaller and smaller and passing to the limit should give an integral in x, which leads to the equality:

$$\int_R f\, dA = \int_a^b \int_c^d f(x, y) dy\, dx.$$

This is useful because the expression on the right-hand side, called an **iterated integral**, can be computed using single variable theory. The theorem that allows us to iterate integrals like this is called "Fubini's theorem," which we'll prove below. We want to point out that there is nothing special about the ordering of first integrating in y and then in x. One can redo the above logic to justify integration in the other order to get:

$$\int_c^d \int_a^b f(x,y)dx\,dy = \int_R f\,dA = \int_a^b \int_c^d f(x,y)dy\,dx.$$

Sometimes in math iterated integrals arise naturally without passing through the multidimensional Riemann integral (i.e. Definition 4.1), and for this reason Fubini's theorem is often thought of as the theorem that justifies interchanging order of integration. We note that some people write the above iterated integrals instead as

$$\int_a^b dx \int_c^d dy\, f(x,y) \quad \text{and} \quad \int_c^d dy \int_a^b dx\, f(x,y).$$

We will avoid this notation.

- **Note 4.17:** Continuing off the previous note, the idea behind the proof of Fubini' Theorem will be the following. Take a very fine partition

$$P = \{a = x_0 < x_1 < \cdots < x_J = b; c = y_0 < y_1 < \cdots < y_K = d\}$$

of R. Then we have that

$$\int_R f\,dA \approx \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} f(x_j,y_k)\underbrace{(x_{j+1}-x_j)(y_{k+1}-y_k)}_{A_{jk}} = \sum_{j=0}^{J-1}(x_{j+1}-x_j)\sum_{k=0}^{K-1} f(x_j,y_k)(y_{k+1}-y_k)$$

$$\approx \sum_{j=0}^{J-1}\int_c^d f(x_j,y)dy\,(x_{j+1}-x_j) \approx \int_a^b\int_c^d f(x,y)dy\,dx.$$

The trick will be to carefully justify all of the "≈." First we need a technical lemma about infimums and supremums:

- **Lemma 4.18:** The following are true:
  - If $f : [a,b]\times[c,d] \to \mathbb{R}$ is a function, then

$$\inf_{(x,y)\in[a,b]\times[c,d]}\{f(x,y)\} = \inf_{x\in[a,b]}\inf_{y\in[c,d]}\{f(x,y)\},$$

$$\sup_{(x,y)\in[a,b]\times[c,d]}\{f(x,y)\} = \sup_{x\in[a,b]}\sup_{y\in[c,d]}\{f(x,y)\}.$$

  - If $\{r_\alpha\}_{\alpha\in\mathrm{A}}$ and $\{s_\alpha\}_{\alpha\in\mathrm{A}}$ are sets of real numbers, then

$$\inf_{\alpha\in A}\{r_\alpha\} + \inf_{\alpha\in A}\{s_\alpha\} \le \inf_{\alpha\in A}\{r_\alpha + s_\alpha\},$$

$$\sup_{\alpha\in A}\{r_\alpha\} + \sup_{\alpha\in A}\{s_\alpha\} \ge \sup_{\alpha\in A}\{r_\alpha + s_\alpha\}.$$

This works for bigger finite sums as well.

Proof: You will prove this on the homework. ■

- **Theorem 4.19: (Fubini's Theorem in 2D)** Suppose that $R = [a,b] \times [c,d] \subseteq \mathbb{R}^2$ is a closed rectangle and that $f : R \to \mathbb{R}$ is an integrable function. Suppose that for any fixed $x \in [a,b]$, the function $f_x : [c,d] \to \mathbb{R}$ given by $f_x(y) = f(x,y)$ is integrable. Then

$$(4.20) \qquad \int_R f\, dA = \int_a^b \int_c^d \underbrace{f(x,y)}_{f_x(y)}\, dy\, dx.$$

Similarly, if for any fixed $y \in [c,d]$ the function $f_y : [a,b] \to \mathbb{R}$ given by $f_y(x) = f(x,y)$ is integrable, then

$$(4.21) \qquad \int_R f\, dA = \int_c^d \int_a^b \underbrace{f(x,y)}_{f_y(x)}\, dx\, dy.$$

Proof: The following is a proof that I learned from Jim Morrow. We will only prove (4.20) since (4.21) is proven similarly. Take any partition $P = \{a = x_0 < x_1 < \cdots < x_J = b\}$ of $[a,b]$ and any partition $Q = \{c = y_0 < y_1 < \cdots < y_K = d\}$ of $[c,d]$ and let $P \times Q$ denote the partition of R given by

$$P \times Q = \{a = x_0 < x_1 < \cdots < x_J = b; c = y_0 < y_1 < \cdots < y_K = d\}.$$

Then

$$s_{P\times Q} f = \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} \inf_{(x,y)\in R_{jk}}\{f(x,y)\}\underbrace{(x_{j+1}-x_j)(y_{k+1}-y_k)}_{A_{jk}}$$

$$\overset{\text{Prev. lemma}}{=} \sum_{j=0}^{J-1}\sum_{k=0}^{K-1} \inf_{x\in[x_j,x_{j+1}]}\ \inf_{y\in[y_k,y_{k+1}]}\{f(x,y)\}\,(x_{j+1}-x_j)(y_{k+1}-y_k)$$

$$\overset{\text{Prev. lemma}}{\le} \sum_{j=0}^{J-1} \inf_{x\in[x_j,x_{j+1}]}\left[\sum_{k=0}^{K-1} \inf_{y\in[y_j,y_{j+1}]}\{f(x,y)\}\,(y_{k+1}-y_k)\right](x_{j+1}-x_j)$$

$$= s_P\left(s_Q(f_x)\right) \le s_P\left(\underline{I}_c^d f_x\right).$$

Following similar logic one can show that

$$S_{P\times Q}f \geq S_P\left(\overline{I}_c^d f_x\right).$$

So in total we get that

$$(4.22) \qquad s_{P\times Q}f \leq s_P\left(\underline{I}_c^d f_x\right) \leq S_P\left(\underline{I}_c^d f_x\right) \leq S_P\left(\overline{I}_c^d f_x\right) \leq S_{P\times Q}f,$$

$$(4.23) \qquad s_{P\times Q}f \leq s_P\left(\underline{I}_c^d f_x\right) \leq s_P\left(\overline{I}_c^d f_x\right) \leq S_P\left(\overline{I}_c^d f_x\right) \leq S_{P\times Q}f.$$

Because f is integrable, for any $\varepsilon > 0$ we can choose P and Q fine enough so that $S_{P\times Q}f - s_{P\times Q}f < \varepsilon$ which by the above inequalities will force both

$$S_P\left(\underline{I}_c^d f_x\right) - s_P\left(\underline{I}_c^d f_x\right) < \varepsilon \quad \text{and} \quad S_P\left(\overline{I}_c^d f_x\right) - s_P\left(\underline{I}_c^d f_x\right) < \varepsilon.$$

Hence both $\underline{I}_c^d f_x$ and $\overline{I}_c^d f_x$ are integrable over $x \in [a,b]$. Moreover, since by (4.22) and (4.23) their Riemann sums are stuck in between $s_{P\times Q}f$ and $S_{P\times Q}f$, we furthermore get that

$$\int_R f\, dA = \int_a^b \underline{I}_c^d f_x\, dx = \int_a^b \overline{I}_c^d f_x\, dx.$$

Since we assumed that each f_x is integrable, we can plug in $\underline{I}_c^d f_x = \overline{I}_c^d f_x = \int_c^d f_x(y)dy = \int_c^d f(x,y)dy$ to prove the theorem.

■

- **Theorem 4.24: (Fubini's Theorem)** Suppose that $R = [a_1,b_1] \times \dots \times [a_m,b_m] \subseteq \mathbb{R}^m$ is a closed box and that $f : R \to \mathbb{R}$ is an integrable function. Suppose also that for any fixed index $i \in \{1,\dots,m-1\}$ and any fixed

$$x_j \in \left[a_j, b_j\right] \quad \text{for} \quad j \leq i,$$

the function $f_{x_1,\dots,x_i}(x_{i+1},\dots,x_m) = f(x_1,\dots,x_m)$ is integrable over $[a_{i+1},b_{i+1}] \times \dots \times [a_m,b_m]$. Then

$$\int_R f\, dV = \int_{a_1}^{b_1} \cdots \int_{a_m}^{b_m} f(x_1,\dots,x_m)dx_1 \dots dx_m.$$

Proof: This is proved by adding induction into the proof of the previous theorem. In particular, in the first step the P will be a partion of $[a_1,b_1]$ and Q will be a partition of $[a_2,b_2] \times \dots \times [a_m,b_m]$.

■

4.2 Change of Variables for Integrals

- **Note 4.25:** Sometimes evaluating multivariable integrals in a given coordinate system isn't convenient (e.g. in Euclidiean coordinates), however passing to a different coordinate system (e.g. polar, cylindrical coordinates, etc.) can make the integration much more manageable. For instance, suppose you want to compute

$$(4.26) \qquad \iint_{B(0;2)} f(x,y)dA\,, \quad \text{where} \quad f(x,y) = 4 - x^2 - y^2$$

where $B(0;2)$ denotes the ball of radius 2 centered at 0. Iterating this integral will prove messy, however if you "change variables" by making the polar coordinates substitution $x = r\cos(\theta)$ and $y = r\sin\theta$ it seems that the above integral should be equal to something like

$$\iint_{[0,2]\times[0,2\pi]} f(r\cos\theta\,, r\sin\theta)(???) = \int_0^{2\pi}\int_0^2 (4 - r^2)(???).$$

This looks like a much simpler integral, but the question is what should go into the $(???)$?

Although you most likely haven't formulated it this way, you actually already did this in single variable integration theory when you studied u-substitution. Suppose you have a one variable function $f : [a,b] \to \mathbb{R}$ and you want to compute:

$$\int_{[a,b]} f\, dx = \int_a^b f(x)dx.$$

Let us "change variables" by substituting $x = g(u)$ where $g : [c,d] \to [a,b]$ is a C^1 bijective function. In your one variable analysis class you most likely had a homework problem proving that either

$$g(c) = a \text{ and } g(d) = b, \text{ and } g' \geq 0 \text{ everywhere}$$
$$\text{or} \quad g(c) = b \text{ and } g(d) = a, \text{ and } g' \leq 0 \text{ everywhere}$$

Thus from calculus we have that

$$\int_c^d f\big(g(u)\big)g'(u)du = \int_{g(c)}^{g(d)} f(x)dx = \begin{cases} \displaystyle\int_a^b f(x)dx & \text{if } \; g' \geq 0 \text{ everywhere} \\ \displaystyle -\int_a^b f(x)dx & \text{if } \; g' \leq 0 \text{ everywhere} \end{cases}$$

Thus

$$\int_{[a,b]} f\, dx = \int_a^b f(x)dx = \int_c^d f\big(g(u)\big)|g'(u)|du = \int_{[c,d]} f\big(g(u)\big)|g'(u)|du$$

$$= \int_{g^{-1}[a,b]} f\big(g(u)\big)|g'(u)|du.$$

Intuitively speaking, as you perform a Riemann sum in u, the horizontal steps in the associated Riemann sum in the x space will be stretched by factors of $|g'(u)|$. This is why the term $|g'(u)|du$ appears in the last integral. Whatever the intuition, this formula generalizes directly to higher dimensions:

- **Theorem 4.27:** Suppose $G : U \to V$ is a C^1 bijective map between two open sets $U, V \subseteq \mathbb{R}^m$ such that DG is invertible everywhere (as a matrix) (this in fact implies that G is a C^1 diffeomorphism by the inverse function theorem). Suppose also that $T \subseteq U$ and $S \subseteq V$ are Jordan measurable sets such that $\bar{T} \subseteq U$ and $T = G^{-1}(S)$. Then for any integrable function $f : S \to \mathbb{R}$, $f \circ G$ is integrable over $T = G^{-1}(S)$ and

$$(4.28) \qquad \int_S f(x)d^m x = \int_{T=G^{-1}(S)} f\big(G(u)\big)|\det DG(u)|d^m u.$$

- **Note 4.29:** We won't prove the above theorem: it would take approximately 1.5 - 2 weeks. You will see a simpler proof in a class on Lebesgue integrals where you'll prove a more general version of the above theorem. However, let us discuss the intuition for where the term $|\det DG(u)|$ comes from on the right-hand side. Suppose $m = 2$ and write $G = \big(G_1(u_1, u_2), G_2(u_1, u_2)\big)$. Similar to the remark made at the end of Note 4.25, as we perform Riemann sums in u, the map G will take a small rectangle in the partition with dimensions du_1 and du_2 to a parallelogram with sides

$$G(u_1 + du_1, u_2) - G(u_1, u_2) = \begin{pmatrix} G_1(u_1 + du_1, u_2) - G_1(u_1, u_2) \\ G_1(u_1 + du_1, u_2) - G_1(u_1, u_2) \end{pmatrix} = \begin{pmatrix} \dfrac{\partial G_1}{\partial u_1} du_1 \\ \dfrac{\partial G_1}{\partial u_1} du_1 \end{pmatrix}$$

$$\text{and} \quad G(u_1, u_2 + du_2) - G(u_1, u_2) = \cdots = \begin{pmatrix} \dfrac{\partial G_1}{\partial u_2} du_2 \\ \dfrac{\partial G_1}{\partial u_2} du_2 \end{pmatrix}.$$

The volume of such a parallelogram is

$$\left| \det \begin{pmatrix} \dfrac{\partial G_1}{\partial u_1} du_1 & \dfrac{\partial G_1}{\partial u_2} du_2 \\ \dfrac{\partial G_1}{\partial u_1} du_1 & \dfrac{\partial G_1}{\partial u_2} du_2 \end{pmatrix} \right| = \left| \det \begin{pmatrix} \dfrac{\partial G_1}{\partial u_1} & \dfrac{\partial G_1}{\partial u_2} \\ \dfrac{\partial G_1}{\partial u_1} & \dfrac{\partial G_1}{\partial u_2} \end{pmatrix} \right| du_1 du_2 = |\det DG| du_1 du_2.$$

Hence this is the "scaled" differential in the x-space that appears on the right-hand side of (4.28).

- **Example 4.30:** Here are some famous and important change of variables:
    1. Polar/spherical coordinates in $\mathbb{R}^2$: $G(r,\theta) = (r\cos\theta\,, r\sin\theta)$, where $|\det DG| = r$.
    2. Cylindrical coordinates in $\mathbb{R}^3$: $G(r,\theta,z) = (r\cos\theta\,, r\sin\theta\,, z)$, where $|\det DG| = r$.
    3. Spherical coordinates in $\mathbb{R}^3$: $G(r,\varphi,\theta) = (r\sin\varphi\cos\theta\,, r\sin\varphi\sin\theta\,, r\cos\varphi)$, where $|\det DG| = r^2\sin\varphi$.
    4. Spherical coordinates in $\mathbb{R}^m$:

$$G(r,\theta_1,\dots,\theta_{m-1}) = \big(r(C^\infty \text{ expression in } \theta_1,\dots,\theta_{m-1}),\dots,r(C^\infty \text{ expression in } \theta_1,\dots,\theta_{m-1})\big)$$

with $|\det DG| = r^{m-1}(C^\infty \text{ expression in } \theta_1,\dots,\theta_{m-1})$.

- **Example 4.31:** Computing the integral (4.26) in polar coordinates gives

$$\iint_{B(0;2)} (4 - x^2 - y^2)dxdy = \iint_{[0,2]\times[0,2\pi]} (4 - r^2)r\,drd\theta = \int_0^{2\pi}\int_0^2 (4 - r^2)rdr\,d\theta = 8\pi.$$

4.3 Improper Integrals

- Sometimes we need to perform integrals over unbounded domains and/or of unbounded functions. These are not classic Riemann integrals: they are called "improper" integrals and are defined via limit operations. Instead of constructing a sophisticated theory of improper integrals, we will take a more simplistic approach that is sufficient virtually for all applications (including all of my research experiences). We will mention in Note 4.37 below what could be refined.
- **Notation 4.32:** We let $B(a;r) \subseteq \mathbb{R}^m$ denote the open ball centered at $a \in \mathbb{R}^m$ of radius $r > 0$.
- First we look at integrals that are improper at infinity:
- **Definition 4.33:** Suppose that $f : \mathbb{R}^m \to \mathbb{R}$ is a function that is integrable over any ball of the form $B(0;r)$. In our class, we define the following **improper integral** as the formal expression:

$$\int_{\mathbb{R}^m} f = \lim_{r\to\infty} \int_{B(0;r)} f$$

If the limit exists, we say that the integral **converges** and define its value to be the limit, otherwise we say that the integral **diverges**. Sometimes people simply write $\int f$ instead of $\int_{\mathbb{R}^m} f$.

- Sometimes you may want to exhaust $\mathbb{R}^m$ in other ways such as by balls centered at some other point or as a limit of integrals over expanding boxes. The following proposition provides a condition for when these give the same answer. In general, the condition that the absolute value of something converges is called **absolute convergence**.

- **Proposition 4.34:** Suppose that $f : \mathbb{R}^m \to \mathbb{R}$ is a function such that the improper integral $\int_{\mathbb{R}^m}|f|$ converges (i.e. $\int_{\mathbb{R}^m} f$ is absolutely convergent). Then the improper integral $\int_{\mathbb{R}^m} f$ converges and for any point $a \in \mathbb{R}^m$

$$\int_{\mathbb{R}^m} f = \lim_{r\to\infty} \int_{B(a;r)} f = \lim_{r\to\infty} \int_{[-r,r]\times\ldots\times[-r,r]} f.$$

  Proof: You will prove this on the homework. ■

- Now we turn to improper integrals that arise from singularities:

- **Definition 4.35:** Suppose that $S \subseteq \mathbb{R}^m$ is a measurable set, $a \in S^{\text{interior}}$, and $f : S \setminus \{a\} \to \mathbb{R}$ is a function that is integrable over $S \setminus B(a;r)$ for any $r > 0$ such that $B(a;r) \subseteq S$. In our class, we define the following **improper integral** as the formal expression

$$\int_{S\setminus\{a\}} f = \lim_{r\to 0^+} \int_{S\setminus B(a;r)} f.$$

  If the limit exists, we say that the integral **converges** and define its value to be the limit, otherwise we say that the integral **diverges**. For technical reasons, which we won't go into, if this integral is not absolutely convergent (i.e. if $\int_S|f|$ is not convergent), then this is instead referred to as the **principal value integral**.

- **Note 4.36:** Sometimes one needs to integrate an integral that is improper at infinity and at a singularity. To illustrate how this is done, suppose $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ is continuous and has a singularity at 0. Fixing some $r_0 > 0$, we define

$$\int_{\mathbb{R}^m\setminus 0} f = \lim_{r\to 0^+} \int_{\overline{B(0;r_0)}\setminus B(0;r)} f + \lim_{r\to\infty} \int_{B(0;r)\setminus\overline{B(0;r_0)}} f.$$

  It's hard to see that this does not depends on the choice of $r_0 > 0$. If there are more singularities, say at points $a_1, \ldots, a_k \in \mathbb{R}^m$, then we do a similar thing by choosing an $r_0 > 0$ such that $B(0;r_0)$ encloses all of the a_j's and using balls $B(a_i, r)$ that "shrink" onto a_i as $r \to 0^+$.

- **Note 4.37:** We mention that one could construct a more refined version of the theory of improper integrals by requiring that all of the above limits exist for any exhaustion of the regions in question and are equal. But then the question arises what conditions on f will ensure that this happens. It turns out that requiring "absolute convergence" is sufficient. We won't pursue this question.

5 Integration on Manifolds

5.1 Tensor Algebra

- We will now combine everything that we studied into this course: multivariable differential calculus, multivariable integration, theory of manifolds, into one unified theory. In particular, we will study how differentiation and integration work on manifolds and how they are intimately connected to one another via Stokes's Theorem – a theory that has had profound influence on electromagnetism and mathematics. In single variable calculus, differentiation and integration were connected via the fundamental theorem of calculus (FTC). Hence Stokes's theorem is the multidimensional generalization of FTC, and on manifolds!

  It turns out that the natural objects for which integration can be defined for on manifolds (i.e. in a coordinate invariant fashion) are top-degree alternating tensor fields. Hence we begin with the study of tensor algebra. We mention that tensors lie at the heart of differential geometry since they define metric tensors, curvature tensors, etc. We also mention that for simplicity we will work over C^∞ manifolds (and C^∞ differential forms) for the rest of the course.

  Please review vector spaces. If you forgot what they are, you can keep in mind $\mathbb{R}^m$ and tangent spaces T_xS brought down to 0 for now.

- **Convention 5.1:** Moving forward, to align our notation with differential geometry, we will write components of vectors as upper indices instead. For instance, we will now write

  $$x = (x^1, \dots, x^m) \in \mathbb{R}^m.$$

  You need to use context to differentiate between indices and raising quantities to powers. This notation will make actions of covectors/tensor on vectors be naturally represented by index positions.

- **Definition 5.2:** Suppose that V is a real vector space. A **linear functional** is a linear map of the form $\omega : V \to \mathbb{R}$. We denote the set of all linear functionals over V by

  $$V^* = \{\omega : V \to \mathbb{R} : \omega \text{ is linear}\}.$$

  You can (and should) check that V^* is also a real vector space. The space V^* is called the **dual space** of V and every $\omega \in V^*$ is called a **covector** – we illustrate the reason for this naming in the following example.

- **Example 5.3:** The map $\omega : \mathbb{R}^4 \to \mathbb{R}$ given by

  $$\omega(x^1, x^2, x^3, x^4) = 2x^1 - 3x^2 + x^3 + 4x^4$$

  is a linear functional. Notice that with respect to the standard basis (see Note 5.4 below to review basis), this map can be written as

$$\omega(x) = \begin{pmatrix} 2 & -3 & 1 & 4 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \end{pmatrix} = (2x^1 - 3x^2 + x^3 + 4x^4).$$

For this reason, with respect to the standard basis we associate ω with the horizontally written vector $\begin{pmatrix} 2 & -3 & 1 & 4 \end{pmatrix}$. In general, we represent vectors as column vectors (i.e. written vertically) and covectors as row vectors (i.e. written horizontally). This example illustrates the more general fact that if a vector space V has dimension m, then its dual space also has dimension m.

- **Note 5.4:** Suppose that V is a real vector space of finite dimension m. Recall that an ordered list of vectors $\beta = \{E_1, \dots, E_m\} \subseteq V$ is called a **basis** if every vector $v \in V$ can be uniquely written as a linear combination $v = v^1 E_1 + \cdots + v^m E_m$ where each $v^i \in \mathbb{R}$. We represent vectors as follows:

$$v = v^1 E_1 + \cdots + v^m E_m \quad \text{is represented by} \quad \begin{pmatrix} v^1 \\ \vdots \\ v^m \end{pmatrix}.$$

Note that the representation of v highly depends on the basis: change the basis β and the column vector representation of v changes. In $\mathbb{R}^m$ we often (but not always) use the standard basis $\{e_1, \dots, e_m\}$ where each

$$e_i = \begin{pmatrix} \left.\begin{matrix} 0 \\ \vdots \\ 0 \\ 1 \end{matrix}\right\} i \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In this case

$$x = x^1 e_1 + \cdots + x^m e_m \quad \text{is represented by} \quad \begin{pmatrix} x^1 \\ \vdots \\ x^m \end{pmatrix}.$$

- **Theorem 5.5:** Suppose that V is a finite-dimensional vector space and that $\{E_1, \dots, E_m\} \subseteq V$ is a basis for V. Let $\{\mathcal{E}^1, \dots, \mathcal{E}^m\} \subseteq V^*$ denote covectors (or "linear functionals") such that each $\mathcal{E}^i : V \to \mathbb{R}$ is given by

$$\mathcal{E}^i(v) = \mathcal{E}^i\left(v^1 E_1 + \cdots v^i E_i + \cdots + v^m E_m\right) = v^i.$$

Then $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ is a basis for V^* and is called the **dual basis** of $\{E_1, \dots, E_m\}$. Hence the dimension of V^* is also m. Furthermore, the unique decomposition of any $\omega \in V^*$ in this dual basis is given by

$$\omega = \omega_1 \mathcal{E}^1 + \cdots + \omega_m \mathcal{E}^m \quad \text{where each} \quad \omega_i = \omega(E_i). \tag{5.6}$$

which we represent by the horizontal vector "$(\omega_1 \quad \cdots \quad \omega_m)$" (observe that we use lower indices on each ω_i).

- *Remark:* For future use we record the following important relations. First (here we introduce the **Kronecker delta notation** δ^i_j):

$$\mathcal{E}^i(E_j) = \mathcal{E}^i(0E_1 + \cdots + 1E_j + \cdots + 0E_m) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} = \delta^i_j.$$

Next, if $\omega = \omega_1\mathcal{E}^1 + \cdots + \omega_m\mathcal{E}^m$ and $v = v^1E_1 + \cdots + v^mE_m$ are a covector and vector respectively, then

$$\omega(v) = (\omega_1\mathcal{E}^1 + \cdots + \omega_m\mathcal{E}^m)(v^1E_1 + \cdots + v^mE_m) = \omega_1v^1 + \cdots + \omega_mv^m.$$

Proof: To prove that $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ is indeed a basis for V^*, let's first show that $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ spans V^*. Take any covector $\omega \in V^*$. For any vector $v \in V$,

$$\omega(v) = \omega(v^1E_1 + \cdots + v^mE_m) = v^1\omega(E_1) + \cdots + v^m\omega(E_m)$$
$$= \omega(E_1)\mathcal{E}^1(v) + \cdots + \omega(E_m)\mathcal{E}^m(v).$$

Hence setting $\omega_i = \omega(E_i)$ we have that

$$\omega = \omega_1\mathcal{E}^1 + \cdots + \omega_m\mathcal{E}^m.$$

So indeed $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ spans V^*. Observe that this also proves (5.6) once we finish proving that $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ is a basis.

To finish proving that $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ is a basis, we just need to show that (5.6) is the unique way to write ω. Suppose $\omega = \widetilde{\omega}_1\mathcal{E}^1 + \cdots + \widetilde{\omega}_m\mathcal{E}^m$ is another way to write ω. We need to show that actually each $\omega_i = \widetilde{\omega}_i$. A powerful technique is to test covectors (and later tensors) on basis elements:

$$\omega(E_i) = \omega_1\mathcal{E}^1(E_1) + \cdots + \omega_m\mathcal{E}^m(E_1) = \omega_i,$$
$$\omega(E_i) = \widetilde{\omega}_1\mathcal{E}^1(E_1) + \cdots + \widetilde{\omega}_m\mathcal{E}^m(E_1) = \widetilde{\omega}_i.$$

Hence each $\omega_i = \widetilde{\omega}_i$.

■

- **Definition 5.7:** Suppose that V is a real vector space. For $k \geq 1$, a **covariant $\boldsymbol{k}$-tensor** on V is a mutli-linear function of the form

$$F : \underbrace{V \times \dots \times V}_{k} \to \mathbb{R},$$

where **multi-linear** (or **bilinear** if $k = 2$) means that (here $a \in \mathbb{R}$ is a constant)

$$F(v_1, \dots, av_i, \dots, v_k) = aF(v_1, \dots, v_i, \dots, v_k),$$
$$F(v_1, \dots, v_i + \tilde{v}_i, \dots, v_k) = F(v_1, \dots, v_i, \dots, v_k) + F(v_1, \dots, \tilde{v}_i, \dots, v_k).$$

The number k is called the **rank** of the tensor F – a completely different usage of the word "rank" used to describe linear maps. We denote the set of all covariant k-tensors by

$$T^k(V^*) = \underbrace{V^* \otimes \dots \otimes V^*}_{k}. \tag{5.8}$$

- The word "covariant" in the above definition won't mean anything special to us in this course. In a more advanced course, it is used to differentiate such tensors from "contravariant" tensors, which are tensors of the form $V^{**} \otimes \dots \otimes V^{**} \cong V \otimes \dots \otimes V$.

- **Example 5.9:** Observe that any covector $\omega \in V^*$ is trivially a tensor of rank 1 since it's a linear function of the form $\omega : V \to \mathbb{R}$. This is an important observation since rank 1 tensors are typically the building blocks for creating higher order tensor – see Example 5.12 below.

- **Definition 5.10:** Suppose that V is a real vector space and that $F \in T^k(V^*)$ and $G \in T^l(V^*)$ are tensors. We define their **tensor product** as the tensor

$$F \otimes G \in T^{k+l}(V^*)$$

given by

$$F \otimes G(v, \dots, v_k, w_1, \dots, w_l) = F(v_1, \dots, v_k)G(w_1, \dots, w_l).$$

This is a different usage of the notation "$\otimes$" from (5.8).

- **Theorem 5.11:** The tensor product has the following properties:

  a. Bilinear (here $a, \hat{a} \in \mathbb{R}$):

$$\left(aF + \hat{a}\hat{F}\right) \otimes G = aF \otimes G + \hat{a}\hat{F} \otimes G,$$

$$F \otimes \left(aG + \hat{a}\hat{G}\right) = aF \otimes G + \hat{a}F \otimes \hat{G}.$$

  b. Associative:

$$F \otimes (G \otimes L) = (F \otimes G) \otimes L.$$

Remark: The tensor product is not commutative!

Proof: We prove b) and leave a) as an exercise. Adopt the context in Definition 5.10. Suppose $L \in T^r(V^*)$ is another tensor. Then

$$F \otimes (G \otimes L)(v_1, \dots, v_k, w_1, \dots, w_l, u_1, \dots, u_r)$$

$$\text{and } (F \otimes G) \otimes L(v_1, \dots, v_k, w_1, \dots, w_l, u_1, \dots, u_r)$$

are both equal to

$$F(v_1, \dots, v_k)G(w_1, \dots, w_l)L(u_1, \dots, u_r),$$

Hence indeed

$$F \otimes (G \otimes L) = (F \otimes G) \otimes L.$$

■

- **Example 5.12:** Consider $V = \mathbb{R}^2$, its standard basis $\{e_1, e_2\}$, and the dual basis $\{\mathcal{E}^1, \mathcal{E}^2\}$ of $\{e_1, e_2\}$. Then $\mathcal{E}^1 \otimes \mathcal{E}^2 \otimes \mathcal{E}^2 \in T^3(V^*)$ is the tensor given by

$$\mathcal{E}^1 \otimes \mathcal{E}^2 \otimes \mathcal{E}^2 \left(\begin{pmatrix} v^1 \\ v^2 \end{pmatrix}, \begin{pmatrix} w^1 \\ w^2 \end{pmatrix}, \begin{pmatrix} u^1 \\ u^2 \end{pmatrix} \right) = v^1 w^2 u^2.$$

- As we can see, tensor products are powerful tools for constructing higher rank tensors. In particular, they allow us to write down a basis for tensors:

- **Theorem 5.13:** Suppose that V is a finite-dimensional vector space and that $\{E_1, \dots, E_m\} \subseteq V$ is a basis for V. Let $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ be the dual basis of $\{E_1, \dots, E_m\}$. Fix $k \geq 1$. Then

$$\left\{\mathcal{E}^{i_1} \otimes \dots \otimes \mathcal{E}^{i_k} : 1 \leq i_1, \dots, i_k \leq m\right\}$$

is a basis for $T^k(V^*)$. Furthermore, the unique decomposition of any $F \in T^k(V^*)$ in this basis is

$$F = \sum_{i_1, \dots, i_k = 1}^{m} F_{i_1, \dots, i_k} \mathcal{E}^{i_1} \otimes \dots \otimes \mathcal{E}^{i_k} \qquad \text{where each} \quad F_{i_1, \dots, i_k} = F\left(E_{i_1}, \dots, E_{i_k}\right).$$

Proof: This is proved very similarly to Theorem 5.5, and so we leave at as HW. ■

- **Example 5.14:** Consider $V = \mathbb{R}^2$, its standard basis $\{e_1, e_2\}$, and the dual basis $\{\mathcal{E}^1, \mathcal{E}^2\}$ of $\{e_1, e_2\}$. Let $L \in T^2(V^*)$ be the 2-tensor that computes the dot product:

$$L\left(\begin{pmatrix} v^1 \\ v^2 \end{pmatrix}, \begin{pmatrix} w^1 \\ w^2 \end{pmatrix} \right) = v^1 w^1 + v^2 w^2.$$

We can decompose this tensor uniquely as

$$L = \mathcal{E}^1 \otimes \mathcal{E}^1 + \mathcal{E}^2 \otimes \mathcal{E}^2.$$

In fact, generalizing this example is the starting point of (pseudo/semi-) Riemannian geometry.

5.2 Alternating Tensors

- **Definition 5.15:** We say that a tensor $F \in T^k(V^*)$ is **symmetric** if its value does not change when interchanging two arguments:

$$F\left(v_1, \dots, v_j, \dots, v_l, \dots, v_k\right) = F\left(v_1, \dots, v_l, \dots, v_j, \dots, v_k\right).$$

We say that $G \in T^k(V^*)$ is **alternating** (or **antisymmetric** or **skew-symmetric**) if its value changes sign when two distinct arguments are interchanged:

$$G\left(v_1, \dots, v_j, \dots, v_l, \dots, v_k\right) = -G\left(v_1, \dots, v_l, \dots, v_j, \dots, v_k\right).$$

Other names for alternating tensors are **exterior forms**, **multi-covectors**, and ***k*-covectors**. We denote the space of all alternating covariant tensors of rank k by $\Lambda^k(V^*)$.

- **Example 5.16:** Consider $V = \mathbb{R}^3$, its standard basis $\{e_1, e_2, e_3\}$, and the dual basis $\{\mathcal{E}^1, \mathcal{E}^2, \mathcal{E}^3\}$ of $\{e_1, e_2, e_3\}$. Then the tensor

$$L\left(\begin{pmatrix} v^1 \\ v^2 \\ v^3 \end{pmatrix}, \begin{pmatrix} w^1 \\ w^2 \\ w^3 \end{pmatrix}\right) = \det\begin{pmatrix} v^1 & w^1 \\ v^2 & w^2 \end{pmatrix}$$

is an alternating tensor of rank 2 (since interchanging two columns in a matrix flips its sign) and hence is in $\Lambda^2(V^*)$.

- **Note 5.17:** Covectors $\omega \in T^1(V^*)$ are both symmetric and alternating.

- For the rest of the course, we will essentially be working only with alternating tensors. In light of this, we remark that if you take two alternating tensors F and G, then their tensor product $F \otimes G$ will not necessarily be alternating (for instance, take a covector ω and consider $\omega \otimes \omega$). So the question arises of whether there exists an analogous "product" operation for alternating tensors that spits out alternating tensors. This is important because as in the case of usual tensors, this would be a powerful way to construct higher rank alternating tensors. The answer is yes and it is simply a slight modification of the tensor product. First we need a definition from group theory:

- **Note 5.18:** A **permutation** of k elements is a bijective map $\sigma : \{1, \ldots, k\} \to \{1, \ldots, k\}$. For instance, an example of a permutation of four elements is

$$\{1,2,3,4\} \xrightarrow{\sigma} \{3,4,1,2\} = \{\sigma(1), \sigma(2), \sigma(3), \sigma(4)\}. \tag{5.19}$$

We denote the set of all permutations by S_k (it's furthermore a group). The **sign** function $\operatorname{sgn} : S_k \to \{\pm 1\}$ assigns a sign for a permutation $\sigma \in S_k$ as follows:

   a. $\operatorname{sgn} \sigma = +1$ if σ can be equivalently obtained by a sequence of an even number of interchanges of only two elements at a time

   b. $\operatorname{sgn} \sigma = -1$ if it can be equivalently obtained by a sequence of an odd number of interchanges of only two elements at a time.

For instance, the sign of the permutation (5.19) is $+1$ (check this!). It's not a trivial fact (and you'll prove this in a class on group theory), that the sign is well-defined. It's not hard to see however that

$$\operatorname{sgn}(\sigma \circ \tau) = \operatorname{sgn} \sigma \cdot \operatorname{sgn} \tau$$

for any two permutations $\sigma, \tau \in S_k$.

- **Definition 5.20:** Suppose that $F \in \Lambda^k(V^*)$ and $G \in \Lambda^l(V^*)$ are alternating tensors. Their **wedge product** (or **exterior product**) $F \wedge G \in \Lambda^{k+l}(V^*)$ is defined to be the alternating tensor given by

$$(F \wedge G)(v_1, \ldots, v_{k+l}) = \frac{1}{k!\, l!} \sum_{\sigma \in S_{k+l}} \operatorname{sgn} \sigma \cdot F\big(v_{\sigma(1)}, \ldots, v_{\sigma(k)}\big) G\big(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}\big).$$

This is indeed alternating since if you interchange any two distinct v_i and v_j on the left-hand side, that causes an additional permutation on the right-hand side, which if you try to reverse will cause a sign flip on each $\operatorname{sgn}\sigma$. The weird coefficient $1/(k!\,l!)$ isn't profound, it's simply there to make part d) of Theorem 5.23 below have a neat form.

- The right-hand side of the above equation is useful in its own right:
- **Definition 5.21:** If $H \in T^j(V^*)$ is a rank j tensor (not necessarily alternating), then we define its **alternation** as

$$\operatorname{Alt}(H) = \frac{1}{j!}\sum_{\sigma \in S_j} \operatorname{sgn}\sigma \cdot H\big(v_{\sigma(1)}, \dots, v_{\sigma(j)}\big).$$

Note that $\operatorname{Alt}(H) \in \Lambda^j(V^*)$ is alternating for similar reasons described in Definition 5.20.

- **Note 5.22:** The alternation notation provides us with a neat way to write down the wedge product in Definition 5.20:

$$F \wedge G = \frac{(k+l)!}{k!\,l!}\operatorname{Alt}(F \otimes G).$$

Although we won't prove this (and hence you can't use it on homework/exams), we mention that the alternation is a projection operator $\operatorname{Alt} : T^j(V^*) \to \Lambda^j(V^*)$. Recall that this means that

$$\operatorname{Alt}(H) \in \Lambda^j(V^*) \quad \text{for any} \quad H \in T^j(V^*),$$

$$\operatorname{Alt}(H) = H \quad \text{for any} \quad H \in \Lambda^j(V^*).$$

- **Theorem 5.23:** The wedge product has the following properties
  a. Bilinearity (here $a, \hat{a} \in \mathbb{R}$):

$$\big(aF + \hat{a}\hat{F}\big) \wedge G = aF \wedge G + \hat{a}\hat{F} \wedge G,$$

$$F \wedge \big(aG + \hat{a}\hat{G}\big) = aF \wedge G + \hat{a}F \wedge \hat{G}.$$

  b. Associativity:

$$F \wedge (G \wedge L) = (F \wedge G) \wedge L.$$

  c. Anticommutativity: If $F \in \Lambda^k(V^*)$ and $G \in \Lambda^l(V^*)$,

$$F \wedge G = (-1)^{kl} G \wedge F$$

  d. If $v_1, \dots, v_j \in V$ are vectors and $\omega^1, \dots, \omega^j \in V^*$ are covectors, then

$$\big(\omega^1 \wedge \dots \wedge \omega^j\big)\big(v_1, \dots, v_j\big) = \det\begin{pmatrix} \omega^1(v_1) & \cdots & \omega^1(v_j) \\ \vdots & \ddots & \vdots \\ \omega^j(v_1) & \cdots & \omega^j(v_j) \end{pmatrix}.$$

Proof: Part a) trivially follows from the definition. The other parts are not trivial whose proofs mostly involve carful arguments about permutations. We'll come back to it if there is time. ■

- We record an important corollary that appears many times in the algebra of alternating tensors:

- **Corollary 5.24:** If $\omega, \eta \in V^*$ covectors, then

$$\omega \wedge \eta = -\eta \wedge \omega \quad \text{and} \quad \omega \wedge \omega = 0.$$

Proof: The first equation directly follows from Theorem 5.23 part c) since in this case both $k = l = 1$. To prove the second equation, take any two vectors $v_1, v_2 \in V$ and observe that by Theorem 5.23 part d)

$$\omega \wedge \omega(v_1, v_2) = \det\begin{pmatrix} \omega(v_1) & \omega(v_2) \\ \omega(v_1) & \omega(v_2) \end{pmatrix} = 0$$

since two of the rows are identical.

■

- **Theorem 5.25:** Suppose that V is a finite-dimensional vector space and that $\{E_1, \dots, E_m\} \subseteq V$ is a basis for V. Let $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ be the dual basis of $\{E_1, \dots, E_m\}$. Fix $k \geq 1$. Then

$$(5.26) \qquad \left\{\mathcal{E}^{i_1} \wedge \dots \wedge \mathcal{E}^{i_k} : 1 \leq i_1 < \dots < i_k \leq m\right\}$$

is a basis for $\Lambda^k(V^*)$. Furthermore, the unique decomposition of any $F \in \Lambda^k(V^*)$ in this basis is

$$F = \sum_{1 \leq i_1 < \dots < i_k \leq m} F_{i_1,\dots,i_k} \mathcal{E}^{i_1} \wedge \dots \wedge \mathcal{E}^{i_k} \qquad \text{where each} \quad F_{i_1,\dots,i_k} = F\left(E_{i_1}, \dots, E_{i_k}\right).$$

If $k > m$, then the set (5.26) and the above sum are empty (i.e. $F = 0$) since in this case it's impossible to fit k indices between 1 and m.

Proof: The proof is very similar to the proof of Theorem 5.13 where the key ingredient is Corollary 5.24. For instance, the reason we need the indices i_j to be <u>strictly</u> increasing is that if two indices match, then the term $\mathcal{E}^{i_1} \wedge \dots \wedge \mathcal{E}^{i_k}$ would be zero by the second equation in Corollary 5.24. You will prove it in the homework. ■

- **Corollary 5.27:** Suppose that V is a finite-dimensional vector space and that $\{E_1, \dots, E_m\} \subseteq V$ is a basis for V (and hence $m = \dim V$). Let $\{\mathcal{E}^1, \dots, \mathcal{E}^m\}$ be the dual basis of $\{E_1, \dots, E_m\}$. Then

$$\{\mathcal{E}^1 \wedge \dots \wedge \mathcal{E}^m\}$$

is a basis for $\Lambda^m(V^*)$ (in particular $\Lambda^m(V^*)$ is 1-dimensional). Furthermore, $\Lambda^k(V^*) = \{0\}$ for $k > m$.

Remark: For this reason we call $\Lambda^m(V^*)$ **top degree alternating tensors**, because m is the last rank before the alternating tensors become zero.

Proof: The first statement follows from the fact that there is only one way to fit m indices between 1 and m in increasing order in (5.26). To prove the second statement, as mentioned at the end of Theorem 5.25, if $k > m$ then any $F \in \Lambda^k(T^*M)$ is zero.

■

5.3 Orientation

- **Note 5.28:** We take a brief aside to define orientability of manifolds. Our end goal is to prove Stokes's Theorem, which as you'll see will require that the manifold is orientable. "Orientable," intuitively speaking, is the property of a manifold that it has two sides. Imagine a 2D surface S sitting in $\mathbb{R}^3$. Its orientability could be defined as the existence of a continuous perpendicular vector field N along the surface "indicating" one of its sides. But the issue with this definition is that it isn't intrinsic. In other words, a 2D creature living on the surface can't use it to test if their world is orientable or not. It also doesn't generalize well to the case of when the difference between the dimension of the manifold S and the Euclidean space that it sits in $\mathbb{R}^m$ is bigger than one. So mathematicians came up with the following definition.
- Haim does an amazing demonstration demonstrating orientability!
- **Definition 5.29:** Suppose that S is a C^∞ j-dimensional manifold possibly with boundary embedded in $\mathbb{R}^m$. An **orientation** on S is a declaration on each parametrization $f : U \subseteq (\mathbb{R}^m \text{ or } \mathbb{H}^m) \to S$, or equivalently chart $\varphi = f^{-1}$, where U is connected whether it is **positively oriented** or **negatively oriented**. In addition, we require that if φ and $\hat{\varphi}$ are two charts whose domains intersect, then

  $$(5.30) \qquad \det D(\varphi \circ \hat{\varphi}^{-1}) > 0 \quad \text{if } \varphi \text{ and } \hat{\varphi} \text{ are of the same orientation,}$$

  $$(5.31) \qquad \det D(\varphi \circ \hat{\varphi}^{-1}) < 0 \quad \text{if } \varphi \text{ and } \hat{\varphi} \text{ are of the opposite orientation.}$$

  If this is possible, then we say that S is **orientable**. If not, then we say that S is not **orientable**.
- **Note 5.32:** To demonstrate the connection between the above definition and the proposed definition in Note 5.28, let the N there be defined as follows. Take any point $p \in M$ and let $f : U \to S$ be a positively oriented parametrization. Then let N be a vector perpendicular to T_pS such that if $f : U \to S$ is a parametrization, then the frame $\{E_1, E_2, N\}$ satisfies the right-hand rule where E_1, E_2 are the basis vectors of T_pS given by (3.34). For higher dimensional surfaces S, you require $\mathcal{E}^1 \wedge \dots \wedge \mathcal{E}^m(E_1, \dots, E_{m-1}, N) > 0$. You will explore this in the homework.

5.4 Differential forms

- **Note 5.33:** Now we push the theory of tensors onto manifolds, where the vector spaces "V" of interest will be tangent planes "T_pS." To do tensor algebra, it's essential to have a fixed basis to work in and so we start by setting notation for a basis for tangent spaces. Suppose that S is C^∞ j-dimensional manifold possibly with boundary embedded in $\mathbb{R}^m$. Let $f : U \subseteq (\mathbb{R}^m \text{ or } \mathbb{H}^m) \to S$ be a parametrization. Fix a point $p \in \operatorname{range} f$ and recall that

$$\left\{\begin{pmatrix}\frac{\partial f^1}{\partial u^1}(p)\\ \vdots \\ \frac{\partial f^m}{\partial u^1}(p)\end{pmatrix},\dots,\begin{pmatrix}\frac{\partial f^1}{\partial u^j}(p)\\ \vdots \\ \frac{\partial f^m}{\partial u^j}(p)\end{pmatrix}\right\} \text{ is a basis for } T_pS.$$

This is clunky to write, and so people instead have come up with the following shorthand notation for the above j tangent vectors:

$$\left\{\left.\frac{\partial}{\partial u^1}\right|_p,\dots,\left.\frac{\partial}{\partial u^j}\right|_p\right\} \text{ is a basis for } T_pS,$$

using context to differentiate these from partial derivatives. The indices here are considered lower indices. The dual basis of this is denoted by:

$$\left\{du^1|_p,\dots,du^j\big|_p\right\} \text{ is a basis for } \left(T_pS\right)^*\text{, which we typically denote by } T_p^*S.$$

The indices here are considered upper indices and we call T_p^*S the **cotangent space** at p. When the point p is clear, we sometime omit writing the " $|_p$."

- **Definition 5.34:** Suppose that S is C^∞ j-dimensional manifold possibly with boundary embedded in $\mathbb{R}^m$. We call

$$T^*S = \bigcup_{p\in S} T_p^*S$$

the **cotangent bundle** of S. For $k \geq 1$, we let

$$\Lambda^k T^*S = \bigcup_{p\in S} \Lambda^k\left(T_p^*S\right).$$

A map $\omega : S \to \Lambda^k T^*S$ is called a C^∞ **differential form** of rank k over S if for every $p \in S$, $\omega|_p \in \Lambda^k\left(T_p^*S\right)$ and for any parametrization $f : U \subseteq \left(\mathbb{R}^j \text{ or } \mathbb{H}^j\right) \to S$ with associated chart φ, ω has the following form over $\operatorname{dom}\varphi$:

$$(5.35) \qquad \omega = \sum_{1\leq i_1<\cdots<i_k\leq j} \omega_{i_1,\dots,i_k} du^{i_1}\wedge \dots \wedge du^{i_k}$$

where each $\omega_{i_1,\dots,i_k} : \operatorname{dom}\varphi \to \mathbb{R}$ is such that $\omega_{i_1,\dots,i_k}\circ f \in C^\infty(U)$. The $\omega_{i_1,\dots,i_k}$ are called the **coordinate components** of ω. We denote the set of all such differential forms by $\Omega^k(S)$.

We say that ω is **top degree** if $\operatorname{rank}\omega = \dim S$ (since $\dim S = \dim T_pS$ for any $p \in S$). Observe that by Corollary 5.27 in our chart a top degree differential form $\omega \in \Omega^j(S)$ is of the form $\omega = h\, du^1 \wedge \dots \wedge du^j$ for some $h : \operatorname{dom}\varphi \to \mathbb{R}$ such that $h \circ f \in C^\infty(U)$.

- Now we will demonstrate that top degree differential forms are natural objects to integrate over manifolds. As expected, we will define such integration in coordinates and hence we will need to show that this definition is coordinate invariant. In order to do this, we start by studying the coordinate transformation laws for differential forms. In this course we will only deal with integration over compact manifolds to avoid convergence issues.

- **Proposition 5.36:** Suppose that S is a C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$. Suppose that $f : U \to S$ and $\hat{f} : \widehat{U} \to S$ are parametrizations and let φ and $\hat{\varphi}$ be their associated charts. Suppose that $\operatorname{dom}\varphi \cap \operatorname{dom}\hat{\varphi} \neq \emptyset$ and take any $p \in \operatorname{dom}\varphi \cap \operatorname{dom}\hat{\varphi}$. Then

  a. The change of basis matrix on T_pS from the basis $\left\{\frac{\partial}{\partial u^1}, \dots, \frac{\partial}{\partial u^j}\right\}$ to the basis $\left\{\frac{\partial}{\partial \hat{u}^1}, \dots, \frac{\partial}{\partial \hat{u}^j}\right\}$ is given by $D(\hat{\varphi} \circ \varphi^{-1})$ evaluated at $\varphi(p)$.

  b. The change of basis matrix on T_p^*S from the basis $\{du^1, \dots, du^j\}$ to the basis $\{d\hat{u}^1, \dots, d\hat{u}^j\}$ is given by $[D(\varphi \circ \hat{\varphi}^{-1})]^\top$ evaluated at $\hat{\varphi}(p)$.

  c. Change of coordinates for top degree alternating tensors at p is given by:

$$h\, du^1 \wedge \dots \wedge du^j = \underbrace{h \det[D(\varphi \circ \hat{\varphi}^{-1})]}_{\hat{h}}\, d\hat{u}^1 \wedge \dots \wedge d\hat{u}^j.$$

Proof: You will prove parts a) and b) in the homework; we will prove part c). The h simply stays out in front, so we don't need to worry about it. Let A denote the matrix $[D(\varphi \circ \hat{\varphi}^{-1})]^\top$ evaluated at $\hat{\varphi}(p)$. We will write the entry of A in the i^{th} row and k^{th} column as A_i^k. By part b) we have that

$$(5.37) \qquad du^1 \wedge \dots \wedge du^j = \left(A_1^1 d\hat{u}^1 + \cdots + A_j^1 d\hat{u}^j\right) \wedge \dots \wedge \left(A_1^j d\hat{u}^1 + \cdots + A_j^j d\hat{u}^j\right).$$

Now we expand this over the "+" sign to get an *enormous* sum. To get an idea of what this looks like, let's do the calculation first in the case $j = 2$ (here we use Corollary 5.24):

$$du^1 \wedge du^2 = (A_1^1 d\hat{u}^1 + A_2^1 d\hat{u}^2) \wedge (A_1^2 d\hat{u}^1 + A_2^2 d\hat{u}^2)$$

$$= A_1^1 A_1^2 \underbrace{d\hat{u}^1 \wedge d\hat{u}^1}_{\text{zero}} + A_1^1 A_2^2 d\hat{u}^1 \wedge d\hat{u}^2 + A_2^1 A_1^2 \underbrace{d\hat{u}^2 \wedge d\hat{u}^1}_{= -d\hat{u}^1 \wedge d\hat{u}^2} + A_2^1 A_2^2 \underbrace{d\hat{u}^2 \wedge d\hat{u}^2}_{\text{zero}}$$

$$(A_1^1 A_2^2 - A_2^1 A_1^2) d\hat{u}^1 \wedge d\hat{u}^2 = \det[A]\, d\hat{u}^1 \wedge d\hat{u}^2.$$

The general case works similarly. When expanding the right-hand side of (5.37) every wedge product term that has a repeat $d\hat{u}^k$ will be zero. Hence we will get a sum of terms of the form

$$\dots + A_{\sigma(1)}^1 \dots A_{\sigma(j)}^j d\hat{u}^{\sigma(1)} \wedge \dots \wedge d\hat{u}^{\sigma(j)} + \cdots$$

where $\sigma \in S_j$ are permutations. We can rearrange the $d\hat{u}^l$ in each term <u>two at a time</u> to get

$$\dots + \operatorname{sgn}(\sigma)\, A_{\sigma(1)}^1 \dots A_{\sigma(j)}^j d\hat{u}^1 \wedge \dots \wedge d\hat{u}^j + \cdots$$

$$= \left(\sum_{\sigma \in S_j} \operatorname{sgn}(\sigma)\, A^1_{\sigma(1)} \dots A^j_{\sigma(j)} \right) d\hat{u}^1 \wedge \dots \wedge d\hat{u}^j$$

In your linear algebra course, you most likely proved that the above sum is equal to $\det A$.[1] Using the fact that $\det A^\top = \det A$, this proves part c).

■

5.5 Integrating Top Degree Differential Forms

- **Definition 5.38:** Suppose that S is a compact oriented C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$. Suppose that $f : U \to S$ is a parametrization, where U is connected, and let φ be its associated chart. Suppose that $\omega \in \Omega^j(S)$ is a top degree differential form such that

  (5.39) $$\overline{\{x \in S : \omega|_x \neq 0\}} \subseteq \operatorname{dom} \varphi.$$

  The left-hand side is called the **support of** ω and is denoted by $\operatorname{supp} \omega$. Hence we can write

  $$\omega = h\, du^1 \wedge \dots \wedge du^j$$

  for some $h : \operatorname{dom} \varphi \to \mathbb{R}$ such that $h \circ f \in C^\infty(U)$. Then we define

  $$\int_S \omega = \int_{\operatorname{dom}\varphi} h\, du^1 \wedge \dots \wedge du^j = \pm \int_U h \circ f\, du^1 \dots du^j$$

  where the sign is "$+$" if the parameterization/chart is positively oriented and "$-$" if it is negatively oriented.

  Remark: Some people write h instead of $h \circ f$ and use context to differentiate this from $h : \operatorname{dom} \varphi \to \mathbb{R}$. This way the above equation demonstrates why integrating top degree differential forms are so natural: to differentiate $h\, du^1 \wedge \dots \wedge du^j$, just erase the wedges!

- **Note 5.40:** Since the usual multivariable integral is linear, it follows immediately that integration of differential forms defined above is also linear:

  $$\int_S (a\omega + b\eta) = a \int_S \omega + b \int_S \eta$$

  where $a, b \in \mathbb{R}$ and $\omega, \eta \in \Omega^j(S)$ that satisfy (5.39).

  We also point out that if you flip the orientation on S (i.e. on each parametrization) then you will flip the sign on the integral $\int_S \omega$.

[1] See for instance (4.2) on page 89 in *Linear Algebra Done Wrong* by Sergei Treil: https://www.math.brown.edu/streil/papers/LADW/LADW.html

- **Note 5.41:** We have to show that the above definition is well-defined. To elaborate, suppose $\hat{f} : \hat{U} \to S$ and $\hat{\varphi}$ are another parametrization and chart that satisfy (5.39). We could have instead equally well defined

$$\int_S \omega = \pm \int_{\hat{U}} \hat{h} \circ \hat{f} \, d\hat{u}^1 \dots d\hat{u}^j \quad \text{where} \quad \omega = \hat{h} \, d\hat{u}^1 \wedge \dots \wedge d\hat{u}^j.$$

Do we get the same number? To show that the above definition is well defined, we need to show that the answer is yes. First suppose that both charts are positively oriented. Then by the change of variables formula (Theorem 4.27),

$$\int_U h \circ f \, du^1 \dots du^j = \int_{\hat{U}} \big(h \circ f \circ (\varphi \circ \hat{\varphi}^{-1})\big) |\det D(\varphi \circ \hat{\varphi}^{-1})| d\hat{u}^1 \dots d\hat{u}^j$$

$$= \int_{\hat{U}} \big(h \circ \hat{f}\big) |\det D(\varphi \circ \hat{\varphi}^{-1})| d\hat{u}^1 \dots d\hat{u}^j$$

Since both φ and $\hat{\varphi}$ have the same orientation, by definition we have that $\det D(\hat{\varphi} \circ \varphi^{-1}) > 0$. Hence we can remove the absolute values and so by Proposition 5.36 part c) the above is indeed equal to

$$\int_{\hat{U}} \big(\hat{h} \circ \hat{f}\big) d\hat{u}^1 \dots d\hat{u}^j.$$

We leave it as an exercise to show that the above calculation works even when φ and $\hat{\varphi}$ don't necessarily have the same orientation. In particular, when they have the opposite orientation, an extra minus sign will come out.

- In general, you cannot cover a manifold with only one chart. Hence at the moment we can't integrate a top degree differential form ω over an entire manifold. The way to do this is to break ω up into smaller pieces, each of which is contained in the domain of a chart, integrate each piece using Definition 5.38, and then sum up the results. Furthermore, to prove Stokes's theorem we will want to break ω into C^∞ pieces so that we can do calculus on them. The tool that allows us to do this is called "partitions of unity," which we study next.

- **Definition 5.42:** Suppose that S is a C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$. We say that a function $h : S \to \mathbb{R}$ is C^k if for any parametrization $f : U \to S$, $h \circ f \in C^k(U)$. In this case we write $h \in C^k(S)$.

- It's easy to check via change of variables that the above definition is well-defined: it boils down to the chain rule and the fact that the change of variables maps $\varphi \circ \hat{\varphi}^{-1}$ are themselves C^∞. The above definition and Definition 5.34 also represent the general philosophy that on C^∞ manifolds objects are C^k if they or their coordinate components are C^k when you compose them with parametrizations.

- **Theorem 5.43:** Suppose that S is a <u>compact</u> C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$. Suppose that $\{V_i = \operatorname{dom}\varphi_i \subseteq S\}_{i=1}^k$ is a finite cover of S by domains of charts $\{\varphi_i\}_{i=1}^k$ (this is possible to arrange since S is compact). Then there exists a set of <u>nonnegative</u> functions $\{\psi_i \in C^\infty(S)\}_{i=1}^k$, called a **smooth partition of unity subordinate to** $\{V_i\}_{i=1}^k$, such that each

$$\operatorname{supp}\psi_i \stackrel{\text{def}}{=} \overline{\{x \in S : \psi_i(x) \neq 0\}} \subseteq V_i$$

and

$$\sum_{i=1}^k \psi_i = 1 \qquad \left(\text{i.e. } \sum_{i=1}^k \psi_i(x) = 1 \quad \forall x \in S\right).$$

Remark: This theorem has generalizations to general covers and noncompact settings, though the statement becomes slightly more delicate in the latter. Note also that each $0 \leq \psi_i \leq 1$.

Proof: Consider the famous function $h : \mathbb{R} \to \mathbb{R}$

$$h(x) = \begin{cases} e^{-\frac{1}{x}} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

(plot it!). This is an amazing function because it is identically zero to the left of zero, positive to the right of zero and, as you may have been asked to prove in your analysis class, it is C^∞ everywhere (proving this at $x = 0$ can be done by repeated applications of L'Hopital's rule). For any $a \in (\mathbb{R}^j \text{ or } \mathbb{H}^j)$ and any fixed $r > 0$, consider the function $\rho_{a,r} : (\mathbb{R}^j \text{ or } \mathbb{H}^j) \to \mathbb{R}$ given by

$$\rho_{a,r} = \frac{h(r - |x-a|)}{h(r - |x-a|) + h(|x-a| - r/2)}$$

(plot it with $j = 2$!). Similar as above, it's a simple exercise to show that this C^∞, is positive on $B(a; r)$, and zero outside of $B(a; r)$. Such classes of functions are called **bump functions**.

Fix one of our charts $\varphi_i : V_i \to U_i = \operatorname{Im}\varphi_i$. For any closed ball $\overline{B(a,r)} \subseteq U_i$ consider the bump functions $\phi_{i,a,r} : S \to \mathbb{R}$ up on the manifold given by

$$\phi_{i,a,r}(x) = \begin{cases} \rho_{a,r} \circ \varphi_i(x) & \text{if } x \in \varphi_i^{-1}[B(a,r)] \\ 0 & \text{if } x \notin \varphi_i^{-1}[B(a,r)] \end{cases}$$

It's an easy exercise to check that $\phi_{i,a,r} \in C^\infty(S)$. Now, cover S by a finite collection $\left\{\varphi_{i_j}^{-1}\left[B\left(a_j; r_j\right)\right]\right\}_{j=1}^l$ of sets of the form considered above and let $\left\{\phi_j = \phi_{i_j,a_j,r_j}\right\}_{j=1}^l$ be their associated bump functions. Since this is a cover,

$$\sum_{j=1}^l \phi_j > 0 \quad \text{everywhere on } S.$$

Now, group the $\phi_1, \dots, \phi_l$ into groups such that

$$\underbrace{\operatorname{supp}\phi_1, \dots, \operatorname{supp}\phi_{j_1} \subseteq U_1}_{\text{Define } \hat{\psi}_1 = \phi_1 + \cdots + \phi_{j_1}}, \underbrace{\operatorname{supp}\phi_{j_1+1}, \dots, \operatorname{supp}\phi_{j_2} \subseteq U_2}_{\text{Define } \hat{\psi}_2 = \phi_{j_1+1} + \cdots + \phi_{j_2}}, \dots,$$

$$\underbrace{\operatorname{supp}\phi_{j_{k-1}+1}, \dots, \operatorname{supp}\phi_{j_k} \subseteq U_k}_{\text{Define } \hat{\psi}_k = \phi_{j_{k-1}} + \cdots + \phi_{j_k}}.$$

We have that

$$\sum_{i=1}^{k} \hat{\psi}_i = \sum_{j=1}^{l} \phi_j > 0 \quad \text{everywhere on } S.$$

You will show in the homework that each $\operatorname{supp}\hat{\psi}_i \subseteq \operatorname{supp}\phi_{j_{i-1}+1} \cup \dots \cup \operatorname{supp}\phi_{j_i} \subseteq V_i$. Hence if we define $\psi_i = \hat{\psi}_i / \sum_{j=1}^{k} \hat{\psi}_j$, then we get that each $\operatorname{supp}\psi_i \subseteq V_i$ and

$$\sum_{i=1}^{k} \psi_i = \sum_{i=1}^{k} \frac{\hat{\psi}_i}{\sum_{j=1}^{k} \hat{\psi}_j} = \frac{1}{\sum_{j=1}^{k} \hat{\psi}_j} \sum_{i=1}^{k} \hat{\psi}_i = 1.$$

Hence $\{\psi_i\}_{i=1}^{k}$ is the partition of unity subordinate to $\{V_i\}_{i=1}^{k}$ that we wanted.

■

- **Definition 5.44:** Suppose that S is a compact oriented C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$. Suppose that $\omega \in \Omega^j(S)$ is a top degree differential form. Suppose that $\{f_i : U_i \to S\}_{i=1}^{k}$ are parametrizations, where each U_i is connected, with associated charts $\{\varphi_i\}_{i=1}^{k}$ such that $\{V_i = \operatorname{dom}\varphi_i\}_{i=1}^{k}$ cover S. Let $\{\psi_i : S \to \mathbb{R}\}_{i=1}^{k}$ be a partition of unity subordinate to $\{V_i\}_{i=1}^{k}$. Then we define

$$\int_S \omega = \sum_{i=1}^{k} \int_S \psi_i \omega$$

  where each $\int_S \psi_i \omega$ is defined as in Definition 5.38, which makes sense since each $\operatorname{supp}(\psi_i \omega) \subseteq V_i = \operatorname{dom}\varphi_i$.

- **Note 5.45:** One has to again show that the above definition is well-defined. In particular, one has to show that it does not depend on the choice of $\{f_i : U_i \to S\}_{i=1}^{k}$ and $\{\psi_i : S \to \mathbb{R}\}_{i=1}^{k}$. You will prove this in the homework (hint: don't do any calculations in coordinates).

- **Note 5.46:** The remark mentioned in Note 5.40 regarding linearity of integration of differential forms and dependence on orientation extends to the global integration defined in Definition 5.44 above as well .

5.6 Exterior Derivatives and Stokes's Theorem

- Now we come to the most magical part of the theory of differential forms. It turns out that there is a coordinate invariant way of defining derivatives of differential forms which beautifully formulates the stunning Stokes's theorem, the latter of which generalizes all of the famous interior-to-boundary integration theorems in mathematics including the fundamental theorem of calculus, Greens theorem, divergence theorem, and the classic Stokes's theorem.

- **Convention 5.47:** Suppose that S is a C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$. We let $\Omega^0(S) = C^\infty(S)$.

- **Definition 5.48:** Suppose that S is a C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$ and that $\varphi : V \subseteq S \to U \subseteq \left(\mathbb{R}^j \text{ or } \mathbb{H}^j\right)$ is a chart. Let $\omega \in \Omega^k(S)$ be a differential form which we write in these coordinates as

$$\omega = \sum_{1 \le i_1 < \cdots < i_k \le j} \omega_{i_1,\dots,i_k} du^{i_1} \wedge \dots \wedge du^{i_k}.$$

The **exterior derivative** of ω is defined as

$$d\omega = \sum_{1 \le i_1 < \cdots < i_k \le j} \sum_{i=1}^{j} \frac{\partial \omega_{i_1,\dots,i_k} \circ \varphi^{-1}}{\partial u^i} \circ \varphi \, du^i \wedge du^{i_1} \wedge \dots \wedge du^{i_k}.$$

This is coordinate invariant, which we'll come back to and prove if we have time. Observe that the exterior derivative maps $d : \Omega^k(S) \to \Omega^{k+1}(S)$.

- **Theorem 5.49:** Suppose S is an oriented C^∞ j-dimensional manifold with boundary embedded in $\mathbb{R}^m$. By Theorem 3.40 we know that ∂S is a $(j-1)$-dimensional submanifold. We will assign the following orientation on ∂S. If $\varphi : V \subseteq S \to U \subseteq \mathbb{H}^j$ is a boundary chart, then by the proof of Theorem 3.40 we have that the restriction $\tilde{\varphi} : V \cap \partial S \to U \cap \partial \mathbb{H}^j$ is a chart of ∂S (we think of $U \cap \partial \mathbb{H}^j$ as an open subset of $\mathbb{R}^{j-1}$). We declare the orientation on $\tilde{\varphi}$ to be the same as that of φ if j is even and the opposite of j is odd. This is called **Stokes's orientation** on ∂S.

  We leave it as an exercise to show that (5.30) and (5.31) will hold for ∂S and so this is indeed an orientation on ∂S (i.e. it satisfies Definition 5.29). The reason for the dependence on the evenness of oddness of j is, as you might guess, that it will allow us to nicely formulate Stokes's Theorem below.

- **Definition 5.50:** Suppose S is a C^∞ j-dimensional manifold with boundary embedded in $\mathbb{R}^m$. If $\omega \in \Omega^k(S)$, the **restriction** $\widetilde{\omega} \in \Omega^k(\partial S)$ of ω is defined as the map $\omega : \partial S \to \Lambda^k T^* \partial S$ such that for any $p \in \partial S$, $\widetilde{\omega}|_p \in \Lambda^k\left(T_p^* \partial S\right)$ and

$$\widetilde{\omega}|_p(v_1, \dots, v_k) = \omega|_p(v_1, \dots, v_k) \qquad \forall v_1, \dots, v_k \in T_p \partial S.$$

In other words, it's the usual restriction except that it's important to note that $\widetilde{\omega}$ can only accept vectors tangent to the boundary. Suppose $\varphi : V \subseteq S \to U \subseteq \mathbb{H}^j$ is a boundary chart and we take the restriction $\tilde{\varphi} : V \cap \partial S \to U \cap \partial \mathbb{H}^j$ which is a chart of ∂S. If we represent ω in local coordinates as in (5.35) with respect to φ, then it's not a hard exercise to show that the local coordinate expression for $\widetilde{\omega}$ with respect to $\tilde{\varphi}$ is given by

$$(5.51) \qquad \widetilde{\omega} = \sum_{1 \le i_1 < \cdots < i_k \le j-1} \omega_{i_1,\dots,i_k} \, du^{i_1} \wedge \dots \wedge du^{i_k}$$

(note the "$j-1$" underneath the Σ) and so indeed $\widetilde{\omega}$ is C^∞ (i.e. $\widetilde{\omega} \in \Omega^k(\partial S)$). We mention that people often don't write the tilde "$\sim$" on $\widetilde{\omega}$ and simply write ω, relying on context to differentiate between ω and its restriction to ∂S – we will do the same as well.

- Now we come to the climax of this class: Stokes's theorem. It's a triumph of mathematics with immeasurable impact on the theory of partial differential equations and differential geometry. It is a trophy that mathematicians proudly display by teaching it in their undergraduate classes:

- **Theorem 5.52: (Stokes's Theorem)** Suppose that S is a compact oriented C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$. Let ∂S have Stokes's orientation. For any $\omega \in \Omega^{j-1}(S)$,

$$\int_{\partial S} \omega = \int_S d\omega.$$

If $\partial S = \emptyset$ (i.e. there is no boundary), then the left-hand side is interpreted to be zero.

- The equation is so short, yet so profound and clever! Notice that it relates the integral of the quantity ω with the integral of its derivative along the whole manifold, just like the classic integration theorems mentioned at the beginning of the section. Before we prove it, let us work out some famous examples of exterior derivatives and study their properties.

- **Note 5.53:** Let $S = \mathbb{R}^m$ where we take the standard parametrization $f(u^1, \dots, u^m) = (u^1, \dots, u^m)$ of $S = \mathbb{R}^m$. For this reason, we simply use $(x^1, \dots, x^m)$ to denote the coordinates of $\mathbb{R}^m$ as usual. Then for any $h \in C^\infty(\mathbb{R}^m) = \Omega^0(\mathbb{R}^m)$

$$dh = \frac{\partial h}{\partial x^1} dx^1 + \cdots + \frac{\partial h}{\partial x^j} dx^j.$$

Finally we arrive at a rigorous interpretation of differentials in (2.21)! Next, suppose that $m = 3$ in which case the exterior derivative displays an analog with some famous operators. Let $\mathfrak{X}(\mathbb{R}^3)$ denote smooth vector fields over $\mathbb{R}^3$. We will associate vector fields and functions over $\mathbb{R}^3$ with elements of $\Omega^k(\mathbb{R}^3)$ as follows:

$$h \in C^\infty(\mathbb{R}^3) \sim_1 h \in \Omega^0(\mathbb{R}^3)$$

$$(P, Q, R) \in \mathfrak{X}(\mathbb{R}^3) \sim_2 P\,dx + Q\,dy + R\,dz \in \Omega^1(\mathbb{R}^3)$$

$$(A, B, C) \in \mathfrak{X}(\mathbb{R}^3) \sim_3 A\,dy \wedge dz + B\,dz \wedge dx + C\,dx \wedge dy \in \Omega^2(\mathbb{R}^3)$$

$$h \in C^\infty(\mathbb{R}^3) \sim_4 h\,dx \wedge dy \wedge dz \in \Omega^3(\mathbb{R}^3).$$

Then observe that (here curl = rot)

$$1)\ h \sim_1 h \xrightarrow{d} \frac{\partial h}{\partial x}dx + \frac{\partial h}{\partial y}dy + \frac{\partial h}{\partial z}dz \sim_2 \nabla h.$$

$$2)\ (P,Q,R) \sim_2 P\ dx + Q\ dy + R\ dz$$

$$\xrightarrow{d} \frac{\partial P}{\partial x}dx\wedge dx + \frac{\partial P}{\partial y}dy\wedge dx + \frac{\partial P}{\partial z}dz\wedge dx$$

$$+\frac{\partial Q}{\partial x}dx\wedge dy + \frac{\partial Q}{\partial y}dy\wedge dy + \frac{\partial Q}{\partial z}dz\wedge dy$$

$$+\frac{\partial R}{\partial x}dx\wedge dz + \frac{\partial R}{\partial y}dy\wedge dz + \frac{\partial R}{\partial z}dz\wedge dz$$

$$= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)dy\wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)dz\wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)dx\wedge dy$$

$$\sim_3 \operatorname{curl}(P,Q,R)$$

$$3)(A,B,C) \sim_3 A\ dy\wedge dz + B\ dz\wedge dx + C\ dx\wedge dy$$

$$\xrightarrow{d} \left(\frac{\partial A}{\partial x}dx + \cdots\right)dy\wedge dz + \left(\ldots + \frac{\partial B}{\partial y}dy + \cdots\right)dz\wedge dx + \left(\ldots + \frac{\partial C}{\partial z}dz\right)dx\wedge dy$$

$$= \left(\frac{\partial A}{\partial x} + \frac{\partial B}{\partial x} + \frac{\partial C}{\partial z}\right)dx\wedge dy\wedge dz \sim_4 \operatorname{div}(A,B,C).$$

So exterior derivatives reformulate gradient, curl, and div in the dual world. This is summarized by the following commutative diagram:

$$\begin{tikzcd}
C^\infty(\mathbb{R}^3) \arrow[r, "\nabla"] \arrow[d, "\sim_1"'] & \mathfrak{X}(\mathbb{R}^3) \arrow[r, "\text{curl}"] \arrow[d, "\sim_2"'] & \mathfrak{X}(\mathbb{R}^3) \arrow[r, "\text{div}"] \arrow[d, "\sim_3"'] & C^\infty(\mathbb{R}^3) \arrow[d, "\sim_4"'] \\
\Omega^0(\mathbb{R}^3) \arrow[r, "d"] & \Omega^1(\mathbb{R}^3) \arrow[r, "d"] & \Omega^2(\mathbb{R}^3) \arrow[r, "d"] & \Omega^3(\mathbb{R}^3)
\end{tikzcd}$$

This is the reason that the fundamental theorem of calculus and the classic formulations of Stokes's and the divergence theorem:

$$\overbrace{h(b) - h(a)}^{\int_{\{a\}\cup\{b\}} h} = \int_{[a,b]} \frac{dh}{dx}dx, \qquad \oint_{\partial S} \vec{V}\cdot d\vec{x} = \iint_S \operatorname{curl}(\vec{V})\cdot\vec{n}\,d\sigma, \qquad \oiint_{\partial U} \vec{V}\cdot\vec{n}\,d\sigma = \iiint_U \operatorname{div}\vec{V}$$

are all special cases of Theorem 5.52 (admittingly we are leaving out some details in this explanation).

- **Theorem 5.54:** Suppose that S is a C^∞ j-dimensional embedded submanifold possibly with boundary in $\mathbb{R}^m$. The exterior derivative $d : \Omega^k(S) \to \Omega^{k+1}(S)$ satisfies

a. (Linearity) For any $a, b \in \mathbb{R}$ and $\omega, \eta \in \Omega^k(S)$,

$$d(a\omega + b\eta) = a\, d\omega + b\, d\eta.$$

b. (Product rule) For any $\omega \in \Omega^k(S)$ and any $\eta \in \Omega^l(S)$,

$$d(\omega \wedge \eta) = d(\omega) \wedge \eta + (-1)^k \omega \wedge d\eta.$$

c. For any chart $\varphi : V \subseteq S \to U \subseteq \left(\mathbb{R}^j \text{ or } \mathbb{H}^j\right)$, the exterior derivative of any basis element

$$d\left(du^{i_1} \wedge \dots \wedge du^{i_k}\right) = 0.$$

d. $d \circ d \equiv 0$. Meaning that for any $\omega \in \Omega^k(S)$,

$$d(d\omega) = 0.$$

Remark 1: It turns out that if in c) you let $u^1, \dots, u^j$ denote the components of φ (i.e. each $u^k = \varphi^k$), then each basis covector du^k is in fact equal to the exterior derivative of u^k:

$$du^k = du^k.$$

This interpretation makes c) a special case of d).

Remark 2: Going back to the commutative diagram in Note 5.53, part d) is the analog of the famous facts that curl $\circ\, \nabla \equiv 0$ and div $\circ$ curl $\equiv 0$.

Proof: You will prove this theorem in the homework, which essentially boils down to computations. ■

- **Proof of Theorem 5.52 (Stokes's Theorem):** Let $\omega \in \Omega^{j-1}(S)$. To avoid confusion, we will denote the restriction of ω to ∂S by $\widetilde{\omega}$. First we will first prove the theorem when $\operatorname{supp}\omega$ is contained in the domain of a chart $\varphi : V \subseteq S \to U \subseteq \left(\mathbb{R}^j \text{ or } \mathbb{H}^j\right)$ and then come back to the general case. Suppose that φ is positively oriented since the proof in the other case is similar. Let $f = \varphi^{-1}$ denote the associated parametrization. In these coordinates we will write ω on S as follows, deviating from our usual convention on how we write its components,

$$(5.55) \qquad \omega = \sum_{i=1}^{j} \omega_i\, du^1 \wedge \dots \wedge du^{i-1} \wedge du^{i+1} \wedge \dots \wedge du^j.$$

Taking the exterior derivative gives (we omit writing "$\circ\, \varphi^{-1}$" and "$\circ\, \varphi$" here)

$$d\omega = \sum_{i=1}^{j} \sum_{k=1}^{j} \frac{\partial \omega_i}{\partial u^k} du^k \wedge du^1 \wedge \dots \wedge du^{i-1} \wedge du^{i+1} \wedge \dots \wedge du^j.$$

If $k \neq i$, then the above wedge product is zero because you will have duplicates in the wedge product (i.e. apply Corollary 5.24). Thus

$$d\omega = \sum_{i=1}^{j} \frac{\partial \omega_i}{\partial u^i} du^i \wedge du^1 \wedge \ldots \wedge du^{i-1} \wedge du^{i+1} \wedge \ldots \wedge du^j$$

$$= \sum_{i=1}^{j} (-1)^{i-1} \frac{\partial \omega_i}{\partial u^i} du^1 \wedge \ldots \wedge du^{i-1} \wedge du^i \wedge du^{i+1} \wedge \ldots \wedge du^j.$$

Having computed this, first suppose that φ is an interior chart and take a box $[-R, R] \times \ldots \times [-R, R]$ containing $\varphi[\operatorname{supp} \omega]$, which is possible since $\operatorname{supp} \omega$ is closed (by definition) and a subset of the compact and thus is compact, and so $\varphi[\operatorname{supp} \omega]$ is compact since φ is continuous. In this case $\omega = 0$ on ∂S and so $\int_{\partial S} \widetilde{\omega} = 0$. On the other hand, in these coordinates (we omit writing "$\circ f$")

$$(5.56) \quad \int_S d\omega = \int_U \sum_{i=1}^{j} (-1)^{i-1} \frac{\partial \omega_i}{\partial u^i} du^1 \ldots du^j = \sum_{i=1}^{j} (-1)^{i-1} \int_{-R}^{R} \cdots \int_{-R}^{R} \frac{\partial \omega_i}{\partial u^i} du^1 \ldots du^j.$$

By the fundamental theorem of calculus,

$$(5.57) \quad \int_{-R}^{R} \frac{\partial \omega_i}{\partial u^i} du^i = \omega_i \circ \varphi^{-1}(\underbrace{u^1, \ldots, R}_{i}, \ldots, u^j) - \omega_i \circ \varphi^{-1}(u^1, \ldots, -R, \ldots, u^j)$$

$$= 0 - 0 = 0.$$

So in each integral in the last quantity in (5.56) we can switch the dx^i integral to be integrated first, conclude from (5.57) that the integral is zero, from which we get that $\int_S d\omega = 0$. So in this case we indeed get that $\int_{\partial S} \widetilde{\omega} = \int_S d\omega$ simply because both sides are zero.

Next suppose that φ is a boundary chart and take a box $[-R, R] \times \ldots \times [0, R]$ containing $\varphi[\operatorname{supp} \omega]$. Then the analog of (5.56) in this case is

$$\int_S d\omega = \sum_{i=1}^{j} (-1)^{i-1} \int_0^R \int_{-R}^{R} \cdots \int_{-R}^{R} \frac{\partial \omega_i}{\partial u^i} du^1 \ldots du^j.$$

Using (5.57) again we conclude that every term of this sum is zero except for the j^{th} summand, in which case (5.57) does not apply since the integral in dx^j is only from 0 to R. Thus we get that (in the second equality below we integrate in du^j):

$$\int_S d\omega = (-1)^{j-1} \int_0^R \int_{-R}^{R} \cdots \int_{-R}^{R} \frac{\partial \omega_j}{\partial u^j} du^1 \ldots du^j$$

$$= (-1)^{j-1} \int_{-R}^{R} \cdots \int_{-R}^{R} [\underbrace{\omega_j \circ \varphi^{-1}(u^1, \ldots, u^{j-1}, R)}_{\text{equal to zero}} - \omega_j \circ \varphi^{-1}(u^1, \ldots, u^{j-1}, 0)] du^1 \ldots du^{j-1}$$

$$= (-1)^j \int_{-R}^{R} \cdots \int_{-R}^{R} \omega_j \circ \varphi^{-1}\left(u^1, \dots, u^{j-1}, 0\right) du^1 \dots du^{j-1}.$$

On the other hand, let's compute $\int_{\partial S} \widetilde{\omega}$. Let $\tilde{\varphi} : V \cap \partial S \to \partial \mathbb{H}^j \cong \mathbb{R}^{j-1}$ be the restriction of φ to ∂S, which is a chart of ∂S whose orientation is $(-1)^j$ since we assumed that φ is positively oriented. By (5.51) and (5.55) we have that in the coordinates of $\tilde{\varphi}$, $\widetilde{\omega}$ is given by:

$$\widetilde{\omega} = \omega_j \, du^1 \wedge \dots \wedge du^{j-1}.$$

Hence

$$\int_{\partial S} \widetilde{\omega} = (-1)^j \int_{-R}^{R} \cdots \int_{-R}^{R} \omega_j \circ \varphi^{-1}\left(u^1, \dots, u^{j-1}, 0\right) du^1 \dots du^{j-1}.$$

So we indeed get that $\int_{\partial S} \widetilde{\omega} = \int_S d\omega$ in this case as well.

Now suppose that $\operatorname{supp} \omega$ is not contained in the domain of any chart. Cover S by a finite collection of domain charts $\{V_i = \operatorname{dom} \varphi_i\}_{i=1}^k$. Let $\{\psi_i : S \to \mathbb{R}\}_{i=1}^k$ be a partition of unity subordinate to $\{V_i\}_{i=1}^k$. Then by linearity of integration and exterior differentiation and the fact that $\sum_{i=1}^k \psi_i = 1$,

$$\int_S d\omega = \int_S d\left(\sum_{i=1}^k \psi_i \omega\right) = \int_S \sum_{i=1}^k d(\psi_i \omega) = \sum_{i=1}^k \int_S d(\psi_i \omega).$$

Each $\operatorname{supp}(\psi_i \omega) \subseteq V_i = \operatorname{dom} \varphi_i$ and so by what we proved before, we have that each $\int_{\partial S} \psi_i \widetilde{\omega} = \int_S d(\psi_i \omega)$. Hence the above is equal to

$$= \sum_{i=1}^k \int_{\partial S} \psi_i \widetilde{\omega} = \int_{\partial S} \sum_{i=1}^k \psi_i \widetilde{\omega} = \int_{\partial S} \widetilde{\omega}.$$

■

What comes next: curvature, abstract manifolds, PDEs, geometric analysis (e.g. spectral theory), inverse problems, etc.