Multivariable Calculus, Applications and Theory

Kenneth Kuttler

August 19, 2011

Contents

0.1 Introduction

Multivariable calculus is just calculus which involves more than one variable. To do it properly, you have to use some linear algebra. Otherwise it is impossible to understand. This book presents the necessary linear algebra and then uses it as a framework upon which to build multivariable calculus. This is not the usual approach in beginning courses but it is the correct approach, leaving open the possibility that at least some students will learn and understand the topics presented. For example, the derivative of a function of many variables is a linear transformation. If you don't know what a linear transformation is, then you can't understand the derivative because that is what it is

and nothing else can be correctly substituted for it. The chain rule is best understood in terms of products of matrices which represent the various derivatives. The concepts involving multiple integrals involve determinants. The understandable version of the second derivative test uses eigenvalues, etc.

The purpose of this book is to present this subject in a way which can be understood by a motivated student. Because of the inherent difficulty, any treatment which is easy for the majority of students will not yield a correct understanding. However, the attempt is being made to make it as easy as possible.

Many applications are presented. Some of these are very difficult but worthwhile.

Hard sections are starred in the table of contents. Most of these sections are enrichment material and can be omitted if one desires nothing more than what is usually done in a standard calculus class. Stunningly difficult sections having substantial mathematical content are also decorated with a picture of a battle between a dragon slayer and a dragon, the outcome of the contest uncertain. These sections are for fearless students who want to understand the subject more than they want to preserve their egos. Sometimes the dragon wins.

Part I

Basic Linear Algebra

Fundamentals

1.0.1 Outcomes

1. Describe $\mathbb{R}^n$ and do algebra with vectors in $\mathbb{R}^n$.
2. Represent a line in 3 space by a vector parameterization, a set of scalar parametric equations or using symmetric form.
3. Find a parameterization of a line given information about
   (a) a point of the line and the direction of the line
   (b) two points contained in the line
4. Determine the direction of a line given its parameterization.

1.1 $\mathbb{R}^n$

The notation, $\mathbb{R}^n$ refers to the collection of ordered lists of n real numbers. More precisely, consider the following definition.

Definition 1.1.1 *Define*

$$\mathbb{R}^n \equiv \{(x_1, \cdots, x_n) : x_j \in \mathbb{R} \text{ for } j = 1, \cdots, n\}.$$

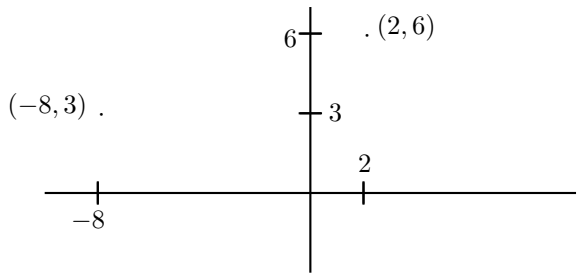
$(x_1, \cdots, x_n) = (y_1, \cdots, y_n)$ *if and only if for all* $j = 1, \cdots, n$, $x_j = y_j$. *When* $(x_1, \cdots, x_n) \in \mathbb{R}^n$, *it is conventional to denote* $(x_1, \cdots, x_n)$ *by the single bold face letter,* $\mathbf{x}$. *The numbers,* x_j *are called the* ***coordinates****. The set*

$$\{(0, \cdots, 0, t, 0, \cdots, 0) : t \in \mathbb{R}\}$$

for t *in the* i^{th} *slot is called the* i^{th} *coordinate axis* ***coordinate axis****, the* x_i *axis for short. The point* $\mathbf{0} \equiv (0, \cdots, 0)$ *is called the* ***origin****.*

Thus $(1, 2, 4) \in \mathbb{R}^3$ and $(2, 1, 4) \in \mathbb{R}^3$ but $(1, 2, 4) \neq (2, 1, 4)$ because, even though the same numbers are involved, they don't match up. In particular, the first entries are not equal.

Why would anyone be interested in such a thing? First consider the case when $n = 1$. Then from the definition, $\mathbb{R}^1 = \mathbb{R}$. Recall that $\mathbb{R}$ is identified with the points of a line. Look at the number line again. Observe that this amounts to identifying a point on this line with a real number. In other words a real number determines where you are on this line. Now suppose $n = 2$ and consider two lines which intersect each other at right angles as shown in the following picture.



Notice how you can identify a point shown in the plane with the ordered pair, $(2,6)$. You go to the right a distance of 2 and then up a distance of 6. Similarly, you can identify another point in the plane with the ordered pair $(-8,3)$. Go to the left a distance of 8 and then up a distance of 3. The reason you go to the left is that there is a $-$ sign on the eight. From this reasoning, every ordered pair determines a unique point in the plane. Conversely, taking a point in the plane, you could draw two lines through the point, one vertical and the other horizontal and determine unique points, x_1 on the horizontal line in the above picture and x_2 on the vertical line in the above picture, such that the point of interest is identified with the ordered pair, (x_1, x_2). In short, points in the plane can be identified with ordered pairs similar to the way that points on the real line are identified with real numbers. Now suppose $n = 3$. As just explained, the first two coordinates determine a point in a plane. Letting the third component determine how far up or down you go, depending on whether this number is positive or negative, this determines a point in space. Thus, $(1, 4, -5)$ would mean to determine the point in the plane that goes with $(1, 4)$ and then to go below this plane a distance of 5 to obtain a unique point in space. You see that the ordered triples correspond to points in space just as the ordered pairs correspond to points in a plane and single real numbers correspond to points on a line.

You can't stop here and say that you are only interested in $n \leq 3$. What if you were interested in the motion of two objects? You would need three coordinates to describe where the first object is and you would need another three coordinates to describe where the other object is located. Therefore, you would need to be considering $\mathbb{R}^6$. If the two objects moved around, you would need a time coordinate as well. As another example, consider a hot object which is cooling and suppose you want the temperature of this object. How many coordinates would be needed? You would need one for the temperature, three for the position of the point in the object and one more for the time. Thus you would need to be considering $\mathbb{R}^5$. Many other examples can be given. Sometimes n is very large. This is often the case in applications to business when they are trying to maximize profit subject to constraints. It also occurs in numerical analysis when people try to solve hard problems on a computer.

There are other ways to identify points in space with three numbers but the one presented is the most basic. In this case, the coordinates are known as **Cartesian coordinates** after Descartes[1] who invented this idea in the first half of the seventeenth century. I will often not bother to draw a distinction between the point in n dimensional space and its Cartesian coordinates.

[1] René Descartes 1596-1650 is often credited with inventing analytic geometry although it seems the ideas were actually known much earlier. He was interested in many different subjects, physiology, chemistry, and physics being some of them. He also wrote a large book in which he tried to explain the book of Genesis scientifically. Descartes ended up dying in Sweden.

1.2 Algebra in $\mathbb{R}^n$

There are two algebraic operations done with elements of $\mathbb{R}^n$. One is addition and the other is multiplication by numbers, called scalars.

Definition 1.2.1 *If $\mathbf{x} \in \mathbb{R}^n$ and a is a number, also called a **scalar**, then $a\mathbf{x} \in \mathbb{R}^n$ is defined by*

$$a\mathbf{x} = a(x_1, \cdots, x_n) \equiv (ax_1, \cdots, ax_n). \tag{1.1}$$

*This is known as **scalar multiplication**. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then $\mathbf{x} + \mathbf{y} \in \mathbb{R}^n$ and is defined by*

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (x_1, \cdots, x_n) + (y_1, \cdots, y_n) \\ &\equiv (x_1 + y_1, \cdots, x_n + y_n) \end{aligned} \tag{1.2}$$

*An element of $\mathbb{R}^n, \mathbf{x} \equiv (x_1, \cdots, x_n)$ is often called a **vector.** The above definition is known as **vector addition.***

With this definition, the algebraic properties satisfy the conclusions of the following theorem.

Theorem 1.2.2 *For $\mathbf{v}, \mathbf{w}$ vectors in $\mathbb{R}^n$ and α, β scalars, (real numbers), the following hold.*

$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}, \tag{1.3}$$

the commutative law of addition,

$$(\mathbf{v} + \mathbf{w}) + \mathbf{z} = \mathbf{v} + (\mathbf{w} + \mathbf{z}), \tag{1.4}$$

the associative law for addition,

$$\mathbf{v} + \mathbf{0} = \mathbf{v}, \tag{1.5}$$

the existence of an additive identity,

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}, \tag{1.6}$$

the existence of an additive inverse, Also

$$\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}, \tag{1.7}$$

$$(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}, \tag{1.8}$$

$$\alpha(\beta\mathbf{v}) = \alpha\beta(\mathbf{v}), \tag{1.9}$$

$$1\mathbf{v} = \mathbf{v}. \tag{1.10}$$

In the above $\mathbf{0} = (0, \cdots, 0)$.

You should verify these properties all hold. For example, consider 1.7

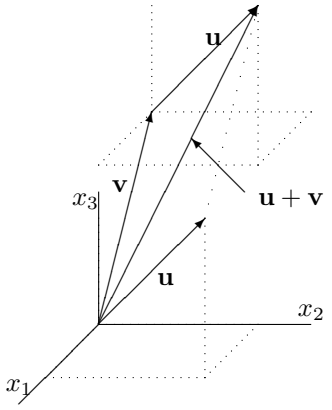
$$\begin{aligned} \alpha(\mathbf{v} + \mathbf{w}) &= \alpha(v_1 + w_1, \cdots, v_n + w_n) \\ &= (\alpha(v_1 + w_1), \cdots, \alpha(v_n + w_n)) \\ &= (\alpha v_1 + \alpha w_1, \cdots, \alpha v_n + \alpha w_n) \\ &= (\alpha v_1, \cdots, \alpha v_n) + (\alpha w_1, \cdots, \alpha w_n) \\ &= \alpha\mathbf{v} + \alpha\mathbf{w}. \end{aligned}$$

As usual subtraction is defined as $\mathbf{x} - \mathbf{y} \equiv \mathbf{x} + (-\mathbf{y})$.

1.3 Geometric Meaning Of Vector Addition In $\mathbb{R}^3$

It was explained earlier that an element of $\mathbb{R}^n$ is an n tuple of numbers and it was also shown that this can be used to determine a point in three dimensional space in the case where $n = 3$ and in two dimensional space, in the case where $n = 2$. This point was specified reletive to some coordinate axes.

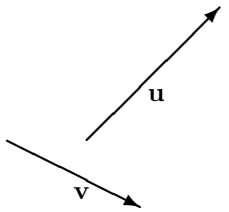
Consider the case where $n = 3$ for now. If you draw an arrow from the point in three dimensional space determined by $(0, 0, 0)$ to the point (a, b, c) with its tail sitting at the point $(0, 0, 0)$ and its point at the point (a, b, c), this arrow is called the **position vector** of the point determined by $\mathbf{u} \equiv (a, b, c)$. One way to get to this point is to start at $(0, 0, 0)$ and move in the direction of the x_1 axis to $(a, 0, 0)$ and then in the direction of the x_2 axis to $(a, b, 0)$ and finally in the direction of the x_3 axis to (a, b, c). It is evident that the same arrow (vector) would result if you began at the point, $\mathbf{v} \equiv (d, e, f)$, moved in the direction of the x_1 axis to $(d + a, e, f)$, then in the direction of the x_2 axis to $(d + a, e + b, f)$, and finally in the x_3 direction to $(d + a, e + b, f + c)$ only this time, the arrow would have its tail sitting at the point determined by $\mathbf{v} \equiv (d, e, f)$ and its point at $(d + a, e + b, f + c)$. It is said to be the same arrow (vector) because it will point in the same direction and have the same length. It is like you took an actual arrow, the sort of thing you shoot with a bow, and moved it from one location to another keeping it pointing the same direction. This is illustrated in the following picture in which $\mathbf{v} + \mathbf{u}$ is illustrated. Note the parallelogram determined in the picture by the vectors $\mathbf{u}$ and $\mathbf{v}$.



Thus the geometric significance of $(d, e, f) + (a, b, c) = (d + a, e + b, f + c)$ is this. You start with the position vector of the point (d, e, f) and at its point, you place the vector determined by (a, b, c) with its tail at (d, e, f). Then the point of this last vector will be $(d + a, e + b, f + c)$. This is the geometric significance of vector addition. Also, as shown in the picture, $\mathbf{u} + \mathbf{v}$ is the directed diagonal of the parallelogram determined by the two vectors $\mathbf{u}$ and $\mathbf{v}$.

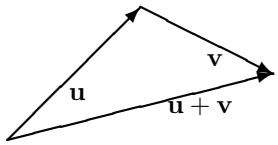
The following example is art.

Exercise 1.3.1 *Here is a picture of two vectors,* $\mathbf{u}$ *and* $\mathbf{v}$.

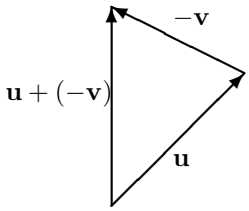


Sketch a picture of $\mathbf{u}+\mathbf{v}, \mathbf{u}-\mathbf{v}$, *and* $\mathbf{u}+2\mathbf{v}$.

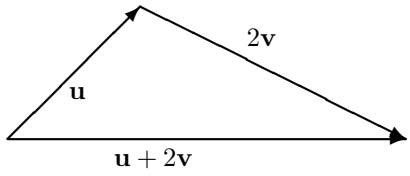
First here is a picture of $\mathbf{u}+\mathbf{v}$. You first draw $\mathbf{u}$ and then at the point of $\mathbf{u}$ you place the tail of $\mathbf{v}$ as shown. Then $\mathbf{u}+\mathbf{v}$ is the vector which results which is drawn in the following pretty picture.



Next consider $\mathbf{u}-\mathbf{v}$. This means $\mathbf{u}+(-\mathbf{v})$. From the above geometric description of vector addition, $-\mathbf{v}$ is the vector which has the same length but which points in the opposite direction to $\mathbf{v}$. Here is a picture.



Finally consider the vector $\mathbf{u}+2\mathbf{v}$. Here is a picture of this one also.

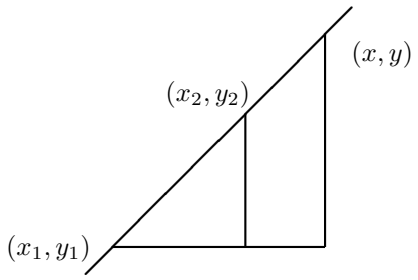


1.4 Lines

To begin with consider the case $n = 1, 2$. In the case where $n = 1$, the only line is just $\mathbb{R}^1 = \mathbb{R}$. Therefore, if x_1 and x_2 are two different points in $\mathbb{R}$, consider

$$x = x_1 + t(x_2 - x_1)$$

where $t \in \mathbb{R}$ and the totality of all such points will give $\mathbb{R}$. You see that you can always solve the above equation for t, showing that every point on $\mathbb{R}$ is of this form. Now consider the plane. Does a similar formula hold? Let (x_1, y_1) and (x_2, y_2) be two different points in $\mathbb{R}^2$ which are contained in a line, l. Suppose that $x_1 \neq x_2$. Then if (x, y) is an arbitrary point on l,



Now by similar triangles,

$$m \equiv \frac{y_2 - y_1}{x_2 - x_1} = \frac{y - y_1}{x - x_1}$$

and so the point slope form of the line, l, is given as

$$y - y_1 = m(x - x_1).$$

If t is defined by

$$x = x_1 + t(x_2 - x_1),$$

you obtain this equation along with

$$\begin{aligned} y &= y_1 + mt(x_2 - x_1) \\ &= y_1 + t(y_2 - y_1). \end{aligned}$$

Therefore,

$$(x, y) = (x_1, y_1) + t(x_2 - x_1, y_2 - y_1).$$

If $x_1 = x_2$, then in place of the point slope form above, $x = x_1$. Since the two given points are different, $y_1 \neq y_2$ and so you still obtain the above formula for the line. Because of this, the following is the definition of a line in $\mathbb{R}^n$.

Definition 1.4.1 *A line in $\mathbb{R}^n$ containing the two different points, $\mathbf{x}^1$ and $\mathbf{x}^2$ is the collection of points of the form*

$$\mathbf{x} = \mathbf{x}^1 + t\left(\mathbf{x}^2 - \mathbf{x}^1\right)$$

*where $t \in \mathbb{R}$. This is known as a **parametric equation** and the variable t is called the **parameter**.*

Often t denotes time in applications to Physics. Note this definition agrees with the usual notion of a line in two dimensions and so this is consistent with earlier concepts.

Lemma 1.4.2 *Let* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ *with* $\mathbf{a} \neq \mathbf{0}$. *Then* $\mathbf{x} = t\mathbf{a} + \mathbf{b}$, $t \in \mathbb{R}$, *is a line.*

Proof: Let $\mathbf{x}^1 = \mathbf{b}$ and let $\mathbf{x}^2 - \mathbf{x}^1 = \mathbf{a}$ so that $\mathbf{x}^2 \neq \mathbf{x}^1$. Then $t\mathbf{a} + \mathbf{b} = \mathbf{x}^1 + t\left(\mathbf{x}^2 - \mathbf{x}^1\right)$ and so $\mathbf{x} = t\mathbf{a} + \mathbf{b}$ is a line containing the two different points, $\mathbf{x}^1$ and $\mathbf{x}^2$. This proves the lemma.

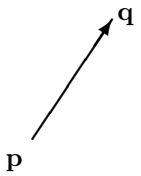
Definition 1.4.3 *The vector* $\mathbf{a}$ *in the above lemma is called a* ***direction vector*** *for the line.*

Definition 1.4.4 *Let* $\mathbf{p}$ *and* $\mathbf{q}$ *be two points in* $\mathbb{R}^n$, $\mathbf{p} \neq \mathbf{q}$. *The* ***directed line segment*** *from* $\mathbf{p}$ *to* $\mathbf{q}$, *denoted by* $\overrightarrow{\mathbf{pq}}$, *is defined to be the collection of points,*

$$\mathbf{x} = \mathbf{p} + t\left(\mathbf{q} - \mathbf{p}\right), \ t \in [0, 1]$$

with the direction corresponding to increasing t. *In the definition, when* $t = 0$, *the point* $\mathbf{p}$ *is obtained and as* t *increases other points on this line segment are obtained until when* $t = 1$, *you get the point,* $\mathbf{q}$. *This is what is meant by saying the direction corresponds to increasing* t.

Think of $\overrightarrow{\mathbf{pq}}$ as an arrow whose point is on $\mathbf{q}$ and whose base is at $\mathbf{p}$ as shown in the following picture.



This line segment is a part of a line from the above Definition.

Example 1.4.5 *Find a parametric equation for the line through the points* $(1, 2, 0)$ *and* $(2, -4, 6)$.

Use the definition of a line given above to write

$$(x, y, z) = (1, 2, 0) + t\left(1, -6, 6\right), \ t \in \mathbb{R}.$$

The vector $(1, -6, 6)$ is obtained by $(2, -4, 6) - (1, 2, 0)$ as indicated above.

The reason for the word, "a", rather than the word, "the" is there are infinitely many different parametric equations for the same line. To see this replace t with $3s$. Then you obtain a parametric equation for the same line because the same set of points is obtained. The difference is they are obtained from different values of the parameter. What happens is this: The line is a set of points but the parametric description gives more information than that. It tells how the set of points are obtained. Obviously, there are many ways to trace out a given set of points and each of these ways corresponds to a different parametric equation for the line.

Example 1.4.6 *Find a parametric equation for the line which contains the point* $(1, 2, 0)$ *and has direction vector,* $(1, 2, 1)$.

From the above this is just

$$(x,y,z) = (1,2,0) + t(1,2,1),\ t \in \mathbb{R}. \tag{1.11}$$

Sometimes people elect to write a line like the above in the form

$$x = 1 + t,\ y = 2 + 2t,\ z = t,\ t \in \mathbb{R}. \tag{1.12}$$

This is a set of scalar parametric equations which amounts to the same thing as 1.11.

There is one other form for a line which is sometimes considered useful. It is the so called symmetric form. Consider the line of 1.12. You can solve for the parameter, t to write

$$t = x - 1, t = \frac{y-2}{2},\ t = z.$$

Therefore,

$$x - 1 = \frac{y-2}{2} = z.$$

This is the symmetric form of the line.

Example 1.4.7 *Suppose the* ***symmetric form of a line*** *is*

$$\frac{x-2}{3} = \frac{y-1}{2} = z + 3.$$

Find the line in parametric form.

Let $t = \frac{x-2}{3}, t = \frac{y-1}{2}$ and $t = z + 3$. Then solving for x, y, z, you get

$$x = 3t + 2,\ y = 2t + 1,\ z = t - 3,\ t \in \mathbb{R}.$$

Written in terms of vectors this is

$$(2,1,-3) + t(3,2,1) = (x,y,z),\ t \in \mathbb{R}.$$

1.5 Distance in $\mathbb{R}^n$

How is distance between two points in $\mathbb{R}^n$ defined?

Definition 1.5.1 *Let* $\mathbf{x} = (x_1, \cdots, x_n)$ *and* $\mathbf{y} = (y_1, \cdots, y_n)$ *be two points in* $\mathbb{R}^n$. *Then* $|\mathbf{x} - \mathbf{y}|$ *to indicates the distance between these points and is defined as*

$$\text{distance between } \mathbf{x} \text{ and } \mathbf{y} \equiv |\mathbf{x} - \mathbf{y}| \equiv \left(\sum_{k=1}^{n} |x_k - y_k|^2 \right)^{1/2}.$$

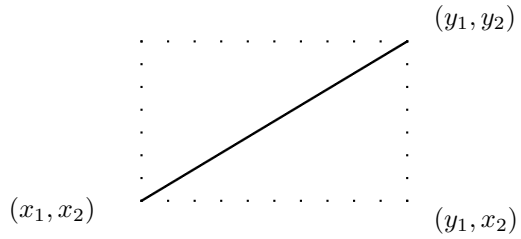
This is called the ***distance formula****. Thus* $|\mathbf{x}| \equiv |\mathbf{x} - \mathbf{0}|$. *The symbol,* $B(\mathbf{a}, r)$ *is defined by*

$$B(\mathbf{a}, r) \equiv \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{a}| < r\}.$$

This is called an ***open ball*** *of radius* r *centered at* $\mathbf{a}$. *It gives all the points in* $\mathbb{R}^n$ *which are closer to* $\mathbf{a}$ *than* r.

First of all note this is a generalization of the notion of distance in $\mathbb{R}$. There the distance between two points, x and y was given by the absolute value of their difference. Thus $|x - y|$ is equal to the distance between these two points on $\mathbb{R}$. Now $|x - y| = \left((x-y)^2\right)^{1/2}$ where the square root is always the positive square root. Thus it is

the same formula as the above definition except there is only one term in the sum. Geometrically, this is the right way to define distance which is seen from the Pythagorean theorem. Consider the following picture in the case that $n = 2$.

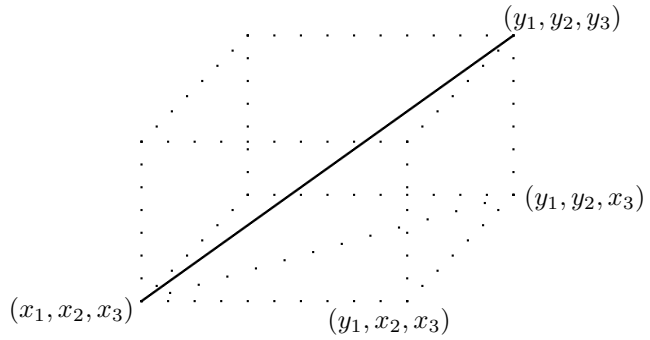


There are two points in the plane whose Cartesian coordinates are (x_1, x_2) and (y_1, y_2) respectively. Then the solid line joining these two points is the hypotenuse of a right triangle which is half of the rectangle shown in dotted lines. What is its length? Note the lengths of the sides of this triangle are $|y_1 - x_1|$ and $|y_2 - x_2|$. Therefore, the Pythagorean theorem implies the length of the hypotenuse equals

$$\left(|y_1 - x_1|^2 + |y_2 - x_2|^2\right)^{1/2} = \left((y_1 - x_1)^2 + (y_2 - x_2)^2\right)^{1/2}$$

which is just the formula for the distance given above.

Now suppose $n = 3$ and let (x_1, x_2, x_3) and (y_1, y_2, y_3) be two points in $\mathbb{R}^3$. Consider the following picture in which one of the solid lines joins the two points and a dotted line joins the points (x_1, x_2, x_3) and (y_1, y_2, x_3).



By the Pythagorean theorem, the length of the dotted line joining (x_1, x_2, x_3) and (y_1, y_2, x_3) equals

$$\left((y_1 - x_1)^2 + (y_2 - x_2)^2\right)^{1/2}$$

while the length of the line joining (y_1, y_2, x_3) to (y_1, y_2, y_3) is just $|y_3 - x_3|$. Therefore, by the Pythagorean theorem again, the length of the line joining the points (x_1, x_2, x_3)

and (y_1, y_2, y_3) equals

$$\left\{\left[\left(\left(y_1-x_1\right)^2+\left(y_2-x_2\right)^2\right)^{1/2}\right]^2+\left(y_3-x_3\right)^2\right\}^{1/2}$$
$$=\left(\left(y_1-x_1\right)^2+\left(y_2-x_2\right)^2+\left(y_3-x_3\right)^2\right)^{1/2},$$

which is again just the distance formula above.

This completes the argument that the above definition is reasonable. Of course you cannot continue drawing pictures in ever higher dimensions but there is no problem with the formula for distance in any number of dimensions. Here is an example.

Example 1.5.2 *Find the distance between the points in* $\mathbb{R}^4$*,*

$$\mathbf{a}=(1,2,-4,6)$$

and $\mathbf{b}=(2,3,-1,0)$

Use the distance formula and write

$$|\mathbf{a}-\mathbf{b}|^2=(1-2)^2+(2-3)^2+(-4-(-1))^2+(6-0)^2=47$$

Therefore, $|\mathbf{a}-\mathbf{b}|=\sqrt{47}$.

All this amounts to defining the distance between two points as the length of a straight line joining these two points. However, there is nothing sacred about using straight lines. One could define the distance to be the length of some other sort of line joining these points. It won't be done in this book but sometimes this sort of thing is done.

Another convention which is usually followed, especially in $\mathbb{R}^2$ and $\mathbb{R}^3$ is to denote the first component of a point in $\mathbb{R}^2$ by x and the second component by y. In $\mathbb{R}^3$ it is customary to denote the first and second components as just described while the third component is called z.

Example 1.5.3 *Describe the points which are at the same distance between* $(1,2,3)$ *and* $(0,1,2)$.

Let (x,y,z) be such a point. Then

$$\sqrt{(x-1)^2+(y-2)^2+(z-3)^2}=\sqrt{x^2+(y-1)^2+(z-2)^2}.$$

Squaring both sides

$$(x-1)^2+(y-2)^2+(z-3)^2=x^2+(y-1)^2+(z-2)^2$$

and so

$$x^2-2x+14+y^2-4y+z^2-6z=x^2+y^2-2y+5+z^2-4z$$

which implies

$$-2x+14-4y-6z=-2y+5-4z$$

and so

$$2x+2y+2z=-9. \tag{1.13}$$

Since these steps are reversible, the set of points which is at the same distance from the two given points consists of the points, (x,y,z) such that 1.13 holds.

The following lemma is fundamental. It is a form of the Cauchy Schwarz inequality.

Lemma 1.5.4 *Let* $\mathbf{x} = (x_1, \cdots, x_n)$ *and* $\mathbf{y} = (y_1, \cdots, y_n)$ *be two points in* $\mathbb{R}^n$. *Then*

$$\left| \sum_{i=1}^{n} x_i y_i \right| \leq |\mathbf{x}| \, |\mathbf{y}| . \tag{1.14}$$

Proof: Let θ be either 1 or -1 such that

$$\theta \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} x_i \left(\theta y_i\right) = \left| \sum_{i=1}^{n} x_i y_i \right|$$

and consider $p(t) \equiv \sum_{i=1}^{n} \left(x_i + t\theta y_i\right)^2$. Then for all $t \in \mathbb{R}$,

$$\begin{aligned} 0 &\leq p(t) = \sum_{i=1}^{n} x_i^2 + 2t \sum_{i=1}^{n} x_i \theta y_i + t^2 \sum_{i=1}^{n} y_i^2 \\ &= |\mathbf{x}|^2 + 2t \sum_{i=1}^{n} x_i \theta y_i + t^2 |\mathbf{y}|^2 \end{aligned}$$

If $|\mathbf{y}| = 0$ then 1.14 is obviously true because both sides equal zero. Therefore, assume $|\mathbf{y}| \neq 0$ and then $p(t)$ is a polynomial of degree two whose graph opens up. Therefore, it either has no zeroes, two zeros or one repeated zero. If it has two zeros, the above inequality must be violated because in this case the graph must dip below the x axis. Therefore, it either has no zeros or exactly one. From the quadratic formula this happens exactly when

$$4 \left(\sum_{i=1}^{n} x_i \theta y_i \right)^2 - 4 |\mathbf{x}|^2 |\mathbf{y}|^2 \leq 0$$

and so

$$\sum_{i=1}^{n} x_i \theta y_i = \left| \sum_{i=1}^{n} x_i y_i \right| \leq |\mathbf{x}| \, |\mathbf{y}|$$

as claimed. This proves the inequality.

There are certain properties of the distance which are obvious. Two of them which follow directly from the definition are

$$|\mathbf{x} - \mathbf{y}| = |\mathbf{y} - \mathbf{x}| ,$$

$$|\mathbf{x} - \mathbf{y}| \geq 0 \text{ and equals 0 only if } \mathbf{y} = \mathbf{x}.$$

The third fundamental property of distance is known as the triangle inequality. Recall that in any triangle the sum of the lengths of two sides is always at least as large as the third side. The following corollary is equivalent to this simple statement.

Corollary 1.5.5 *Let* $\mathbf{x}, \mathbf{y}$ *be points of* $\mathbb{R}^n$. *Then*

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}| .$$

Proof: Using the Cauchy Schwarz inequality, Lemma 1.5.4,

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &\equiv \sum_{i=1}^{n} \left(x_i + y_i\right)^2 \\ &= \sum_{i=1}^{n} x_i^2 + 2 \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 \\ &\leq |\mathbf{x}|^2 + 2 |\mathbf{x}| \, |\mathbf{y}| + |\mathbf{y}|^2 \\ &= \left(|\mathbf{x}| + |\mathbf{y}|\right)^2 \end{aligned}$$

and so upon taking square roots of both sides,

$$|\mathbf{x}+\mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$$

and this proves the corollary.

1.6 Geometric Meaning Of Scalar Multiplication In $\mathbb{R}^3$

As discussed earlier, $\mathbf{x} = (x_1, x_2, x_3)$ determines a vector. You draw the line from $\mathbf{0}$ to $\mathbf{x}$ placing the point of the vector on $\mathbf{x}$. What is the length of this vector? The length of this vector is defined to equal $|\mathbf{x}|$ as in Definition 1.5.1. Thus the length of $\mathbf{x}$ equals $\sqrt{x_1^2 + x_2^2 + x_3^2}$. When you multiply $\mathbf{x}$ by a scalar, α, you get $(\alpha x_1, \alpha x_2, \alpha x_3)$ and the length of this vector is defined as $\sqrt{\left((\alpha x_1)^2 + (\alpha x_2)^2 + (\alpha x_3)^2\right)} = |\alpha| \sqrt{x_1^2 + x_2^2 + x_3^2}$. Thus the following holds.

$$|\alpha \mathbf{x}| = |\alpha| \, |\mathbf{x}| .$$

In other words, multiplication by a scalar magnifies the length of the vector. What about the direction? You should convince yourself by drawing a picture that if α is negative, it causes the resulting vector to point in the opposite direction while if $\alpha > 0$ it preserves the direction the vector points. One way to see this is to first observe that if $\alpha \neq 1$, then $\mathbf{x}$ and $\alpha\mathbf{x}$ are both points on the same line.

1.7 Exercises

1. Verify all the properties 1.3-1.10.
2. Compute the following
   (a) $5(1,2,3,-2) + 6(2,1,-2,7)$
   (b) $5(1,2,-2) - 6(2,1,-2)$
   (c) $-3(1,0,3,-2) + (2,0,-2,1)$
   (d) $-3(1,-2,-3,-2) - 2(2,-1,-2,7)$
   (e) $-(2,2,-3,-2) + 2(2,4,-2,7)$
3. Find symmetric equations for the line through the points $(2,2,4)$ and $(-2,3,1)$.
4. Find symmetric equations for the line through the points $(1,2,4)$ and $(-2,1,1)$.
5. Symmetric equations for a line are given. Find parametric equations of the line.
   (a) ♠ $\frac{x+1}{3} = \frac{2y+3}{2} = z+7$
   (b) ♠ $\frac{2x-1}{3} = \frac{2y+3}{6} = z-7$
   (c) ♠ $\frac{x+1}{3} = 2y+3 = 2z-1$
   (d) $\frac{1-2x}{3} = \frac{3-2y}{2} = z+1$
   (e) $\frac{x-1}{3} = \frac{2y-3}{5} = z+2$
   (f) $\frac{x+1}{3} = \frac{3-y}{5} = z+1$
6. Parametric equations for a line are given. Find symmetric equations for the line if possible. If it is not possible to do it explain why.

(a) ♠$x = 1 + 2t, y = 3 - t, z = 5 + 3t$

(b) ♠$x = 1 + t, y = 3 - t, z = 5 - 3t$

(c) ♠$x = 1 + 2t, y = 3 + t, z = 5 + 3t$

(d) $x = 1 - 2t, y = 1, z = 1 + t$

(e) $x = 1 - t, y = 3 + 2t, z = 5 - 3t$

(f) $x = t, y = 3 - t, z = 1 + t$

7. The first point given is a point containing the line. The second point given is a direction vector for the line. Find parametric equations for the line determined by this information.

(a) ♠$(1, 2, 1), (2, 0, 3)$

(b) ♠$(1, 0, 1), (1, 1, 3)$

(c) ♠$(1, 2, 0), (1, 1, 0)$

(d) $(1, 0, -6), (-2, -1, 3)$

(e) $(-1, -2, -1), (2, 1, -1)$

(f) $(0, 0, 0), (2, -3, 1)$

8. Parametric equations for a line are given. Determine a direction vector for this line.

(a) ♠$x = 1 + 2t, y = 3 - t, z = 5 + 3t$

(b) ♠$x = 1 + t, y = 3 + 3t, z = 5 - t$

(c) ♠$x = 7 + t, y = 3 + 4t, z = 5 - 3t$

(d) $x = 2t, y = -3t, z = 3t$

(e) $x = 2t, y = 3 + 2t, z = 5 + t$

(f) $x = t, y = 3 + 3t, z = 5 + t$

9. A line contains the given two points. Find parametric equations for this line. Identify the direction vector.

(a) ♠$(0, 1, 0), (2, 1, 2)$

(b) ♠$(0, 1, 1), (2, 5, 0)$

(c) $(1, 1, 0), (0, 1, 2)$

(d) $(0, 1, 3), (0, 3, 0)$

(e) $(0, 1, 0), (0, 6, 2)$

(f) $(0, 1, 2), (2, 0, 2)$

10. Draw a picture of the points in $\mathbb{R}^2$ which are determined by the following ordered pairs.

(a) $(1, 2)$

(b) $(-2, -2)$

(c) $(-2, 3)$

(d) $(2, -5)$

11. Does it make sense to write $(1, 2) + (2, 3, 1)$? Explain.

12. Draw a picture of the points in $\mathbb{R}^3$ which are determined by the following ordered triples.

    (a) $(1,2,0)$

    (b) $(-2,-2,1)$

    (c) $(-2,3,-2)$

13. You are given two points in $\mathbb{R}^3, (4,5,-4)$ and $(2,3,0)$. Show the distance from the point, $(3,4,-2)$ to the first of these points is the same as the distance from this point to the second of the original pair of points. Note that $3 = \frac{4+2}{2}, 4 = \frac{5+3}{2}$. Obtain a theorem which will be valid for general pairs of points, (x,y,z) and (x_1,y_1,z_1) and prove your theorem using the distance formula.

14. A sphere is the set of all points which are at a given distance from a single given point. Find an equation for the sphere which is the set of all points that are at a distance of 4 from the point $(1,2,3)$ in $\mathbb{R}^3$.

15. A parabola is the set of all points (x,y) in the plane such that the distance from the point (x,y) to a given point, (x_0,y_0) equals the distance from (x,y) to a given line. The point, (x_0,y_0) is called the **focus** and the line is called the **directrix**. Find the equation of the parabola which results from the line $y = l$ and (x_0,y_0) a given focus with $y_0 < l$. Repeat for $y_0 > l$.

16. A sphere centered at the point $(x_0,y_0,z_0) \in \mathbb{R}^3$ having radius r consists of all points, (x,y,z) whose distance to (x_0,y_0,z_0) equals r. Write an equation for this sphere in $\mathbb{R}^3$.

17. Suppose the distance between (x,y) and (x',y') were defined to equal the larger of the two numbers $|x-x'|$ and $|y-y'|$. Draw a picture of the sphere centered at the point, $(0,0)$ if this notion of distance is used.

18. Repeat the same problem except this time let the distance between the two points be $|x-x'| + |y-y'|$.

19. If (x_1,y_1,z_1) and (x_2,y_2,z_2) are two points such that $|(x_i,y_i,z_i)| = 1$ for $i = 1,2$, show that in terms of the usual distance, $\left|\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}, \frac{z_1+z_2}{2}\right)\right| < 1$. What would happen if you used the way of measuring distance given in Problem 17 ($|(x,y,z)| =$ maximum of $|z|, |x|, |y|$.)?

20. Give a simple description using the distance formula of the set of points which are at an equal distance between the two points (x_1,y_1,z_1) and (x_2,y_2,z_2).

21. Suppose you are given two points, $(-a,0)$ and $(a,0)$ in $\mathbb{R}^2$ and a number, $r > 2a$. The set of points described by

$$\left\{(x,y) \in \mathbb{R}^2 : |(x,y)-(-a,0)| + |(x,y)-(a,0)| = r\right\}$$

    is known as an ellipse. The two given points are known as the **focus points** of the ellipse. Simplify this to the form $\left(\frac{x-A}{\alpha}\right)^2 + \left(\frac{y}{\beta}\right)^2 = 1$. This is a nice exercise in messy algebra.

22. Suppose you are given two points, $(-a,0)$ and $(a,0)$ in $\mathbb{R}^2$ and a number, $r > 2a$. The set of points described by

$$\left\{(x,y) \in \mathbb{R}^2 : |(x,y)-(-a,0)| - |(x,y)-(a,0)| = r\right\}$$

is known as **hyperbola**. The two given points are known as the **focus points** of the hyperbola. Simplify this to the form $\left(\frac{x-A}{\alpha}\right)^2 - \left(\frac{y}{\beta}\right)^2 = 1$. This is a nice exercise in messy algebra.

23. Let (x_1, y_1) and (x_2, y_2) be two points in $\mathbb{R}^2$. Give a simple description using the distance formula of the perpendicular bisector of the line segment joining these two points. Thus you want all points, (x, y) such that $|(x, y) - (x_1, y_1)| = |(x, y) - (x_2, y_2)|$.

1.8 Exercises With Answers

1. Compute the following

   (a) $5(1, -2, 3, -2) + 4(2, 1, -2, 7) = (13, -6, 7, 18)$

   (b) $2(1, 2, 1) + 6(2, 9, -2) = (14, 58, -10)$

   (c) $-3(1, 0, -4, -2) + 3(2, 4, -2, 1) = (3, 12, 6, 9)$

2. Find symmetric equations for the line through the points $(2, 7, 4)$ and $(-1, 3, 1)$.

   First find parametric equations. These are $(x, y, z) = (2, 7, 4) + t(-3, -4, -3)$. Therefore,

   $$t = \frac{x-2}{-3} = \frac{y-7}{-4} = \frac{z-4}{-3}.$$

   The symmetric equations of this line are therefore,

   $$\frac{x-2}{-3} = \frac{y-7}{-4} = \frac{z-4}{-3}.$$

3. Symmetric equations for a line are given as $\frac{1-2x}{3} = \frac{3+2y}{2} = z+1$. Find parametric equations of the line.

   Let $t = \frac{1-2x}{3} = \frac{3+2y}{2} = z+1$. Then $x = \frac{3t-1}{-2}, y = \frac{2t-3}{2}, z = t-1$.

4. Parametric equations for a line are $x = 1-t, y = 3+2t, z = 5-3t$. Find symmetric equations for the line if possible. If it is not possible to do it explain why.

   Solve the parametric equations for t. This gives

   $$t = 1 - x = \frac{y-3}{2} = \frac{5-z}{3}.$$

   Thus symmetric equations for this line are

   $$1 - x = \frac{y-3}{2} = \frac{5-z}{3}.$$

5. Parametric equations for a line are $x = 1, y = 3+2t, z = 5-3t$. Find symmetric equations for the line if possible. If it is not possible to do it explain why.

   In this case you can't do it. The second two equations give $\frac{y-3}{2} = \frac{5-z}{3}$ but you can't have these equal to an expression of the form $\frac{ax+b}{c}$ because x is always equal to 1. Thus any expression of this form must be constant but the other two are not constant.

6. The first point given is a point containing the line. The second point given is a direction vector for the line. Find parametric equations for the line determined by this information. $(1,1,2),(2,1,-3)$. Parametric equations are equivalent to $(x,y,z)=(1,1,2)+t(2,1,-3)$. Written parametrically, $x=1+2t, y=1+t, z=2-3t$.

7. Parametric equations for a line are given. Determine a direction vector for this line. $x=t, y=3+2t, z=5+t$

   A direction vector is $(1,2,1)$. You just form the vector which has components equal to the coefficients of t in the parametric equations for x, y, and z respectively.

8. A line contains the given two points. Find parametric equations for this line. Identify the direction vector. $(1,-2,0),(2,1,2)$.

   A direction vector is $(1,3,2)$ and so parametric equations are equivalent to $(x,y,z)=(2,1,2)+(1,3,2)$. Of course you could also have written $(x,y,z)=(1,-2,0)+t(1,3,2)$ or $(x,y,z)=(1,-2,0)-t(1,3,2)$ or $(x,y,z)=(1,-2,0)+t(2,6,4)$, etc. As explained above, there are always infinitely many parameterizations for a given line.

Matrices And Linear Transformations

2.0.1 Outcomes

1. Perform the basic matrix operations of matrix addition, scalar multiplication, transposition and matrix multiplication. Identify when these operations are not defined. Represent the basic operations in terms of double subscript notation.

2. Recall and prove algebraic properties for matrix addition, scalar multiplication, transposition, and matrix multiplication. Apply these properties to manipulate an algebraic expression involving matrices.

3. Evaluate the inverse of a matrix using Gauss Jordan elimination.

4. Recall the cancellation laws for matrix multiplication. Demonstrate when cancellation laws do not apply.

5. Recall and prove identities involving matrix inverses.

6. Understand the relationship between linear transformations and matrices.

2.1 Matrix Arithmetic

2.1.1 Addition And Scalar Multiplication Of Matrices

When people speak of vectors and matrices, it is common to refer to numbers as **scalars**. In this book, scalars will always be real numbers.

A **matrix** is a rectangular array of numbers. Several of them are referred to as **matrices**. For example, here is a matrix.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{pmatrix}$$

The size or dimension of a matrix is defined as $m \times n$ where m is the number of rows and n is the number of columns. The above matrix is a 3×4 matrix because there are three rows and four columns. The first row is $(1\ 2\ 3\ 4)$, the second row is $(5\ 2\ 8\ 7)$ and so forth. The first column is $\begin{pmatrix} 1 \\ 5 \\ 6 \end{pmatrix}$. When specifying the size of a matrix, you always list the number of rows before the number of columns. Also, you can remember the columns are like columns in a Greek temple. They stand upright while the rows

just lay there like rows made by a tractor in a plowed field. Elements of the matrix are identified according to position in the matrix. For example, 8 is in position $2,3$ because it is in the second row and the third column. You might remember that you always list the rows before the columns by using the phrase **Row**man **C**atholic. The symbol, (a_{ij}) refers to a matrix. The entry in the i^{th} row and the j^{th} column of this matrix is denoted by a_{ij}. Using this notation on the above matrix, $a_{23} = 8, a_{32} = -9, a_{12} = 2$, etc.

There are various operations which are done on matrices. Matrices can be added multiplied by a scalar, and multiplied by other matrices. To illustrate scalar multiplication, consider the following example in which a matrix is being multiplied by the scalar, 3.

$$3\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 9 & 12 \\ 15 & 6 & 24 & 21 \\ 18 & -27 & 3 & 6 \end{pmatrix}.$$

The new matrix is obtained by multiplying every entry of the original matrix by the given scalar. If A is an $m \times n$ matrix, $-A$ is defined to equal $(-1)\, A$.

Two matrices must be the same size to be added. The sum of two matrices is a matrix which is obtained by adding the corresponding entries. Thus

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 2 \end{pmatrix} + \begin{pmatrix} -1 & 4 \\ 2 & 8 \\ 6 & -4 \end{pmatrix} = \begin{pmatrix} 0 & 6 \\ 5 & 12 \\ 11 & -2 \end{pmatrix}.$$

Two matrices are equal exactly when they are the same size and the corresponding entries areidentical. Thus

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

because they are different sizes. As noted above, you write (c_{ij}) for the matrix C whose ij^{th} entry is c_{ij}. In doing arithmetic with matrices you must define what happens in terms of the c_{ij} sometimes called the **entries** of the matrix or the **components** of the matrix.

The above discussion stated for general matrices is given in the following definition.

Definition 2.1.1 *(Scalar Multiplication) If $A = (a_{ij})$ and k is a scalar, then $kA = (ka_{ij})$.*

Example 2.1.2 $7\begin{pmatrix} 2 & 0 \\ 1 & -4 \end{pmatrix} = \begin{pmatrix} 14 & 0 \\ 7 & -28 \end{pmatrix}.$

Definition 2.1.3 *(Addition) If $A = (a_{ij})$ and $B = (b_{ij})$ are two $m \times n$ matrices. Then $A + B = C$ where*

$$C = (c_{ij})$$

for $c_{ij} = a_{ij} + b_{ij}$.

Example 2.1.4

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 2 & 3 \\ -6 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 4 & 6 \\ -5 & 2 & 5 \end{pmatrix}$$

To save on notation, A_{ij} will refer to the ij^{th} entry of the matrix, A.

Definition 2.1.5 *(The **zero matrix**) The $m \times n$ zero matrix is the $m \times n$ matrix having every entry equal to zero. It is denoted by* 0.

Example 2.1.6 *The* 2×3 *zero matrix is* $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Note there are 2×3 zero matrices, 3×4 zero matrices, etc. In fact there is a zero matrix for every size.

Definition 2.1.7 *(Equality of matrices) Let* A *and* B *be two matrices. Then* $A = B$ *means that the two matrices are of the same size and for* $A = (a_{ij})$ *and* $B = (b_{ij})$, $a_{ij} = b_{ij}$ *for all* $1 \leq i \leq m$ *and* $1 \leq j \leq n$.

The following properties of matrices can be easily verified. You should do so.

- Commutative Law Of Addition.

$$A + B = B + A, \tag{2.1}$$

- Associative Law for Addition.

$$(A + B) + C = A + (B + C), \tag{2.2}$$

- Existence of an Additive Identity

$$A + 0 = A, \tag{2.3}$$

- Existence of an Additive Inverse

$$A + (-A) = 0, \tag{2.4}$$

Also for α, β scalars, the following additional properties hold.

- Distributive law over Matrix Addition.

$$\alpha (A + B) = \alpha A + \alpha B, \tag{2.5}$$

- Distributive law over Scalar Addition

$$(\alpha + \beta) A = \alpha A + \beta A, \tag{2.6}$$

- Associative law for Scalar Multiplication

$$\alpha (\beta A) = \alpha \beta (A), \tag{2.7}$$

- Rule for Multiplication by 1.

$$1A = A. \tag{2.8}$$

As an example, consider the Commutative Law of Addition. Let $A + B = C$ and $B + A = D$. Why is $D = C$?

$$C_{ij} = A_{ij} + B_{ij} = B_{ij} + A_{ij} = D_{ij}.$$

Therefore, $C = D$ because the ij^{th} entries are the same. Note that the conclusion follows from the commutative law of addition of numbers.

2.1.2 Multiplication Of Matrices

Definition 2.1.8 *Matrices which are $n \times 1$ or $1 \times n$ are called **vectors** and are often denoted by a bold letter. Thus the $n \times 1$ matrix*

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

*is also called a **column vector.** The $1 \times n$ matrix*

$$(x_1 \cdots x_n)$$

*is called a **row vector**.*

Although the following description of matrix multiplication may seem strange, it is in fact the most important and useful of the matrix operations. To begin with consider the case where a matrix is multiplied by a column vector. First consider a special case.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = ?$$

One way to remember this is as follows. Slide the vector, placing it on top the two rows as shown and then do the indicated operation.

$$\begin{pmatrix} \mathbf{7} & \mathbf{8} & \mathbf{9} \\ 1 & 2 & 3 \\ \mathbf{7} & \mathbf{8} & \mathbf{9} \\ 4 & 5 & 6 \end{pmatrix} \rightarrow \begin{pmatrix} 7 \times 1 + 8 \times 2 + 9 \times 3 \\ 7 \times 4 + 8 \times 5 + 9 \times 6 \end{pmatrix} = \begin{pmatrix} 50 \\ 122 \end{pmatrix}.$$

multiply the numbers on the top by the numbers on the bottom and add them up to get a single number for each row of the matrix as shown above.

In more general terms,

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{pmatrix}.$$

Another way to think of this is

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} + x_3 \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}$$

Thus you take x_1 times the first column, add to x_2 times the second column, and finally x_3 times the third column. In general, here is the definition of how to multiply an $(m \times n)$ matrix times a $(n \times 1)$ matrix.

Definition 2.1.9 *Let $A = A_{ij}$ be an $m \times n$ matrix and let $\mathbf{v}$ be an $n \times 1$ matrix,*

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

Then $A\mathbf{v}$ is an $m \times 1$ matrix and the i^{th} component of this matrix is

$$(A\mathbf{v})_i = A_{i1}v_1 + A_{i2}v_2 + \cdots + A_{in}v_n = \sum_{j=1}^{n} A_{ij}v_j.$$

Thus

$$A\mathbf{v} = \begin{pmatrix} \sum_{j=1}^{n} A_{1j}v_j \\ \vdots \\ \sum_{j=1}^{n} A_{mj}v_j \end{pmatrix}. \tag{2.9}$$

In other words, if

$$A = (\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

where the $\mathbf{a}_k$ *are the columns,*

$$A\mathbf{v} = \sum_{k=1}^{n} v_k \mathbf{a}_k$$

This follows from 2.9 and the observation that the j^{th} *column of* A *is*

$$\begin{pmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{pmatrix}$$

so 2.9 reduces to

$$v_1 \begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} + v_2 \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} + \cdots + v_k \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix}$$

Note also that multiplication by an $m \times n$ *matrix takes an* $n \times 1$ *matrix, and produces an* $m \times 1$ *matrix.*

Here is another example.

Example 2.1.10 *Compute*

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

First of all this is of the form $(3 \times 4)(4 \times 1)$ and so the result should be a (3×1). Note how the inside numbers cancel. To get the element in the second row and first and only column, compute

$$\begin{aligned} \sum_{k=1}^{4} a_{2k}v_k &= a_{21}v_1 + a_{22}v_2 + a_{23}v_3 + a_{24}v_4 \\ &= 0 \times 1 + 2 \times 2 + 1 \times 0 + (-2) \times 1 = 2. \end{aligned}$$

You should do the rest of the problem and verify

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \\ 5 \end{pmatrix}.$$

The next task is to multiply an $m \times n$ matrix times an $n \times p$ matrix. Before doing so, the following may be helpful.

For A and B matrices, in order to form the product, AB the number of columns of A must equal the number of rows of B.

$$(m \times \overbrace{\quad n}^{\text{these must match!}}) (n \times p \quad) = m \times p$$

Note the two outside numbers give the size of the product. Remember:

If the two middle numbers don't match, you can't multiply the matrices!

Definition 2.1.11 *When the number of columns of A equals the number of rows of B the two matrices are said to be* ***conformable*** *and the product, AB is obtained as follows. Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. Then B is of the form*

$$B = (\mathbf{b}_1, \cdots, \mathbf{b}_p)$$

where $\mathbf{b}_k$ is an $n \times 1$ matrix or column vector. Then the $m \times p$ matrix, AB is defined as follows:

$$AB \equiv (A\mathbf{b}_1, \cdots, A\mathbf{b}_p) \tag{2.10}$$

where $A\mathbf{b}_k$ is an $m \times 1$ matrix or column vector which gives the k^{th} column of AB.

Example 2.1.12 *Multiply the following.*

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix}$$

The first thing you need to check before doing anything else is whether it is possible to do the multiplication. The first matrix is a 2×3 and the second matrix is a 3×3. Therefore, is it possible to multiply these matrices. According to the above discussion it should be a 2×3 matrix of the form

$$\left(\overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}}^{\text{First column}}, \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}}^{\text{Second column}}, \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}}^{\text{Third column}} \right)$$

You know how to multiply a matrix times a vector and so you do so to obtain each of the three columns. Thus

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 9 & 3 \\ -2 & 7 & 3 \end{pmatrix}.$$

Example 2.1.13 *Multiply the following.*

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}$$

First check if it is possible. This is of the form $(3 \times 3)(2 \times 3)$. The inside numbers do not match and so you can't do this multiplication. This means that anything you write will be absolute nonsense because it is impossible to multiply these matrices in this order. Aren't they the same two matrices considered in the previous example? Yes they are. It is just that here they are in a different order. This shows something you must always remember about matrix multiplication.

Order Matters!

Matrix Multiplication Is Not Commutative!

This is very different than multiplication of numbers!

2.1.3 The ij^{th} Entry Of A Product

It is important to describe matrix multiplication in terms of entries of the matrices. What is the ij^{th} entry of AB? It would be the i^{th} entry of the j^{th} column of AB. Thus it would be the i^{th} entry of $A\mathbf{b}_j$. Now

$$\mathbf{b}_j = \begin{pmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

and from the above definition, the i^{th} entry is

$$\sum_{k=1}^{n} A_{ik} B_{kj}. \tag{2.11}$$

In terms of pictures of the matrix, you are doing

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{pmatrix}$$

Then as explained above, the j^{th} column is of the form

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

which is a $m \times 1$ matrix or column vector which equals

$$\begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} B_{1j} + \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} B_{2j} + \cdots + \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix} B_{nj}.$$

The second entry of this $m \times 1$ matrix is

$$A_{21}B_{ij} + A_{22}B_{2j} + \cdots + A_{2n}B_{nj} = \sum_{k=1}^{m} A_{2k} B_{kj}.$$

Similarly, the i^{th} entry of this $m \times 1$ matrix is

$$A_{i1}B_{ij} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj} = \sum_{k=1}^{m} A_{ik} B_{kj}.$$

This shows the following definition for matrix multiplication in terms of the ij^{th} entries of the product coincides with Definition 2.1.11.

This shows the following definition for matrix multiplication in terms of the ij^{th} entries of the product coincides with Definition 2.1.11.

Definition 2.1.14 *Let $A = (A_{ij})$ be an $m \times n$ matrix and let $B = (B_{ij})$ be an $n \times p$ matrix. Then AB is an $m \times p$ matrix and*

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}. \tag{2.12}$$

Example 2.1.15 *Multiply if possible* $\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \end{pmatrix}$.

First check to see if this is possible. It is of the form $(3 \times 2)(2 \times 3)$ and since the inside numbers match, the two matrices are conformable and it is possible to do the multiplication. The result should be a 3×3 matrix. The answer is of the form

$$\left(\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 \\ 7 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$$

where the commas separate the columns in the resulting product. Thus the above product equals

$$\begin{pmatrix} 16 & 15 & 5 \\ 13 & 15 & 5 \\ 46 & 42 & 14 \end{pmatrix},$$

a 3×3 matrix as desired. In terms of the ij^{th} entries and the above definition, the entry in the third row and second column of the product should equal

$$\begin{aligned} \sum_j a_{3k} b_{k2} &= a_{31} b_{12} + a_{32} b_{22} \\ &= 2 \times 3 + 6 \times 6 = 42. \end{aligned}$$

You should try a few more such examples to verify the above definition in terms of the ij^{th} entries works for other entries.

Example 2.1.16 *Multiply if possible* $\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix}$.

This is not possible because it is of the form $(3 \times 2)(3 \times 3)$ and the middle numbers don't match. In other words the two matrices are not conformable in the indicated order.

Example 2.1.17 *Multiply if possible* $\begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix}$.

This is possible because in this case it is of the form $(3 \times 3)(3 \times 2)$ and the middle numbers do match so the matrices are conformable. When the multiplication is done it equals

$$\begin{pmatrix} 13 & 13 \\ 29 & 32 \\ 0 & 0 \end{pmatrix}.$$

Check this and be sure you come up with the same answer.

Example 2.1.18 *Multiply if possible* $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 0 \end{pmatrix}$.

In this case you are trying to do $(3 \times 1)(1 \times 4)$. The inside numbers match so you can do it. Verify

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 4 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{pmatrix}$$

2.1.4 Properties Of Matrix Multiplication

As pointed out above, sometimes it is possible to multiply matrices in one order but not in the other order. What if it makes sense to multiply them in either order? Will the two products be equal then?

Example 2.1.19 *Compare* $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *and* $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

The first product is

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}.$$

The second product is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}.$$

You see these are not equal. Again you cannot conclude that $AB = BA$ for matrix multiplication even when multiplication is defined in both orders. However, there are some properties which do hold.

Proposition 2.1.20 *If all multiplications and additions make sense, the following hold for matrices,* A, B, C *and* a, b *scalars.*

$$A(aB + bC) = a(AB) + b(AC) \tag{2.13}$$

$$(B + C)A = BA + CA \tag{2.14}$$

$$A(BC) = (AB)C \tag{2.15}$$

Proof: Using Definition 2.1.14,

$$\begin{aligned} (A(aB + bC))_{ij} &= \sum_k A_{ik} (aB + bC)_{kj} \\ &= \sum_k A_{ik} (aB_{kj} + bC_{kj}) \\ &= a \sum_k A_{ik} B_{kj} + b \sum_k A_{ik} C_{kj} \\ &= a(AB)_{ij} + b(AC)_{ij} \\ &= (a(AB) + b(AC))_{ij}. \end{aligned}$$

Thus $A(B + C) = AB + AC$ as claimed. Formula 2.14 is entirely similar.

Formula 2.15 is the associative law of multiplication. Using Definition 2.1.14,

$$\begin{aligned} (A(BC))_{ij} &= \sum_k A_{ik} (BC)_{kj} \\ &= \sum_k A_{ik} \sum_l B_{kl} C_{lj} \\ &= \sum_l (AB)_{il} C_{lj} \\ &= ((AB)C)_{ij}. \end{aligned}$$

This proves 2.15.

2.1.5 The Transpose

Another important operation on matrices is that of taking the **transpose**. The following example shows what is meant by this operation, denoted by placing a T as an exponent on the matrix.

$$\begin{pmatrix} 1 & 4 \\ 3 & 1 \\ 2 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 1 & 6 \end{pmatrix}$$

What happened? The first column became the first row and the second column became the second row. Thus the 3×2 matrix became a 2×3 matrix. The number 3 was in the second row and the first column and it ended up in the first row and second column. Here is the definition.

Definition 2.1.21 *Let A be an $m \times n$ matrix. Then A^T denotes the $n \times m$ matrix which is defined as follows.*

$$\left(A^T\right)_{ij} = A_{ji}$$

Example 2.1.22

$$\begin{pmatrix} 1 & 2 & -6 \\ 3 & 5 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ -6 & 4 \end{pmatrix}.$$

The transpose of a matrix has the following important properties.

Lemma 2.1.23 *Let A be an $m \times n$ matrix and let B be a $n \times p$ matrix. Then*

$$\left(AB\right)^T = B^T A^T \tag{2.16}$$

and if α and β are scalars,

$$\left(\alpha A + \beta B\right)^T = \alpha A^T + \beta B^T \tag{2.17}$$

Proof: From the definition,

$$\begin{aligned} \left(\left(AB\right)^T\right)_{ij} &= \left(AB\right)_{ji} \\ &= \sum_k A_{jk} B_{ki} \\ &= \sum_k \left(B^T\right)_{ik} \left(A^T\right)_{kj} \\ &= \left(B^T A^T\right)_{ij} \end{aligned}$$

The proof of Formula 2.17 is left as an exercise and this proves the lemma.

Definition 2.1.24 *An $n \times n$ matrix, A is said to be **symmetric** if $A = A^T$. It is said to be **skew symmetric** if $A = -A^T$.*

Example 2.1.25 *Let*

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 5 & -3 \\ 3 & -3 & 7 \end{pmatrix}.$$

Then A is symmetric.

Example 2.1.26 *Let*

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & 0 & 2 \\ -3 & -2 & 0 \end{pmatrix}$$

Then A is skew symmetric.

2.1.6 The Identity And Inverses

There is a special matrix called I and referred to as the identity matrix. It is always a square matrix, meaning the number of rows equals the number of columns and it has the property that there are ones down the main diagonal and zeroes elsewhere. Here are some identity matrices of various sizes.

$$(1), \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The first is the 1×1 identity matrix, the second is the 2×2 identity matrix, the third is the 3×3 identity matrix, and the fourth is the 4×4 identity matrix. By extension, you can likely see what the $n \times n$ identity matrix would be. It is so important that there is a special symbol to denote the ij^{th} entry of the identity matrix

$$I_{ij} = \delta_{ij}$$

where δ_{ij} is the **Kroneker symbol** defined by

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

It is called the **identity matrix** because it is a **multiplicative identity** in the following sense.

Lemma 2.1.27 *Suppose A is an $m \times n$ matrix and I_n is the $n \times n$ identity matrix. Then $AI_n = A$. If I_m is the $m \times m$ identity matrix, it also follows that $I_m A = A$.*

Proof:

$$\begin{aligned} (AI_n)_{ij} &= \sum_k A_{ik}\delta_{kj} \\ &= A_{ij} \end{aligned}$$

and so $AI_n = A$. The other case is left as an exercise for you.

Definition 2.1.28 *An $n \times n$ matrix, A has an **inverse**, A^{-1} if and only if $AA^{-1} = A^{-1}A = I$. Such a matrix is called **invertible**.*

It is very important to observe that the inverse of a matrix, if it exists, is unique. Another way to think of this is that if it acts like the inverse, then it is the inverse.

Theorem 2.1.29 *Suppose A^{-1} exists and $AB = BA = I$. Then $B = A^{-1}$.*

Proof:

$$A^{-1} = A^{-1}I = A^{-1}(AB) = \left(A^{-1}A\right)B = IB = B.$$

Unlike ordinary multiplication of numbers, it can happen that $A \neq 0$ but A may fail to have an inverse. This is illustrated in the following example.

Example 2.1.30 *Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Does A have an inverse?*

One might think A would have an inverse because it does not equal zero. However,

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and if A^{-1} existed, this could not happen because you could write

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = A^{-1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = A^{-1}\left(A\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right) =$$

$$= (A^{-1}A)\begin{pmatrix} -1 \\ 1 \end{pmatrix} = I\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

a contradiction. Thus the answer is that A does not have an inverse.

Example 2.1.31 *Let* $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. *Show* $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ *is the inverse of* A.

To check this, multiply

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

showing that this matrix is indeed the inverse of A.

There are various ways of finding the inverse of a matrix. One way will be presented in the discussion on determinants. You can also find them directly from the definition provided they exist.

In the last example, how would you find A^{-1}? You wish to find a matrix, $\begin{pmatrix} x & z \\ y & w \end{pmatrix}$ such that

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x & z \\ y & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This requires the solution of the systems of equations,

$$x + y = 1, x + 2y = 0$$

and

$$z + w = 0, z + 2w = 1.$$

The first pair of equations has the solution $y = -1$ and $x = 2$. The second pair of equations has the solution $w = 1$, $z = -1$. Therefore, from the definition of the inverse,

$$A^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

To be sure it is the inverse, you should multiply on both sides of the original matrix. It turns out that if it works on one side, it will always work on the other. The consideration of this and as well as a more detailed treatment of inverses is a good topic for a linear algebra course.

2.2 Linear Transformations

An $m \times n$ matrix can be used to transform vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$ through the use of matrix multiplication.

Example 2.2.1 *Consider the matrix,* $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix}$. *Think of it as a function which takes vectors in* $\mathbb{R}^3$ *and makes them in to vectors in* $\mathbb{R}^2$ *as follows. For* $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ *a vector in* $\mathbb{R}^3$, *multiply on the left by the given matrix to obtain the vector in* $\mathbb{R}^2$. *Here are some numerical examples.*

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 5 \\ -3 \end{pmatrix} = \begin{pmatrix} 20 \\ 25 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 7 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 7 \end{pmatrix},$$

More generally,

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + 2y \\ 2x + y \end{pmatrix}$$

The idea is to define a function which takes vectors in $\mathbb{R}^3$ *and delivers new vectors in* $\mathbb{R}^2$.

This is an example of something called a linear transformation.

Definition 2.2.2 *Let* $T : \mathbb{R}^n \to \mathbb{R}^m$ *be a function. Thus for each* $\mathbf{x} \in \mathbb{R}^n, T\mathbf{x} \in \mathbb{R}^m$. *Then* T *is a* ***linear transformation*** *if whenever* α, β *are scalars and* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *are vectors in* $\mathbb{R}^n$,

$$T(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) = \alpha_1 T\mathbf{x}_1 + \beta T\mathbf{x}_2.$$

In words, linear transformations distribute across + and allow you to factor out scalars. At this point, recall the properties of matrix multiplication. The pertinent property is 2.14 on Page 37. Recall it states that for a and b scalars,

$$A(aB + bC) = aAB + bAC$$

In particular, for A an $m \times n$ matrix and B and $C, n \times 1$ matrices (column vectors) the above formula holds which is nothing more than the statement that matrix multiplication gives an example of a linear transformation.

Definition 2.2.3 *A linear transformation is called* ***one to one*** *(often written as* $1-1$*) if it never takes two different vectors to the same vector. Thus* T *is one to one if whenever* $\mathbf{x} \neq \mathbf{y}$

$$T\mathbf{x} \neq T\mathbf{y}.$$

Equivalently, if $T(\mathbf{x}) = T(\mathbf{y})$, *then* $\mathbf{x} = \mathbf{y}$.

In the case that a linear transformation comes from matrix multiplication, it is common usage to refer to the matrix as a one to one matrix when the linear transformation it determines is one to one.

Definition 2.2.4 *A linear transformation mapping $\mathbb{R}^n$ to $\mathbb{R}^m$ is called* ***onto*** *if whenever $\mathbf{y} \in \mathbb{R}^m$ there exists $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{y}$.*

Thus T is onto if everything in $\mathbb{R}^m$ gets hit. In the case that a linear transformation comes from matrix multiplication, it is common to refer to the matrix as onto when the linear transformation it determines is onto. Also it is common usage to write $T\mathbb{R}^n$, $T(\mathbb{R}^n)$,or $\mathrm{Im}(T)$ as the set of vectors of $\mathbb{R}^m$ which are of the form $T\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. In the case that T is obtained from multiplication by an $m \times n$ matrix, A, it is standard to simply write $A(\mathbb{R}^n)$, $A\mathbb{R}^n$, or $\mathrm{Im}(A)$ to denote those vectors in $\mathbb{R}^m$ which are obtained in the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$.

2.3 Constructing The Matrix Of A Linear Transformation

It turns out that if T is any linear transformation which maps $\mathbb{R}^n$ to $\mathbb{R}^m$, there is always an $m \times n$ matrix, A with the property that

$$A\mathbf{x} = T\mathbf{x} \tag{2.18}$$

for all $\mathbf{x} \in \mathbb{R}^n$. Here is why. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and you want to find the matrix defined by this linear transformation as described in 2.18. Then if $\mathbf{x} \in \mathbb{R}^n$ it follows

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

where as implied above, $\mathbf{e}_i$ is the vector which has zeros in every slot but the i^{th} and a 1 in this slot. Then since T is linear,

$$\begin{aligned} T\mathbf{x} &= \sum_{i=1}^{n} x_i T(\mathbf{e}_i) \\ &= \begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &\equiv A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \end{aligned}$$

and so you see that the matrix desired is obtained from letting the i^{th} column equal $T(\mathbf{e}_i)$. This yields the following theorem.

Theorem 2.3.1 *Let T be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then the matrix, A satisfying 2.18 is given by*

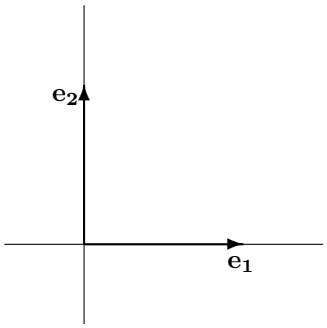
$$\begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix}$$

where $T\mathbf{e}_i$ is the i^{th} column of A.

Sometimes you need to find a matrix which represents a given linear transformation which is described in geometrical terms. The idea is to produce a matrix which you can multiply a vector by to get the same thing as some geometrical description. A good example of this is the problem of rotation of vectors.

Example 2.3.2 *Determine the matrix which represents the linear transformation defined by rotating every vector through an angle of θ.*

Let $\mathbf{e}_1 \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_2 \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. These identify the geometric vectors which point along the positive x axis and positive y axis as shown.



From the above, you only need to find $T\mathbf{e}_1$ and $T\mathbf{e}_2$, the first being the first column of the desired matrix, A and the second being the second column. From drawing a picture and doing a little geometry, you see that

$$T\mathbf{e}_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, T\mathbf{e}_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Therefore, from Theorem 2.3.1,

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Example 2.3.3 *Find the matrix of the linear transformation which is obtained by first rotating all vectors through an angle of ϕ and then through an angle θ. Thus you want the linear transformation which rotates all angles through an angle of $\theta + \phi$.*

Let $T_{\theta+\phi}$ denote the linear transformation which rotates every vector through an angle of $\theta + \phi$. Then to get $T_{\theta+\phi}$, you could first do T_ϕ and then do T_θ where T_ϕ is the linear transformation which rotates through an angle of ϕ and T_θ is the linear transformation which rotates through an angle of θ. Denoting the corresponding matrices by $A_{\theta+\phi}$, A_ϕ, and A_θ, you must have for every $\mathbf{x}$

$$A_{\theta+\phi}\mathbf{x} = T_{\theta+\phi}\mathbf{x} = T_\theta T_\phi \mathbf{x} = A_\theta A_\phi \mathbf{x}.$$

Consequently, you must have

$$\begin{aligned} A_{\theta+\phi} &= \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} = A_\theta A_\phi \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}. \end{aligned}$$

You know how to multiply matrices. Do so to the pair on the right. This yields

$$\begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} = \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \sin\theta\cos\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{pmatrix}.$$

Don't these look familiar? They are the usual trig. identities for the sum of two angles derived here using linear algebra concepts.

You do not have to stop with two dimensions. You can consider rotations and other geometric concepts in any number of dimensions. This is one of the major advantages of linear algebra. You can break down a difficult geometrical procedure into small steps, each corresponding to multiplication by an appropriate matrix. Then by multiplying the matrices, you can obtain a single matrix which can give you numerical information on the results of applying the given sequence of simple procedures. That which you could never visualize can still be understood to the extent of finding exact numerical answers. The following is a more routine example quite typical of what will be important in the calculus of several variables.

Example 2.3.4 *Let* $T(x_1,x_2) = \begin{pmatrix} x_1+3x_2 \\ x_1-x_2 \\ x_1 \\ 3x_2+5x_1 \end{pmatrix}$. *Thus* $T:\mathbb{R}^2 \to \mathbb{R}^4$. *Explain why* T *is a linear transformation and write* $T(x_1,x_2)$ *in the form* $A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *where* A *is an appropriate matrix.*

From the definition of matrix multiplication,

$$T(x_1,x_2) = \begin{pmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 0 \\ 5 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Since $T\mathbf{x}$ is of the form $A\mathbf{x}$ for A a matrix, it follows T is a linear transformation. You could also verify directly that $T(\alpha\mathbf{x}+\beta\mathbf{y}) = \alpha T(\mathbf{x}) + \beta T(\mathbf{y})$.

2.4 Exercises

1. Here are some matrices:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 7 \end{pmatrix}, B = \begin{pmatrix} 3 & -1 & 2 \\ -3 & 2 & 1 \end{pmatrix}, \\ C &= \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}, D = \begin{pmatrix} -1 & 2 \\ 2 & -3 \end{pmatrix}, E = \begin{pmatrix} 2 \\ 3 \end{pmatrix}. \end{aligned}$$

   Find if possible $-3A, 3B-A, AC, CB, AE, EA$. If it is not possible explain why.

2. Here are some matrices:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix}, B = \begin{pmatrix} 2 & -5 & 2 \\ -3 & 2 & 1 \end{pmatrix}, \\ C &= \begin{pmatrix} 1 & 2 \\ 5 & 0 \end{pmatrix}, D = \begin{pmatrix} -1 & 1 \\ 4 & -3 \end{pmatrix}, E = \begin{pmatrix} 1 \\ 3 \end{pmatrix}. \end{aligned}$$

   Find if possible $-3A, 3B-A, AC, CA, AE, EA, BE, DE$. If it is not possible explain why.

3. Here are some matrices:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix}, B = \begin{pmatrix} 2 & -5 & 2 \\ -3 & 2 & 1 \end{pmatrix}, \\ C &= \begin{pmatrix} 1 & 2 \\ 5 & 0 \end{pmatrix}, D = \begin{pmatrix} -1 & 1 \\ 4 & -3 \end{pmatrix}, E = \begin{pmatrix} 1 \\ 3 \end{pmatrix}. \end{aligned}$$

Find if possible $-3A^T, 3B - A^T, AC, CA, AE, E^TB, BE, DE, EE^T, E^TE$. If it is not possible explain why.

4. Here are some matrices:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix}, B = \begin{pmatrix} 2 & -5 & 2 \\ -3 & 2 & 1 \end{pmatrix}, \\ C &= \begin{pmatrix} 1 & 2 \\ 5 & 0 \end{pmatrix}, D = \begin{pmatrix} -1 \\ 4 \end{pmatrix}, E = \begin{pmatrix} 1 \\ 3 \end{pmatrix}. \end{aligned}$$

Find the following if possible and explain why it is not possible if this is the case. $AD, DA, D^TB, D^TBE, E^TD, DE^T$.

5. Let $A = \begin{pmatrix} 1 & 1 \\ -2 & -1 \\ 1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & -1 & -2 \\ 2 & 1 & -2 \end{pmatrix}$, and $C = \begin{pmatrix} 1 & 1 & -3 \\ -1 & 2 & 0 \\ -3 & -1 & 0 \end{pmatrix}$. Find if possible.

   (a) AB
   (b) BA
   (c) AC
   (d) CA
   (e) CB
   (f) BC

6. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, B = \begin{pmatrix} 1 & 2 \\ 3 & k \end{pmatrix}$. Is it possible to choose k such that $AB = BA$? If so, what should k equal?

7. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, B = \begin{pmatrix} 1 & 2 \\ 1 & k \end{pmatrix}$. Is it possible to choose k such that $AB = BA$? If so, what should k equal?

8. Let $\mathbf{x} = (-1, -1, 1)$ and $\mathbf{y} = (0, 1, 2)$. Find $\mathbf{x}^T\mathbf{y}$ and $\mathbf{x}\mathbf{y}^T$ if possible.

9. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/4$.

10. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $-\pi/3$.

11. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $2\pi/3$.

12. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/12$. **Hint:** Note that $\pi/12 = \pi/3 - \pi/4$.

13. Let $T(x_1,x_2)=\begin{pmatrix} x_1+4x_2 \\ x_2+2x_1 \end{pmatrix}$. Thus $T:\mathbb{R}^2\to\mathbb{R}^2$. Explain why T is a linear transformation and write $T(x_1,x_2)$ in the form $A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ where A is an appropriate matrix.

14. Let $T(x_1,x_2)=\begin{pmatrix} x_1-x_2 \\ x_1 \\ 3x_2+x_1 \\ 3x_2+5x_1 \end{pmatrix}$. Thus $T:\mathbb{R}^2\to\mathbb{R}^4$. Explain why T is a linear transformation and write $T(x_1,x_2)$ in the form $A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ where A is an appropriate matrix.

15. Let $T(x_1,x_2,x_3,x_4)=\begin{pmatrix} x_1-x_2+2x_3 \\ 2x_3+x_1 \\ 3x_3 \\ 3x_4+3x_2+x_1 \end{pmatrix}$. Thus $T:\mathbb{R}^4\to\mathbb{R}^4$. Explain why T is a linear transformation and write $T(x_1,x_2,x_3,x_4)$ in the form $A\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ where A is an appropriate matrix.

16. Let $T(x_1,x_2)=\begin{pmatrix} x_1^2+4x_2 \\ x_2+2x_1 \end{pmatrix}$. Thus $T:\mathbb{R}^2\to\mathbb{R}^2$. Explain why T cannot possibly be a linear transformation.

17. Suppose A and B are square matrices of the same size. Which of the following are correct?

    (a) $(A-B)^2=A^2-2AB+B^2$
    (b) $(AB)^2=A^2B^2$
    (c) $(A+B)^2=A^2+2AB+B^2$
    (d) $(A+B)^2=A^2+AB+BA+B^2$
    (e) $A^2B^2=A(AB)B$
    (f) $(A+B)^3=A^3+3A^2B+3AB^2+B^3$
    (g) $(A+B)(A-B)=A^2-B^2$

18. Let $A=\begin{pmatrix} -1 & -1 \\ 3 & 3 \end{pmatrix}$. Find $\boxed{\underline{\textbf{all}}}$ 2×2 matrices, B such that $AB=0$.

19. In 2.1 - 2.8 describe $-A$ and 0.

20. Let A be an $n\times n$ matrix. Show A equals the sum of a symmetric and a skew symmetric matrix. **Hint:** Consider the matrix $\frac{1}{2}\left(A+A^T\right)$. Is this matrix symmetric?

21. If A is a skew symmetric matrix, what can be concluded about A^n where $n=1,2,3,\cdots$?

22. Show every skew symmetric matrix has all zeros down the main diagonal. The main diagonal consists of every entry of the matrix which is of the form a_{ii}. It runs from the upper left down to the lower right.

23. Using only the properties 2.1 - 2.8 show $-A$ is unique.

24. Using only the properties 2.1 - 2.8 show 0 is unique.

25. Using only the properties 2.1 - 2.8 show $0A = 0$. Here the 0 on the left is the scalar 0 and the 0 on the right is the zero for $m \times n$ matrices.

26. Using only the properties 2.1 - 2.8 and previous problems show $(-1)\, A = -A$.

27. Prove 2.17.

28. Prove that $I_m A = A$ where A is an $m \times n$ matrix.

29. Let
$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.

30. Let
$$A = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}.$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.

31. Let
$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.

32. Give an example of matrices, A, B, C such that $B \neq C$, $A \neq 0$, and yet $AB = AC$.

33. Suppose $AB = AC$ and A is an invertible $n \times n$ matrix. Does it follow that $B = C$? Explain why or why not. What if A were a non invertible $n \times n$ matrix?

34. Find your own examples:

    (a) ♠2×2 matrices, A and B such that $A \neq 0, B \neq 0$ with $AB \neq BA$.

    (b) ♠2×2 matrices, A and B such that $A \neq 0, B \neq 0$, but $AB = 0$.

    (c) 2×2 matrices, A, D, and C such that $A \neq 0, C \neq D$, but $AC = AD$.

35. Explain why if $AB = AC$ and A^{-1} exists, then $B = C$.

36. Give an example of a matrix, A such that $A^2 = I$ and yet $A \neq I$ and $A \neq -I$.

37. Give an example of matrices, A, B such that neither A nor B equals zero and yet $AB = 0$.

38. Give another example other than the one given in this section of two square matrices, A and B such that $AB \neq BA$.

39. Show that if A^{-1} exists for an $n \times n$ matrix, then it is unique. That is, if $BA = I$ and $AB = I$, then $B = A^{-1}$.

40. Show $(AB)^{-1} = B^{-1} A^{-1}$.

41. Show that if A is an invertible $n \times n$ matrix, then so is A^T and $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

42. Show that if A is an $n \times n$ invertible matrix and $\mathbf{x}$ is a $n \times 1$ matrix such that $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b}$ an $n \times 1$ matrix, then $\mathbf{x} = A^{-1}\mathbf{b}$.

43. Prove that if A^{-1} exists and $A\mathbf{x} = \mathbf{0}$ then $\mathbf{x} = \mathbf{0}$.

44. Show that $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ by verifying that

$$(ABC)\left(C^{-1}B^{-1}A^{-1}\right) = \left(C^{-1}B^{-1}A^{-1}\right)(ABC) = I.$$

2.5 Exercises With Answers

1. Here are some matrices:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 7 \end{pmatrix}, B = \begin{pmatrix} 0 & -1 & 2 \\ -3 & 2 & 1 \end{pmatrix}, \\ C &= \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}, D = \begin{pmatrix} 0 & 2 \\ 2 & -3 \end{pmatrix}, E = \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \end{aligned}$$

Find if possible $-3A, 3B - A, AC, CB, AE, EA$. If it is not possible explain why.

$$-3A = (-3)\begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 7 \end{pmatrix} = \begin{pmatrix} -3 & -6 & -3 \\ -6 & 0 & -21 \end{pmatrix}.$$

$$3B - A = 3\begin{pmatrix} 0 & -1 & 2 \\ -3 & 2 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 7 \end{pmatrix} = \begin{pmatrix} -1 & -5 & 5 \\ -11 & 6 & -4 \end{pmatrix}$$

AC makes no sense because A is a 2×3 and C is a 2×2. You can't do $(2 \times 3)(2 \times 2)$ because the inside numbers don't match.

$$CB = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 & 2 \\ -3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} -6 & 3 & 4 \\ -3 & -1 & 7 \end{pmatrix}.$$

You can't multiply AE because it is of the form $(2 \times 3)(2 \times 1)$ and the inside numbers don't match. EA also cannot be multiplied because it is of the form $(2 \times 1)(2 \times 3)$.

2. Here are some matrices:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix}, B = \begin{pmatrix} 0 & -5 & 2 \\ -3 & 1 & 1 \end{pmatrix}, \\ C &= \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}, D = \begin{pmatrix} -1 & 1 \\ 4 & -2 \end{pmatrix}, E = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{aligned}$$

Find if possible $-3A^T, 3B - A^T, AC, CA, AE, E^TB, EE^T, E^TE$. If it is not possible explain why.

$$-3A^T = -3\begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix}^T = \begin{pmatrix} -3 & -9 & -3 \\ -6 & -6 & 3 \end{pmatrix}$$

$$3B - A^T = 3\begin{pmatrix} 0 & -5 & 2 \\ -3 & 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix}^T = \begin{pmatrix} -1 & -18 & 5 \\ -11 & 1 & 4 \end{pmatrix}$$

$$AC = \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 7 & 4 \\ 9 & 8 \\ -2 & 1 \end{pmatrix}$$

$CA = (2\times 2)(3\times 2)$ so this makes no sense.

$$AE = \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 0 \end{pmatrix}$$

$$E^TB = \begin{pmatrix} 1 \\ 1 \end{pmatrix}^T \begin{pmatrix} 0 & -5 & 2 \\ -3 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -3 & -4 & 3 \end{pmatrix}$$

Note in this case you have a $(1\times 2)(2\times 3) = 1\times 3$.

$$E^TE = \begin{pmatrix} 1 \\ 1 \end{pmatrix}^T \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2$$

Note in this case you have $(1\times 2)(2\times 1) = 1\times 1$

$$EE^T = \begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}^T = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

In this case you have $(2\times 1)(1\times 2) = 2\times 2$.

3. Let $A = \begin{pmatrix} 4 & 1 \\ -2 & 0 \\ 1 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 & -2 \\ 2 & 1 & 2 \end{pmatrix}$, and $C = \begin{pmatrix} 1 & 1 & -3 \\ 0 & 2 & 0 \\ -3 & -1 & 0 \end{pmatrix}$. Find if possible.

   (a) $AB = \begin{pmatrix} 4 & 1 \\ -2 & 0 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 & -2 \\ 2 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 6 & 1 & -6 \\ -2 & 0 & 4 \\ 5 & 2 & 2 \end{pmatrix}$

   (b) $BA = \begin{pmatrix} 1 & 0 & -2 \\ 2 & 1 & 2 \end{pmatrix}\begin{pmatrix} 4 & 1 \\ -2 & 0 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ 8 & 6 \end{pmatrix}$

   (c) $AC = \begin{pmatrix} 4 & 1 \\ -2 & 0 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 & -3 \\ 0 & 2 & 0 \\ -3 & -1 & 0 \end{pmatrix} = (3\times 2)(3\times 3) = \text{nonsense}$

   (d) $CA = \begin{pmatrix} 1 & 1 & -3 \\ 0 & 2 & 0 \\ -3 & -1 & 0 \end{pmatrix}\begin{pmatrix} 4 & 1 \\ -2 & 0 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} -1 & -5 \\ -4 & 0 \\ -10 & -3 \end{pmatrix}$

   (e) $CB = \begin{pmatrix} 1 & 1 & -3 \\ 0 & 2 & 0 \\ -3 & -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & -2 \\ 2 & 1 & 2 \end{pmatrix} = (3\times 3)(2\times 3) = \text{nonsense}$

4. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, B = \begin{pmatrix} 1 & 2 \\ 0 & k \end{pmatrix}$. Is it possible to choose k such that $AB = BA$? If so, what should k equal?

   $AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & k \end{pmatrix} = \begin{pmatrix} 1 & 2+2k \\ 3 & 6+4k \end{pmatrix}$ while $BA = \begin{pmatrix} 1 & 2 \\ 0 & k \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 7 & 10 \\ 3k & 4k \end{pmatrix}$ If $AB = BA$, then from what was just shown, you would need to have $1 = 7$ and this is not true. Therefore, there is no way to choose k such that these two matrices commute.

5. Write $\begin{pmatrix} x_1 - x_2 + 2x_3 \\ 2x_3 - x_1 \\ 3x_3 + x_1 + x_4 \\ 3x_4 + 3x_2 + x_1 \end{pmatrix}$ in the form $A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ where A is an appropriate matrix.

$$\begin{pmatrix} 1 & -1 & 2 & 0 \\ -1 & 0 & 2 & 0 \\ 1 & 0 & 3 & 1 \\ 1 & 3 & 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

6. Suppose A and B are square matrices of the same size. Which of the following are correct?

   (a) $(A-B)^2 = A^2 - 2AB + B^2$ Is matrix multiplication commutative?
   (b) $(AB)^2 = A^2B^2$ Is matrix multiplication commutative?
   (c) $(A+B)^2 = A^2 + 2AB + B^2$ Is matrix multiplication commutative?
   (d) $(A+B)^2 = A^2 + AB + BA + B^2$
   (e) $A^2B^2 = A(AB)B$
   (f) $(A+B)^3 = A^3 + 3A^2B + 3AB^2 + B^3$ Is matrix multiplication commutative?
   (g) $(A+B)(A-B) = A^2 - B^2$ Is matrix multiplication commutative?

7. Let $A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$. Find $\boxed{\textbf{all}}$ 2×2 matrices, B such that $AB = 0$.

   You need a matrix, $\begin{pmatrix} x & y \\ z & w \end{pmatrix}$ such that

$$\begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} x+z & y+w \\ 2x+2z & 2y+2w \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

   Thus you need $x = -z$ and $y = -w$. It appears you can pick z and w at random and any matrix of the form $\begin{pmatrix} -z & -w \\ z & w \end{pmatrix}$ will work.

8. Let

$$A = \begin{pmatrix} 3 & 2 & 3 \\ 2 & 1 & 2 \\ 1 & 0 & 2 \end{pmatrix}.$$

   Find A^{-1} if possible. If A^{-1} does not exist, determine why.

$$\begin{pmatrix} 3 & 2 & 3 \\ 2 & 1 & 2 \\ 1 & 0 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & 4 & -1 \\ 2 & -3 & 0 \\ 1 & -2 & 1 \end{pmatrix}$$

9. Let

$$A = \begin{pmatrix} 0 & 0 & 3 \\ 2 & 4 & 4 \\ 1 & 0 & 1 \end{pmatrix}.$$

   Find A^{-1} if possible. If A^{-1} does not exist, determine why.

$$\begin{pmatrix} 0 & 0 & 3 \\ 2 & 4 & 4 \\ 1 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} -\frac{1}{3} & 0 & 1 \\ -\frac{1}{6} & \frac{1}{4} & -\frac{1}{2} \\ \frac{1}{3} & 0 & 0 \end{pmatrix}$$

10. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 3 & 7 \end{pmatrix}.$$

Find A^{-1} if possible. If A^{-1} does not exist, determine why. In this case there is no inverse.

$$\begin{pmatrix} 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 3 & 7 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}.$$

If A^{-1} existed then you could multiply on the right side in the above equations and find

$$\begin{pmatrix} 1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$$

which is not true.

Determinants

3.0.1 Outcomes

1. Evaluate the determinant of a square matrix by applying
   (a) the cofactor formula or
   (b) row operations.
2. Recall the general properties of determinants.
3. Recall that the determinant of a product of matrices is the product of the determinants. Recall that the determinant of a matrix is equal to the determinant of its transpose.
4. Apply Cramer's Rule to solve a 2×2 or a 3×3 linear system.
5. Use determinants to determine whether a matrix has an inverse.
6. Evaluate the inverse of a matrix using cofactors.

3.1 Basic Techniques And Properties

3.1.1 Cofactors And 2×2 Determinants

Let A be an $n \times n$ matrix. The **determinant** of A, denoted as $\det(A)$ is a number. If the matrix is a 2×2 matrix, this number is very easy to find.

Definition 3.1.1 *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *Then*

$$\det(A) \equiv ad - cb.$$

The determinant is also often denoted by enclosing the matrix with two vertical lines. Thus

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Example 3.1.2 *Find* $\det \begin{pmatrix} 2 & 4 \\ -1 & 6 \end{pmatrix}$.

From the definition this is just $(2)(6) - (-1)(4) = 16$.

Having defined what is meant by the determinant of a 2×2 matrix, what about a 3×3 matrix?

Definition 3.1.3 *Suppose A is a 3×3 matrix. The ij^{th}* ***minor,*** *denoted as $\text{minor}(A)_{ij}$, is the determinant of the 2×2 matrix which results from deleting the i^{th} row and the j^{th} column.*

Example 3.1.4 *Consider the matrix,*

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

The $(1,2)$ minor is the determinant of the 2×2 matrix which results when you delete the first row and the second column. This minor is therefore

$$\det\begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = -2.$$

The $(2,3)$ minor is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

$$\det\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = -4.$$

Definition 3.1.5 *Suppose A is a 3×3 matrix. The ij^{th}* ***cofactor*** *is defined to be $(-1)^{i+j}\times\left(ij^{th}\text{ minor}\right)$. In words, you multiply $(-1)^{i+j}$ times the ij^{th} minor to get the ij^{th} cofactor. The cofactors of a matrix are so important that special notation is appropriate when referring to them. The ij^{th} cofactor of a matrix, A will be denoted by $\text{cof}(A)_{ij}$. It is also convenient to refer to the cofactor of an entry of a matrix as follows. For a_{ij} an entry of the matrix, its cofactor is just $\text{cof}(A)_{ij}$. Thus the cofactor of the ij^{th} entry is just the ij^{th} cofactor.*

Example 3.1.6 *Consider the matrix,*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

The $(1,2)$ minor is the determinant of the 2×2 matrix which results when you delete the first row and the second column. This minor is therefore

$$\det\begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = -2.$$

It follows

$$\text{cof}(A)_{12} = (-1)^{1+2}\det\begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = (-1)^{1+2}(-2) = 2$$

The $(2,3)$ minor is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

$$\det\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = -4.$$

Therefore,

$$\text{cof}(A)_{23} = (-1)^{2+3}\det\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = (-1)^{2+3}(-4) = 4.$$

Similarly,

$$\text{cof}(A)_{22} = (-1)^{2+2}\det\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} = -8.$$

Definition 3.1.7 *The determinant of a 3×3 matrix, A, is obtained by picking a row (column) and taking the product of each entry in that row (column) with its cofactor and adding these up. This process when applied to the i^{th} row (column) is known as expanding the determinant along the i^{th} row (column).*

Example 3.1.8 *Find the determinant of*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Here is how it is done by "**expanding along the first column**".

$$\overbrace{1(-1)^{1+1}\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{11}} + \overbrace{4(-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{21}} + \overbrace{3(-1)^{3+1}\begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix}}^{\mathrm{cof}(A)_{31}} = 0.$$

You see, I just followed the rule in the above definition. I took the 1 in the first column and multiplied it by its cofactor, the 4 in the first column and multiplied it by its cofactor, and the 3 in the first column and multiplied it by its cofactor. Then I added these numbers together.

You could also expand the determinant along the second row as follows.

$$\overbrace{4(-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{21}} + \overbrace{3(-1)^{2+2}\begin{vmatrix} 1 & 3 \\ 3 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{22}} + \overbrace{2(-1)^{2+3}\begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix}}^{\mathrm{cof}(A)_{23}} = 0.$$

Observe this gives the same number. You should try expanding along other rows and columns. If you don't make any mistakes, you will always get the same answer.

What about a 4×4 matrix? You know now how to find the determinant of a 3×3 matrix. The pattern is the same.

Definition 3.1.9 *Suppose A is a 4×4 matrix. The ij^{th} **minor** is the determinant of the 3×3 matrix you obtain when you delete the i^{th} row and the j^{th} column. The ij^{th} **cofactor,** $\mathrm{cof}(A)_{ij}$ is defined to be $(-1)^{i+j} \times \left(ij^{th} \text{ minor}\right)$. In words, you multiply $(-1)^{i+j}$ times the ij^{th} minor to get the ij^{th} cofactor.*

Definition 3.1.10 *The determinant of a 4×4 matrix, A, is obtained by picking a row (column) and taking the product of each entry in that row (column) with its cofactor and adding these up. This process when applied to the i^{th} row (column) is known as expanding the determinant along the i^{th} row (column).*

Example 3.1.11 *Find* $\det(A)$ *where*

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 4 & 2 & 3 \\ 1 & 3 & 4 & 5 \\ 3 & 4 & 3 & 2 \end{pmatrix}$$

As in the case of a 3×3 matrix, you can expand this along any row or column. Lets pick the third column. $\det(A) =$

$$3(-1)^{1+3}\begin{vmatrix} 5 & 4 & 3 \\ 1 & 3 & 5 \\ 3 & 4 & 2 \end{vmatrix} + 2(-1)^{2+3}\begin{vmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \\ 3 & 4 & 2 \end{vmatrix} +$$

$$4\left(-1\right)^{3+3}\begin{vmatrix}1&2&4\\5&4&3\\3&4&2\end{vmatrix}+3\left(-1\right)^{4+3}\begin{vmatrix}1&2&4\\5&4&3\\1&3&5\end{vmatrix}.$$

Now you know how to expand each of these 3×3 matrices along a row or a column. If you do so, you will get -12 assuming you make no mistakes. You could expand this matrix along any row or any column and assuming you make no mistakes, you will always get the same thing which is defined to be the determinant of the matrix, A. This method of evaluating a determinant by expanding along a row or a column is called the **method of Laplace expansion**.

Note that each of the four terms above involves three terms consisting of determinants of 2×2 matrices and each of these will need 2 terms. Therefore, there will be $4\times 3\times 2=24$ terms to evaluate in order to find the determinant using the method of Laplace expansion. Suppose now you have a 10×10 matrix and you follow the above pattern for evaluating determinants. By analogy to the above, there will be $10!=3,628,800$ terms involved in the evaluation of such a determinant by Laplace expansion along a row or column. This is a lot of terms.

In addition to the difficulties just discussed, you should regard the above claim that you always get the same answer by picking any row or column with considerable skepticism. It is incredible and not at all obvious. However, it requires a little effort to establish it. This is done in the section on the theory of the determinant The above examples motivate the following definitions, the second of which is incredible.

Definition 3.1.12 *Let $A=(a_{ij})$ be an $n\times n$ matrix and suppose the determinant of a $(n-1)\times(n-1)$ matrix has been defined. Then a new matrix called the* ***cofactor matrix****, $\operatorname{cof}(A)$ is defined by $\operatorname{cof}(A)=(c_{ij})$ where to obtain c_{ij} delete the i^{th} row and the j^{th} column of A, take the determinant of the $(n-1)\times(n-1)$ matrix which results, (This is called the ij^{th}* ***minor*** *of A.) and then multiply this number by $(-1)^{i+j}$. Thus $(-1)^{i+j}\times\left(\text{the } ij^{th} \text{ minor}\right)$ equals the ij^{th} cofactor. To make the formulas easier to remember, $\operatorname{cof}(A)_{ij}$ will denote the ij^{th} entry of the cofactor matrix.*

With this definition of the cofactor matrix, here is how to define the determinant of an $n\times n$ matrix.

Definition 3.1.13 *Let A be an $n\times n$ matrix where $n\geq 2$ and suppose the determinant of an $(n-1)\times(n-1)$ has been defined. Then*

$$\det(A)=\sum_{j=1}^{n}a_{ij}\operatorname{cof}(A)_{ij}=\sum_{i=1}^{n}a_{ij}\operatorname{cof}(A)_{ij}. \tag{3.1}$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column. This is called the method of Laplace expansion.

Theorem 3.1.14 *Expanding the $n\times n$ matrix along any row or column always gives the same answer so the above definition is a good definition.*

3.1.2 The Determinant Of A Triangular Matrix

Notwithstanding the difficulties involved in using the method of Laplace expansion, certain types of matrices are very easy to deal with.

Definition 3.1.15 *A matrix M, is upper triangular if $M_{ij} = 0$ whenever $i > j$. Thus such a matrix equals zero below the main diagonal, the entries of the form M_{ii}, as shown.*

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

You should verify the following using the above theorem on Laplace expansion.

Corollary 3.1.16 *Let M be an upper (lower) triangular matrix. Then* $\det(M)$ *is obtained by taking the product of the entries on the main diagonal.*

Example 3.1.17 *Let*

$$A = \begin{pmatrix} 1 & 2 & 3 & 77 \\ 0 & 2 & 6 & 7 \\ 0 & 0 & 3 & 33.7 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Find $\det(A)$.

From the above corollary, it suffices to take the product of the diagonal elements. Thus $\det(A) = 1 \times 2 \times 3 \times (-1) = -6$. Without using the corollary, you could expand along the first column. This gives

$$1\begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} + 0\,(-1)^{2+1}\begin{vmatrix} 2 & 3 & 77 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} +$$
$$0\,(-1)^{3+1}\begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 0 & -1 \end{vmatrix} + 0\,(-1)^{4+1}\begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 3 & 33.7 \end{vmatrix}$$

and the only nonzero term in the expansion is

$$1\begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix}.$$

Now expand this along the first column to obtain

$$1 \times \left(2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix} + 0\,(-1)^{2+1}\begin{vmatrix} 6 & 7 \\ 0 & -1 \end{vmatrix} + 0\,(-1)^{3+1}\begin{vmatrix} 6 & 7 \\ 3 & 33.7 \end{vmatrix}\right)$$
$$= 1 \times 2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix}$$

Next expand this last determinant along the first column to obtain the above equals

$$1 \times 2 \times 3 \times (-1) = -6$$

which is just the product of the entries down the main diagonal of the original matrix.

3.1.3 Properties Of Determinants

There are many properties satisfied by determinants. Some of these properties have to do with row operations which are described below.

Definition 3.1.18 *The* ***row operations*** *consist of the following*

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by a multiple of another row added to itself.*

Theorem 3.1.19 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from multiplying some row of A by a scalar, c. Then $c \det(A) = \det(A_1)$.*

Example 3.1.20 *Let* $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, A_1 = \begin{pmatrix} 2 & 4 \\ 3 & 4 \end{pmatrix}$. $\det(A) = -2$, $\det(A_1) = -4$.

Theorem 3.1.21 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from switching two rows of A. Then $\det(A) = -\det(A_1)$. Also, if one row of A is a multiple of another row of A, then $\det(A) = 0$.*

Example 3.1.22 *Let* $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ *and let* $A_1 = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$. $\det A = -2$, $\det(A_1) = 2$.

Theorem 3.1.23 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from applying row operation 3. That is you replace some row by a multiple of another row added to itself. Then $\det(A) = \det(A_1)$.*

Example 3.1.24 *Let* $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ *and let* $A_1 = \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix}$. *Thus the second row of A_1 is one times the first row added to the second row.* $\det(A) = -2$ *and* $\det(A_1) = -2$.

Theorem 3.1.25 *In Theorems 3.1.19 - 3.1.23 you can replace the word, "row" with the word "column".*

There are two other major properties of determinants which do not involve row operations.

Theorem 3.1.26 *Let A and B be two $n \times n$ matrices. Then*

$$\boxed{\det(AB) = \det(A)\det(B).}$$

Also,

$$\boxed{\det(A) = \det\left(A^T\right).}$$

Example 3.1.27 *Compare* $\det(AB)$ *and* $\det(A)\det(B)$ *for*

$$A = \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix}, B = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}.$$

First

$$AB = \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 11 & 4 \\ -1 & -4 \end{pmatrix}$$

and so

$$\det(AB) = \det \begin{pmatrix} 11 & 4 \\ -1 & -4 \end{pmatrix} = -40.$$

Now

$$\det(A) = \det \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix} = 8$$

and

$$\det(B) = \det \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix} = -5.$$

Thus $\det(A)\det(B) = 8 \times (-5) = -40$.

3.1.4 Finding Determinants Using Row Operations

Theorems 3.1.23 - 3.1.25 can be used to find determinants using row operations. As pointed out above, the method of Laplace expansion will not be practical for any matrix of large size. Here is an example in which all the row operations are used.

Example 3.1.28 *Find the determinant of the matrix,*

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 1 & 2 & 3 \\ 4 & 5 & 4 & 3 \\ 2 & 2 & -4 & 5 \end{pmatrix}$$

Replace the second row by (-5) times the first row added to it. Then replace the third row by (-4) times the first row added to it. Finally, replace the fourth row by (-2) times the first row added to it. This yields the matrix,

$$B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -9 & -13 & -17 \\ 0 & -3 & -8 & -13 \\ 0 & -2 & -10 & -3 \end{pmatrix}$$

and from Theorem 3.1.23, it has the same determinant as A. Now using other row operations, $\det(B) = \left(\frac{-1}{3}\right)\det(C)$ where

$$C = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 11 & 22 \\ 0 & -3 & -8 & -13 \\ 0 & 6 & 30 & 9 \end{pmatrix}.$$

The second row was replaced by (-3) times the third row added to the second row. By Theorem 3.1.23 this didn't change the value of the determinant. Then the last row was multiplied by (-3) . By Theorem 3.1.19 the resulting matrix has a determinant which is (-3) times the determinant of the unmultiplied matrix. Therefore, I multiplied by $-1/3$ to retain the correct value. Now replace the last row with 2 times the third added to it. This does not change the value of the determinant by Theorem 3.1.23. Finally switch

the third and second rows. This causes the determinant to be multiplied by (-1). Thus $\det(C) = -\det(D)$ where

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -3 & -8 & -13 \\ 0 & 0 & 11 & 22 \\ 0 & 0 & 14 & -17 \end{pmatrix}$$

You could do more row operations or you could note that this can be easily expanded along the first column followed by expanding the 3×3 matrix which results along its first column. Thus

$$\det(D) = 1(-3)\begin{vmatrix} 11 & 22 \\ 14 & -17 \end{vmatrix} = 1485$$

and so $\det(C) = -1485$ and $\det(A) = \det(B) = \left(\frac{-1}{3}\right)(-1485) = 495$.

Example 3.1.29 *Find the determinant of the matrix*

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & -3 & 2 & 1 \\ 2 & 1 & 2 & 5 \\ 3 & -4 & 1 & 2 \end{pmatrix}$$

Replace the second row by (-1) times the first row added to it. Next take -2 times the first row and add to the third and finally take -3 times the first row and add to the last row. This yields

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -1 & -1 \\ 0 & -3 & -4 & 1 \\ 0 & -10 & -8 & -4 \end{pmatrix}.$$

By Theorem 3.1.23 this matrix has the same determinant as the original matrix. Remember you can work with the columns also. Take -5 times the last column and add to the second column. This yields

$$\begin{pmatrix} 1 & -8 & 3 & 2 \\ 0 & 0 & -1 & -1 \\ 0 & -8 & -4 & 1 \\ 0 & 10 & -8 & -4 \end{pmatrix}$$

By Theorem 3.1.25 this matrix has the same determinant as the original matrix. Now take (-1) times the third row and add to the top row. This gives.

$$\begin{pmatrix} 1 & 0 & 7 & 1 \\ 0 & 0 & -1 & -1 \\ 0 & -8 & -4 & 1 \\ 0 & 10 & -8 & -4 \end{pmatrix}$$

which by Theorem 3.1.23 has the same determinant as the original matrix. Lets expand it now along the first column. This yields the following for the determinant of the original matrix.

$$\det \begin{pmatrix} 0 & -1 & -1 \\ -8 & -4 & 1 \\ 10 & -8 & -4 \end{pmatrix}$$

which equals

$$8 \det \begin{pmatrix} -1 & -1 \\ -8 & -4 \end{pmatrix} + 10 \det \begin{pmatrix} -1 & -1 \\ -4 & 1 \end{pmatrix} = -82$$

Do not try to be fancy in using row operations. That is, stick mostly to the one which replaces a row or column with a multiple of another row or column added to it. Also note there is no way to check your answer other than working the problem more than one way. To be sure you have gotten it right you must do this.

3.2 Applications

3.2.1 A Formula For The Inverse

The definition of the determinant in terms of Laplace expansion along a row or column also provides a way to give a formula for the inverse of a matrix. Recall the definition of the inverse of a matrix in Definition 2.1.28 on Page 39. Also recall the definition of the cofactor matrix given in Definition 3.1.12 on Page 56. This cofactor matrix was just the matrix which results from replacing the ij^{th} entry of the matrix with the ij^{th} cofactor.

The following theorem says that to find the inverse, take the transpose of the cofactor matrix and divide by the determinant. The transpose of the cofactor matrix is called the **adjugate** or sometimes the **classical adjoint** of the matrix A. In other words, A^{-1} is equal to one divided by the determinant of A times the adjugate matrix of A. This is what the following theorem says with more precision.

Theorem 3.2.1 *A^{-1} exists if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$, then $A^{-1} = \left(a_{ij}^{-1}\right)$ where*

$$a_{ij}^{-1} = \det(A)^{-1} \operatorname{cof}(A)_{ji}$$

for $\operatorname{cof}(A)_{ij}$ the ij^{th} cofactor of A.

Example 3.2.2 *Find the inverse of the matrix,*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

First find the determinant of this matrix. Using Theorems 3.1.23 - 3.1.25 on Page 58, the determinant of this matrix equals the determinant of the matrix,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -8 \\ 0 & 0 & -2 \end{pmatrix}$$

which equals 12. The cofactor matrix of A is

$$\begin{pmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{pmatrix}.$$

Each entry of A was replaced by its cofactor. Therefore, from the above theorem, the inverse of A should equal

$$\frac{1}{12}\begin{pmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{pmatrix}^T = \begin{pmatrix} -\frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}.$$

Does it work? You should check to see if it does. When the matrices are multiplied

$$\begin{pmatrix} -\frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so it is correct.

Example 3.2.3 *Find the inverse of the matrix,*

$$A = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ -\frac{5}{6} & \frac{2}{3} & -\frac{1}{2} \end{pmatrix}$$

First find its determinant. This determinant is $\frac{1}{6}$. The inverse is therefore equal to

$$6\begin{pmatrix} \begin{vmatrix} \frac{1}{3} & -\frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} -\frac{1}{6} & -\frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} -\frac{1}{6} & \frac{1}{3} \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ -\begin{vmatrix} 0 & \frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ \begin{vmatrix} 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{3} \end{vmatrix} \end{pmatrix}^T.$$

Expanding all the 2×2 determinants this yields

$$6\begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}^T = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix}$$

Always check your work.

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ -\frac{5}{6} & \frac{2}{3} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so it is correct. If the result of multiplying these matrices had been something other than the identity matrix, you would know there was an error. When this happens, you need to search for the mistake if you are interested in getting the right answer. A common mistake is to forget to take the transpose of the cofactor matrix.

Proof of Theorem 3.2.1: From the definition of the determinant in terms of expansion along a column, and letting $(a_{ir}) = A$, if $\det(A) \neq 0$,

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ir} \det(A)^{-1} = \det(A) \det(A)^{-1} = 1.$$

Now consider

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

when $k \neq r$. Replace the k^{th} column with the r^{th} column to obtain a matrix, B_k whose determinant equals zero by Theorem 3.1.21. However, expanding this matrix, B_k along the k^{th} column yields

$$0 = \det(B_k) \det(A)^{-1} = \sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

Summarizing,

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1} = \delta_{rk} \equiv \begin{cases} 1 \text{ if } r = k \\ 0 \text{ if } r \neq k \end{cases}.$$

Now

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} = \sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)^{T}_{ki}$$

which is the kr^{th} entry of $\operatorname{cof}(A)^{T} A$. Therefore,

$$\frac{\operatorname{cof}(A)^{T}}{\det(A)} A = I. \tag{3.2}$$

Using the other formula in Definition 3.1.13, and similar reasoning,

$$\sum_{j=1}^{n} a_{rj} \operatorname{cof}(A)_{kj} \det(A)^{-1} = \delta_{rk}$$

Now

$$\sum_{j=1}^{n} a_{rj} \operatorname{cof}(A)_{kj} = \sum_{j=1}^{n} a_{rj} \operatorname{cof}(A)^{T}_{jk}$$

which is the rk^{th} entry of $A \operatorname{cof}(A)^{T}$. Therefore,

$$A \frac{\operatorname{cof}(A)^{T}}{\det(A)} = I, \tag{3.3}$$

and it follows from 3.2 and 3.3 that $A^{-1} = \left(a_{ij}^{-1}\right)$, where

$$a_{ij}^{-1} = \operatorname{cof}(A)_{ji} \det(A)^{-1}.$$

In other words,

$$A^{-1} = \frac{\operatorname{cof}(A)^{T}}{\det(A)}.$$

Now suppose A^{-1} exists. Then by Theorem 3.1.26,

$$1 = \det(I) = \det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right)$$

so $\det(A) \neq 0$. This proves the theorem.

This way of finding inverses is especially useful in the case where it is desired to find the inverse of a matrix whose entries are functions.

Example 3.2.4 *Suppose*

$$A(t) = \begin{pmatrix} e^t & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{pmatrix}$$

Show that $A(t)^{-1}$ exists and then find it.

First note $\det(A(t)) = e^t \neq 0$ so $A(t)^{-1}$ exists. The cofactor matrix is

$$C(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t\cos t & e^t\sin t \\ 0 & -e^t\sin t & e^t\cos t \end{pmatrix}$$

and so the inverse is

$$\frac{1}{e^t}\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t\cos t & e^t\sin t \\ 0 & -e^t\sin t & e^t\cos t \end{pmatrix}^T = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}.$$

3.2.2 Cramer's Rule

This formula for the inverse also implies a famous procedure known as **Cramer's rule**. Cramer's rule gives a formula for the solutions, $\mathbf{x}$, to a system of equations, $A\mathbf{x} = \mathbf{y}$ in the special case that A is a square matrix. Note this rule does not apply if you have a system of equations in which there is a different number of equations than variables.

In case you are solving a system of equations, $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, it follows that if A^{-1} exists,

$$\mathbf{x} = \left(A^{-1}A\right)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{y}$$

thus solving the system. Now in the case that A^{-1} exists, there is a formula for A^{-1} given above. Using this formula,

$$x_i = \sum_{j=1}^{n} a_{ij}^{-1} y_j = \sum_{j=1}^{n} \frac{1}{\det(A)} \operatorname{cof}(A)_{ji}\, y_j.$$

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det\begin{pmatrix} * & \cdots & y_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & y_n & \cdots & * \end{pmatrix},$$

where here the i^{th} column of A is replaced with the column vector, $(y_1 \cdots, y_n)^T$, and the determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

Procedure 3.2.5 *Suppose A is an $n \times n$ matrix and it is desired to solve the system $A\mathbf{x} = \mathbf{y}, \mathbf{y} = (y_1, \cdots, y_n)^T$ for $\mathbf{x} = (x_1, \cdots, x_n)^T$. Then Cramer's rule says*

$$x_i = \frac{\det A_i}{\det A}$$

where A_i is obtained from A by replacing the i^{th} column of A with the column $(y_1, \cdots, y_n)^T$.

Example 3.2.6 *Find x, y if*

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

From Cramer's rule,

$$x = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 3 & -3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{1}{2}$$

Now to find y,

$$y = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = -\frac{1}{7}$$

$$z = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 2 \\ 2 & -3 & 3 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{11}{14}$$

You see the pattern. For large systems Cramer's rule is less than useful if you want to find an answer. This is because to use it you must evaluate determinants. However, you have no practical way to evaluate determinants for large matrices other than row operations and if you are using row operations, you might just as well use them to solve the system to begin with. It will be a lot less trouble. Nevertheless, there are situations in which Cramer's rule is useful.

Example 3.2.7 *Solve for z if*

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix}$$

You could do it by row operations but it might be easier in this case to use Cramer's rule because the matrix of coefficients does not consist of numbers but of functions.

Thus

$$z = \frac{\begin{vmatrix} 1 & 0 & 1 \\ 0 & e^t \cos t & t \\ 0 & -e^t \sin t & t^2 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{vmatrix}} = t\left((\cos t)\, t + \sin t\right) e^{-t}.$$

You end up doing this sort of thing sometimes in ordinary differential equations in the method of variation of parameters.

3.3 Exercises

1. Find the determinants of the following matrices.

   (a) ♠ $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 2 \\ 0 & 9 & 8 \end{pmatrix}$ (The answer is 31.)

   (b) ♠ $\begin{pmatrix} 4 & 3 & 2 \\ 1 & 7 & 8 \\ 3 & -9 & 3 \end{pmatrix}$ (The answer is 375.)

   (c) $\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 2 & 3 \\ 4 & 1 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{pmatrix}$, (The answer is -2.)

2. Find the following determinant by expanding along the first row and second column.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

3. Find the following determinant by expanding along the first column and third row.

$$\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix}$$

4. Find the following determinant by expanding along the second row and first column.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

5. Compute the determinant by cofactor expansion. Pick the easiest row or column to use.

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 2 & 1 & 3 & 1 \end{vmatrix}$$

6. Find the determinant using row operations.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ -4 & 1 & 2 \end{vmatrix}$$

7. Find the determinant using row operations.

$$\begin{vmatrix} 2 & 1 & 3 \\ 2 & 4 & 2 \\ 1 & 4 & -5 \end{vmatrix}$$

8. Find the determinant using row operations.

$$\begin{vmatrix} 1 & 2 & 1 & 2 \\ 3 & 1 & -2 & 3 \\ -1 & 0 & 3 & 1 \\ 2 & 3 & 2 & -2 \end{vmatrix}$$

9. Find the determinant using row operations.

$$\begin{vmatrix} 1 & 4 & 1 & 2 \\ 3 & 2 & -2 & 3 \\ -1 & 0 & 3 & 3 \\ 2 & 1 & 2 & -2 \end{vmatrix}$$

10. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

11. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} c & d \\ a & b \end{pmatrix}$$

12. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & b \\ a+c & b+d \end{pmatrix}$$

13. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & b \\ 2c & 2d \end{pmatrix}$$

14. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} b & a \\ d & c \end{pmatrix}$$

15. Tell whether the statement is true or false.

    (a) If A is a 3×3 matrix with a zero determinant, then one column must be a multiple of some other column.

    (b) If any two columns of a square matrix are equal, then the determinant of the matrix equals zero.

    (c) For A and B two $n \times n$ matrices, $\det(A+B) = \det(A) + \det(B)$.

    (d) For A an $n \times n$ matrix, $\det(3A) = 3\det(A)$

    (e) If A^{-1} exists then $\det\left(A^{-1}\right) = \det(A)^{-1}$.

    (f) If B is obtained by multiplying a single row of A by 4 then $\det(B) = 4\det(A)$.

    (g) For A an $n \times n$ matrix, $\det(-A) = (-1)^n \det(A)$.

    (h) If A is a real $n \times n$ matrix, then $\det\left(A^T A\right) \geq 0$.

    (i) Cramer's rule is useful for finding solutions to systems of linear equations in which there is an infinite set of solutions.

    (j) If $A^k = 0$ for some positive integer, k, then $\det(A) = 0$.

    (k) If $A\mathbf{x} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$, then $\det(A) = 0$.

16. Verify an example of each property of determinants found in Theorems 3.1.23 - 3.1.25 for 2×2 matrices.

17. A matrix is said to be **orthogonal** if $A^T A = I$. Thus the inverse of an orthogonal matrix is just its transpose. What are the possible values of $\det(A)$ if A is an orthogonal matrix?

18. Fill in the missing entries to make the matrix orthogonal as in Problem 17.

$$\begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{\sqrt{12}}{6} \\ \frac{1}{\sqrt{2}} & - & - \\ - & \frac{\sqrt{6}}{3} & - \end{pmatrix}.$$

19. If A^{-1} exist, what is the relationship between $\det(A)$ and $\det\left(A^{-1}\right)$. Explain your answer.

20. Is it true that $\det(A+B) = \det(A) + \det(B)$? If this is so, explain why it is so and if it is not so, give a counter example.

21. Let A be an $r \times r$ matrix and suppose there are $r-1$ rows (columns) such that all rows (columns) are linear combinations of these $r-1$ rows (columns). Show $\det(A) = 0$.

22. Show $\det(aA) = a^n \det(A)$ where here A is an $n \times n$ matrix and a is a scalar.

23. Suppose A is an upper triangular matrix. Show that A^{-1} exists if and only if all elements of the main diagonal are non zero. Is it true that A^{-1} will also be upper triangular? Explain. Is everything the same for lower triangular matrices?

24. Let A and B be two $n \times n$ matrices. $A \sim B$ (A is **similar** to B) means there exists an invertible matrix, S such that $A = S^{-1}BS$. Show that if $A \sim B$, then $B \sim A$. Show also that $A \sim A$ and that if $A \sim B$ and $B \sim C$, then $A \sim C$.

25. In the context of Problem 24 show that if $A \sim B$, then $\det(A) = \det(B)$.

26. Two $n \times n$ matrices, A and B, are similar if $B = S^{-1}AS$ for some invertible $n \times n$ matrix, S. Show that if two matrices are similar, they have the same characteristic polynomials. The characteristic polynomial of an $n \times n$ matrix, M is the polynomial, $\det(\lambda I - M)$.

27. Prove by doing computations that $\det(AB) = \det(A)\det(B)$ if A and B are 2×2 matrices.

28. Illustrate with an example of 2×2 matrices that the determinant of a product equals the product of the determinants.

29. An $n \times n$ matrix is called **nilpotent** if for some positive integer, k it follows $A^k = 0$. If A is a nilpotent matrix and k is the smallest possible integer such that $A^k = 0$, what are the possible values of $\det(A)$?

30. Use Cramer's rule to find the solution to

$$\begin{array}{c} x + 2y = 1 \\ 2x - y = 2 \end{array}$$

31. Use Cramer's rule to find the solution to

$$\begin{array}{c} x + 2y + z = 1 \\ 2x - y - z = 2 \\ x + z = 1 \end{array}$$

32. Here is a matrix,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 3 & 1 & 0 \end{pmatrix}$$

Determine whether the matrix has an inverse by finding whether the determinant is non zero.

33. Here is a matrix,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

34. Here is a matrix,

$$\begin{pmatrix} 1 & t & t^2 \\ 0 & 1 & 2t \\ t & 0 & 2 \end{pmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

35. Here is a matrix,

$$\begin{pmatrix} e^t & e^{-t}\cos t & e^{-t}\sin t \\ e^t & -e^{-t}\cos t - e^{-t}\sin t & -e^{-t}\sin t + e^{-t}\cos t \\ e^t & 2e^{-t}\sin t & -2e^{-t}\cos t \end{pmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

36. Here is a matrix,
$$\begin{pmatrix} e^t & \cosh t & \sinh t \\ e^t & \sinh t & \cosh t \\ e^t & \cosh t & \sinh t \end{pmatrix}$$
Does there exist a value of t for which this matrix fails to have an inverse? Explain.

37. Use the formula for the inverse in terms of the cofactor matrix to find if possible the inverses of the matrices
$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 4 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$
If it is not possible to take the inverse, explain why.

38. Use the formula for the inverse in terms of the cofactor matrix to find the inverse of the matrix,
$$A = \begin{pmatrix} e^t & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & e^t \cos t - e^t \sin t & e^t \cos t + e^t \sin t \end{pmatrix}.$$

39. Find the inverse if it exists of the matrix,
$$\begin{pmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{pmatrix}.$$

40. Let $F(t) = \det \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix}$. Verify
$$F'(t) = \det \begin{pmatrix} a'(t) & b'(t) \\ c(t) & d(t) \end{pmatrix} + \det \begin{pmatrix} a(t) & b(t) \\ c'(t) & d'(t) \end{pmatrix}.$$
Now suppose
$$F(t) = \det \begin{pmatrix} a(t) & b(t) & c(t) \\ d(t) & e(t) & f(t) \\ g(t) & h(t) & i(t) \end{pmatrix}.$$
Use Laplace expansion and the first part to verify $F'(t) =$
$$\det \begin{pmatrix} a'(t) & b'(t) & c'(t) \\ d(t) & e(t) & f(t) \\ g(t) & h(t) & i(t) \end{pmatrix} + \det \begin{pmatrix} a(t) & b(t) & c(t) \\ d'(t) & e'(t) & f'(t) \\ g(t) & h(t) & i(t) \end{pmatrix}$$
$$+ \det \begin{pmatrix} a(t) & b(t) & c(t) \\ d(t) & e(t) & f(t) \\ g'(t) & h'(t) & i'(t) \end{pmatrix}.$$
Conjecture a general result valid for $n \times n$ matrices and explain why it will be true. Can a similar thing be done with the columns?

41. Let $Ly = y^{(n)} + a_{n-1}(x) y^{(n-1)} + \cdots + a_1(x) y' + a_0(x) y$ where the a_i are given continuous functions defined on a closed interval, (a, b) and y is some function

which has n derivatives so it makes sense to write Ly. Suppose $Ly_k = 0$ for $k = 1, 2, \cdots, n$. The **Wronskian** of these functions, y_i is defined as

$$W(y_1, \cdots, y_n)(x) \equiv \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{pmatrix}$$

Show that for $W(x) = W(y_1, \cdots, y_n)(x)$ to save space,

$$W'(x) = \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n)}(x) & \cdots & y_n^{(n)}(x) \end{pmatrix}.$$

Now use the differential equation, $Ly = 0$ which is satisfied by each of these functions, y_i and properties of determinants presented above to verify that $W' + a_{n-1}(x) W = 0$. Give an explicit solution of this linear differential equation, **Abel's formula,** and use your answer to verify that the Wronskian of these solutions to the equation, $Ly = 0$ either vanishes identically on (a, b) or never. **Hint:** To solve the differential equation, let $A'(x) = a_{n-1}(x)$ and multiply both sides of the differential equation by $e^{A(x)}$ and then argue the left side is the derivative of something.

3.4 Exercises With Answers

1. Find the following determinant by expanding along the first row and second column.

$$\begin{vmatrix} 1 & 2 & 1 \\ 0 & 4 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

Expanding along the first row you would have

$$1\begin{vmatrix} 4 & 3 \\ 1 & 1 \end{vmatrix} - 2\begin{vmatrix} 0 & 3 \\ 2 & 1 \end{vmatrix} + 1\begin{vmatrix} 0 & 4 \\ 2 & 1 \end{vmatrix} = 5.$$

Expanding along the second column you would have

$$-2\begin{vmatrix} 0 & 3 \\ 2 & 1 \end{vmatrix} + 4\begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} - 1\begin{vmatrix} 1 & 1 \\ 0 & 3 \end{vmatrix} = 5$$

Be sure you understand how you must multiply by $(-1)^{i+j}$ to get the term which goes with the ij^{th} entry. For example, in the above, there is a -2 because the 2 is in the first row and the second column.

2. Find the following determinant by expanding along the first column and third row.

$$\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix}$$

Expanding along the first column you get

$$1\begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} - 1\begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} + 2\begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 2$$

Expanding along the third row you get

$$2\begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} - 1\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} + 1\begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix} = 2$$

3. Compute the determinant by cofactor expansion. Pick the easiest row or column to use.

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \\ 2 & 1 & 3 & 1 \end{vmatrix}$$

Probably it is easiest to expand along the third row. This gives

$$(-1)\,3\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 3 \end{vmatrix} = -3 \times 1 \times \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} = -6$$

Notice how I expanded the three by three matrix along the top row.

4. Find the determinant using row operations.

$$\begin{vmatrix} 11 & 2 & 1 \\ 2 & 7 & 2 \\ -4 & 1 & 2 \end{vmatrix}$$

The following is a sequence of numbers which according to the theorems on row operations have the same value as the original determinant.

$$\begin{vmatrix} 11 & 2 & 1 \\ 2 & 7 & 2 \\ 0 & 15 & 6 \end{vmatrix}$$

To get this one, I added 2 times the second row to the last row. This gives a matrix which has the same determinant as the original matrix. Next I will multiply the second row by 11 and the top row by 2. This has the effect of producing a matrix whose determinant is 22 times too large. Therefore, I need to divide the result by 22.

$$\begin{vmatrix} 22 & 4 & 2 \\ 22 & 77 & 22 \\ 0 & 15 & 6 \end{vmatrix} \frac{1}{22}.$$

Next I will add the (-1) times the top row to the second row. This leaves things unchanged.

$$\begin{vmatrix} 22 & 4 & 2 \\ 0 & 73 & 20 \\ 0 & 15 & 6 \end{vmatrix} \frac{1}{22}$$

That 73 looks pretty formidable so I shall take -3 times the third column and add to the second column. This will leave the number unchanged.

$$\begin{vmatrix} 22 & -2 & 2 \\ 0 & 13 & 20 \\ 0 & -3 & 6 \end{vmatrix} \frac{1}{22}$$

Now I will divide the bottom row by 3. To compensate for the damage inflicted, I must then multiply by 3.

$$\begin{vmatrix} 22 & -2 & 2 \\ 0 & 13 & 20 \\ 0 & -1 & 2 \end{vmatrix} \frac{3}{22}$$

I don't like the 13 so I will take 13 times the bottom row and add to the middle. This will leave the number unchanged.

$$\begin{vmatrix} 22 & -2 & 2 \\ 0 & 0 & 46 \\ 0 & -1 & 2 \end{vmatrix} \frac{3}{22}$$

Finally, I will switch the two bottom rows. This will change the sign. Therefore, after doing this row operation, I need to multiply the result by (-1) to compensate for the damage done by the row operation.

$$-\begin{vmatrix} 22 & -2 & 2 \\ 0 & -1 & 2 \\ 0 & 0 & 46 \end{vmatrix} \frac{3}{22} = 138$$

The final matrix is upper triangular so to get its determinant, just multiply the entries on the main diagonal.

5. Find the determinant using row operations.

$$\begin{vmatrix} 1 & 2 & 1 & 2 \\ 3 & 1 & -2 & 3 \\ -1 & 0 & 3 & 1 \\ 2 & 3 & 2 & -2 \end{vmatrix}$$

In this case, you can do row operations on the matrix which are of the sort where a row is replaced with itself added to another row without switching any rows and eventually end up with

$$\begin{pmatrix} 1 & 2 & 1 & 2 \\ 0 & -5 & -5 & -3 \\ 0 & 0 & 2 & \frac{9}{5} \\ 0 & 0 & 0 & -\frac{63}{10} \end{pmatrix}$$

Each of these row operations does not change the value of the determinant of the matrix and so the determinant is 63 which is obtained by multiplying the entries which are down the main diagonal.

6. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

In this case the transpose of the matrix on the left was taken. The new matrix will have the same determinant as the original matrix.

7. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & b \\ a+c & b+d \end{pmatrix}$$

This simply replaced the second row with the first row added to the second row. The new matrix will have the same determinant as the original one.

8. A matrix is said to be **orthogonal** if $A^T A = I$. Thus the inverse of an orthogonal matrix is just its transpose. What are the possible values of $\det(A)$ if A is an orthogonal matrix?

   $\det(I) = 1$ and so $\det\left(A^T\right)\det(A) = 1$. Now how does $\det(A)$ relate to $\det\left(A^T\right)$? You finish the argument.

9. Show $\det(aA) = a^n \det(A)$ where here A is an $n \times n$ matrix and a is a scalar.

   Each time you multiply a row by a the new matrix has determinant equal to a times the determinant of the matrix you multiplied by a. aA can be obtained by multiplying a succession of n rows by a and so aA has determinant equal to a^n times the determinant of A.

10. Let A and B be two $n \times n$ matrices. $A \sim B$ (A is **similar** to B) means there exists an invertible matrix, S such that $A = S^{-1}BS$. Show that if $A \sim B$, then $B \sim A$. Show also that $A \sim A$ and that if $A \sim B$ and $B \sim C$, then $A \sim C$.

11. In the context of Problem 24 show that if $A \sim B$, then $\det(A) = \det(B)$.

    $A = S^{-1}BS$ and so

$$\begin{aligned} \det(A) &= \det\left(S^{-1}\right)\det(B)\det(S) = \\ \det\left(S^{-1}S\right)\det(B) &= \det(I)\det(B) = \det(B) \end{aligned}$$

12. An $n \times n$ matrix is called **nilpotent** if for some positive integer, k it follows $A^k = 0$. If A is a nilpotent matrix and k is the smallest possible integer such that $A^k = 0$, what are the possible values of $\det(A)$?

    Remember the determinant of a product equals the product of the determinants.

13. Use Cramer's rule to find the solution to

$$\begin{aligned} x + 5y + z &= 1 \\ 2x - y - z &= 2 \\ x + z &= 1 \end{aligned}$$

    To find y, you can use Cramer's rule.

$$y = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 2 & 2 & -1 \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & 5 & 1 \\ 2 & -1 & -1 \\ 1 & 0 & 1 \end{vmatrix}} = 0$$

    You can find the other variables in the same way.

14. Here is a matrix,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}$$

    Does there exist a value of t for which this matrix fails to have an inverse? Explain.

    You should take the determinant and remember the identity that $\cos^2(t) + \sin^2(t) = 1$.

15. Here is a matrix,
$$\begin{pmatrix} 1 & t & t^2 \\ 0 & 1 & 2t \\ t^3 & 0 & 2 \end{pmatrix}$$
Does there exist a value of t for which this matrix fails to have an inverse? Explain.

$$\begin{vmatrix} 1 & t & t^2 \\ 0 & 1 & 2t \\ t^3 & 0 & 2 \end{vmatrix} = 2 + t^5$$
and so if $t = \sqrt[5]{-2}$, then this matrix has fails to have an inverse. However, it has an inverse for all other values of t.

16. Use the formula for the inverse in terms of the cofactor matrix to find if possible the inverses of the matrices
$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 4 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$
If it is not possible to take the inverse, explain why.

Part II

Vectors In $\mathbb{R}^n$

Vectors And Points In $\mathbb{R}^n$

4.0.1 Outcomes

1. Evaluate the distance between two points in $\mathbb{R}^n$.
2. Be able to represent a vector in each of the following ways for $n = 2, 3$
   (a) as a directed arrow in n space
   (b) as an ordered n tuple
   (c) as a linear combination of unit coordinate vectors
3. Carry out the vector operations:
   (a) addition
   (b) scalar multiplication
   (c) find magnitude (norm or length)
   (d) Find the vector of unit length in the direction of a given vector.
4. Represent the operations of vector addition, scalar multiplication and norm geometrically.
5. Recall and apply the basic properties of vector addition, scalar multiplication and norm.
6. Model and solve application problems using vectors.
7. Describe an open ball in $\mathbb{R}^n$.
8. Determine whether a set in $\mathbb{R}^n$ is open, closed, or neither.

4.1 Open And Closed Sets

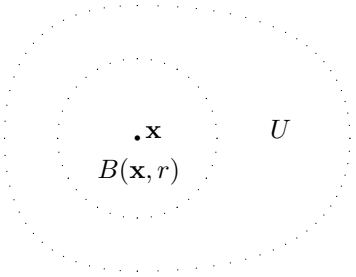
Eventually, one must consider functions which are defined on subsets of $\mathbb{R}^n$ and their properties. The next definition will end up being quite important. It describe a type of subset of $\mathbb{R}^n$ with the property that if $\mathbf{x}$ is in this set, then so is $\mathbf{y}$ whenever $\mathbf{y}$ is close enough to $\mathbf{x}$.

Definition 4.1.1 *Let $U \subseteq \mathbb{R}^n$. U is an **open set** if whenever $\mathbf{x} \in U$, there exists $r > 0$ such that $B(\mathbf{x}, r) \subseteq U$. More generally, if U is any subset of $\mathbb{R}^n$, $\mathbf{x} \in U$ is an **interior point** of U if there exists $r > 0$ such that $\mathbf{x} \in B(\mathbf{x}, r) \subseteq U$. In other words U is an open set exactly when every point of U is an interior point of U.*

If there is something called an open set, surely there should be something called a closed set and here is the definition of one.

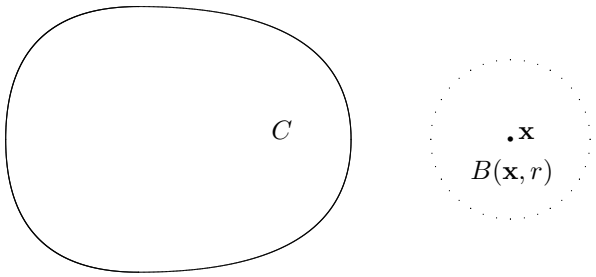
Definition 4.1.2 *A subset, C, of $\mathbb{R}^n$ is called a **closed set** if $\mathbb{R}^n \setminus C$ is an open set. They symbol, $\mathbb{R}^n \setminus C$ denotes everything in $\mathbb{R}^n$ which is not in C. It is also called the **complement** of C. The symbol, S^C is a short way of writing $\mathbb{R}^n \setminus S$.*

To illustrate this definition, consider the following picture.



You see in this picture how the edges are dotted. This is because an open set, can not include the edges or the set would fail to be open. For example, consider what would happen if you picked a point out on the edge of U in the above picture. Every open ball centered at that point would have in it some points which are outside U. Therefore, such a point would violate the above definition. You also see the edges of $B(\mathbf{x}, r)$ dotted suggesting that $B(\mathbf{x}, r)$ ought to be an open set. This is intuitively clear but does require a proof. This will be done in the next theorem and will give examples of open sets. Also, you can see that if $\mathbf{x}$ is close to the edge of U, you might have to take r to be very small.

It is roughly the case that open sets don't have their skins while closed sets do. Here is a picture of a closed set, C.



Note that $\mathbf{x} \notin C$ and since $\mathbb{R}^n \setminus C$ is open, there exists a ball, $B(\mathbf{x}, r)$ contained entirely in $\mathbb{R}^n \setminus C$. If you look at $\mathbb{R}^n \setminus C$, what would be its skin? It can't be in $\mathbb{R}^n \setminus C$ and so it must be in C. This is a rough heuristic explanation of what is going on with these definitions. Also note that $\mathbb{R}^n$ and $\emptyset$ are both open and closed. Here is why. If $\mathbf{x} \in \emptyset$, then there must be a ball centered at $\mathbf{x}$ which is also contained in $\emptyset$. This must be considered to be true because there is nothing in $\emptyset$ so there can be no example to show it false[1]. Therefore, from the definition, it follows $\emptyset$ is open. It is also closed because

[1]To a mathematician, the statment: Whenever a pig is born with wings it can fly must be taken as true. We do not consider biological or aerodynamic considerations in such statements. There is no such thing as a winged pig and therefore, all winged pigs must be superb flyers since there can be no example of one which is not. On the other hand we would also consider the statement: Whenever a pig is born with wings it can't possibly fly, as equally true. The point is, you can say anything you

if $\mathbf{x} \notin \emptyset$, then $B(\mathbf{x}, 1)$ is also contained in $\mathbb{R}^n \setminus \emptyset = \mathbb{R}^n$. Therefore, $\emptyset$ is both open and closed. From this, it follows $\mathbb{R}^n$ is also both open and closed.

Theorem 4.1.3 *Let $\mathbf{x} \in \mathbb{R}^n$ and let $r \geq 0$. Then $B(\mathbf{x}, r)$ is an open set. Also,*

$$D(\mathbf{x}, r) \equiv \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{y} - \mathbf{x}| \leq r\}$$

is a closed set.

Proof: Suppose $\mathbf{y} \in B(\mathbf{x}, r)$. It is necessary to show there exists $r_1 > 0$ such that $B(\mathbf{y}, r_1) \subseteq B(\mathbf{x}, r)$. Define $r_1 \equiv r - |\mathbf{x} - \mathbf{y}|$. Then if $|\mathbf{z} - \mathbf{y}| < r_1$, it follows from the above triangle inequality that

$$\begin{aligned} |\mathbf{z} - \mathbf{x}| &= |\mathbf{z} - \mathbf{y} + \mathbf{y} - \mathbf{x}| \\ &\leq |\mathbf{z} - \mathbf{y}| + |\mathbf{y} - \mathbf{x}| \\ &< r_1 + |\mathbf{y} - \mathbf{x}| = r - |\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{x}| = r. \end{aligned}$$

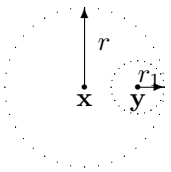
Note that if $r = 0$ then $B(\mathbf{x}, r) = \emptyset$, the empty set. This is because if $\mathbf{y} \in \mathbb{R}^n$, $|\mathbf{x} - \mathbf{y}| \geq 0$ and so $\mathbf{y} \notin B(\mathbf{x}, 0)$. Since $\emptyset$ has no points in it, it must be open because every point in it, (There are none.) satisfies the desired property of being an interior point.

Now suppose $\mathbf{y} \notin D(\mathbf{x}, r)$. Then $|\mathbf{x} - \mathbf{y}| > r$ and defining $\delta \equiv |\mathbf{x} - \mathbf{y}| - r$, it follows that if $\mathbf{z} \in B(\mathbf{y}, \delta)$, then by the triangle inequality,

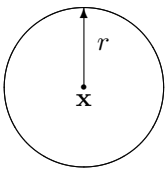
$$\begin{aligned} |\mathbf{x} - \mathbf{z}| &\geq |\mathbf{x} - \mathbf{y}| - |\mathbf{y} - \mathbf{z}| > |\mathbf{x} - \mathbf{y}| - \delta \\ &= |\mathbf{x} - \mathbf{y}| - (|\mathbf{x} - \mathbf{y}| - r) = r \end{aligned}$$

and this shows that $B(\mathbf{y}, \delta) \subseteq \mathbb{R}^n \setminus D(\mathbf{x}, r)$. Since $\mathbf{y}$ was an arbitrary point in $\mathbb{R}^n \setminus D(\mathbf{x}, r)$, it follows $\mathbb{R}^n \setminus D(\mathbf{x}, r)$ is an open set which shows from the definition that $D(\mathbf{x}, r)$ is a closed set as claimed.

A picture which is descriptive of the conclusion of the above theorem which also implies the manner of proof is the following.



$B(\mathbf{x}, r)$



$D(\mathbf{x}, r)$

Recall $\mathbb{R}^2$ consists of ordered pairs, (x, y) such that $x \in \mathbb{R}$ and $y \in \mathbb{R}$. $\mathbb{R}^2$ is also written as $\mathbb{R} \times \mathbb{R}$. In general, the following definition holds.

Definition 4.1.4 *The **Cartesian product** of two sets, $A \times B$, means $\{(a, b) : a \in A,\ b \in B\}$. If you have n sets, $A_1, A_2, \cdots, A_n$*

$$\prod_{i=1}^{n} A_i = \{(x_1, x_2, \cdots, x_n) : \text{ each } x_i \in A_i\}.$$

want about the elements of the empty set and no one can gainsay your statement. Therefore, such statements are considered as true by default. You may say this is a very strange way of thinking about truth and ultimately this is because mathematics is not about truth. It is more about consistency and logic.

Now suppose $A \subseteq \mathbb{R}^m$ and $B \subseteq \mathbb{R}^n$. Then if $(\mathbf{x}, \mathbf{y}) \in A \times B$, $\mathbf{x} = (x_1, \cdots, x_m)$ and $\mathbf{y} = (y_1, \cdots, y_n)$, the following identification will be made.

$$(\mathbf{x}, \mathbf{y}) = (x_1, \cdots, x_m, y_1, \cdots, y_n) \in \mathbb{R}^{n+m}.$$

Similarly, starting with something in $\mathbb{R}^{n+m}$, you can write it in the form $(\mathbf{x}, \mathbf{y})$ where $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$. The following theorem has to do with the Cartesian product of two closed sets or two open sets. Also here is an important definition.

Definition 4.1.5 *A set, $A \subseteq \mathbb{R}^n$ is said to be* ***bounded*** *if there exist finite intervals, $[a_i, b_i]$ such that $A \subseteq \prod_{i=1}^n [a_i, b_i]$.*

Theorem 4.1.6 *Let U be an open set in $\mathbb{R}^m$ and let V be an open set in $\mathbb{R}^n$. Then $U \times V$ is an open set in $\mathbb{R}^{n+m}$. If C is a closed set in $\mathbb{R}^m$ and H is a closed set in $\mathbb{R}^n$, then $C \times H$ is a closed set in $\mathbb{R}^{n+m}$. If C and H are bounded, then so is $C \times H$.*

Proof: Let $(\mathbf{x}, \mathbf{y}) \in U \times V$. Since U is open, there exists $r_1 > 0$ such that $B(\mathbf{x}, r_1) \subseteq U$. Similarly, there exists $r_2 > 0$ such that $B(\mathbf{y}, r_2) \subseteq V$. Now

$$B((\mathbf{x}, \mathbf{y}), \delta) \equiv$$

$$\left\{ (\mathbf{s}, \mathbf{t}) \in \mathbb{R}^{n+m} : \sum_{k=1}^{m} |x_k - s_k|^2 + \sum_{j=1}^{n} |y_j - t_j|^2 < \delta^2 \right\}$$

Therefore, if $\delta \equiv \min(r_1, r_2)$ and $(\mathbf{s}, \mathbf{t}) \in B((\mathbf{x}, \mathbf{y}), \delta)$, then it follows that $\mathbf{s} \in B(\mathbf{x}, r_1) \subseteq U$ and that $\mathbf{t} \in B(\mathbf{y}, r_2) \subseteq V$ which shows that $B((\mathbf{x}, \mathbf{y}), \delta) \subseteq U \times V$. Hence $U \times V$ is open as claimed.

Next suppose $(\mathbf{x}, \mathbf{y}) \notin C \times H$. It is necessary to show there exists $\delta > 0$ such that $B((\mathbf{x}, \mathbf{y}), \delta) \subseteq \mathbb{R}^{n+m} \setminus (C \times H)$. Either $\mathbf{x} \notin C$ or $\mathbf{y} \notin H$ since otherwise $(\mathbf{x}, \mathbf{y})$ would be a point of $C \times H$. Suppose therefore, that $\mathbf{x} \notin C$. Since C is closed, there exists $r > 0$ such that $B(\mathbf{x}, r) \subseteq \mathbb{R}^m \setminus C$. Consider $B((\mathbf{x}, \mathbf{y}), r)$. If $(\mathbf{s}, \mathbf{t}) \in B((\mathbf{x}, \mathbf{y}), r)$, it follows that $\mathbf{s} \in B(\mathbf{x}, r)$ which is contained in $\mathbb{R}^m \setminus C$. Therefore, $B((\mathbf{x}, \mathbf{y}), r) \subseteq \mathbb{R}^{n+m} \setminus (C \times H)$ showing $C \times H$ is closed. A similar argument holds if $\mathbf{y} \notin H$.

If C is bounded, there exist $[a_i, b_i]$ such that $C \subseteq \prod_{i=1}^m [a_i, b_i]$ and if H is bounded, $H \subseteq \prod_{i=m+1}^{m+n} [a_i, b_i]$ for intervals $[a_{m+1}, b_{m+1}], \cdots, [a_{m+n}, b_{m+n}]$. Therefore, $C \times H \subseteq \prod_{i=1}^{m+n} [a_i, b_i]$ and this establishes the last part of this theorem.

4.2 Physical Vectors

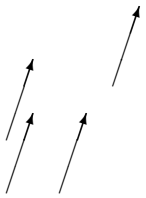
Suppose you push on something. What is important? There are really two things which are important, how hard you push and the direction you push. This illustrates the concept of force.

Definition 4.2.1 ***Force*** *is a vector. The magnitude of this vector is a measure of how hard it is pushing. It is measured in units such as Newtons or pounds or tons. Its direction is the direction in which the push is taking place.*

Of course this is a little vague and will be left a little vague until the presentation of Newton's second law later.

Vectors are used to model force and other physical vectors like velocity. What was just described would be called a force vector. It has two essential ingredients, its magnitude and its direction. Geometrically think of vectors as directed line segments or arrows as shown in the following picture in which all the directed line segments are

considered to be the same vector because they have the same direction, the direction in which the arrows point, and the same magnitude (length).



Because of this fact that only direction and magnitude are important, it is always possible to put a vector in a certain particularly simple form. Let $\overrightarrow{\mathbf{pq}}$ be a directed line segment or vector. Then from Definition 1.4.4 it follows that $\overrightarrow{\mathbf{pq}}$ consists of the points of the form

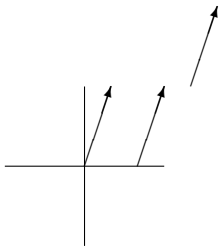
$$\mathbf{p} + t(\mathbf{q} - \mathbf{p})$$

where $t \in [0,1]$. Subtract $\mathbf{p}$ from all these points to obtain the directed line segment consisting of the points

$$\mathbf{0} + t(\mathbf{q} - \mathbf{p}),\ t \in [0,1].$$

The point in $\mathbb{R}^n, \mathbf{q} - \mathbf{p}$, will represent the vector.

Geometrically, the arrow, $\overrightarrow{\mathbf{pq}}$, was slid so it points in the same direction and the base is at the origin, $\mathbf{0}$. For example, see the following picture.



In this way vectors can be identified with elements of $\mathbb{R}^n$.

Definition 4.2.2 *Let $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$. The **position vector** of this point is the vector whose point is at $\mathbf{x}$ and whose tail is at the origin, $(0, \cdots, 0)$. If $\mathbf{x} = (x_1, \cdots, x_n)$ is called a vector, the vector which is meant is this position vector just described. Another term associated with this is **standard position.** A vector is in standard position if the tail is placed at the origin.*

It is customary to identify the point in $\mathbb{R}^n$ with its position vector.

The magnitude of a vector determined by a directed line segment $\overrightarrow{\mathbf{pq}}$ is just the distance between the point $\mathbf{p}$ and the point $\mathbf{q}$. By the distance formula this equals

$$\left(\sum_{k=1}^{n} (q_k - p_k)^2 \right)^{1/2} = |\mathbf{p} - \mathbf{q}|$$

and for $\mathbf{v}$ any vector in $\mathbb{R}^n$ the magnitude of $\mathbf{v}$ equals $\left(\sum_{k=1}^{n} v_k^2 \right)^{1/2} = |\mathbf{v}|$.

Example 4.2.3 *Consider the vector, $\mathbf{v} \equiv (1,2,3)$ in $\mathbb{R}^n$. Find $|\mathbf{v}|$.*

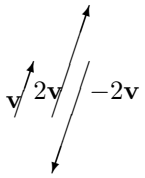
First, the vector is the directed line segment (arrow) which has its base at $\mathbf{0} \equiv (0,0,0)$ and its point at $(1,2,3)$. Therefore,

$$|\mathbf{v}| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}.$$

What is the geometric significance of scalar multiplication? If $\mathbf{a}$ represents the vector, $\mathbf{v}$ in the sense that when it is slid to place its tail at the origin, the element of $\mathbb{R}^n$ at its point is $\mathbf{a}$, what is $r\mathbf{v}$?

$$|r\mathbf{v}| = \left(\sum_{k=1}^{n} (ra_i)^2\right)^{1/2} = \left(\sum_{k=1}^{n} r^2 (a_i)^2\right)^{1/2}$$
$$= \left(r^2\right)^{1/2} \left(\sum_{k=1}^{n} a_i^2\right)^{1/2} = |r|\,|\mathbf{v}|\,.$$

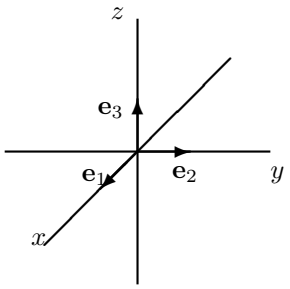
Thus the magnitude of $r\mathbf{v}$ equals $|r|$ times the magnitude of $\mathbf{v}$. If r is positive, then the vector represented by $r\mathbf{v}$ has the same direction as the vector, $\mathbf{v}$ because multiplying by the scalar, r, only has the effect of scaling all the distances. Thus the unit distance along any coordinate axis now has length r and in this rescaled system the vector is represented by $\mathbf{a}$. If $r < 0$ similar considerations apply except in this case all the a_i also change sign. From now on, $\mathbf{a}$ will be referred to as a vector instead of an element of $\mathbb{R}^n$ representing a vector as just described. The following picture illustrates the effect of scalar multiplication.



Note there are n special vectors which point along the coordinate axes. These are

$$\mathbf{e}_i \equiv (0, \cdots, 0, 1, 0, \cdots, 0)$$

where the 1 is in the i^{th} slot and there are zeros in all the other spaces. See the picture in the case of $\mathbb{R}^3$.



The direction of $\mathbf{e}_i$ is referred to as the i^{th} direction. Given a vector, $\mathbf{v} = (a_1, \cdots, a_n)$, $a_i\mathbf{e}_i$ is the i^{th} component of the vector. Thus $a_i\mathbf{e}_i = (0, \cdots, 0, a_i, 0, \cdots, 0)$ and so this vector gives something possibly nonzero only in the i^{th} direction. Also, knowledge of the i^{th} component of the vector is equivalent to knowledge of the vector because it gives the entry in the i^{th} slot and for $\mathbf{v} = (a_1, \cdots, a_n)$,

$$\mathbf{v} = \sum_{k=1}^{n} a_i\mathbf{e}_i.$$

What does addition of vectors mean physically? Suppose two forces are applied to some object. Each of these would be represented by a force vector and the two forces acting together would yield an overall force acting on the object which would also be a force vector known as the resultant. Suppose the two vectors are $\mathbf{a} = \sum_{k=1}^{n} a_i\mathbf{e}_i$ and

$\mathbf{b} = \sum_{k=1}^{n} b_i \mathbf{e}_i$. Then the vector, $\mathbf{a}$ involves a component in the i^{th} direction, $a_i \mathbf{e}_i$ while the component in the i^{th} direction of $\mathbf{b}$ is $b_i \mathbf{e}_i$. Then it seems physically reasonable that the resultant vector should have a component in the i^{th} direction equal to $(a_i + b_i)\,\mathbf{e}_i$. This is exactly what is obtained when the vectors, $\mathbf{a}$ and $\mathbf{b}$ are added.

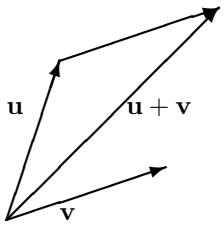
$$\begin{aligned} \mathbf{a} + \mathbf{b} &= (a_1 + b_1, \cdots, a_n + b_n)\,. \\ &= \sum_{i=1}^{n} (a_i + b_i)\,\mathbf{e}_i. \end{aligned}$$

Thus the addition of vectors according to the rules of addition in $\mathbb{R}^n$ which were presented earlier, yields the appropriate vector which duplicates the cumulative effect of all the vectors in the sum.

What is the geometric significance of vector addition? Suppose $\mathbf{u}, \mathbf{v}$ are vectors,

$$\mathbf{u} = (u_1, \cdots, u_n)\,, \mathbf{v} = (v_1, \cdots, v_n)$$

Then $\mathbf{u} + \mathbf{v} = (u_1 + v_1, \cdots, u_n + v_n)$. How can one obtain this geometrically? Consider the directed line segment, $\overrightarrow{\mathbf{0u}}$ and then, starting at the end of this directed line segment, follow the directed line segment $\overrightarrow{\mathbf{u}\,(\mathbf{u}+\mathbf{v})}$ to its end, $\mathbf{u} + \mathbf{v}$. In other words, place the vector $\mathbf{u}$ in standard position with its base at the origin and then slide the vector $\mathbf{v}$ till its base coincides with the point of $\mathbf{u}$. The point of this slid vector, determines $\mathbf{u} + \mathbf{v}$. To illustrate, see the following picture



Note the vector $\mathbf{u} + \mathbf{v}$ is the diagonal of a parallelogram determined from the two vectors $\mathbf{u}$ and $\mathbf{v}$ and that identifying $\mathbf{u} + \mathbf{v}$ with the directed diagonal of the parallelogram determined by the vectors $\mathbf{u}$ and $\mathbf{v}$ amounts to the same thing as the above procedure.

An item of notation should be mentioned here. In the case of $\mathbb{R}^n$ where $n \leq 3$, it is standard notation to use $\mathbf{i}$ for $\mathbf{e}_1, \mathbf{j}$ for $\mathbf{e}_2$, and $\mathbf{k}$ for $\mathbf{e}_3$. Now here are some applications of vector addition to some problems.

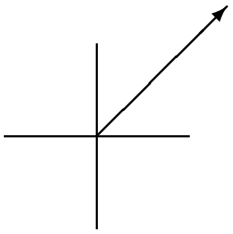
Example 4.2.4 *There are three ropes attached to a car and three people pull on these ropes. The first exerts a force of* $2\mathbf{i}+3\mathbf{j}-2\mathbf{k}$ *Newtons, the second exerts a force of* $3\mathbf{i}+5\mathbf{j}+\mathbf{k}$ *Newtons and the third exerts a force of* $5\mathbf{i}-\mathbf{j}+2\mathbf{k}$. *Newtons. Find the total force in the direction of* $\mathbf{i}$.

To find the total force add the vectors as described above. This gives $10\mathbf{i}+7\mathbf{j}+\mathbf{k}$ Newtons. Therefore, the force in the $\mathbf{i}$ direction is 10 Newtons.

As mentioned earlier, the Newton is a unit of force like pounds.

Example 4.2.5 *An airplane flies North East at 100 miles per hour. Write this as a vector.*

A picture of this situation follows.



The vector has length 100. Now using that vector as the hypotenuse of a right triangle having equal sides, the sides should be each of length $100/\sqrt{2}$. Therefore, the vector would be $100/\sqrt{2}\mathbf{i} + 100/\sqrt{2}\mathbf{j}$.

This example also motivates the concept of **velocity**.

Definition 4.2.6 *The **speed** of an object is a measure of how fast it is going. It is measured in units of length per unit time. For example, miles per hour, kilometers per minute, feet per second. The **velocity** is a vector having the speed as the magnitude but also specifing the direction.*

Thus the velocity vector in the above example is $100/\sqrt{2}\mathbf{i} + 100/\sqrt{2}\mathbf{j}$.

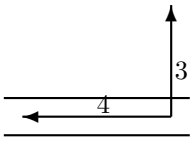
Example 4.2.7 *The velocity of an airplane is* $100\mathbf{i} + \mathbf{j} + \mathbf{k}$ *measured in kilometers per hour and at a certain instant of time its position is* $(1, 2, 1)$. *Here imagine a Cartesian coordinate system in which the third component is altitude and the first and second components are measured on a line from West to East and a line from South to North. Find the position of this airplane one minute later.*

Consider the vector $(1, 2, 1)$, is the initial position vector of the airplane. As it moves, the position vector changes. After one minute the airplane has moved in the $\mathbf{i}$ direction a distance of $100 \times \frac{1}{60} = \frac{5}{3}$ kilometer. In the $\mathbf{j}$ direction it has moved $\frac{1}{60}$ kilometer during this same time, while it moves $\frac{1}{60}$ kilometer in the $\mathbf{k}$ direction. Therefore, the new displacement vector for the airplane is

$$(1, 2, 1) + \left(\frac{5}{3}, \frac{1}{60}, \frac{1}{60}\right) = \left(\frac{8}{3}, \frac{121}{60}, \frac{121}{60}\right)$$

Example 4.2.8 *A certain river is one half mile wide with a current flowing at 4 miles per hour from East to West. A man swims directly toward the opposite shore from the South bank of the river at a speed of 3 miles per hour. How far down the river does he find himself when he has swam across? How far does he end up swimming?*

Consider the following picture.



You should write these vectors in terms of components. The velocity of the swimmer in still water would be $3\mathbf{j}$ while the velocity of the river would be $-4\mathbf{i}$. Therefore, the velocity of the swimmer is $-4\mathbf{i} + 3\mathbf{j}$. Since the component of velocity in the direction across the river is 3, it follows the trip takes 1/6 hour or 10 minutes. The speed at which he travels is $\sqrt{4^2 + 3^2} = 5$ miles per hour and so he travels $5 \times \frac{1}{6} = \frac{5}{6}$ miles. Now to find the distance downstream he finds himself, note that if x is this distance, x and 1/2 are two legs of a right triangle whose hypotenuse equals 5/6 miles. Therefore, by the Pythagorean theorem the distance downstream is $\sqrt{(5/6)^2 - (1/2)^2} = \frac{2}{3}$ miles.

4.3 Exercises

1. The wind blows from West to East at a speed of 50 miles per hour and an airplane which travels at 300 miles per hour in still air is heading North West. What is the velocity of the airplane relative to the ground? What is the component of this velocity in the direction North?

2. In the situation of Problem 1 how many degrees to the West of North should the airplane head in order to fly exactly North. What will be the speed of the airplane relative to the ground?

3. In the situation of 2 suppose the airplane uses 34 gallons of fuel every hour at that air speed and that it needs to fly North a distance of 600 miles. Will the airplane have enough fuel to arrive at its destination given that it has 63 gallons of fuel?

4. An airplane is flying due north at 150 miles per hour. A wind is pushing the airplane due east at 40 miles per hour. After 1 hour, the plane starts flying 30° East of North. Assuming the plane starts at $(0,0)$, where is it after 2 hours? Let North be the direction of the positive y axis and let East be the direction of the positive x axis.

5. City A is located at the origin while city B is located at $(300,500)$ where distances are in miles. An airplane flies at 250 miles per hour in still air. This airplane wants to fly from city A to city B but the wind is blowing in the direction of the positive y axis at a speed of 50 miles per hour. Find a unit vector such that if the plane heads in this direction, it will end up at city B having flown the shortest possible distance. How long will it take to get there?

6. A certain river is one half mile wide with a current flowing at 2 miles per hour from East to West. A man swims directly toward the opposite shore from the South bank of the river at a speed of 3 miles per hour. How far down the river does he find himself when he has swam across? How far does he end up swimming?

7. A certain river is one half mile wide with a current flowing at 2 miles per hour from East to West. A man can swim at 3 miles per hour in still water. In what direction should he swim in order to travel directly across the river? What would the answer to this problem be if the river flowed at 3 miles per hour and the man could swim only at the rate of 2 miles per hour?

8. Three forces are applied to a point which does not move. Two of the forces are $2\mathbf{i}+\mathbf{j}+3\mathbf{k}$ Newtons and $\mathbf{i}-3\mathbf{j}+2\mathbf{k}$ Newtons. Find the third force.

9. The total force acting on an object is to be $2\mathbf{i}+\mathbf{j}+\mathbf{k}$ Newtons. A force of $-\mathbf{i}+\mathbf{j}+\mathbf{k}$ Newtons is being applied. What other force should be applied to achieve the desired total force?

10. A bird flies from its nest 5 km. in the direction 60° north of east where it stops to rest on a tree. It then flies 10 km. in the direction due southeast and lands atop a telephone pole. Place an xy coordinate system so that the origin is the bird's nest, and the positive x axis points east and the positive y axis points north. Find the displacement vector from the nest to the telephone pole.

11. A car is stuck in the mud. There is a cable stretched tightly from this car to a tree which is 20 feet long. A person grasps the cable in the middle and pulls with a force of 100 pounds perpendicular to the stretched cable. The center of the cable moves two feet and remains still. What is the tension in the cable? The tension

in the cable is the force exerted on this point by the part of the cable nearer the car as well as the force exerted on this point by the part of the cable nearer the tree.

12. Let $U = \{(x,y,z) \text{ such that } z > 0\}$. Determine whether U is open, closed or neither.

13. Let $U = \{(x,y,z) \text{ such that } z \geq 0\}$. Determine whether U is open, closed or neither.

14. Let $U = \left\{(x,y,z) \text{ such that } \sqrt{x^2+y^2+z^2} < 1\right\}$. Determine whether U is open, closed or neither.

15. Let $U = \left\{(x,y,z) \text{ such that } \sqrt{x^2+y^2+z^2} \leq 1\right\}$. Determine whether U is open, closed or neither.

16. Show carefully that $\mathbb{R}^n$ is both open and closed.

17. Show that every open set in $\mathbb{R}^n$ is the union of open balls contained in it.

18. Show the intersection of any two open sets is an open set.

19. If S is a nonempty subset of $\mathbb{R}^p$, a point, $\mathbf{x}$ is said to be a **limit point** of S if $B(\mathbf{x},r)$ contains infinitely many points of S for each $r > 0$. Show this is equivalent to saying that $B(\mathbf{x},r)$ contains a point of S different than $\mathbf{x}$ for each $r > 0$.

20. Closed sets were defined to be those sets which are complements of open sets. Show that a set is closed if and only if it contains all its limit points.

4.4 Exercises With Answers

1. The wind blows from West to East at a speed of 30 kilometers per hour and an airplane which travels at 300 Kilometers per hour in still air is heading North West. What is the velocity of the airplane relative to the ground? What is the component of this velocity in the direction North?

   Let the positive y axis point in the direction North and let the positive x axis point in the direction East. The velocity of the wind is $30\mathbf{i}$. The plane moves in the direction $\mathbf{i}+\mathbf{j}$. A unit vector in this direction is $\frac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j})$. Therefore, the velocity of the plane relative to the ground is

$$30\mathbf{i}+\frac{300}{\sqrt{2}}(\mathbf{i}+\mathbf{j}) = 150\sqrt{2}\mathbf{j}+\left(30+150\sqrt{2}\right)\mathbf{i}.$$

   The component of velocity in the direction North is $150\sqrt{2}$.

2. In the situation of Problem 1 how many degrees to the West of North should the airplane head in order to fly exactly North. What will be the speed of the airplane relative to the ground?

   In this case the unit vector will be $-\sin(\theta)\mathbf{i}+\cos(\theta)\mathbf{j}$. Therefore, the velocity of the plane will be

$$300\left(-\sin(\theta)\mathbf{i}+\cos(\theta)\mathbf{j}\right)$$

   and this is supposed to satisfy

$$300\left(-\sin(\theta)\mathbf{i}+\cos(\theta)\mathbf{j}\right)+30\mathbf{i} = 0\mathbf{i}+?\mathbf{j}.$$

Therefore, you need to have $\sin\theta = 1/10$, which means $\theta = .100\,17$ radians. Therefore, the degrees should be $\frac{.1\times 180}{\pi} = 5.729\,6$ degrees. In this case the velocity vector of the plane relative to the ground is $300\left(\frac{\sqrt{99}}{10}\right)\mathbf{j}$.

3. In the situation of 2 suppose the airplane uses 34 gallons of fuel every hour at that air speed and that it needs to fly North a distance of 600 kilometers. Will the airplane have enough fuel to arrive at its destination given that it has 63 gallons of fuel?

   The airplane needs to fly 600 kilometers at a speed of $300\left(\frac{\sqrt{99}}{10}\right)$. Therefore, it takes $\frac{600}{\left(300\left(\frac{\sqrt{99}}{10}\right)\right)} = 2.010\,1$ hours to get there. Therefore, the plane will need to use about 68 gallons of gas. It won't make it.

4. A certain river is one half mile wide with a current flowing at 3 miles per hour from East to West. A man swims directly toward the opposite shore from the South bank of the river at a speed of 2 miles per hour. How far down the river does he find himself when he has swam across? How far does he end up swimming?

   The velocity of the man relative to the earth is then $-3\mathbf{i}+2\mathbf{j}$. Since the component of $\mathbf{j}$ equals 2 it follows he takes 1/8 of an hour to get across. Durring this time he is swept downstream at the rate of 3 miles per hour and so he ends up 3/8 of a mile down stream. He has gone $\sqrt{\left(\frac{3}{8}\right)^2+\left(\frac{1}{2}\right)^2} = .625$ miles in all.

5. Three forces are applied to a point which does not move. Two of the forces are $2\mathbf{i}-\mathbf{j}+3\mathbf{k}$ Newtons and $\mathbf{i}-3\mathbf{j}-2\mathbf{k}$ Newtons**.** Find the third force.

   Call it $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$ Then you need $a+2+1=0, b-1-3=0$, and $c+3-2=0$. Therefore, the force is $-3\mathbf{i}+4\mathbf{j}-\mathbf{k}$.

Vector Products

5.0.1 Outcomes

1. Evaluate a dot product from the angle formula or the coordinate formula.
2. Interpret the dot product geometrically.
3. Evaluate the following using the dot product:
   (a) the angle between two vectors
   (b) the magnitude of a vector
   (c) the work done by a constant force on an object
4. Evaluate a cross product from the angle formula or the coordinate formula.
5. Interpret the cross product geometrically.
6. Evaluate the following using the cross product:
   (a) the area of a parallelogram
   (b) the area of a triangle
   (c) physical quantities such as the torque and angular velocity.
7. Find the volume of a parallelepiped using the box product.
8. Recall, apply and derive the algebraic properties of the dot and cross products.

5.1 The Dot Product

There are two ways of multiplying vectors which are of great importance in applications. The first of these is called the **dot product**, also called the **scalar product** and sometimes the **inner product**.

Definition 5.1.1 *Let* $\mathbf{a}, \mathbf{b}$ *be two vectors in* $\mathbb{R}^n$ *define* $\mathbf{a} \cdot \mathbf{b}$ *as*

$$\mathbf{a} \cdot \mathbf{b} \equiv \sum_{k=1}^{n} a_k b_k.$$

With this definition, there are several important properties satisfied by the dot product. In the statement of these properties, α and β will denote scalars and $\mathbf{a}, \mathbf{b}, \mathbf{c}$ will denote vectors.

Proposition 5.1.2 *The dot product satisfies the following properties.*

$$\mathbf{a}\cdot\mathbf{b}=\mathbf{b}\cdot\mathbf{a} \tag{5.1}$$

$$\mathbf{a}\cdot\mathbf{a}\geq 0 \text{ and equals zero if and only if } \mathbf{a}=\mathbf{0} \tag{5.2}$$

$$(\alpha\mathbf{a}+\beta\mathbf{b})\cdot\mathbf{c}=\alpha(\mathbf{a}\cdot\mathbf{c})+\beta(\mathbf{b}\cdot\mathbf{c}) \tag{5.3}$$

$$\mathbf{c}\cdot(\alpha\mathbf{a}+\beta\mathbf{b})=\alpha(\mathbf{c}\cdot\mathbf{a})+\beta(\mathbf{c}\cdot\mathbf{b}) \tag{5.4}$$

$$|\mathbf{a}|^2=\mathbf{a}\cdot\mathbf{a} \tag{5.5}$$

You should verify these properties. Also be sure you understand that 5.4 follows from the first three and is therefore redundant. It is listed here for the sake of convenience.

Example 5.1.3 *Find* $(1,2,0,-1)\cdot(0,1,2,3)$.

This equals $0+2+0+-3=-1$.

Example 5.1.4 *Find the magnitude of* $\mathbf{a}=(2,1,4,2)$. *That is, find* $|\mathbf{a}|$.

This is $\sqrt{(2,1,4,2)\cdot(2,1,4,2)}=5$.

The dot product satisfies a fundamental inequality known as the **Cauchy Schwarz inequality.**

Theorem 5.1.5 *The dot product satisfies the inequality*

$$|\mathbf{a}\cdot\mathbf{b}|\leq|\mathbf{a}||\mathbf{b}|. \tag{5.6}$$

Furthermore equality is obtained if and only if one of $\mathbf{a}$ *or* $\mathbf{b}$ *is a scalar multiple of the other.*

Proof: First note that if $\mathbf{b}=\mathbf{0}$ both sides of 5.6 equal zero and so the inequality holds in this case. Therefore, it will be assumed in what follows that $\mathbf{b}\neq\mathbf{0}$.

Define a function of $t\in\mathbb{R}$

$$f(t)=(\mathbf{a}+t\mathbf{b})\cdot(\mathbf{a}+t\mathbf{b}).$$

Then by 5.2, $f(t)\geq 0$ for all $t\in\mathbb{R}$. Also from 5.3,5.4,5.1, and 5.5

$$\begin{aligned}f(t)&=\mathbf{a}\cdot(\mathbf{a}+t\mathbf{b})+t\mathbf{b}\cdot(\mathbf{a}+t\mathbf{b})\\&=\mathbf{a}\cdot\mathbf{a}+t(\mathbf{a}\cdot\mathbf{b})+t\mathbf{b}\cdot\mathbf{a}+t^2\mathbf{b}\cdot\mathbf{b}\\&=|\mathbf{a}|^2+2t(\mathbf{a}\cdot\mathbf{b})+|\mathbf{b}|^2t^2.\end{aligned}$$

Now

$$\begin{aligned}f(t)&=|\mathbf{b}|^2\left(t^2+2t\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}+\frac{|\mathbf{a}|^2}{|\mathbf{b}|^2}\right)\\&=|\mathbf{b}|^2\left(t^2+2t\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}+\left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\right)^2-\left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\right)^2+\frac{|\mathbf{a}|^2}{|\mathbf{b}|^2}\right)\\&=|\mathbf{b}|^2\left(\left(t+\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\right)^2+\left(\frac{|\mathbf{a}|^2}{|\mathbf{b}|^2}-\left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\right)^2\right)\right)\geq 0\end{aligned}$$

for all $t \in \mathbb{R}$. In particular $f(t) \geq 0$ when $t = -\left(\mathbf{a}\cdot\mathbf{b}/|\mathbf{b}|^2\right)$ which implies

$$\frac{|\mathbf{a}|^2}{|\mathbf{b}|^2} - \left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\right)^2 \geq 0. \tag{5.7}$$

Multiplying both sides by $|\mathbf{b}|^4$,

$$|\mathbf{a}|^2 |\mathbf{b}|^2 \geq (\mathbf{a}\cdot\mathbf{b})^2$$

which yields 5.6.

From Theorem 5.1.5, equality holds in 5.6 whenever one of the vectors is a scalar multiple of the other. It only remains to verify this is the only way equality can occur. If either vector equals zero, then equality is obtained in 5.6 so it can be assumed both vectors are non zero and that equality is obtained in 5.7. This implies that $f(t) = 0$ when $t = -\left(\mathbf{a}\cdot\mathbf{b}/|\mathbf{b}|^2\right)$ and so from 5.2, it follows that for this value of t, $\mathbf{a}+t\mathbf{b} = \mathbf{0}$ showing $\mathbf{a} = -t\mathbf{b}$. This proves the theorem.

You should note that the entire argument was based only on the properties of the dot product listed in 5.1 - 5.5. This means that whenever something satisfies these properties, the Cauchy Schwartz inequality holds. There are many other instances of these properties besides vectors in $\mathbb{R}^n$.

The Cauchy Schwartz inequality allows a proof of the **triangle inequality** for distances in $\mathbb{R}^n$ in much the same way as the triangle inequality for the absolute value.

Theorem 5.1.6 *(Triangle inequality) For* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$

$$|\mathbf{a}+\mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}| \tag{5.8}$$

and equality holds if and only if one of the vectors is a nonnegative scalar multiple of the other. Also

$$||\mathbf{a}| - |\mathbf{b}|| \leq |\mathbf{a}-\mathbf{b}| \tag{5.9}$$

Proof: By properties of the dot product and the Cauchy Schwartz inequality,

$$\begin{aligned}
|\mathbf{a}+\mathbf{b}|^2 &= (\mathbf{a}+\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) \\
&= (\mathbf{a}\cdot\mathbf{a}) + (\mathbf{a}\cdot\mathbf{b}) + (\mathbf{b}\cdot\mathbf{a}) + (\mathbf{b}\cdot\mathbf{b}) \\
&= |\mathbf{a}|^2 + 2(\mathbf{a}\cdot\mathbf{b}) + |\mathbf{b}|^2 \\
&\leq |\mathbf{a}|^2 + 2|\mathbf{a}\cdot\mathbf{b}| + |\mathbf{b}|^2 \\
&\leq |\mathbf{a}|^2 + 2|\mathbf{a}||\mathbf{b}| + |\mathbf{b}|^2 \\
&= (|\mathbf{a}| + |\mathbf{b}|)^2.
\end{aligned}$$

Taking square roots of both sides you obtain 5.8.

It remains to consider when equality occurs. If either vector equals zero, then that vector equals zero times the other vector and the claim about when equality occurs is verified. Therefore, it can be assumed both vectors are nonzero. To get equality in the second inequality above, Theorem 5.1.5 implies one of the vectors must be a multiple of the other. Say $\mathbf{b} = \alpha\mathbf{a}$. If $\alpha < 0$ then equality cannot occur in the first inequality because in this case

$$(\mathbf{a}\cdot\mathbf{b}) = \alpha|\mathbf{a}|^2 < 0 < |\alpha||\mathbf{a}|^2 = |\mathbf{a}\cdot\mathbf{b}|$$

Therefore, $\alpha \geq 0$.

To get the other form of the triangle inequality,

$$\mathbf{a} = \mathbf{a} - \mathbf{b} + \mathbf{b}$$

so

$$\begin{aligned} |\mathbf{a}| &= |\mathbf{a} - \mathbf{b} + \mathbf{b}| \\ &\leq |\mathbf{a} - \mathbf{b}| + |\mathbf{b}| . \end{aligned}$$

Therefore,

$$|\mathbf{a}| - |\mathbf{b}| \leq |\mathbf{a} - \mathbf{b}| \tag{5.10}$$

Similarly,

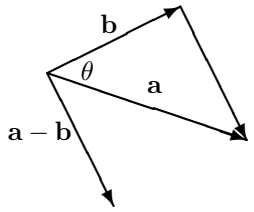
$$|\mathbf{b}| - |\mathbf{a}| \leq |\mathbf{b} - \mathbf{a}| = |\mathbf{a} - \mathbf{b}| . \tag{5.11}$$

It follows from 5.10 and 5.11 that 5.9 holds. This is because $||\mathbf{a}| - |\mathbf{b}||$ equals the left side of either 5.10 or 5.11 and either way, $||\mathbf{a}| - |\mathbf{b}|| \leq |\mathbf{a} - \mathbf{b}|$. This proves the theorem.

5.2 The Geometric Significance Of The Dot Product

5.2.1 The Angle Between Two Vectors

Given two vectors, $\mathbf{a}$ and $\mathbf{b}$, the included angle is the angle between these two vectors which is less than or equal to 180 degrees. The dot product can be used to determine the included angle between two vectors. To see how to do this, consider the following picture.



By the law of cosines,

$$|\mathbf{a} - \mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2 |\mathbf{a}| |\mathbf{b}| \cos \theta.$$

Also from the properties of the dot product,

$$\begin{aligned} |\mathbf{a} - \mathbf{b}|^2 &= (\mathbf{a} - \mathbf{b}) \cdot (\mathbf{a} - \mathbf{b}) \\ &= |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2\mathbf{a} \cdot \mathbf{b} \end{aligned}$$

and so comparing the above two formulas,

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta. \tag{5.12}$$

In words, the dot product of two vectors equals the product of the magnitude of the two vectors multiplied by the cosine of the included angle. Note this gives a geometric description of the dot product which does not depend explicitly on the coordinates of the vectors.

Example 5.2.1 *Find the angle between the vectors* $2\mathbf{i} + \mathbf{j} - \mathbf{k}$ *and* $3\mathbf{i} + 4\mathbf{j} + \mathbf{k}$.

The dot product of these two vectors equals $6+4-1=9$ and the norms are $\sqrt{4+1+1}=\sqrt{6}$ and $\sqrt{9+16+1}=\sqrt{26}$. Therefore, from 5.12 the cosine of the included angle equals

$$\cos\theta = \frac{9}{\sqrt{26}\sqrt{6}} = .720\,58$$

Now the cosine is known, the angle can be determines by solving the equation, $\cos\theta = .720\,58$. This will involve using a calculator or a table of trigonometric functions. The answer is $\theta = .766\,16$ radians or in terms of degrees, $\theta = .766\,16 \times \frac{360}{2\pi} = 43.898°$. Recall how this last computation is done. Set up a proportion, $\frac{x}{.766\,16} = \frac{360}{2\pi}$ because $360°$ corresponds to 2π radians. However, in calculus, you should get used to thinking in terms of radians and not degrees. This is because all the important calculus formulas are defined in terms of radians.

Example 5.2.2 *Let* $\mathbf{u}, \mathbf{v}$ *be two vectors whose magnitudes are equal to 3 and 4 respectively and such that if they are placed in standard position with their tails at the origin, the angle between* $\mathbf{u}$ *and the positive* x *axis equals 30° and the angle between* $\mathbf{v}$ *and the positive* x *axis is -30°. Find* $\mathbf{u}\cdot\mathbf{v}$.

From the geometric description of the dot product in 5.12

$$\mathbf{u}\cdot\mathbf{v} = 3\times 4\times\cos(60°) = 3\times 4\times 1/2 = 6.$$

Observation 5.2.3 *Two vectors are said to be* ***perpendicular*** *if the included angle is* $\pi/2$ *radians (90°). You can tell if two nonzero vectors are perpendicular by simply taking their dot product. If the answer is zero, this means they are are perpendicular because* $\cos\theta = 0$.

Example 5.2.4 *Determine whether the two vectors,* $2\mathbf{i}+\mathbf{j}-\mathbf{k}$ *and* $1\mathbf{i}+3\mathbf{j}+5\mathbf{k}$ *are perpendicular.*

When you take this dot product you get $2+3-5=0$ and so these two are indeed perpendicular.

Definition 5.2.5 *When two lines intersect, the angle between the two lines is the smaller of the two angles determined.*

Example 5.2.6 *Find the angle between the two lines,* $(1,2,0)+t(1,2,3)$ *and* $(0,4,-3)+t(-1,2,-3)$.

These two lines intersect, when $t=0$ in the first and $t=-1$ in the second. It is only a matter of finding the angle between the direction vectors. One angle determined is given by

$$\cos\theta = \frac{-6}{14} = \frac{-3}{7}. \tag{5.13}$$

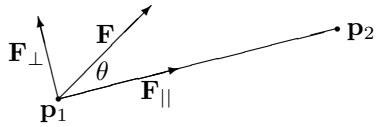
We don't want this angle because it is obtuse. The angle desired is the acute angle given by

$$\cos\theta = \frac{3}{7}.$$

It is obtained by using replacing one of the direction vectors with -1 times it.

5.2.2 Work And Projections

Our first application will be to the concept of work. The physical concept of work does not in any way correspond to the notion of work employed in ordinary conversation. For example, if you were to slide a 150 pound weight off a table which is three feet high and shuffle along the floor for 50 yards, sweating profusely and exerting all your strength to keep the weight from falling on your feet, keeping the height always three feet and then deposit this weight on another three foot high table, the physical concept of work would indicate that the force exerted by your arms did no work during this project even though the muscles in your hands and arms would likely be very tired. The reason for such an unusual definition is that even though your arms exerted considerable force on the weight, enough to keep it from falling, the direction of motion was at right angles to the force they exerted. The only part of a force which does work in the sense of physics is the component of the force in the direction of motion (This is made more precise below.). The work is defined to be the magnitude of the component of this force times the distance over which it acts in the case where this component of force points in the direction of motion and (-1) times the magnitude of this component times the distance in case the force tends to impede the motion. Thus the work done by a force on an object as the object moves from one point to another is a measure of the extent to which the force contributes to the motion. This is illustrated in the following picture in the case where the given force contributes to the motion.



In this picture the force, $\mathbf{F}$ is applied to an object which moves on the straight line from $\mathbf{p}_1$ to $\mathbf{p}_2$. There are two vectors shown, $\mathbf{F}_{||}$ and $\mathbf{F}_{\perp}$ and the picture is intended to indicate that when you add these two vectors you get $\mathbf{F}$ while $\mathbf{F}_{||}$ acts in the direction of motion and $\mathbf{F}_{\perp}$ acts perpendicular to the direction of motion. Only $\mathbf{F}_{||}$ contributes to the work done by $\mathbf{F}$ on the object as it moves from $\mathbf{p}_1$ to $\mathbf{p}_2$. $\mathbf{F}_{||}$ is called the **component of the force** in the direction of motion. From trigonometry, you see the magnitude of $\mathbf{F}_{||}$ should equal $|\mathbf{F}||\cos\theta|$. Thus, since $\mathbf{F}_{||}$ points in the direction of the vector from $\mathbf{p}_1$ to $\mathbf{p}_2$, the total work done should equal

$$|\mathbf{F}|\left|\overrightarrow{\mathbf{p}_1\mathbf{p}_2}\right|\cos\theta = |\mathbf{F}||\mathbf{p}_2-\mathbf{p}_1|\cos\theta$$

If the included angle had been obtuse, then the work done by the force, $\mathbf{F}$ on the object would have been negative because in this case, the force tends to impede the motion from $\mathbf{p}_1$ to $\mathbf{p}_2$ but in this case, $\cos\theta$ would also be negative and so it is still the case that the work done would be given by the above formula. Thus from the geometric description of the dot product given above, the work equals

$$|\mathbf{F}||\mathbf{p}_2-\mathbf{p}_1|\cos\theta = \mathbf{F}\cdot(\mathbf{p}_2-\mathbf{p}_1).$$

This explains the following definition.

Definition 5.2.7 *Let $\mathbf{F}$ be a force acting on an object which moves from the point, $\mathbf{p}_1$ to the point $\mathbf{p}_2$. Then the **work** done on the object by the given force equals $\mathbf{F}\cdot(\mathbf{p}_2-\mathbf{p}_1)$.*

The concept of writing a given vector, $\mathbf{F}$ in terms of two vectors, one which is parallel to a given vector, $\mathbf{D}$ and the other which is perpendicular can also be explained with no reliance on trigonometry, completely in terms of the algebraic properties of the dot

product. As before, this is mathematically more significant than any approach involving geometry or trigonometry because it extends to more interesting situations. This is done next.

Theorem 5.2.8 *Let* $\mathbf{F}$ *and* $\mathbf{D}$ *be nonzero vectors. Then there exist unique vectors* $\mathbf{F}_{||}$ *and* $\mathbf{F}_{\perp}$ *such that*

$$\mathbf{F} = \mathbf{F}_{||} + \mathbf{F}_{\perp} \tag{5.14}$$

where $\mathbf{F}_{||}$ *is a scalar multiple of* $\mathbf{D}$*, also referred to as*

$$\mathrm{proj}_{\mathbf{D}}(\mathbf{F}),$$

and $\mathbf{F}_{\perp} \cdot \mathbf{D} = 0$*. The vector* $\mathrm{proj}_{\mathbf{D}}(\mathbf{F})$ *is called the* ***projection*** *of* $\mathbf{F}$ *onto* $\mathbf{D}$*.*

Proof: Suppose 5.14 and $\mathbf{F}_{||} = \alpha\mathbf{D}$. Taking the dot product of both sides with $\mathbf{D}$ and using $\mathbf{F}_{\perp} \cdot \mathbf{D} = 0$, this yields

$$\mathbf{F} \cdot \mathbf{D} = \alpha |\mathbf{D}|^2$$

which requires $\alpha = \mathbf{F} \cdot \mathbf{D} / |\mathbf{D}|^2$. Thus there can be no more than one vector, $\mathbf{F}_{||}$. It follows $\mathbf{F}_{\perp}$ must equal $\mathbf{F} - \mathbf{F}_{||}$. This verifies there can be no more than one choice for both $\mathbf{F}_{||}$ and $\mathbf{F}_{\perp}$.

Now let

$$\mathbf{F}_{||} \equiv \frac{\mathbf{F} \cdot \mathbf{D}}{|\mathbf{D}|^2}\mathbf{D}$$

and let

$$\mathbf{F}_{\perp} = \mathbf{F} - \mathbf{F}_{||} = \mathbf{F} - \frac{\mathbf{F} \cdot \mathbf{D}}{|\mathbf{D}|^2}\mathbf{D}$$

Then $\mathbf{F}_{||} = \alpha\, \mathbf{D}$ where $\alpha = \frac{\mathbf{F} \cdot \mathbf{D}}{|\mathbf{D}|^2}$. It only remains to verify $\mathbf{F}_{\perp} \cdot \mathbf{D} = 0$. But

$$\begin{aligned} \mathbf{F}_{\perp} \cdot \mathbf{D} &= \mathbf{F} \cdot \mathbf{D} - \frac{\mathbf{F} \cdot \mathbf{D}}{|\mathbf{D}|^2}\mathbf{D} \cdot \mathbf{D} \\ &= \mathbf{F} \cdot \mathbf{D} - \mathbf{F} \cdot \mathbf{D} = 0. \end{aligned}$$

This proves the theorem.

Example 5.2.9 *Let* $\mathbf{F} = 2\mathbf{i} + 7\mathbf{j} - 3\mathbf{k}$ *Newtons. Find the work done by this force in moving from the point* $(1, 2, 3)$ *to the point* $(-9, -3, 4)$ *along the straight line segment joining these points where distances are measured in meters.*

According to the definition, this work is

$$\begin{aligned} (2\mathbf{i} + 7\mathbf{j} - 3\mathbf{k}) \cdot (-10\mathbf{i} - 5\mathbf{j} + \mathbf{k}) &= -20 + (-35) + (-3) \\ &= -58 \text{ Newton meters.} \end{aligned}$$

Note that if the force had been given in pounds and the distance had been given in feet, the units on the work would have been foot pounds. In general, work has units equal to units of a force times units of a length. Instead of writing Newton meter, people write joule because a joule is by definition a Newton meter. That word is pronounced "jewel" and it is the unit of work in the metric system of units. Also be sure you observe that the work done by the force can be negative as in the above example. In fact, work can be either positive, negative, or zero. You just have to do the computations to find out.

Example 5.2.10 *Find* $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ *if* $\mathbf{u} = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$ *and* $\mathbf{v} = \mathbf{i} - 2\mathbf{j} + \mathbf{k}$.

From the above discussion in Theorem 5.2.8, this is just

$$\begin{aligned} & \frac{1}{4+9+16}(\mathbf{i} - 2\mathbf{j} + \mathbf{k}) \cdot (2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k})(2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}) \\ = & \frac{-8}{29}(2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}) = -\frac{16}{29}\mathbf{i} - \frac{24}{29}\mathbf{j} + \frac{32}{29}\mathbf{k}. \end{aligned}$$

Example 5.2.11 *Suppose* $\mathbf{a}$, *and* $\mathbf{b}$ *are vectors and* $\mathbf{b}_\perp = \mathbf{b} - \mathrm{proj}_{\mathbf{a}}(\mathbf{b})$. *What is the magnitude of* $\mathbf{b}_\perp$ *in terms of the included angle*?

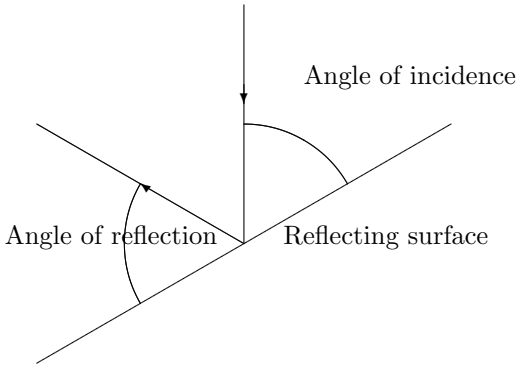
$$\begin{aligned} |\mathbf{b}_\perp|^2 &= (\mathbf{b} - \mathrm{proj}_{\mathbf{a}}(\mathbf{b})) \cdot (\mathbf{b} - \mathrm{proj}_{\mathbf{a}}(\mathbf{b})) \\ &= \left(\mathbf{b} - \frac{\mathbf{b}\cdot\mathbf{a}}{|\mathbf{a}|^2}\mathbf{a}\right) \cdot \left(\mathbf{b} - \frac{\mathbf{b}\cdot\mathbf{a}}{|\mathbf{a}|^2}\mathbf{a}\right) \\ &= |\mathbf{b}|^2 - 2\frac{(\mathbf{b}\cdot\mathbf{a})^2}{|\mathbf{a}|^2} + \left(\frac{\mathbf{b}\cdot\mathbf{a}}{|\mathbf{a}|^2}\right)^2 |\mathbf{a}|^2 \\ &= |\mathbf{b}|^2 \left(1 - \frac{(\mathbf{b}\cdot\mathbf{a})^2}{|\mathbf{a}|^2 |\mathbf{b}|^2}\right) \\ &= |\mathbf{b}|^2 \left(1 - \cos^2\theta\right) = |\mathbf{b}|^2 \sin^2(\theta) \end{aligned}$$

where θ is the included angle between $\mathbf{a}$ and $\mathbf{b}$ which is less than π radians. Therefore, taking square roots,

$$|\mathbf{b}_\perp| = |\mathbf{b}| \sin\theta.$$

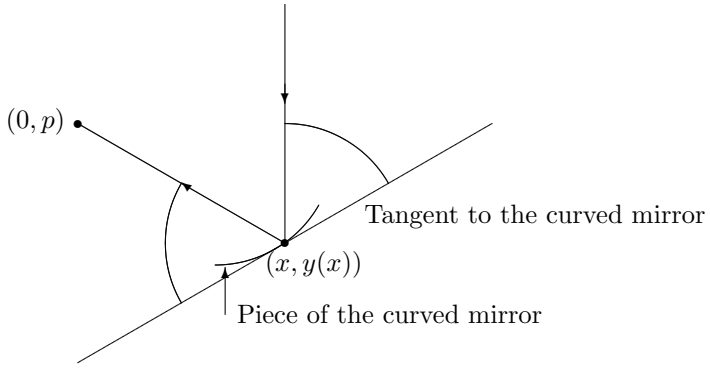
5.2.3 The Parabolic Mirror, An Application

When light is reflected the angle of incidence is always equal to the angle of reflection. This is illustrated in the following picture in which a ray of light reflects off something like a mirror.



An interesting problem is to design a curved mirror which has the property that it will direct all rays of light coming from a long distance away (essentially parallel rays of light) to a single point. You might be interested in a reflecting telescope for example or some sort of scheme for achieving high temperatures by reflecting the rays of the sun to a small area. Turning things around, you could place a source of light at the single point and desire to have the mirror reflect this in a beam of light consisting of parallel rays. How can you design such a mirror?

It turns out this is pretty easy given the above techniques for finding the angle between vectors. Consider the following picture.



It suffices to consider this in a plane for $x > 0$ and then let the mirror be obtained as a surface of revolution. In the above picture, let $(0, p)$ be the special point at which all the parallel rays of light will be directed. This is set up so the rays of light are parallel to the y axis. The two indicated angles will be equal and the equation of the indicated curve will be $y = y(x)$ while the reflection is taking place at the point $(x, y(x))$ as shown. To say the two angles are equal is to say their cosines are equal. Thus from the above,

$$\frac{(0,1)\cdot(1,y'(x))}{\sqrt{1+y'(x)^2}} = \frac{(-x,p-y)\cdot(-1,-y'(x))}{\sqrt{x^2+(y-p)^2}\sqrt{1+y'(x)^2}}.$$

This follows because the vectors forming the sides of one of the angles are $(0,1)$ and $(1,y'(x))$ while the vectors forming the other angle are $(-x,p-y)$ and $(-1,-y'(x))$. Therefore, this yields the differential equation,

$$y'(x) = \frac{-y'(x)(p-y)+x}{\sqrt{x^2+(y-p)^2}}$$

which is written more simply as

$$\left(\sqrt{x^2+(y-p)^2}+(p-y)\right)y' = x$$

Now let $y - p = xv$ so that $y' = xv' + v$. Then in terms of v the differential equation is

$$xv' = \frac{1}{\sqrt{1+v^2}-v} - v.$$

This reduces to

$$\left(\frac{1}{\sqrt{1+v^2}-v} - v\right)\frac{dv}{dx} = \frac{1}{x}.$$

If $G \in \int \left(\frac{1}{\sqrt{1+v^2}-v} - v\right) dv$, then a solution to the differential equation is of the form

$$G(v) - \ln x = C$$

where C is a constant. This is because if you differentiate both sides with respect to x,

$$G'(v)\frac{dv}{dx} - \frac{1}{x} = \left(\frac{1}{\sqrt{1+v^2}-v} - v\right)\frac{dv}{dx} - \frac{1}{x} = 0.$$

To find $G \in \int \left(\frac{1}{\sqrt{1+v^2}-v} - v \right) dv$, use a trig. substitution, $v = \tan\theta$. Then in terms of θ, the antiderivative becomes

$$\begin{aligned} \int \left(\frac{1}{\sec\theta - \tan\theta} - \tan\theta \right) \sec^2\theta \, d\theta &= \int \sec\theta \, d\theta \\ &= \ln|\sec\theta + \tan\theta| + C. \end{aligned}$$

Now in terms of v, this is

$$\ln\left(v + \sqrt{1+v^2} \right) = \ln x + c.$$

There is no loss of generality in letting $c = \ln C$ because ln maps onto $\mathbb{R}$. Therefore, from laws of logarithms,

$$\begin{aligned} \ln\left| v + \sqrt{1+v^2} \right| &= \ln x + c = \ln x + \ln C \\ &= \ln Cx. \end{aligned}$$

Therefore,

$$v + \sqrt{1+v^2} = Cx$$

and so

$$\sqrt{1+v^2} = Cx - v.$$

Now square both sides to get

$$1 + v^2 = C^2x^2 + v^2 - 2Cxv$$

which shows

$$1 = C^2x^2 - 2Cx\frac{y-p}{x} = C^2x^2 - 2C\left(y - p\right).$$

Solving this for y yields

$$y = \frac{C}{2}x^2 + \left(p - \frac{1}{2C} \right)$$

and for this to correspond to reflection as described above, it must be that $C > 0$. As described in an earlier section, this is just the equation of a parabola. Note it is possible to choose C as desired adjusting the shape of the mirror.

5.2.4 The Dot Product And Distance In $\mathbb{C}^n$

It is necessary to give a generalization of the dot product for vectors in $\mathbb{C}^n$. This definition reduces to the usual one in the case the components of the vector are real.

Definition 5.2.12 *Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. Thus $\mathbf{x} = (x_1, \cdots, x_n)$ where each $x_k \in \mathbb{C}$ and a similar formula holding for $\mathbf{y}$. Then the dot product of these two vectors is defined to be*

$$\mathbf{x} \cdot \mathbf{y} \equiv \sum_j x_j \overline{y_j} \equiv x_1\overline{y_1} + \cdots + x_n\overline{y_n}.$$

Notice how you put the conjugate on the entries of the vector, $\mathbf{y}$. It makes no difference if the vectors happen to be real vectors but with complex vectors you must do it this way. The reason for this is that when you take the dot product of a vector with itself, you want to get the square of the length of the vector, a positive number.

Placing the conjugate on the components of $\mathbf{y}$ in the above definition assures this will take place. Thus

$$\mathbf{x}\cdot\mathbf{x}=\sum_j x_j\overline{x_j}=\sum_j |x_j|^2\geq 0.$$

If you didn't place a conjugate as in the above definition, things wouldn't work out correctly. For example,

$$(1+i)^2+2^2=4+2i$$

and this is not a positive number.

The following properties of the dot product follow immediately from the definition and you should verify each of them.

Properties of the dot product:

1. $\mathbf{u}\cdot\mathbf{v}=\overline{\mathbf{v}\cdot\mathbf{u}}$.
2. If a,b are numbers and $\mathbf{u},\mathbf{v},\mathbf{z}$ are vectors then $(a\mathbf{u}+b\mathbf{v})\cdot\mathbf{z}=a(\mathbf{u}\cdot\mathbf{z})+b(\mathbf{v}\cdot\mathbf{z})$.
3. $\mathbf{u}\cdot\mathbf{u}\geq 0$ and it equals 0 if and only if $\mathbf{u}=\mathbf{0}$.

The norm is defined in the usual way.

Definition 5.2.13 *For* $\mathbf{x}\in\mathbb{C}^n$,

$$|\mathbf{x}|\equiv\left(\sum_{k=1}^n |x_k|^2\right)^{1/2}=(\mathbf{x}\cdot\mathbf{x})^{1/2}$$

Here is a fundamental inequality called the **Cauchy Schwarz inequality** which is stated here in $\mathbb{C}^n$. First here is a simple lemma.

Lemma 5.2.14 *If* $z\in\mathbb{C}$ *there exists* $\theta\in\mathbb{C}$ *such that* $\theta z=|z|$ *and* $|\theta|=1$.

Proof: Let $\theta=1$ if $z=0$ and otherwise, let $\theta=\dfrac{\overline{z}}{|z|}$. Recall that for $z=x+iy,\overline{z}=x-iy$ and $\overline{z}z=|z|^2$.

Theorem 5.2.15 *(Cauchy Schwarz)The following inequality holds for* x_i *and* $y_i\in\mathbb{C}$.

$$|(\mathbf{x}\cdot\mathbf{y})|=\left|\sum_{i=1}^n x_i\overline{y}_i\right|\leq\left(\sum_{i=1}^n |x_i|^2\right)^{1/2}\left(\sum_{i=1}^n |y_i|^2\right)^{1/2}=|\mathbf{x}|\,|\mathbf{y}| \tag{5.15}$$

Proof: Let $\theta\in\mathbb{C}$ such that $|\theta|=1$ and

$$\theta\sum_{i=1}^n x_i\overline{y}_i=\left|\sum_{i=1}^n x_i\overline{y}_i\right|$$

Thus

$$\theta\sum_{i=1}^n x_i\overline{y}_i=\sum_{i=1}^n x_i\overline{(\theta y_i)}=\left|\sum_{i=1}^n x_i\overline{y}_i\right|.$$

Consider $p(t)\equiv\sum_{i=1}^n\left(x_i+t\overline{\theta}y_i\right)\left(\overline{x_i+t\overline{\theta}y_i}\right)$ where $t\in\mathbb{R}$.

$$\begin{aligned}0 &\leq p(t)=\sum_{i=1}^n |x_i|^2+2t\,\mathrm{Re}\left(\theta\sum_{i=1}^n x_i\overline{y}_i\right)+t^2\sum_{i=1}^n |y_i|^2\\ &= |\mathbf{x}|^2+2t\left|\sum_{i=1}^n x_i\overline{y}_i\right|+t^2|\mathbf{y}|^2\end{aligned}$$

If $|\mathbf{y}| = 0$ then 5.15 is obviously true because both sides equal zero. Therefore, assume $|\mathbf{y}| \neq 0$ and then $p(t)$ is a polynomial of degree two whose graph opens up. Therefore, it either has no zeroes, two zeros or one repeated zero. If it has two zeros, the above inequality must be violated because in this case the graph must dip below the x axis. Therefore, it either has no zeros or exactly one. From the quadratic formula this happens exactly when

$$4\left|\sum_{i=1}^{n} x_i\overline{y}_i\right|^2 - 4\left|\mathbf{x}\right|^2\left|\mathbf{y}\right|^2 \leq 0$$

and so

$$\left|\sum_{i=1}^{n} x_i\overline{y}_i\right| \leq \left|\mathbf{x}\right|\left|\mathbf{y}\right|$$

as claimed. This proves the inequality.

By analogy to the case of $\mathbb{R}^n$, length or magnitude of vectors in $\mathbb{C}^n$ can be defined.

Definition 5.2.16 *Let $\mathbf{z} \in \mathbb{C}^n$. Then $|\mathbf{z}| \equiv (\mathbf{z}\cdot\mathbf{z})^{1/2}$.*

Theorem 5.2.17 *For length defined in Definition 5.2.16, the following hold.*

$$|\mathbf{z}| \geq 0 \text{ and } |\mathbf{z}| = 0 \text{ if and only if } \mathbf{z} = \mathbf{0} \tag{5.16}$$

$$\text{If } \alpha \text{ is a scalar, } |\alpha\mathbf{z}| = |\alpha|\,|\mathbf{z}| \tag{5.17}$$

$$|\mathbf{z}+\mathbf{w}| \leq |\mathbf{z}| + |\mathbf{w}|\,. \tag{5.18}$$

Proof: The first two claims are left as exercises. To establish the third, you use the same argument which was used in $\mathbb{R}^n$.

$$\begin{aligned}
|\mathbf{z}+\mathbf{w}|^2 &= (\mathbf{z}+\mathbf{w}, \mathbf{z}+\mathbf{w}) \\
&= \mathbf{z}\cdot\mathbf{z} + \mathbf{w}\cdot\mathbf{w} + \mathbf{w}\cdot\mathbf{z} + \mathbf{z}\cdot\mathbf{w} \\
&= |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2\,\mathrm{Re}\,\mathbf{w}\cdot\mathbf{z} \\
&\leq |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2\,|\mathbf{w}\cdot\mathbf{z}| \\
&\leq |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2\,|\mathbf{w}|\,|\mathbf{z}| = (|\mathbf{z}| + |\mathbf{w}|)^2\,.
\end{aligned}$$

All other considerations such as open and closed sets and the like are identical in this more general context with the corresponding definition in $\mathbb{R}^n$. The main difference is that here the scalars are complex numbers.

Definition 5.2.18 *Suppose you have a vector space, V and for $\mathbf{z}, \mathbf{w} \in V$ and α a scalar a norm is a way of measuring distance or magnitude which satisfies the properties 5.16 - 5.18. Thus a norm is something which does the following.*

$$||\mathbf{z}|| \geq 0 \text{ and } ||\mathbf{z}|| = 0 \text{ if and only if } \mathbf{z} = \mathbf{0} \tag{5.19}$$

$$\text{If } \alpha \text{ is a scalar, } ||\alpha\mathbf{z}|| = |\alpha|\,||\mathbf{z}|| \tag{5.20}$$

$$||\mathbf{z}+\mathbf{w}|| \leq ||\mathbf{z}|| + ||\mathbf{w}||\,. \tag{5.21}$$

Here is is understood that for all $\mathbf{z} \in V, ||\mathbf{z}|| \in [0,\infty)$.

5.3 Exercises

1. Use formula 5.12 to verify the Cauchy Schwartz inequality and to show that equality occurs if and only if one of the vectors is a scalar multiple of the other.

2. For $\mathbf{u}, \mathbf{v}$ vectors in $\mathbb{R}^3$, define the product, $\mathbf{u} * \mathbf{v} \equiv u_1 v_1 + 2u_2 v_2 + 3u_3 v_3$. Show the axioms for a dot product all hold for this funny product. Prove

$$|\mathbf{u} * \mathbf{v}| \leq (\mathbf{u} * \mathbf{u})^{1/2} (\mathbf{v} * \mathbf{v})^{1/2}.$$

   Hint: Do not try to do this with methods from trigonometry.

3. Find the angle between the vectors $3\mathbf{i} - \mathbf{j} - \mathbf{k}$ and $\mathbf{i} + 4\mathbf{j} + 2\mathbf{k}$.

4. Find the angle between the vectors $\mathbf{i} - 2\mathbf{j} + \mathbf{k}$ and $\mathbf{i} + 2\mathbf{j} - 7\mathbf{k}$.

5. Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{v} = (1, 0, -2)$ and $\mathbf{u} = (1, 2, 3)$.

6. Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{v} = (1, 2, -2)$ and $\mathbf{u} = (1, 0, 3)$.

7. Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{v} = (1, 2, -2, 1)$ and $\mathbf{u} = (1, 2, 3, 0)$.

8. Does it make sense to speak of $\text{proj}_{\mathbf{0}}(\mathbf{v})$?

9. Prove that $T\mathbf{v} \equiv \text{proj}_{\mathbf{u}}(\mathbf{v})$ is a linear transformation and find the matrix of T where $T\mathbf{v} = \text{proj}_{\mathbf{u}}(\mathbf{v})$ for $\mathbf{u} = (1, 2, 3)$.

10. If $\mathbf{F}$ is a force and $\mathbf{D}$ is a vector, show $\text{proj}_{\mathbf{D}}(\mathbf{F}) = (|\mathbf{F}| \cos\theta)\,\mathbf{u}$ where $\mathbf{u}$ is the unit vector in the direction of $\mathbf{D}$, $\mathbf{u} = \mathbf{D}/|\mathbf{D}|$ and θ is the included angle between the two vectors, $\mathbf{F}$ and $\mathbf{D}$. $|\mathbf{F}| \cos\theta$ is sometimes called the component of the force, $\mathbf{F}$ in the direction, $\mathbf{D}$.

11. A boy drags a sled for 100 feet along the ground by pulling on a rope which is 20 degrees from the horizontal with a force of 10 pounds. How much work does this force do?

12. A boy drags a sled for 200 feet along the ground by pulling on a rope which is 30 degrees from the horizontal with a force of 20 pounds. How much work does this force do?

13. How much work in Newton meters does it take to slide a crate 20 meters along a loading dock by pulling on it with a 200 Newton force at an angle of 30° from the horizontal?

14. An object moves 10 meters in the direction of $\mathbf{j}$. There are two forces acting on this object, $\mathbf{F}_1 = \mathbf{i} + \mathbf{j} + 2\mathbf{k}$, and $\mathbf{F}_2 = -5\mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$. Find the total work done on the object by the two forces. **Hint:** You can take the work done by the resultant of the two forces or you can add the work done by each force.

15. An object moves 10 meters in the direction of $\mathbf{j} + \mathbf{i}$. There are two forces acting on this object, $\mathbf{F}_1 = \mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$, and $\mathbf{F}_2 = 5\mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$. Find the total work done on the object by the two forces. **Hint:** You can take the work done by the resultant of the two forces or you can add the work done by each force.

16. An object moves 20 meters in the direction of $\mathbf{k} + \mathbf{j}$. There are two forces acting on this object, $\mathbf{F}_1 = \mathbf{i} + \mathbf{j} + 2\mathbf{k}$, and $\mathbf{F}_2 = \mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$. Find the total work done on the object by the two forces. **Hint:** You can take the work done by the resultant of the two forces or you can add the work done by each force.

17. If $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ are vectors. Show that $(\mathbf{b}+\mathbf{c})_{\perp} = \mathbf{b}_{\perp}+\mathbf{c}_{\perp}$ where $\mathbf{b}_{\perp} = \mathbf{b} - \text{proj}_{\mathbf{a}}(\mathbf{b})$.

18. In the discussion of the reflecting mirror which directs all rays to a particular point, $(0,p)$. Show that for any choice of positive C this point is the focus of the parabola and the directrix is $y = p - \frac{1}{C}$.

19. Suppose you wanted to make a solar powered oven to cook food. Are there reasons for using a mirror which is not parabolic? Also describe how you would design a good flash light with a beam which does not spread out too quickly.

20. Find $(1,2,3,4)\cdot(2,0,1,3)$.

21. Show that $(\mathbf{a}\cdot\mathbf{b}) = \frac{1}{4}\left[|\mathbf{a}+\mathbf{b}|^2 - |\mathbf{a}-\mathbf{b}|^2\right]$.

22. Prove from the axioms of the dot product the parallelogram identity, $|\mathbf{a}+\mathbf{b}|^2 + |\mathbf{a}-\mathbf{b}|^2 = 2|\mathbf{a}|^2 + 2|\mathbf{b}|^2$.

23. Let A and be a real $m\times n$ matrix and let $\mathbf{x}\in\mathbb{R}^n$ and $\mathbf{y}\in\mathbb{R}^m$. Show $(A\mathbf{x},\mathbf{y})_{\mathbb{R}^m} = \left(\mathbf{x},A^T\mathbf{y}\right)_{\mathbb{R}^n}$ where $(\cdot,\cdot)_{\mathbb{R}^k}$ denotes the dot product in $\mathbb{R}^k$. In the notation above, $A\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A^T\mathbf{y}$. Use the definition of matrix multiplication to do this.

24. Use the result of Problem 23 to verify directly that $(AB)^T = B^TA^T$ without making any reference to subscripts.

25. Suppose f, g are two continuous functions defined on $[0,1]$. Define

$$(f\cdot g) = \int_0^1 f(x)\,g(x)\,dx.$$

Show this dot product satisfies conditons 5.1 - 5.5. Explain why the Cauchy Schwarz inequality continues to hold in this context and state the Cauchy Schwarz inequality in terms of integrals.

5.4 Exercises With Answers

1. Find the angle between the vectors $3\mathbf{i}-\mathbf{j}-\mathbf{k}$ and $\mathbf{i}+4\mathbf{j}+2\mathbf{k}$.

   $\cos\theta = \frac{3-4-2}{\sqrt{9+1+1}\sqrt{1+16+4}} = -.19739$. Therefore, you have to solve the equation $\cos\theta = -.19739$, Solution is : $\theta = 1.7695$ radians. You need to use a calculator or table to solve this.

2. Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{v} = (1,3,-2)$ and $\mathbf{u} = (1,2,3)$.

   Remember to find this you take $\frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. Thus the answer is $\frac{1}{14}(1,2,3)$.

3. If $\mathbf{F}$ is a force and $\mathbf{D}$ is a vector, show $\text{proj}_{\mathbf{D}}(\mathbf{F}) = (|\mathbf{F}|\cos\theta)\,\mathbf{u}$ where $\mathbf{u}$ is the unit vector in the direction of $\mathbf{D}$, $\mathbf{u} = \mathbf{D}/|\mathbf{D}|$ and θ is the included angle between the two vectors, $\mathbf{F}$ and $\mathbf{D}$. $|\mathbf{F}|\cos\theta$ is sometimes called the component of the force, $\mathbf{F}$ in the direction, $\mathbf{D}$.

   $\text{proj}_{\mathbf{D}}(\mathbf{F}) = \frac{\mathbf{F}\cdot\mathbf{D}}{\mathbf{D}\cdot\mathbf{D}}\mathbf{D} = |\mathbf{F}||\mathbf{D}|\cos\theta\frac{1}{|\mathbf{D}|^2}\mathbf{D} = |\mathbf{F}|\cos\theta\frac{\mathbf{D}}{|\mathbf{D}|}$.

4. A boy drags a sled for 100 feet along the ground by pulling on a rope which is 40 degrees from the horizontal with a force of 10 pounds. How much work does this force do?

   The component of force is $10\cos\left(\frac{40}{180}\pi\right)$ and it acts for 100 feet so the work done is

$$10\cos\left(\frac{40}{180}\pi\right)\times 100 = 766.04$$

5. If $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ are vectors. Show that $(\mathbf{b}+\mathbf{c})_\perp = \mathbf{b}_\perp + \mathbf{c}_\perp$ where $\mathbf{b}_\perp = \mathbf{b} - \text{proj}_{\mathbf{a}}(\mathbf{b})$.

6. Find $(1,0,3,4)\cdot(2,7,1,3)$. $(1,0,3,4)\cdot(2,7,1,3) = 17$.

7. Show that $(\mathbf{a}\cdot\mathbf{b}) = \frac{1}{4}\left[|\mathbf{a}+\mathbf{b}|^2 - |\mathbf{a}-\mathbf{b}|^2\right]$.

   This follows from the axioms of the dot product and the definition of the norm. Thus
   $$|\mathbf{a}+\mathbf{b}|^2 = (\mathbf{a}+\mathbf{b},\mathbf{a}+\mathbf{b}) = |\mathbf{a}|^2 + |\mathbf{b}|^2 + (\mathbf{a}\cdot\mathbf{b})$$
   Do something similar for $|\mathbf{a}-\mathbf{b}|^2$.

8. Prove from the axioms of the dot product the parallelogram identity, $|\mathbf{a}+\mathbf{b}|^2 + |\mathbf{a}-\mathbf{b}|^2 = 2|\mathbf{a}|^2 + 2|\mathbf{b}|^2$.

   Use the properties of the dot product and the definition of the norm in terms of the dot product.

9. Let A and be a real $m\times n$ matrix and let $\mathbf{x}\in\mathbb{R}^n$ and $\mathbf{y}\in\mathbb{R}^m$. Show $(A\mathbf{x},\mathbf{y})_{\mathbb{R}^m} = \left(\mathbf{x},A^T\mathbf{y}\right)_{\mathbb{R}^n}$ where $(\cdot,\cdot)_{\mathbb{R}^k}$ denotes the dot product in $\mathbb{R}^k$. In the notation above, $A\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A^T\mathbf{y}$. Use the definition of matrix multiplication to do this.

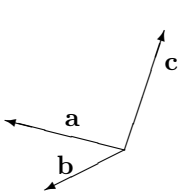
   Remember the ij^{th} entry of $A\mathbf{x} = \sum_j A_{ij}x_j$. Therefore,
   $$A\mathbf{x}\cdot\mathbf{y} = \sum_i (A\mathbf{x})_i\, y_i = \sum_i\sum_j A_{ij}x_j y_i.$$
   Recall now that $\left(A^T\right)_{ij} = A_{ji}$. Use this to write a formula for $\left(\mathbf{x},A^T\mathbf{y}\right)_{\mathbb{R}^n}$.

5.5 The Cross Product

The cross product is the other way of multiplying two vectors in $\mathbb{R}^3$. It is very different from the dot product in many ways. First the geometric meaning is discussed and then a description in terms of coordinates is given. Both descriptions of the cross product are important. The geometric description is essential in order to understand the applications to physics and geometry while the coordinate description is the only way to practically compute the cross product.

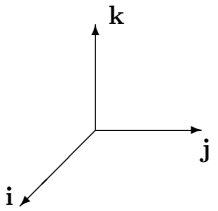
Definition 5.5.1 *Three vectors,* $\mathbf{a},\mathbf{b},\mathbf{c}$ *form a right handed system if when you extend the fingers of your right hand along the vector,* $\mathbf{a}$ *and close them in the direction of* $\mathbf{b}$*, the thumb points roughly in the direction of* $\mathbf{c}$.

For an example of a right handed system of vectors, see the following picture.



In this picture the vector $\mathbf{c}$ points upwards from the plane determined by the other two vectors. You should consider how a right hand system would differ from a left hand system. Try using your left hand and you will see that the vector, $\mathbf{c}$ would need to point in the opposite direction as it would for a right hand system.

From now on, the vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ will always form a right handed system. To repeat, if you extend the fingers of our right hand along $\mathbf{i}$ and close them in the direction $\mathbf{j}$, the thumb points in the direction of $\mathbf{k}$.

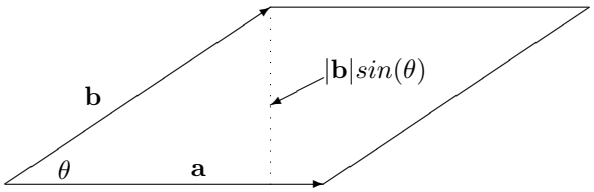


The following is the geometric description of the cross product. It gives both the direction and the magnitude and therefore specifies the vector.

Definition 5.5.2 *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be two vectors in* $\mathbb{R}^3$. *Then* $\mathbf{a} \times \mathbf{b}$ *is defined by the following two rules.*

1. $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}| \sin\theta$ *where* θ *is the included angle.*
2. $\mathbf{a} \times \mathbf{b} \cdot \mathbf{a} = 0$, $\mathbf{a} \times \mathbf{b} \cdot \mathbf{b} = 0$, *and* $\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}$ *forms a right hand system.*

Note that $|\mathbf{a} \times \mathbf{b}|$ is the area of the parallelogram spanned by $\mathbf{a}$ and $\mathbf{b}$.



The cross product satisfies the following properties.

$$\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a}) \ , \ \mathbf{a} \times \mathbf{a} = \mathbf{0}, \tag{5.22}$$

For α a scalar,

$$(\alpha \mathbf{a}) \times \mathbf{b} = \alpha(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times (\alpha \mathbf{b}), \tag{5.23}$$

For $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ vectors, one obtains the distributive laws,

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}, \tag{5.24}$$

$$(\mathbf{b} + \mathbf{c}) \times \mathbf{a} = \mathbf{b} \times \mathbf{a} + \mathbf{c} \times \mathbf{a}. \tag{5.25}$$

Formula 5.22 follows immediately from the definition. The vectors $\mathbf{a} \times \mathbf{b}$ and $\mathbf{b} \times \mathbf{a}$ have the same magnitude, $|\mathbf{a}|\,|\mathbf{b}| \sin\theta$, and an application of the right hand rule shows they have opposite direction. Formula 5.23 is also fairly clear. If α is a nonnegative scalar, the direction of $(\alpha \mathbf{a}) \times \mathbf{b}$ is the same as the direction of $\mathbf{a} \times \mathbf{b}$,$\alpha(\mathbf{a} \times \mathbf{b})$ and $\mathbf{a} \times (\alpha \mathbf{b})$ while the magnitude is just α times the magnitude of $\mathbf{a} \times \mathbf{b}$ which is the same as the magnitude of $\alpha(\mathbf{a} \times \mathbf{b})$ and $\mathbf{a} \times (\alpha \mathbf{b})$. Using this yields equality in 5.23. In the case where $\alpha < 0$, everything works the same way except the vectors are all pointing in the opposite direction and you must multiply by $|\alpha|$ when comparing their magnitudes. The distributive laws are much harder to establish but the second follows from the first quite easily. Thus, assuming the first, and using 5.22,

$$\begin{aligned}(\mathbf{b} + \mathbf{c}) \times \mathbf{a} &= -\mathbf{a} \times (\mathbf{b} + \mathbf{c}) \\ &= -(\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}) \\ &= \mathbf{b} \times \mathbf{a} + \mathbf{c} \times \mathbf{a}.\end{aligned}$$

A proof of the distributive law is given in a later section for those who are interested. Now from the definition of the cross product,

$$\begin{array}{ll} \mathbf{i}\times\mathbf{j}=\mathbf{k} & \mathbf{j}\times\mathbf{i}=-\mathbf{k} \\ \mathbf{k}\times\mathbf{i}=\mathbf{j} & \mathbf{i}\times\mathbf{k}=-\mathbf{j} \\ \mathbf{j}\times\mathbf{k}=\mathbf{i} & \mathbf{k}\times\mathbf{j}=-\mathbf{i} \end{array}$$

With this information, the following gives the coordinate description of the cross product.

Proposition 5.5.3 *Let* $\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$ *and* $\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}$ *be two vectors. Then*

$$\begin{aligned}\mathbf{a}\times\mathbf{b}=&\left(a_2b_3-a_3b_2\right)\mathbf{i}+\left(a_3b_1-a_1b_3\right)\mathbf{j}+\\&+\left(a_1b_2-a_2b_1\right)\mathbf{k}.\end{aligned}\tag{5.26}$$

Proof: From the above table and the properties of the cross product listed,

$$\begin{aligned}&\left(a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}\right)\times\left(b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}\right)=\\&a_1b_2\mathbf{i}\times\mathbf{j}+a_1b_3\mathbf{i}\times\mathbf{k}+a_2b_1\mathbf{j}\times\mathbf{i}+a_2b_3\mathbf{j}\times\mathbf{k}+\\&+a_3b_1\mathbf{k}\times\mathbf{i}+a_3b_2\mathbf{k}\times\mathbf{j}\\&=a_1b_2\mathbf{k}-a_1b_3\mathbf{j}-a_2b_1\mathbf{k}+a_2b_3\mathbf{i}+a_3b_1\mathbf{j}-a_3b_2\mathbf{i}\\&=\left(a_2b_3-a_3b_2\right)\mathbf{i}+\left(a_3b_1-a_1b_3\right)\mathbf{j}+\left(a_1b_2-a_2b_1\right)\mathbf{k}\end{aligned}\tag{5.27}$$

This proves the proposition.

It is probably impossible for most people to remember 5.26. Fortunately, there is a somewhat easier way to remember it.

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\a_1&a_2&a_3\\b_1&b_2&b_3\end{vmatrix}\tag{5.28}$$

where you expand the determinant along the top row. This yields

$$\left(a_2b_3-a_3b_2\right)\mathbf{i}-\left(a_1b_3-a_3b_1\right)\mathbf{j}+\left(a_1b_2-a_2b_1\right)\mathbf{k}\tag{5.29}$$

which is the same as 5.27.

Example 5.5.4 *Find* $\left(\mathbf{i}-\mathbf{j}+2\mathbf{k}\right)\times\left(3\mathbf{i}-2\mathbf{j}+\mathbf{k}\right)$.

Use 5.28 to compute this.

$$\begin{aligned}\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\1&-1&2\\3&-2&1\end{vmatrix}&=\begin{vmatrix}-1&2\\-2&1\end{vmatrix}\mathbf{i}-\begin{vmatrix}1&2\\3&1\end{vmatrix}\mathbf{j}+\begin{vmatrix}1&-1\\3&-2\end{vmatrix}\mathbf{k}\\&=3\mathbf{i}+5\mathbf{j}+\mathbf{k}.\end{aligned}$$

Example 5.5.5 *Find the area of the parallelogram determined by the vectors,* $\left(\mathbf{i}-\mathbf{j}+2\mathbf{k}\right)$ *and* $\left(3\mathbf{i}-2\mathbf{j}+\mathbf{k}\right)$. *These are the same two vectors in Example 5.5.4.*

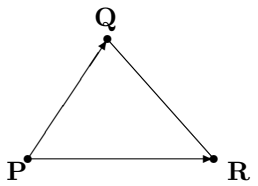
From Example 5.5.4 and the geometric description of the cross product, the area is just the norm of the vector obtained in Example 5.5.4. Thus the area is $\sqrt{9+25+1}=\sqrt{35}$.

Example 5.5.6 *Find the area of the triangle determined by* $(1,2,3)$, $(0,2,5)$, *and* $(5,1,2)$.

This triangle is obtained by connecting the three points with lines. Picking $(1,2,3)$ as a starting point, there are two displacement vectors, $(-1,0,2)$ and $(4,-1,-1)$ such that the given vector added to these displacement vectors gives the other two vectors. The area of the triangle is half the area of the parallelogram determined by $(-1,0,2)$ and $(4,-1,-1)$. Thus $(-1,0,2)\times(4,-1,-1)=(2,7,1)$ and so the area of the triangle is $\frac{1}{2}\sqrt{4+49+1}=\frac{3}{2}\sqrt{6}$.

Observation 5.5.7 *In general, if you have three points (vectors) in* $\mathbb{R}^3, \mathbf{P}, \mathbf{Q}, \mathbf{R}$ *the area of the triangle is given by*

$$\frac{1}{2}\left|(\mathbf{Q}-\mathbf{P})\times(\mathbf{R}-\mathbf{P})\right|.$$

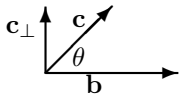


5.5.1 The Distributive Law For The Cross Product

This section gives a proof for 5.24, a fairly difficult topic. It is included here for the interested student. If you are satisfied with taking the distributive law on faith, it is not necessary to read this section. The proof given here is quite clever and follows the one given in [7]. Another approach, based on volumes of parallelepipeds is found in [25] and is discussed a little later.

Lemma 5.5.8 *Let* $\mathbf{b}$ *and* $\mathbf{c}$ *be two vectors. Then* $\mathbf{b}\times\mathbf{c}=\mathbf{b}\times\mathbf{c}_\perp$ *where* $\mathbf{c}_{||}+\mathbf{c}_\perp=\mathbf{c}$ *and* $\mathbf{c}_\perp\cdot\mathbf{b}=0$.

Proof: Consider the following picture.



Now $\mathbf{c}_\perp=\mathbf{c}-\mathbf{c}\cdot\frac{\mathbf{b}}{|\mathbf{b}|}\frac{\mathbf{b}}{|\mathbf{b}|}$ and so $\mathbf{c}_\perp$ is in the plane determined by $\mathbf{c}$ and $\mathbf{b}$. Therefore, from the geometric definition of the cross product, $\mathbf{b}\times\mathbf{c}$ and $\mathbf{b}\times\mathbf{c}_\perp$ have the same direction. Now, referring to the picture,

$$\begin{aligned}|\mathbf{b}\times\mathbf{c}_\perp| &= |\mathbf{b}|\,|\mathbf{c}_\perp| \\ &= |\mathbf{b}|\,|\mathbf{c}|\sin\theta \\ &= |\mathbf{b}\times\mathbf{c}|.\end{aligned}$$

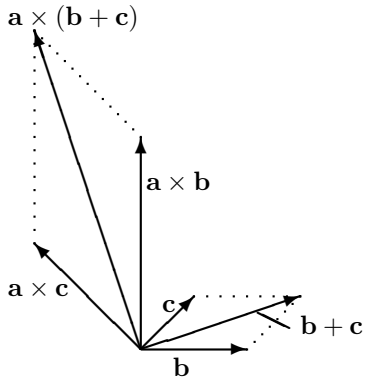
Therefore, $\mathbf{b}\times\mathbf{c}$ and $\mathbf{b}\times\mathbf{c}_\perp$ also have the same magnitude and so they are the same vector.

With this, the proof of the distributive law is in the following theorem.

Theorem 5.5.9 *Let* $\mathbf{a},\mathbf{b}$, *and* $\mathbf{c}$ *be vectors in* $\mathbb{R}^3$. *Then*

$$\mathbf{a}\times(\mathbf{b}+\mathbf{c})=\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c} \tag{5.30}$$

Proof: Suppose first that $\mathbf{a}\cdot\mathbf{b}=\mathbf{a}\cdot\mathbf{c}=0$. Now imagine $\mathbf{a}$ is a vector coming out of the page and let $\mathbf{b},\mathbf{c}$ and $\mathbf{b}+\mathbf{c}$ be as shown in the following picture.



Then $\mathbf{a}\times\mathbf{b},\mathbf{a}\times(\mathbf{b}+\mathbf{c})$, and $\mathbf{a}\times\mathbf{c}$ are each vectors in the same plane, perpendicular to $\mathbf{a}$ as shown. Thus $\mathbf{a}\times\mathbf{c}\cdot\mathbf{c}=0,\mathbf{a}\times(\mathbf{b}+\mathbf{c})\cdot(\mathbf{b}+\mathbf{c})=0$, and $\mathbf{a}\times\mathbf{b}\cdot\mathbf{b}=0$. This implies that to get $\mathbf{a}\times\mathbf{b}$ you move counterclockwise through an angle of $\pi/2$ radians from the vector, $\mathbf{b}$. Similar relationships exist between the vectors $\mathbf{a}\times(\mathbf{b}+\mathbf{c})$ and $\mathbf{b}+\mathbf{c}$ and the vectors $\mathbf{a}\times\mathbf{c}$ and $\mathbf{c}$. Thus the angle between $\mathbf{a}\times\mathbf{b}$ and $\mathbf{a}\times(\mathbf{b}+\mathbf{c})$ is the same as the angle between $\mathbf{b}+\mathbf{c}$ and $\mathbf{b}$ and the angle between $\mathbf{a}\times\mathbf{c}$ and $\mathbf{a}\times(\mathbf{b}+\mathbf{c})$ is the same as the angle between $\mathbf{c}$ and $\mathbf{b}+\mathbf{c}$. In addition to this, since $\mathbf{a}$ is perpendicular to these vectors,

$$|\mathbf{a}\times\mathbf{b}|=|\mathbf{a}||\mathbf{b}|,|\mathbf{a}\times(\mathbf{b}+\mathbf{c})|=|\mathbf{a}||\mathbf{b}+\mathbf{c}|,\text{ and}$$
$$|\mathbf{a}\times\mathbf{c}|=|\mathbf{a}||\mathbf{c}|.$$

Therefore,

$$\frac{|\mathbf{a}\times(\mathbf{b}+\mathbf{c})|}{|\mathbf{b}+\mathbf{c}|}=\frac{|\mathbf{a}\times\mathbf{c}|}{|\mathbf{c}|}=\frac{|\mathbf{a}\times\mathbf{b}|}{|\mathbf{b}|}=|\mathbf{a}|$$

and so

$$\frac{|\mathbf{a}\times(\mathbf{b}+\mathbf{c})|}{|\mathbf{a}\times\mathbf{c}|}=\frac{|\mathbf{b}+\mathbf{c}|}{|\mathbf{c}|},\frac{|\mathbf{a}\times(\mathbf{b}+\mathbf{c})|}{|\mathbf{a}\times\mathbf{b}|}=\frac{|\mathbf{b}+\mathbf{c}|}{|\mathbf{b}|}$$

showing the triangles making up the parallelogram on the right and the four sided figure on the left in the above picture are similar. It follows the four sided figure on the left is in fact a parallelogram and this implies the diagonal is the vector sum of the vectors on the sides, yielding 5.30.

Now suppose it is not necessarily the case that $\mathbf{a}\cdot\mathbf{b}=\mathbf{a}\cdot\mathbf{c}=0$. Then write $\mathbf{b}=\mathbf{b}_{||}+\mathbf{b}_{\perp}$ where $\mathbf{b}_{\perp}\cdot\mathbf{a}=0$. Similarly $\mathbf{c}=\mathbf{c}_{||}+\mathbf{c}_{\perp}$. By the above lemma and what was just shown,

$$\begin{aligned}\mathbf{a}\times(\mathbf{b}+\mathbf{c})&=\mathbf{a}\times(\mathbf{b}+\mathbf{c})_{\perp}\\&=\mathbf{a}\times(\mathbf{b}_{\perp}+\mathbf{c}_{\perp})\\&=\mathbf{a}\times\mathbf{b}_{\perp}+\mathbf{a}\times\mathbf{c}_{\perp}\\&=\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}.\end{aligned}$$

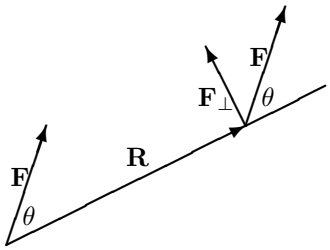
This proves the theorem.

The result of Problem 17 of the exercises 5.3 is used to go from the first to the second line.

5.5.2 Torque

Imagine you are using a wrench to loosen a nut. The idea is to turn the nut by applying a force to the end of the wrench. If you push or pull the wrench directly toward or away

from the nut, it should be obvious from experience that no progress will be made in turning the nut. The important thing is the component of force perpendicular to the wrench. It is this component of force which will cause the nut to turn. For example see the following picture.



In the picture a force, $\mathbf{F}$ is applied at the end of a wrench represented by the position vector, $\mathbf{R}$ and the angle between these two is θ. Then the tendency to turn will be $|\mathbf{R}|\,|\mathbf{F}_\perp| = |\mathbf{R}|\,|\mathbf{F}|\sin\theta$, which you recognize as the magnitude of the cross product of $\mathbf{R}$ and $\mathbf{F}$. If there were just one force acting at one point whose position vector is $\mathbf{R}$, perhaps this would be sufficient, but what if there are numerous forces acting at many different points with neither the position vectors nor the force vectors in the same plane; what then? To keep track of this sort of thing, define for each $\mathbf{R}$ and $\mathbf{F}$, the torque vector,

$$\tau \equiv \mathbf{R} \times \mathbf{F}.$$

This is also called the moment of the force, $\mathbf{F}$. That way, if there are several forces acting at several points, the total torque can be obtained by simply adding up the torques associated with the different forces and positions.

Example 5.5.10 *Suppose* $\mathbf{R}_1 = 2\mathbf{i} - \mathbf{j} + 3\mathbf{k}, \mathbf{R}_2 = \mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$ *meters and at the points determined by these vectors there are forces,* $\mathbf{F}_1 = \mathbf{i} - \mathbf{j} + 2\mathbf{k}$ *and* $\mathbf{F}_2 = \mathbf{i} - 5\mathbf{j} + \mathbf{k}$ *Newtons respectively. Find the total torque about the origin produced by these forces acting at the given points.*

It is necessary to take $\mathbf{R}_1 \times \mathbf{F}_1 + \mathbf{R}_2 \times \mathbf{F}_2$. Thus the total torque equals

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & -1 & 3 \\ 1 & -1 & 2 \end{vmatrix} + \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & -6 \\ 1 & -5 & 1 \end{vmatrix} = -27\mathbf{i} - 8\mathbf{j} - 8\mathbf{k} \text{ Newton meters}$$

Example 5.5.11 *Find if possible a single force vector,* $\mathbf{F}$ *which if applied at the point* $\mathbf{i} + \mathbf{j} + \mathbf{k}$ *will produce the same torque as the above two forces acting at the given points.*

This is fairly routine. The problem is to find $\mathbf{F} = F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k}$ which produces the above torque vector. Therefore,

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 1 \\ F_1 & F_2 & F_3 \end{vmatrix} = -27\mathbf{i} - 8\mathbf{j} - 8\mathbf{k}$$

which reduces to $(F_3 - F_2)\,\mathbf{i} + (F_1 - F_3)\,\mathbf{j} + (F_2 - F_1)\,\mathbf{k} = -27\mathbf{i} - 8\mathbf{j} - 8\mathbf{k}$. This amounts to solving the system of three equations in three unknowns, F_1, F_2, and F_3,

$$\begin{aligned} F_3 - F_2 &= -27 \\ F_1 - F_3 &= -8 \\ F_2 - F_1 &= -8 \end{aligned}$$

However, there is no solution to these three equations. (Why?) Therefore no single force acting at the point $\mathbf{i} + \mathbf{j} + \mathbf{k}$ will produce the given torque.

5.5.3 Center Of Mass

The mass of an object is a measure of how much stuff there is in the object. An object has mass equal to one kilogram, a unit of mass in the metric system, if it would exactly balance a known one kilogram object when placed on a balance. The known object is one kilogram by definition. The mass of an object does not depend on where the balance is used. It would be one kilogram on the moon as well as on the earth. The weight of an object is something else. It is the force exerted on the object by gravity and has magnitude gm where g is a constant called the acceleration of gravity. Thus the weight of a one kilogram object would be different on the moon which has much less gravity, smaller g, than on the earth. An important idea is that of the center of mass. This is the point at which an object will balance no matter how it is turned.

Definition 5.5.12 *Let an object consist of p point masses, $m_1, \cdots, m_p$ with the position of the k^{th} of these at $\mathbf{R}_k$. The center of mass of this object, $\mathbf{R}_0$ is the point satisfying*

$$\sum_{k=1}^{p} (\mathbf{R}_k - \mathbf{R}_0) \times g m_k \mathbf{u} = \mathbf{0}$$

for all unit vectors, $\mathbf{u}$.

The above definition indicates that no matter how the object is suspended, the total torque on it due to gravity is such that no rotation occurs. Using the properties of the cross product,

$$\left(\sum_{k=1}^{p} \mathbf{R}_k g m_k - \mathbf{R}_0 \sum_{k=1}^{p} g m_k \right) \times \mathbf{u} = \mathbf{0} \tag{5.31}$$

for any choice of unit vector, $\mathbf{u}$. You should verify that if $\mathbf{a} \times \mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$, then it must be the case that $\mathbf{a} = \mathbf{0}$. Then the above formula requires that

$$\sum_{k=1}^{p} \mathbf{R}_k g m_k - \mathbf{R}_0 \sum_{k=1}^{p} g m_k = \mathbf{0}.$$

dividing by g, and then by $\sum_{k=1}^{p} m_k$,

$$\mathbf{R}_0 = \frac{\sum_{k=1}^{p} \mathbf{R}_k m_k}{\sum_{k=1}^{p} m_k}. \tag{5.32}$$

This is the formula for the center of mass of a collection of point masses. To consider the center of mass of a solid consisting of continuously distributed masses, you need the methods of calculus.

Example 5.5.13 *Let $m_1 = 5, m_2 = 6$, and $m_3 = 3$ where the masses are in kilograms. Suppose m_1 is located at $2\mathbf{i} + 3\mathbf{j} + \mathbf{k}$, m_2 is located at $\mathbf{i} - 3\mathbf{j} + 2\mathbf{k}$ and m_3 is located at $2\mathbf{i} - \mathbf{j} + 3\mathbf{k}$. Find the center of mass of these three masses.*

Using 5.32

$$\begin{aligned} \mathbf{R}_0 &= \frac{5\left(2\mathbf{i} + 3\mathbf{j} + \mathbf{k}\right) + 6\left(\mathbf{i} - 3\mathbf{j} + 2\mathbf{k}\right) + 3\left(2\mathbf{i} - \mathbf{j} + 3\mathbf{k}\right)}{5 + 6 + 3} \\ &= \frac{11}{7}\mathbf{i} - \frac{3}{7}\mathbf{j} + \frac{13}{7}\mathbf{k} \end{aligned}$$

5.5.4 Angular Velocity

Definition 5.5.14 *In a rotating body, a vector,* $\boldsymbol{\Omega}$ *is called an* ***angular velocity vector*** *if the velocity of a point having position vector,* $\mathbf{u}$ *relative to the body is given by* $\boldsymbol{\Omega} \times \mathbf{u}$.

The existence of an angular velocity vector is the key to understanding motion in a moving system of coordinates. It is used to explain the motion on the surface of the rotating earth. For example, have you ever wondered why low pressure areas rotate counter clockwise in the upper hemisphere but clockwise in the lower hemisphere? To quantify these things, you will need the concept of an angular velocity vector. Details are presented later for interesting examples. In the above example, think of a coordinate system fixed in the rotating body. Thus if you were riding on the rotating body, you would observe this coordinate system as fixed even though it is not.

Example 5.5.15 *A wheel rotates counter clockwise about the vector* $\mathbf{i}+\mathbf{j}+\mathbf{k}$ *at 60 revolutions per minute. This means that if the thumb of your right hand were to point in the direction of* $\mathbf{i}+\mathbf{j}+\mathbf{k}$ *your fingers of this hand would wrap in the direction of rotation. Find the angular velocity vector for this wheel. Assume the unit of distance is meters and the unit of time is minutes.*

Let $\omega = 60 \times 2\pi = 120\pi$. This is the number of radians per minute corresponding to 60 revolutions per minute. Then the angular velocity vector is $\frac{120\pi}{\sqrt{3}}(\mathbf{i}+\mathbf{j}+\mathbf{k})$. Note this gives what you would expect in the case the position vector to the point is perpendicular to $\mathbf{i}+\mathbf{j}+\mathbf{k}$ and at a distance of r. This is because of the geometric description of the cross product. The magnitude of the vector is $r120\pi$ meters per minute and corresponds to the speed and an exercise with the right hand shows the direction is correct also. However, if this body is rigid, this will work for every other point in it, even those for which the position vector is not perpendicular to the given vector. A complete analysis of this is given later.

Example 5.5.16 *A wheel rotates counter clockwise about the vector* $\mathbf{i}+\mathbf{j}+\mathbf{k}$ *at 60 revolutions per minute exactly as in Example 5.5.15. Let* $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ *denote an orthogonal right handed system attached to the rotating wheel in which* $\mathbf{u}_3 = \frac{1}{\sqrt{3}}(\mathbf{i}+\mathbf{j}+\mathbf{k})$. *Thus* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *depend on time. Find the velocity of the point of the wheel located at the point* $2\mathbf{u}_1 + 3\mathbf{u}_2 - \mathbf{u}_3$. *Note this point is not fixed in space. It is moving.*

Since $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a right handed system like $\mathbf{i}, \mathbf{j}, \mathbf{k}$, everything applies to this system in the same way as with $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Thus the cross product is given by

$$\begin{aligned} & (a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3) \times (d\mathbf{u}_1 + e\mathbf{u}_2 + f\mathbf{u}_3) \\ = & \begin{vmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \\ a & b & c \\ d & e & f \end{vmatrix} \end{aligned}$$

Therefore, in terms of the given vectors $\mathbf{u}_i$, the angular velocity vector is

$$120\pi\mathbf{u}_3$$

the velocity of the given point is

$$\begin{aligned} & \begin{vmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \\ 0 & 0 & 120\pi \\ 2 & 3 & -1 \end{vmatrix} \\ = & -360\pi\mathbf{u}_1 + 240\pi\mathbf{u}_2 \end{aligned}$$

in meters per minute. Note how this gives the answer in terms of these vectors which are fixed in the body, not in space. Since $\mathbf{u}_i$ depends on t, this shows the answer in this case does also. Of course this is right. Just think of what is going on with the wheel rotating. Those vectors which are fixed in the wheel are moving in space. The velocity of a point in the wheel should be constantly changing. However, its speed will not change. The speed will be the magnitude of the velocity and this is

$$\sqrt{(-360\pi\mathbf{u}_1 + 240\pi\mathbf{u}_2) \cdot (-360\pi\mathbf{u}_1 + 240\pi\mathbf{u}_2)}$$

which from the properties of the dot product equals

$$\sqrt{(-360\pi)^2 + (240\pi)^2} = 120\sqrt{13}\pi$$

because the $\mathbf{u}_i$ are given to be orthogonal.

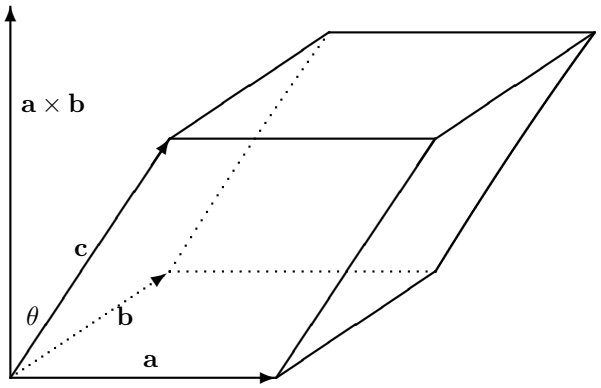
5.5.5 The Box Product

Definition 5.5.17 *A parallelepiped determined by the three vectors,* $\mathbf{a}, \mathbf{b}$, *and* $\mathbf{c}$ *consists of*

$$\{r\mathbf{a}+s\mathbf{b}+t\mathbf{c} : r, s, t \in [0,1]\}.$$

That is, if you pick three numbers, r, s, *and* t *each in* $[0,1]$ *and form* $r\mathbf{a}+s\mathbf{b}+t\mathbf{c}$, *then the collection of all such points is what is meant by the parallelepiped determined by these three vectors.*

The following is a picture of such a thing.



You notice the area of the base of the parallelepiped, the parallelogram determined by the vectors, $\mathbf{a}$ and $\mathbf{b}$ has area equal to $|\mathbf{a} \times \mathbf{b}|$ while the altitude of the parallelepiped is $|\mathbf{c}| \cos\theta$ where θ is the angle shown in the picture between $\mathbf{c}$ and $\mathbf{a} \times \mathbf{b}$. Therefore, the volume of this parallelepiped is the area of the base times the altitude which is just

$$|\mathbf{a} \times \mathbf{b}|\,|\mathbf{c}| \cos\theta = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c}.$$

This expression is known as the box product and is sometimes written as $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$. You should consider what happens if you interchange the $\mathbf{b}$ with the $\mathbf{c}$ or the $\mathbf{a}$ with the $\mathbf{c}$. You can see geometrically from drawing pictures that this merely introduces a minus sign. In any case the box product of three vectors always equals either the volume of the parallelepiped determined by the three vectors or else minus this volume.

Example 5.5.18 *Find the volume of the parallelepiped determined by the vectors,* $\mathbf{i} + 2\mathbf{j} - 5\mathbf{k}, \mathbf{i} + 3\mathbf{j} - 6\mathbf{k}, 3\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.

According to the above discussion, pick any two of these, take the cross product and then take the dot product of this with the third of these vectors. The result will be either the desired volume or minus the desired volume.

$$\begin{aligned}(\mathbf{i}+2\mathbf{j}-5\mathbf{k})\times(\mathbf{i}+3\mathbf{j}-6\mathbf{k}) &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & -5 \\ 1 & 3 & -6 \end{vmatrix} \\ &= 3\mathbf{i}+\mathbf{j}+\mathbf{k}\end{aligned}$$

Now take the dot product of this vector with the third which yields

$$(3\mathbf{i}+\mathbf{j}+\mathbf{k})\cdot(3\mathbf{i}+2\mathbf{j}+3\mathbf{k}) = 9+2+3 = 14.$$

This shows the volume of this parallelepiped is 14 cubic units.

There is a fundamental observation which comes directly from the geometric definitions of the cross product and the dot product.

Lemma 5.5.19 *Let* $\mathbf{a},\mathbf{b}$, *and* $\mathbf{c}$ *be vectors. Then* $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$.

Proof: This follows from observing that either $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$ and $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$ both give the volume of the parallellepiped or they both give -1 times the volume.

An Alternate Proof Of The Distributive Law

Here is another proof of the distributive law for the cross product. Let $\mathbf{x}$ be a vector. From the above observation,

$$\begin{aligned}\mathbf{x}\cdot\mathbf{a}\times(\mathbf{b}+\mathbf{c}) &= (\mathbf{x}\times\mathbf{a})\cdot(\mathbf{b}+\mathbf{c}) \\ &= (\mathbf{x}\times\mathbf{a})\cdot\mathbf{b}+(\mathbf{x}\times\mathbf{a})\cdot\mathbf{c} \\ &= \mathbf{x}\cdot\mathbf{a}\times\mathbf{b}+\mathbf{x}\cdot\mathbf{a}\times\mathbf{c} \\ &= \mathbf{x}\cdot(\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}).\end{aligned}$$

Therefore,

$$\mathbf{x}\cdot[\mathbf{a}\times(\mathbf{b}+\mathbf{c})-(\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c})] = 0$$

for all $\mathbf{x}$. In particular, this holds for $\mathbf{x} = \mathbf{a}\times(\mathbf{b}+\mathbf{c})-(\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c})$ showing that $\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}$ and this proves the distributive law for the cross product another way.

Observation 5.5.20 *Suppose you have three vectors,* $\mathbf{u}=(a,b,c)$, $\mathbf{v}=(d,e,f)$, *and* $\mathbf{w}=(g,h,i)$. *Then* $\mathbf{u}\cdot\mathbf{v}\times\mathbf{w}$ *is given by the following.*

$$\begin{aligned}\mathbf{u}\cdot\mathbf{v}\times\mathbf{w} &= (a,b,c)\cdot\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ d & e & f \\ g & h & i \end{vmatrix} \\ &= a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= \det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}.\end{aligned}$$

The message is that to take the box product, you can simply take the determinant of the matrix which results by letting the rows be the rectangular components of the given vectors in the order in which they occur in the box product.

5.6 Vector Identities And Notation

To begin with consider $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})$ and it is desired to simplify this quantity. It turns out this is an important quantity which comes up in many different contexts. Let $\mathbf{u}=(u_1,u_2,u_3)$ and let $\mathbf{v}$ and $\mathbf{w}$ be defined similarly.

$$\begin{aligned}\mathbf{v}\times\mathbf{w} &= \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ v_1 & v_2 & v_3\\ w_1 & w_2 & w_3\end{vmatrix}\\ &= (v_2w_3-v_3w_2)\,\mathbf{i}+(w_1v_3-v_1w_3)\,\mathbf{j}+(v_1w_2-v_2w_1)\,\mathbf{k}\end{aligned}$$

Next consider $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})$ which is given by

$$\mathbf{u}\times(\mathbf{v}\times\mathbf{w})=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ u_1 & u_2 & u_3\\ (v_2w_3-v_3w_2) & (w_1v_3-v_1w_3) & (v_1w_2-v_2w_1)\end{vmatrix}.$$

When you multiply this out, you get

$$\begin{aligned}&\mathbf{i}\,(v_1u_2w_2+u_3v_1w_3-w_1u_2v_2-u_3w_1v_3)+\mathbf{j}\,(v_2u_1w_1+v_2w_3u_3-w_2u_1v_1-u_3w_2v_3)\\ &+\mathbf{k}\,(u_1w_1v_3+v_3w_2u_2-u_1v_1w_3-v_2w_3u_2)\end{aligned}$$

and if you are clever, you see right away that

$$(\mathbf{i}v_1+\mathbf{j}v_2+\mathbf{k}v_3)\,(u_1w_1+u_2w_2+u_3w_3)-(\mathbf{i}w_1+\mathbf{j}w_2+\mathbf{k}w_3)\,(u_1v_1+u_2v_2+u_3v_3)\,.$$

Thus

$$\mathbf{u}\times(\mathbf{v}\times\mathbf{w})=\mathbf{v}\,(\mathbf{u}\cdot\mathbf{w})-\mathbf{w}\,(\mathbf{u}\cdot\mathbf{v})\,. \tag{5.33}$$

A related formula is

$$\begin{aligned}(\mathbf{u}\times\mathbf{v})\times\mathbf{w} &= -[\mathbf{w}\times(\mathbf{u}\times\mathbf{v})]\\ &= -[\mathbf{u}\,(\mathbf{w}\cdot\mathbf{v})-\mathbf{v}\,(\mathbf{w}\cdot\mathbf{u})]\\ &= \mathbf{v}\,(\mathbf{w}\cdot\mathbf{u})-\mathbf{u}\,(\mathbf{w}\cdot\mathbf{v})\,.\end{aligned} \tag{5.34}$$

This derivation is simply wretched and it does nothing for other identities which may arise in applications. Actually, the above two formulas, 5.33 and 5.34 are sufficient for most applications if you are creative in using them, but there is another way. This other way allows you to discover such vector identities as the above without any creativity or any cleverness. Therefore, it is far superior to the above nasty computation. It is a vector identity discovering machine and it is this which is the main topic in what follows.

There are two special symbols, δ_{ij} and ε_{ijk} which are very useful in dealing with vector identities. To begin with, here is the definition of these symbols.

Definition 5.6.1 *The symbol, δ_{ij}, called the Kroneker delta symbol is defined as follows.*

$$\delta_{ij}\equiv\begin{cases}1 \text{ if } i=j\\ 0 \text{ if } i\neq j\end{cases}.$$

With the Kroneker symbol, i and j can equal any integer in $\{1,2,\cdots,n\}$ for any $n\in\mathbb{N}$.

Definition 5.6.2 *For i,j, and k integers in the set, $\{1,2,3\}$, ε_{ijk} is defined as follows.*

$$\varepsilon_{ijk}\equiv\begin{cases}1 \text{ if } (i,j,k)=(1,2,3),(2,3,1), \text{ or } (3,1,2)\\ -1 \text{ if } (i,j,k)=(2,1,3),(1,3,2), \text{ or } (3,2,1)\\ 0 \text{ if there are any repeated integers}\end{cases}.$$

The subscripts ijk and ij in the above are called indices. A single one is called an index. This symbol, ε_{ijk} is also called the permutation symbol.

The way to think of ε_{ijk} is that $\varepsilon_{123} = 1$ and if you switch any two of the numbers in the list i, j, k, it changes the sign. Thus $\varepsilon_{ijk} = -\varepsilon_{jik}$ and $\varepsilon_{ijk} = -\varepsilon_{kji}$ etc. You should check that this rule reduces to the above definition. For example, it immediately implies that if there is a repeated index, the answer is zero. This follows because $\varepsilon_{iij} = -\varepsilon_{iij}$ and so $\varepsilon_{iij} = 0$.

It is useful to use the Einstein summation convention when dealing with these symbols. Simply stated, the convention is that you sum over the repeated index. Thus $a_i b_i$ means $\sum_i a_i b_i$. Also, $\delta_{ij} x_j$ means $\sum_j \delta_{ij} x_j = x_i$. When you use this convention, there is one very important thing to never forget. It is this: Never have an index be repeated more than once. Thus $a_i b_i$ is all right but $a_{ii} b_i$ is not. The reason for this is that you end up getting confused about what is meant. If you want to write $\sum_i a_i b_i c_i$ it is best to simply use the summation notation. There is a very important reduction identity connecting these two symbols.

Lemma 5.6.3 *The following holds.*

$$\varepsilon_{ijk}\varepsilon_{irs} = (\delta_{jr}\delta_{ks} - \delta_{kr}\delta_{js}).$$

Proof: If $\{j, k\} \neq \{r, s\}$ then every term in the sum on the left must have either ε_{ijk} or ε_{irs} contains a repeated index. Therefore, the left side equals zero. The right side also equals zero in this case. To see this, note that if the two sets are not equal, then there is one of the indices in one of the sets which is not in the other set. For example, it could be that j is not equal to either r or s. Then the right side equals zero.

Therefore, it can be assumed $\{j, k\} = \{r, s\}$. If $i = r$ and $j = s$ for $s \neq r$, then there is exactly one term in the sum on the left and it equals 1. The right also reduces to 1 in this case. If $i = s$ and $j = r$, there is exactly one term in the sum on the left which is nonzero and it must equal -1. The right side also reduces to -1 in this case. If there is a repeated index in $\{j, k\}$, then every term in the sum on the left equals zero. The right also reduces to zero in this case because then $j = k = r = s$ and so the right side becomes $(1)(1) - (-1)(-1) = 0$.

Proposition 5.6.4 *Let* $\mathbf{u}, \mathbf{v}$ *be vectors in* $\mathbb{R}^n$ *where the Cartesian coordinates of* $\mathbf{u}$ *are* $(u_1, \cdots, u_n)$ *and the Cartesian coordinates of* $\mathbf{v}$ *are* $(v_1, \cdots, v_n)$. *Then* $\mathbf{u} \cdot \mathbf{v} = u_i v_i$. *If* $\mathbf{u}, \mathbf{v}$ *are vectors in* $\mathbb{R}^3$, *then*

$$(\mathbf{u} \times \mathbf{v})_i = \varepsilon_{ijk} u_j v_k.$$

Also, $\delta_{ik} a_k = a_i$.

Proof: The first claim is obvious from the definition of the dot product. The second is verified by simply checking it works. For example,

$$\mathbf{u} \times \mathbf{v} \equiv \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

and so

$$(\mathbf{u} \times \mathbf{v})_1 = (u_2 v_3 - u_3 v_2).$$

From the above formula in the proposition,

$$\varepsilon_{1jk} u_j v_k \equiv u_2 v_3 - u_3 v_2,$$

the same thing. The cases for $(\mathbf{u} \times \mathbf{v})_2$ and $(\mathbf{u} \times \mathbf{v})_3$ are verified similarly. The last claim follows directly from the definition.

With this notation, you can easily discover vector identities and simplify expressions which involve the cross product.

Example 5.6.5 *Discover a formula which simplifies* $(\mathbf{u}\times\mathbf{v})\times\mathbf{w}$.

From the above reduction formula,

$$\begin{aligned}((\mathbf{u}\times\mathbf{v})\times\mathbf{w})_i &= \varepsilon_{ijk}(\mathbf{u}\times\mathbf{v})_j w_k \\ &= \varepsilon_{ijk}\varepsilon_{jrs}u_r v_s w_k \\ &= -\varepsilon_{jik}\varepsilon_{jrs}u_r v_s w_k \\ &= -(\delta_{ir}\delta_{ks}-\delta_{is}\delta_{kr})u_r v_s w_k \\ &= -(u_i v_k w_k - u_k v_i w_k) \\ &= \mathbf{u}\cdot\mathbf{w}v_i - \mathbf{v}\cdot\mathbf{w}u_i \\ &= ((\mathbf{u}\cdot\mathbf{w})\mathbf{v}-(\mathbf{v}\cdot\mathbf{w})\mathbf{u})_i .\end{aligned}$$

Since this holds for all i, it follows that

$$(\mathbf{u}\times\mathbf{v})\times\mathbf{w} = (\mathbf{u}\cdot\mathbf{w})\mathbf{v}-(\mathbf{v}\cdot\mathbf{w})\mathbf{u}.$$

This is good notation and it will be used in the rest of the book whenever convenient.

5.7 Exercises

1. Show that if $\mathbf{a}\times\mathbf{u}=\mathbf{0}$ for all unit vectors, $\mathbf{u}$, then $\mathbf{a}=\mathbf{0}$.
2. If you only assume 5.31 holds for $\mathbf{u}=\mathbf{i},\mathbf{j},\mathbf{k}$, show that this implies 5.31 holds for all unit vectors, $\mathbf{u}$.
3. Let $m_1=5, m_2=1$, and $m_3=4$ where the masses are in kilograms and the distance is in meters. Suppose m_1 is located at $2\mathbf{i}-3\mathbf{j}+\mathbf{k}$, m_2 is located at $\mathbf{i}-3\mathbf{j}+6\mathbf{k}$ and m_3 is located at $2\mathbf{i}+\mathbf{j}+3\mathbf{k}$. Find the center of mass of these three masses.
4. Let $m_1=2, m_2=3$, and $m_3=1$ where the masses are in kilograms and the distance is in meters. Suppose m_1 is located at $2\mathbf{i}-\mathbf{j}+\mathbf{k}$, m_2 is located at $\mathbf{i}-2\mathbf{j}+\mathbf{k}$ and m_3 is located at $4\mathbf{i}+\mathbf{j}+3\mathbf{k}$. Find the center of mass of these three masses.
5. Find the angular velocity vector of a rigid body which rotates counter clockwise about the vector $\mathbf{i}-2\mathbf{j}+\mathbf{k}$ at 40 revolutions per minute. Assume distance is measured in meters.
6. Let $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ be a right handed system with $\mathbf{u}_3$ pointing in the direction of $\mathbf{i}-2\mathbf{j}+\mathbf{k}$ and $\mathbf{u}_1$ and $\mathbf{u}_2$ being fixed with the body which is rotating at 40 revolutions per minute. Assuming all distances are in meters, find the constant speed of the point of the body located at $3\mathbf{u}_1+\mathbf{u}_2-\mathbf{u}_3$ in meters per minute.
7. Find the area of the triangle determined by the three points, $(1,2,3),(4,2,0)$ and $(-3,2,1)$.
8. Find the area of the triangle determined by the three points, $(1,0,3),(4,1,0)$ and $(-3,1,1)$.
9. Find the area of the triangle determined by the three points, $(1,2,3),(2,3,1)$ and $(0,1,2)$. Did something interesting happen here? What does it mean geometrically?
10. Find the area of the parallelogram determined by the vectors, $(1,2,3)$ and $(3,-2,1)$.

11. Find the area of the parallelogram determined by the vectors, $(1,0,3)$ and $(4,-2,1)$.

12. Find the area of the parallelogram determined by the vectors, $(1,-2,2)$ and $(3,1,1)$.

13. Find the volume of the parallelepiped determined by the vectors, $\mathbf{i}-7\mathbf{j}-5\mathbf{k}, \mathbf{i}-2\mathbf{j}-6\mathbf{k}$,$3\mathbf{i}+2\mathbf{j}+3\mathbf{k}$.

14. Find the volume of the parallelepiped determined by the vectors, $\mathbf{i}+\mathbf{j}-5\mathbf{k}, \mathbf{i}+5\mathbf{j}-6\mathbf{k}$,$3\mathbf{i}+\mathbf{j}+3\mathbf{k}$.

15. Find the volume of the parallelepiped determined by the vectors, $\mathbf{i}+6\mathbf{j}+5\mathbf{k}, \mathbf{i}+5\mathbf{j}-6\mathbf{k}$,$3\mathbf{i}+\mathbf{j}+\mathbf{k}$.

16. Suppose $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ are three vectors whose components are all integers. Can you conclude the volume of the parallelepiped determined from these three vectors will always be an integer?

17. What does it mean geometrically if the box product of three vectors gives zero?

18. It is desired to find an equation of a plane containing the two vectors, $\mathbf{a}$ and $\mathbf{b}$ and the point $\mathbf{0}$. Using Problem 17, show an equation for this plane is

$$\begin{vmatrix} x & y & z \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = 0$$

That is, the set of all (x,y,z) such that the above expression equals zero.

19. Using the notion of the box product yielding either plus or minus the volume of the parallelepiped determined by the given three vectors, show that

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$$

In other words, the dot and the cross can be switched as long as the order of the vectors remains the same. **Hint:** There are two ways to do this, by the coordinate description of the dot and cross product and by geometric reasoning.

20. Is $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\times\mathbf{c}$? What is the meaning of $\mathbf{a}\times\mathbf{b}\times\mathbf{c}$? Explain. **Hint:** Try $(\mathbf{i}\times\mathbf{j})\times\mathbf{j}$.

21. Verify directly that the coordinate description of the cross product, $\mathbf{a}\times\mathbf{b}$ has the property that it is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$. Then show by direct computation that this coordinate description satisfies

$$\begin{aligned} |\mathbf{a}\times\mathbf{b}|^2 &= |\mathbf{a}|^2|\mathbf{b}|^2 - (\mathbf{a}\cdot\mathbf{b})^2 \\ &= |\mathbf{a}|^2|\mathbf{b}|^2\left(1-\cos^2(\theta)\right) \end{aligned}$$

where θ is the angle included between the two vectors. Explain why $|\mathbf{a}\times\mathbf{b}|$ has the correct magnitude. All that is missing is the material about the right hand rule. Verify directly from the coordinate description of the cross product that the right thing happens with regards to the vectors $\mathbf{i},\mathbf{j},\mathbf{k}$. Next verify that the distributive law holds for the coordinate description of the cross product. This gives another way to approach the cross product. First define it in terms of coordinates and then get the geometric properties from this.

22. Discover a vector identity for $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})$.

23. Discover a vector identity for $(\mathbf{u}\times\mathbf{v})\cdot(\mathbf{z}\times\mathbf{w})$.

24. Discover a vector identity for $(\mathbf{u}\times\mathbf{v})\times(\mathbf{z}\times\mathbf{w})$ in terms of box products.

25. Simplify $(\mathbf{u}\times\mathbf{v})\cdot(\mathbf{v}\times\mathbf{w})\times(\mathbf{w}\times\mathbf{z})$.

26. Simplify $|\mathbf{u}\times\mathbf{v}|^2+(\mathbf{u}\cdot\mathbf{v})^2-|\mathbf{u}|^2|\mathbf{v}|^2$.

27. Prove that $\varepsilon_{ijk}\varepsilon_{ijr}=2\delta_{kr}$.

28. If A is a 3×3 matrix such that $A=\begin{pmatrix}\mathbf{u} & \mathbf{v} & \mathbf{w}\end{pmatrix}$ where these are the columns of the matrix, A. Show that $\det(A)=\varepsilon_{ijk}u_iv_jw_k$.

29. If A is a 3×3 matrix, show $\varepsilon_{rps}\det(A)=\varepsilon_{ijk}A_{ri}A_{pj}A_{sk}$.

30. Suppose A is a 3×3 matrix and $\det(A)\neq0$. Show using 29 and 27 that
$$\left(A^{-1}\right)_{ks}=\frac{1}{2\det(A)}\varepsilon_{rps}\varepsilon_{ijk}A_{pj}A_{ri}.$$

31. When you have a rotating rigid body with angular velocity vector, $\boldsymbol{\Omega}$ then the velocity, $\mathbf{u}'$ is given by $\mathbf{u}'=\boldsymbol{\Omega}\times\mathbf{u}$. It turns out that all the usual calculus rules such as the product rule hold. Also, $\mathbf{u}''$ is the acceleration. Show using the product rule that for $\boldsymbol{\Omega}$ a constant vector,
$$\mathbf{u}''=\boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{u}).$$
It turns out this is the centripetal acceleration. Note how it involves cross products. Things get really interesting when you move about on the rotating body. weird forces are felt. This is in the section on moving coordinate systems.

5.8 Exercises With Answers

1. If you only assume 5.31 holds for $\mathbf{u}=\mathbf{i},\mathbf{j},\mathbf{k}$, show that this implies 5.31 holds for all unit vectors, $\mathbf{u}$.

   Suppose than that $(\sum_{k=1}^p\mathbf{R}_kgm_k-\mathbf{R}_0\sum_{k=1}^pgm_k)\times\mathbf{u}=\mathbf{0}$ for $\mathbf{u}=\mathbf{i},\mathbf{j},\mathbf{k}$. Then if $\mathbf{u}$ is an arbitrary unit vector, $\mathbf{u}$ must be of the form $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$. Now from the distributive property of the cross product and letting $\mathbf{w}=(\sum_{k=1}^p\mathbf{R}_kgm_k-\mathbf{R}_0\sum_{k=1}^pgm_k)$, this says
$$\begin{aligned}&(\textstyle\sum_{k=1}^p\mathbf{R}_kgm_k-\mathbf{R}_0\sum_{k=1}^pgm_k)\times\mathbf{u}\\&=\mathbf{w}\times(a\mathbf{i}+b\mathbf{j}+c\mathbf{k})\\&=a\mathbf{w}\times\mathbf{i}+b\mathbf{w}\times\mathbf{j}+c\mathbf{w}\times\mathbf{k}\\&=\mathbf{0}+\mathbf{0}+\mathbf{0}=\mathbf{0}.\end{aligned}$$

2. Let $m_1=4,m_2=3$, and $m_3=1$ where the masses are in kilograms and the distance is in meters. Suppose m_1 is located at $2\mathbf{i}-\mathbf{j}+\mathbf{k}$, m_2 is located at $2\mathbf{i}-3\mathbf{j}+\mathbf{k}$ and m_3 is located at $2\mathbf{i}+\mathbf{j}+3\mathbf{k}$. Find the center of mass of these three masses.

   Let the center of mass be located at $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$. Then $(4+3+1)(a\mathbf{i}+b\mathbf{j}+c\mathbf{k})=4(2\mathbf{i}-\mathbf{j}+\mathbf{k})+3(2\mathbf{i}-3\mathbf{j}+\mathbf{k})+1(2\mathbf{i}+\mathbf{j}+3\mathbf{k})=16\mathbf{i}-12\mathbf{j}+10\mathbf{k}$. Therefore, $a=2,b=\frac{-3}{2}$ and $c=\frac{5}{4}$. The center of mass is then $2\mathbf{i}-\frac{3}{2}\mathbf{j}+\frac{5}{4}\mathbf{k}$.

3. Find the angular velocity vector of a rigid body which rotates counter clockwise about the vector $\mathbf{i}-\mathbf{j}+\mathbf{k}$ at 20 revolutions per minute. Assume distance is measured in meters.

   The angular velocity is $20\times2\pi=40\pi$. Then $\boldsymbol{\Omega}=40\pi\frac{1}{\sqrt{3}}(\mathbf{i}-\mathbf{j}+\mathbf{k})$.

4. Find the area of the triangle determined by the three points, $(1,2,3)$, $(1,2,6)$ and $(-3,2,1)$.

   The three points determine two displacement vectors from the point $(1,2,3)$, $(0,0,3)$ and $(-4,0,-2)$. To find the area of the parallelogram determined by these two displacement vectors, you simply take the norm of their cross product. To find the area of the triangle, you take one half of that. Thus the area is

$$(1/2)\left|(0,0,3)\times(-4,0,-2)\right| = \frac{1}{2}\left|(0,-12,0)\right| = 6.$$

5. Find the area of the parallelogram determined by the vectors, $(1,0,3)$ and $(4,-2,1)$.

   $|(1,0,3)\times(4,-2,1)| = |(6,11,-2)| = \sqrt{26+121+4} = \sqrt{151}$.

6. Find the volume of the parallelepiped determined by the vectors, $\mathbf{i}-7\mathbf{j}-5\mathbf{k}, \mathbf{i}+2\mathbf{j}-6\mathbf{k}$,$3\mathbf{i}-3\mathbf{j}+\mathbf{k}$.

   Remember you just need to take the absolute value of the determinant having the given vectors as rows. Thus the volume is the absolute value of

$$\begin{vmatrix} 1 & -7 & -5 \\ 1 & 2 & -6 \\ 3 & -3 & 1 \end{vmatrix} = 162$$

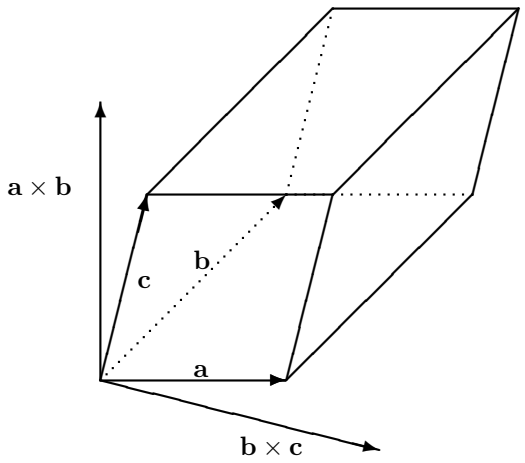
7. Suppose $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ are three vectors whose components are all integers. Can you conclude the volume of the parallelepiped determined from these three vectors will always be an integer?

   Hint: Consider what happens when you take the determinant of a matrix which has all integers.

8. Using the notion of the box product yielding either plus or minus the volume of the parallelepiped determined by the given three vectors, show that

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$$

   In other words, the dot and the cross can be switched as long as the order of the vectors remains the same. **Hint:** There are two ways to do this, by the coordinate description of the dot and cross product and by geometric reasoning. It is best if you use the geometric reasoning. Here is a picture which might help.



In this picture there is an angle between $\mathbf{a}\times\mathbf{b}$ and $\mathbf{c}$. Call it θ. Now if you take $|\mathbf{a}\times\mathbf{b}|\,|\mathbf{c}|\cos\theta$ this gives the area of the base of the parallelepiped determined by $\mathbf{a}$ and $\mathbf{b}$ times the altitude of the parallelepiped, $|\mathbf{c}|\cos\theta$. This is what is meant by the volume of the parallelepiped. It also equals $\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}$ by the geometric description of the dot product. Similarly, there is an angle between $\mathbf{b}\times\mathbf{c}$ and $\mathbf{a}$. Call it α. Then if you take $|\mathbf{b}\times\mathbf{c}|\,|\mathbf{a}|\cos\alpha$ this would equal the area of the face determined by the vectors $\mathbf{b}$ and $\mathbf{c}$ times the altitude measured from this face, $|\mathbf{a}|\cos\alpha$. Thus this also is the volume of the parallelepiped. and it equals $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}$. The picture is not completely representative. If you switch the labels of two of these vectors, say $\mathbf{b}$ and $\mathbf{c}$, explain why it is still the case that $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}=\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}$. You should draw a similar picture and explain why in this case you get -1 times the volume of the parallelepiped.

9. Discover a vector identity for$(\mathbf{u}\times\mathbf{v})\times\mathbf{w}$.

$$\begin{aligned}((\mathbf{u}\times\mathbf{v})\times\mathbf{w})_i &= \varepsilon_{ijk}\left(\mathbf{u}\times\mathbf{v}\right)_j w_k = \varepsilon_{ijk}\varepsilon_{jrs}u_r v_s w_k = \left(\delta_{is}\delta_{kr}-\delta_{ir}\delta_{ks}\right)u_r v_s w_k \\ &= u_k w_k v_i - u_i v_k w_k = \left(\mathbf{u}\cdot\mathbf{w}\right)v_i - \left(\mathbf{v}\cdot\mathbf{w}\right)u_i.\end{aligned}$$

Therefore, $(\mathbf{u}\times\mathbf{v})\times\mathbf{w}=\left(\mathbf{u}\cdot\mathbf{w}\right)\mathbf{v}-\left(\mathbf{v}\cdot\mathbf{w}\right)\mathbf{u}$.

10. Discover a vector identity for $\left(\mathbf{u}\times\mathbf{v}\right)\cdot\left(\mathbf{z}\times\mathbf{w}\right)$.

Start with $\varepsilon_{ijk}u_j v_k\varepsilon_{irs}z_r w_s$ and then go to work on it using the reduction identities for the permutation symbol.

11. Discover a vector identity for $\left(\mathbf{u}\times\mathbf{v}\right)\times\left(\mathbf{z}\times\mathbf{w}\right)$ in terms of box products.

You will save time if you use the identity for $\left(\mathbf{u}\times\mathbf{v}\right)\times\mathbf{w}$ or $\mathbf{u}\times\left(\mathbf{v}\times\mathbf{w}\right)$.

12. If A is a 3×3 matrix such that $A=\begin{pmatrix}\mathbf{u} & \mathbf{v} & \mathbf{w}\end{pmatrix}$ where these are the columns of the matrix, A. Show that $\det(A)=\varepsilon_{ijk}u_i v_j w_k$.

You can do this directly by expanding the determinant along the first column and then writing out the sum of nine terms occurring in $\varepsilon_{ijk}u_i v_j w_k$. That is, $\varepsilon_{ijk}u_i v_j w_k = \varepsilon_{123}u_1 v_2 w_3 + \varepsilon_{213}u_2 v_1 w_3 + \cdots$. Fill in the correct values for the permutation symbol and you will have an expression which can be compared with what you get when you expand the given determinant along the first column.

13. If A is a 3×3 matrix, show $\varepsilon_{rps}\det(A)=\varepsilon_{ijk}A_{ri}A_{pj}A_{sk}$.

From Problem 12, $\varepsilon_{123}\det(A)=\varepsilon_{ijk}A_{1i}A_{2j}A_{3k}$. Now by switching any pair of columns, you know from properties of determinants that this will change the sign of the determinant. Switching the same two indices in the permutation symbol on the left, also changes the sign of the expression on the left. Therefore, making a succession of these switches, you get the result desired.

14. Suppose A is a 3×3 matrix and $\det(A)\neq0$. Show using 13 that

$$\left(A^{-1}\right)_{ks}=\frac{1}{2\det(A)}\varepsilon_{rps}\varepsilon_{ijk}A_{pj}A_{ri}.$$

Just show the expression on the right acts like the ks^{th} entry of the inverse. Using the repeated index summation convention this amounts to showing

$$\frac{1}{2\det(A)}\varepsilon_{rps}\varepsilon_{ijk}A_{ri}A_{pj}A_{sl}=\delta_{kl}.$$

From Problem 13, $\varepsilon_{rps}\det(A) = \varepsilon_{ijl}A_{ri}A_{pj}A_{sl}$. Therefore,

$$6\det(A) = \varepsilon_{rps}\varepsilon_{rps}\det(A) = \varepsilon_{rps}\varepsilon_{ijl}A_{ri}A_{pj}A_{sl}$$

and so $\det(A) = \det\left(A^T\right)$.

Hence

$$\frac{1}{2\det(A)}\varepsilon_{rps}\varepsilon_{ijk}A_{ri}A_{pj}A_{sl} = \frac{1}{2\det(A)}\varepsilon_{ijl}\varepsilon_{ijk}\det(A) = \delta_{kl}$$

by the identity, $\varepsilon_{ijl}\varepsilon_{ijk} = 2\delta_{kl}$ done in an earlier problem.

Planes And Surfaces In $\mathbb{R}^n$

6.0.1 Outcomes

1. Find the angle between two lines.

2. Determine a point of intersection between a line and a surface.

3. Find the equation of a plane in 3 space given a point and a normal vector, three points, a sketch of a plane or a geometric description of the plane.

4. Determine the normal vector and the intercepts of a given plane.

5. Sketch the graph of a plane given its equation.

6. Determine the cosine of the angle between two planes.

7. Find the equation of a plane determined by lines.

8. Identify standard quadric surfaces given their functions or graphs.

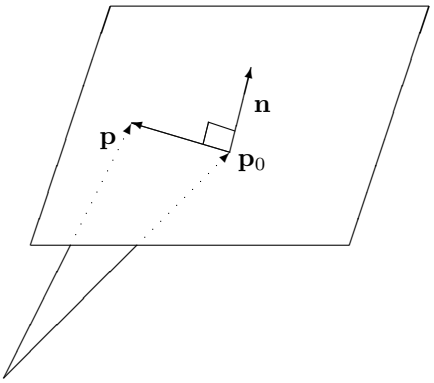
9. Sketch the graph of a quadric surface by identifying the intercepts, traces, sections, symmetry and boundedness or unboundedness of the surface.

6.1 Planes

You have an idea of what a plane is already. It is the span of some vectors. However, it can also be considered geometrically in terms of a dot product. To find the equation of a plane, you need two things, a point contained in the plane and a vector normal to the plane. Let $\mathbf{p}_0 = (x_0, y_0, z_0)$ denote the position vector of a point in the plane, let $\mathbf{p} = (x, y, z)$ be the position vector of an arbitrary point in the plane, and let $\mathbf{n}$ denote a vector normal to the plane. This means that

$$\mathbf{n} \cdot (\mathbf{p} - \mathbf{p}_0) = 0$$

whenever $\mathbf{p}$ is the position vector of a point in the plane. The following picture illustrates the geometry of this idea.



Expressed equivalently, the plane is just the set of all points $\mathbf{p}$ such that the vector, $\mathbf{p}-\mathbf{p}_0$ is perpendicular to the given normal vector, $\mathbf{n}$.

Example 6.1.1 *Find the equation of the plane with normal vector,* $\mathbf{n}=(1,2,3)$ *containing the point* $(2,-1,5)$.

From the above, the equation of this plane is just

$$(1,2,3)\cdot(x-2,y+1,z-3)=x-9+2y+3z=0$$

Example 6.1.2 $2x+4y-5z=11$ *is the equation of a plane. Find the normal vector and a point on this plane.*

You can write this in the form $2\left(x-\frac{11}{2}\right)+4(y-0)+(-5)(z-0)=0$. Therefore, a normal vector to the plane is $2\mathbf{i}+4\mathbf{j}-5\mathbf{k}$ and a point in this plane is $\left(\frac{11}{2},0,0\right)$. Of course there are many other points in the plane.

Definition 6.1.3 *Suppose two planes intersect. The angle between the planes is defined to be the angle between their normal vectors.*

Example 6.1.4 *Find the equation of the plane which contains the three points,*

$$(1,2,1),(3,-1,2),(4,2,1).$$

You just need to get a normal vector to this plane. This can be done by taking the cross products of the two vectors,

$$(3,-1,2)-(1,2,1) \text{ and } (4,2,1)-(1,2,1)$$

Thus a normal vector is $(2,-3,1)\times(3,0,0)=(0,3,9)$. Therefore, the equation of the plane is

$$0(x-1)+3(y-2)+9(z-1)=0$$

or $3y+9z=15$ which is the same as $y+3z=5$.

Example 6.1.5 *Find the equation of the plane which contains the three points,*

$$(1,2,1),(3,-1,2),(4,2,1)$$

another way.

Letting (x,y,z) be a point on the plane, the volume of the parallelepiped spanned by $(x,y,z)-(1,2,1)$ and the two vectors, $(2,-3,1)$ and $(3,0,0)$ must be equal to zero. Thus the equation of the plane is

$$\det\begin{pmatrix} 3 & 0 & 0 \\ 2 & -3 & 1 \\ x-1 & y-2 & z-1 \end{pmatrix} = 0.$$

Hence $-9z+15-3y=0$ and dividing by 3 yields the same answer as the above.

Proposition 6.1.6 *If $(a,b,c)\neq(0,0,0)$, then $ax+by+cz=d$ is the equation of a plane with normal vector $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$. Conversely, any plane can be written in this form.*

Proof: One of a,b,c is nonzero. Suppose for example that $c\neq 0$. Then the equation can be written as

$$a(x-0)+b(y-0)+c\left(z-\frac{d}{c}\right)=0$$

Therefore, $\left(0,0,\frac{d}{c}\right)$ is a point on the plane and a normal vector is $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$. The converse follows from the above discussion involving the point and a normal vector. This proves the proposition.

Example 6.1.7 *Find the equation of the plane which contains the three points,*

$$(1,2,1),(3,-1,2),(4,2,1)$$

another way.

You need to find numbers, a,b,c,d not all zero such that each of the given three points satisfies the equation, $ax+by+cz=d$. Then you must have for (x,y,z) a point on this plane,

$$\begin{gathered} a+2b+c-d=0, \\ 3a-b+2c-d=0, \\ 4a+2b+c-d=0, \\ xa+yb+zc-d=0. \end{gathered}$$

You need a nonzero solution to the above system of four equations for the unknowns, a,b,c, and d. Therefore,

$$\det\begin{pmatrix} 1 & 2 & 1 & -1 \\ 3 & -1 & 2 & -1 \\ 4 & 2 & 1 & -1 \\ x & y & z & -1 \end{pmatrix} = 0$$

because the matrix sends a nonzero vector, $(a,b,c,-d)$ to zero and is therefore, not one to one. Consequently from Theorem 3.2.1 on Page 61, its determinant equals zero. Hence upon evaluating the determinant, $-15+9z+3y=0$ which reduces to $3z+y=5$.

Example 6.1.8 *Find the equation of the plane containing the points $(1,2,3)$ and the line $(0,1,1)+t(2,1,2)=(x,y,z)$.*

There are several ways to do this. One is to find three points and use any of the above procedures. Let $t=0$ and then let $t=1$ to get two points on the line. This yields $(1,2,3),(0,1,1)$, and $(2,2,3)$. Then the equation of the plane is

$$\det\begin{pmatrix} x & y & z & -1 \\ 1 & 2 & 3 & -1 \\ 0 & 1 & 1 & -1 \\ 2 & 2 & 3 & -1 \end{pmatrix} = 2y-z-1=0.$$

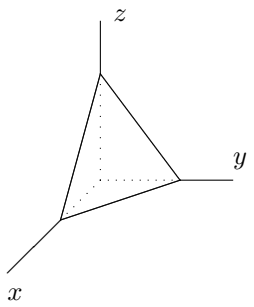
Example 6.1.9 *Find the equation of the plane which contains the two lines, given by the following parametric expressions in which* $t \in \mathbb{R}$.

$$(2t, 1+t, 1+2t) = (x, y, z)\,,\ (2t+2, 1, 3+2t) = (x, y, z)$$

Note first that you don't know there even is such a plane. However, if there is, you could find it by obtaining three points, two on one line and one on another and then using any of the above procedures for finding the plane. From the first line, two points are $(0, 1, 1)$ and $(2, 2, 3)$ while a third point can be obtained from second line, $(2, 1, 3)$. You need a normal vector and then use any of these points. To get a normal vector, form $(2, 0, 2) \times (2, 1, 2) = (-2, 0, 2)$. Therefore, the plane is $-2x + 0\,(y-1) + 2\,(z-1) = 0$. This reduces to $z - x = 1$. If there is a plane, this is it. Now you can simply verify that both of the lines are really in this plane. From the first, $(1+2t) - 2t = 1$ and the second, $(3+2t) - (2t+2) = 1$ so both lines lie in the plane.

One way to understand how a plane looks is to connect the points where it intercepts the x, y, and z axes. This allows you to visualize the plane somewhat and is a good way to sketch the plane. Not surprisingly these points are called intercepts.

Example 6.1.10 *Sketch the plane which has intercepts* $(2, 0, 0)\,, (0, 3, 0)\,,$ *and* $(0, 0, 4)$.



You see how connecting the intercepts gives a fairly good geometric description of the plane. These lines which connect the intercepts are also called the traces of the plane. Thus the line which joins $(0, 3, 0)$ to $(0, 0, 4)$ is the intersection of the plane with the yz plane. It is the trace on the yz plane.

Example 6.1.11 *Identify the intercepts of the plane,* $3x - 4y + 5z = 11$.

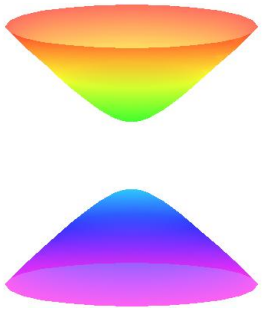
The easy way to do this is to divide both sides by 11.

$$\frac{x}{(11/3)} + \frac{y}{(-11/4)} + \frac{z}{(11/5)} = 1$$

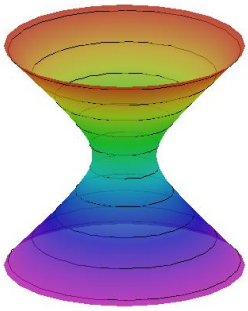
The intercepts are $(11/3, 0, 0)\,, (0, -11/4, 0)$ and $(0, 0, 11/5)$. You can see this by letting both y and z equal to zero to find the point on the x axis which is intersected by the plane. The other axes are handled similarly.

6.2 Quadric Surfaces

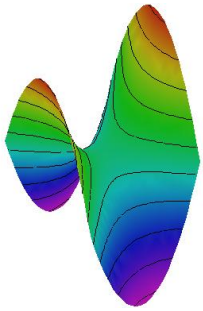
In the above it was shown that the equation of an arbitrary plane is an equation of the form $ax + by + cz = d$. Such equations are called level surfaces. There are some standard level surfaces which involve certain variables being raised to a power of 2 which are sufficiently important that they are given names, usually involving the portentous semi-word "oid". These are graphed below using Maple, a computer algebra system.



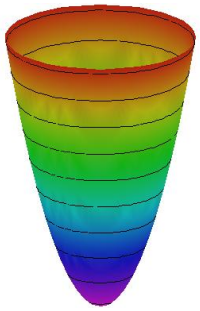
$z^2/a^2 - x^2/b^2 - y^2/c^2 = 1$
hyperboloid of two sheets



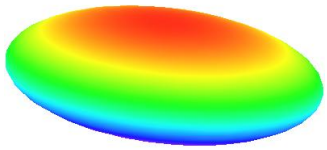
$x^2/b^2 + y^2/c^2 - z^2/a^2 = 1$
hyperboloid of one sheet



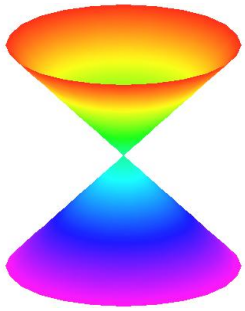
$z = x^2/a^2 - y^2/b^2$
hyperbolic paraboloid



$z = x^2/a^2 + y^2/b^2$
elliptic paraboloid



$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$
ellipsoid



$z^2/a^2 = x^2/b^2 + y^2/c^2$
elliptic cone

Why do the graphs of these level surfaces look the way they do? Consider first the hyperboloid of two sheets. The equation defining this surface can be written in the form

$$\frac{z^2}{a^2} - 1 = \frac{x^2}{b^2} + \frac{y^2}{c^2}.$$

Suppose you fix a value for z. What ordered pairs, (x, y) will satisfy the equation? If $\frac{z^2}{a^2} < 1$, there is no such ordered pair because the above equation would require a negative number to equal a nonnegative one. This is why there is a gap and there are two sheets. If $\frac{z^2}{a^2} > 1$, then the above equation is the equation for an ellipse. That is why if you slice the graph by letting $z = z_0$ the result is an ellipse in the plane $z = z_0$.

Consider the hyperboloid of one sheet.

$$\frac{x^2}{b^2} + \frac{y^2}{c^2} = 1 + \frac{z^2}{a^2}.$$

This time, it doesn't matter what value z takes. The resulting equation for (x, y) is an ellipse.

Similar considerations apply to the elliptic paraboloid as long as $z > 0$ and the ellipsoid. The elliptic cone is like the hyperboloid of two sheets without the 1. Therefore, z can have any value. In case $z = 0$, $(x, y) = (0, 0)$. Viewed from the side, it appears straight, not curved like the hyperboloid of two sheets.This is because if (x, y, z) is a point on the surface, then if t is a scalar, it follows (tx, ty, tz) is also on this surface.

The most interesting of these graphs is the hyperbolic paraboloid[1], $z = \frac{x^2}{a^2} - \frac{y^2}{b^2}$. If $z > 0$ this is the equation of a hyperbola which opens to the right and left while if $z < 0$ it is a hyperbola which opens up and down. As z passes from positive to negative, the hyperbola changes type and this is what yields the shape shown in the picture.

Not surprisingly, you can find intercepts and traces of quadric surfaces just as with planes.

Example 6.2.1 *Find the trace on the xy plane of the hyperbolic paraboloid,* $z = x^2 - y^2$.

This occurs when $z = 0$ and so this reduces to $y^2 = x^2$. In other words, this trace is just the two straight lines, $y = x$ and $y = -x$.

Example 6.2.2 *Find the intercepts of the ellipsoid,* $x^2 + 2y^2 + 4z^2 = 9$.

To find the intercept on the x axis, let $y = z = 0$ and this yields $x = \pm 3$. Thus there are two intercepts, $(3, 0, 0)$ and $(-3, 0, 0)$. The other intercepts are left for you to find. You can see this is an aid in graphing the quadric surface. The surface is said to be bounded if there is some number, C such that whenever, (x, y, z) is a point on the surface, $\sqrt{x^2 + y^2 + z^2} < C$. The surface is called unbounded if no such constant, C exists. Ellipsoids are bounded but the other quadric surfaces are not bounded.

Example 6.2.3 *Why is the hyperboloid of one sheet,* $x^2 + 2y^2 - z^2 = 1$ *unbounded?*

Let z be very large. Does there correspond (x, y) such that (x, y, z) is a point on the hyperboloid of one sheet? Certainly. Simply pick any (x, y) on the ellipse $x^2 + 2y^2 = 1 + z^2$. Then $\sqrt{x^2 + y^2 + z^2}$ is large, at lest as large as z. Thus it is unbounded.

You can also find intersections between lines and surfaces.

Example 6.2.4 *Find the points of intersection of the line* $(x, y, z) = (1 + t, 1 + 2t, 1 + t)$ *with the surface,* $z = x^2 + y^2$.

First of all, there is no guarantee there is any intersection at all. But if it exists, you have only to solve the equation for t

$$1 + t = (1 + t)^2 + (1 + 2t)^2$$

[1]It is traditional to refer to this as a hyperbolic paraboloid. Not a parabolic hyperboloid.

This occurs at the two values of $t = -\frac{1}{2} + \frac{1}{10}\sqrt{5}, t = -\frac{1}{2} - \frac{1}{10}\sqrt{5}$. Therefore, the two points are

$$(1,1,1) + \left(-\frac{1}{2} + \frac{1}{10}\sqrt{5}\right)(1,2,1), \text{ and } (1,1,1) + \left(-\frac{1}{2} - \frac{1}{10}\sqrt{5}\right)(1,2,1)$$

That is

$$\left(\frac{1}{2} + \frac{1}{10}\sqrt{5}, \frac{1}{5}\sqrt{5}, \frac{1}{2} + \frac{1}{10}\sqrt{5}\right), \left(\frac{1}{2} - \frac{1}{10}\sqrt{5}, -\frac{1}{5}\sqrt{5}, \frac{1}{2} - \frac{1}{10}\sqrt{5}\right).$$

6.3 Exercises

1. Determine whether the lines $(1,1,2) + t(1,0,3)$ and $(4,1,3) + t(3,0,1)$ have a point of intersection. If they do, find the cosine of the angle between the two lines. If they do not intersect, explain why they do not.

2. Determine whether the lines $(1,1,2) + t(1,0,3)$ and $(4,2,3) + t(3,0,1)$ have a point of intersection. If they do, find the cosine of the angle between the two lines. If they do not intersect, explain why they do not.

3. Find where the line $(1,0,1) + t(1,2,1)$ intersects the surface $x^2 + y^2 + z^2 = 9$ if possible. If there is no intersection, explain why.

4. Find a parametric equation for the line through the points $(2,3,4,5)$ and $(-2,3,0,1)$.

5. Find the equation of a line through $(1,2,3,0)$ which has direction vector, $(2,1,3,1)$.

6. Let $(x,y) = (2\cos(t), 2\sin(t))$ where $t \in [0, 2\pi]$. Describe the set of points encountered as t changes.

7. Let $(x,y,z) = (2\cos(t), 2\sin(t), t)$ where $t \in \mathbb{R}$. Describe the set of points encountered as t changes.

8. If there is a plane which contains the two lines, $(2t+2, 1+t, 3+2t) = (x,y,z)$ and $(4+t, 3+2t, 4+t) = (x,y,z)$ find it. If there is no such plane tell why.

9. If there is a plane which contains the two lines, $(2t+4, 1+t, 3+2t) = (x,y,z)$ and $(4+t, 3+2t, 4+t) = (x,y,z)$ find it. If there is no such plane tell why.

10. Find the equation of the plane which contains the three points $(1,-2,3), (2,3,4)$, and $(3,1,2)$.

11. Find the equation of the plane which contains the three points $(1,2,3), (2,0,4)$, and $(3,1,2)$.

12. Find the equation of the plane which contains the three points $(0,2,3), (2,3,4)$, and $(3,5,2)$.

13. Find the equation of the plane which contains the three points $(1,2,3), (0,3,4)$, and $(3,6,2)$.

14. Find the equation of the plane having a normal vector, $5\mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$ which contains the point $(2,1,3)$.

15. Find the equation of the plane having a normal vector, $\mathbf{i} + 2\mathbf{j} - 4\mathbf{k}$ which contains the point $(2,0,1)$.

16. Find the equation of the plane having a normal vector, $2\mathbf{i}+\mathbf{j}-6\mathbf{k}$ which contains the point $(1,1,2)$.

17. Find the equation of the plane having a normal vector, $\mathbf{i}+2\mathbf{j}-3\mathbf{k}$ which contains the point $(1,0,3)$.

18. Find the cosine of the angle between the two planes $2x+3y-z=11$ and $3x+y+2z=9$.

19. Find the cosine of the angle between the two planes $x+3y-z=11$ and $2x+y+2z=9$.

20. Find the cosine of the angle between the two planes $2x+y-z=11$ and $3x+5y+2z=9$.

21. Find the cosine of the angle between the two planes $x+3y+z=11$ and $3x+2y+2z=9$.

22. Determine the intercepts and sketch the plane $3x-2y+z=4$.

23. Determine the intercepts and sketch the plane $x-2y+z=2$.

24. Determine the intercepts and sketch the plane $x+y+z=3$.

25. Based on an analogy with the above pictures, sketch or otherwise describe the graph of $y=\frac{x^2}{a^2}-\frac{z^2}{b^2}$.

26. Based on an analogy with the above pictures, sketch or otherwise describe the graph of $\frac{z^2}{b^2}+\frac{y^2}{c^2}=1+\frac{x^2}{a^2}$.

27. The equation of a cone is $z^2=x^2+y^2$. Suppose this cone is intersected with the plane, $z=ay+1$. Consider the projection of the intersection of the cone with this plane. This means $\left\{(x,y):(ay+1)^2=x^2+y^2\right\}$. Show this sometimes results in a parabola, sometimes a hyperbola, and sometimes an ellipse depending on a.

28. Find the intercepts of the quadric surface, $x^2+4y^2-z^2=4$ and sketch the surface.

29. Find the intercepts of the quadric surface, $x^2-\left(4y^2+z^2\right)=4$ and sketch the surface.

30. Find the intersection of the line $(x,y,z)=(1+t,t,3t)$ with the surface, $x^2/9+y^2/4+z^2/16=1$ if possible.

Part III

Vector Calculus

Vector Valued Functions

7.0.1 Outcomes

1. Identify the domain of a vector function.
2. Represent combinations of multivariable functions algebraically.
3. Evaluate the limit of a function of several variables or show that it does not exist.
4. Determine whether a function is continuous at a given point. Give examples of continuous functions.
5. Recall and apply the extreme value theorem.

7.1 Vector Valued Functions

Vector valued functions have values in $\mathbb{R}^p$ where p is an integer at least as large as 1. Here is a simple example which is obviously of interest.

Example 7.1.1 *A rocket is launched from the rotating earth. You could define a function having values in $\mathbb{R}^3$ as $(r(t), \theta(t), \phi(t))$ where $r(t)$ is the distance of the center of mass of the rocket from the center of the earth, $\theta(t)$ is the longitude, and $\phi(t)$ is the latitude of the rocket.*

Example 7.1.2 *Let $\mathbf{f}(x,y) = \left(\sin xy, y^3 + x, x^4\right)$. Then $\mathbf{f}$ is a function defined on $\mathbb{R}^2$ which has values in $\mathbb{R}^3$. For example, $\mathbf{f}(1,2) = (\sin 2, 9, 16)$.*

As usual, $D(\mathbf{f})$ denotes the domain of the function, $\mathbf{f}$ which is written in bold face because it will possibly have values in $\mathbb{R}^p$. When $D(\mathbf{f})$ is not specified, it will be understood that the domain of $\mathbf{f}$ consists of those things for which $\mathbf{f}$ makes sense.

Example 7.1.3 *Let $\mathbf{f}(x,y,z) = \left(\frac{x+y}{z}, \sqrt{1-x^2}, y\right)$. Then $D(\mathbf{f})$ would consist of the set of all (x,y,z) such that $|x| \leq 1$ and $z \neq 0$.*

There are many ways to make new functions from old ones.

Definition 7.1.4 *Let $\mathbf{f}, \mathbf{g}$ be functions with values in $\mathbb{R}^p$. Let a, b be elements of $\mathbb{R}$ (scalars). Then $a\mathbf{f} + b\mathbf{g}$ is the name of a function whose domain is $D(\mathbf{f}) \cap D(\mathbf{g})$ which is defined as*

$$(a\mathbf{f} + b\mathbf{g})(\mathbf{x}) = a\mathbf{f}(\mathbf{x}) + b\mathbf{g}(\mathbf{x}).$$

$\mathbf{f} \cdot \mathbf{g}$ or $(\mathbf{f}, \mathbf{g})$ is the name of a function whose domain is $D(\mathbf{f}) \cap D(\mathbf{g})$ which is defined as

$$(\mathbf{f}, \mathbf{g})(\mathbf{x}) \equiv \mathbf{f} \cdot \mathbf{g}(\mathbf{x}) \equiv \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x}).$$

If $\mathbf{f}$ *and* $\mathbf{g}$ *have values in* $\mathbb{R}^3$*, define a new function,* $\mathbf{f} \times \mathbf{g}$ *by*

$$\mathbf{f} \times \mathbf{g}(t) \equiv \mathbf{f}(t) \times \mathbf{g}(t).$$

If $\mathbf{f} : D(\mathbf{f}) \to X$ *and* $\mathbf{g} : X \to Y$*, then* $\mathbf{g} \circ \mathbf{f}$ *is the name of a function whose domain is*

$$\{\mathbf{x} \in D(\mathbf{f}) : \mathbf{f}(\mathbf{x}) \in D(\mathbf{g})\}$$

which is defined as

$$\mathbf{g} \circ \mathbf{f}(\mathbf{x}) \equiv \mathbf{g}(\mathbf{f}(\mathbf{x})).$$

This is called the composition of the two functions.

You should note that $\mathbf{f}(\mathbf{x})$ is not a function. It is the value of the function at the point, $\mathbf{x}$. The name of the function is $\mathbf{f}$. Nevertheless, people often write $\mathbf{f}(\mathbf{x})$ to denote a function and it doesn't cause too many problems in beginning courses. When this is done, the variable, $\mathbf{x}$ should be considered as a generic variable free to be anything in $D(\mathbf{f})$. I will use this slightly sloppy abuse of notation whenever convenient.

Example 7.1.5 *Let* $\mathbf{f}(t) \equiv (t, 1+t, 2)$ *and* $\mathbf{g}(t) \equiv \left(t^2, t, t\right)$*. Then* $\mathbf{f} \cdot \mathbf{g}$ *is the name of the function satisfying*

$$\mathbf{f} \cdot \mathbf{g}(t) = \mathbf{f}(t) \cdot \mathbf{g}(t) = t^3 + t + t^2 + 2t = t^3 + t^2 + 3t$$

Note that in this case is was assumed the domains of the functions consisted of all of $\mathbb{R}$ because this was the set on which the two both made sense. Also note that $\mathbf{f}$ and $\mathbf{g}$ map $\mathbb{R}$ into $\mathbb{R}^3$ but $\mathbf{f} \cdot \mathbf{g}$ maps $\mathbb{R}$ into $\mathbb{R}$.

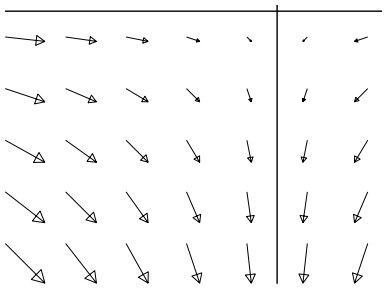
Example 7.1.6 *Suppose* $\mathbf{f}(t) = \left(2t, 1+t^2\right)$ *and* $g{:}\mathbb{R}^2 \to \mathbb{R}$ *is given by* $g(x,y) \equiv x + y$*. Then* $g \circ \mathbf{f} : \mathbb{R} \to \mathbb{R}$ *and*

$$g \circ \mathbf{f}(t) = g(\mathbf{f}(t)) = g\left(2t, 1+t^2\right) = 1 + 2t + t^2.$$

7.2 Vector Fields

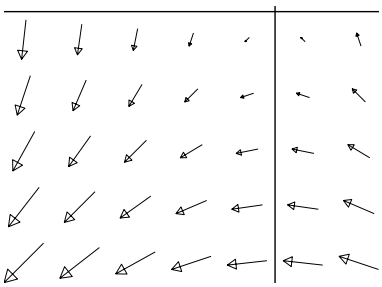
Some people find it useful to try and draw pictures to illustrate a vector valued function. This can be a very useful idea in the case where the function takes points in $D \subseteq \mathbb{R}^2$ and delivers a vector in $\mathbb{R}^2$. For many points, $(x,y) \in D$, you draw an arrow of the appropriate length and direction with its tail at (x,y). The picture of all these arrows can give you an understanding of what is happening. For example if the vector valued function gives the velocity of a fluid at the point, (x,y), the picture of these arrows can give an idea of the motion of the fluid. When they are long the fluid is moving fast, when they are short, the fluid is moving slowly the direction of these arrows is an indication of the direction of motion. The only sensible way to produce such a picture is with a computer. Otherwise, it becomes a worthless exercise in busy work. Furthermore, it is of limited usefulness in three dimensions because in three dimensions such pictures are too cluttered to convey much insight.

Example 7.2.1 *Draw a picture of the vector field,* $(-x, y)$ *which gives the velocity of a fluid flowing in two dimensions.*



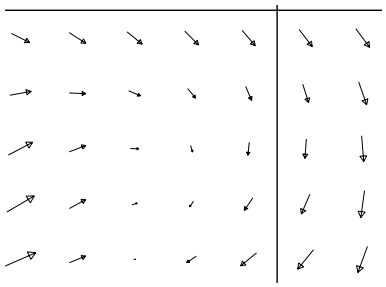
In this example, drawn by Maple, you can see how the arrows indicate the motion of this fluid.

Example 7.2.2 *Draw a picture of the vector field* (y, x) *for the velocity of a fluid flowing in two dimensions.*



So much for art. Get the computer to do it and it can be useful. If you try to do it, you will mainly waste time.

Example 7.2.3 *Draw a picture of the vector field* $(y\cos(x)+1, x\sin(y)-1)$ *for the velocity of a fluid flowing in two dimensions.*



7.3 Continuous Functions

What was done in beginning calculus for scalar functions is generalized here to include the case of a vector valued function.

Definition 7.3.1 *A function* $\mathbf{f} : D(\mathbf{f}) \subseteq \mathbb{R}^p \to \mathbb{R}^q$ *is continuous at* $\mathbf{x} \in D(\mathbf{f})$ *if for each* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that whenever* $\mathbf{y} \in D(\mathbf{f})$ *and*

$$|\mathbf{y} - \mathbf{x}| < \delta$$

it follows that

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon.$$

$\mathbf{f}$ *is continuous if it is continuous at every point of* $D(\mathbf{f})$.

Note the total similarity to the scalar valued case.

7.3.1 Sufficient Conditions For Continuity

The next theorem is a fundamental result which will allow us to worry less about the $\varepsilon\ \delta$ definition of continuity.

Theorem 7.3.2 *The following assertions are valid.*

1. *The function,* $a\mathbf{f} + b\mathbf{g}$ *is continuous at* $\mathbf{x}$ *whenever* $\mathbf{f}$, $\mathbf{g}$ *are continuous at* $\mathbf{x} \in D(\mathbf{f}) \cap D(\mathbf{g})$ *and* $a, b \in \mathbb{R}$.

2. *If* $\mathbf{f}$ *is continuous at* $\mathbf{x}$, $\mathbf{f}(\mathbf{x}) \in D(\mathbf{g}) \subseteq \mathbb{R}^p$, *and* $\mathbf{g}$ *is continuous at* $\mathbf{f}(\mathbf{x})$, *then* $\mathbf{g} \circ \mathbf{f}$ *is continuous at* $\mathbf{x}$.

3. *If* $\mathbf{f} = (f_1, \cdots, f_q) : D(\mathbf{f}) \to \mathbb{R}^q$, *then* $\mathbf{f}$ *is continuous if and only if each* f_k *is a continuous real valued function.*

4. *The function* $f : \mathbb{R}^p \to \mathbb{R}$, *given by* $f(\mathbf{x}) = |\mathbf{x}|$ *is continuous.*

The proof of this theorem is in the last section of this chapter. Its conclusions are not surprising. For example the first claim says that $(a\mathbf{f} + b\mathbf{g})(\mathbf{y})$ is close to $(a\mathbf{f} + b\mathbf{g})(\mathbf{x})$ when $\mathbf{y}$ is close to $\mathbf{x}$ provided the same can be said about $\mathbf{f}$ and $\mathbf{g}$. For the second claim, if $\mathbf{y}$ is close to $\mathbf{x}$, $\mathbf{f}(\mathbf{x})$ is close to $\mathbf{f}(\mathbf{y})$ and so by continuity of $\mathbf{g}$ at $\mathbf{f}(\mathbf{x})$, $\mathbf{g}(\mathbf{f}(\mathbf{y}))$ is close to $\mathbf{g}(\mathbf{f}(\mathbf{x}))$. To see the third claim is likely, note that closeness in $\mathbb{R}^p$ is the same as closeness in each coordinate. The fourth claim is immediate from the triangle inequality.

For functions defined on $\mathbb{R}^n$, there is a notion of polynomial just as there is for functions defined on $\mathbb{R}$.

Definition 7.3.3 *Let* α *be an* n *dimensional multi-index. This means*

$$\alpha = (\alpha_1, \cdots, \alpha_n)$$

where each α_i *is a natural number or zero. Also, let*

$$|\alpha| \equiv \sum_{i=1}^{n} |\alpha_i|$$

The symbol, $\mathbf{x}^\alpha$ *means*

$$\mathbf{x}^\alpha \equiv x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_3^{\alpha_n}.$$

An n *dimensional polynomial of degree* m *is a function of the form*

$$p(\mathbf{x}) = \sum_{|\alpha| \le m} d_\alpha \mathbf{x}^\alpha.$$

where the d_α *are real numbers.*

The above theorem implies that polynomials are all continuous.

7.4 Limits Of A Function

As in the case of scalar valued functions of one variable, a concept closely related to continuity is that of the **limit of a function**. The notion of limit of a function makes sense at points, $\mathbf{x}$, which are limit points of $D(\mathbf{f})$ and this concept is defined next.

Definition 7.4.1 *Let $A \subseteq \mathbb{R}^m$ be a set. A point, $\mathbf{x}$, is a limit point of A if $B(\mathbf{x}, r)$ contains infinitely many points of A for every $r > 0$.*

Definition 7.4.2 *Let $\mathbf{f} : D(\mathbf{f}) \subseteq \mathbb{R}^p \to \mathbb{R}^q$ be a function and let $\mathbf{x}$ be a **limit point** of $D(\mathbf{f})$. Then*

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$$

if and only if the following condition holds. For all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < |\mathbf{y} - \mathbf{x}| < \delta, \text{ and } \mathbf{y} \in D(\mathbf{f})$$

then,

$$|\mathbf{L} - \mathbf{f}(\mathbf{y})| < \varepsilon.$$

Theorem 7.4.3 *If $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$ and $\lim_{y\to x} \mathbf{f}(\mathbf{y}) = \mathbf{L}_1$, then $\mathbf{L} = \mathbf{L}_1$.*

Proof: Let $\varepsilon > 0$ be given. There exists $\delta > 0$ such that if $0 < |\mathbf{y} - \mathbf{x}| < \delta$ and $\mathbf{y} \in D(\mathbf{f})$, then

$$|\mathbf{f}(\mathbf{y}) - \mathbf{L}| < \varepsilon, \ |\mathbf{f}(\mathbf{y}) - \mathbf{L}_1| < \varepsilon.$$

Pick such a $\mathbf{y}$. There exists one because $\mathbf{x}$ is a limit point of $D(\mathbf{f})$. Then

$$|\mathbf{L} - \mathbf{L}_1| \leq |\mathbf{L} - \mathbf{f}(\mathbf{y})| + |\mathbf{f}(\mathbf{y}) - \mathbf{L}_1| < \varepsilon + \varepsilon = 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this shows $\mathbf{L} = \mathbf{L}_1$.

As in the case of functions of one variable, one can define what it means for $\lim_{\mathbf{y}\to\mathbf{x}} f(\mathbf{x}) = \pm\infty$.

Definition 7.4.4 *If $f(\mathbf{x}) \in \mathbb{R}$, $\lim_{\mathbf{y}\to\mathbf{x}} f(\mathbf{x}) = \infty$ if for every number l, there exists $\delta > 0$ such that whenever $|\mathbf{y} - \mathbf{x}| < \delta$ and $\mathbf{y} \in D(\mathbf{f})$, then $f(\mathbf{x}) > l$.*

The following theorem is just like the one variable version of calculus.

Theorem 7.4.5 *Suppose $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$ and $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{g}(\mathbf{y}) = \mathbf{K}$ where $\mathbf{K}, \mathbf{L} \in \mathbb{R}^q$. Then if a, $b \in \mathbb{R}$,*

$$\lim_{\mathbf{y}\to\mathbf{x}} (a\mathbf{f}(\mathbf{y}) + b\mathbf{g}(\mathbf{y})) = a\mathbf{L} + b\mathbf{K}, \tag{7.1}$$

$$\lim_{y\to x} \mathbf{f} \cdot \mathbf{g}(y) = \mathbf{L} \cdot \mathbf{K} \tag{7.2}$$

and if g is scalar valued with $\lim_{\mathbf{y}\to\mathbf{x}} g(\mathbf{y}) = K \neq 0$,

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y})\, g(\mathbf{y}) = \mathbf{L}K. \tag{7.3}$$

Also, if $\mathbf{h}$ is a continuous function defined near $\mathbf{L}$, then

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{h} \circ \mathbf{f}(\mathbf{y}) = \mathbf{h}(\mathbf{L}). \tag{7.4}$$

Suppose $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$. If $|\mathbf{f}(\mathbf{y}) - \mathbf{b}| \leq r$ for all $\mathbf{y}$ sufficiently close to $\mathbf{x}$, then $|\mathbf{L} - \mathbf{b}| \leq r$ also.

Proof: The proof of 7.1 is left for you. It is like a corresponding theorem for continuous functions. Now 7.2is to be verified. Let $\varepsilon > 0$ be given. Then by the triangle inequality,

$$\begin{aligned}|\mathbf{f}\cdot\mathbf{g}(\mathbf{y})-\mathbf{L}\cdot\mathbf{K}| &\leq |\mathbf{fg}(\mathbf{y})-\mathbf{f}(\mathbf{y})\cdot\mathbf{K}|+|\mathbf{f}(\mathbf{y})\cdot\mathbf{K}-\mathbf{L}\cdot\mathbf{K}|\\ &\leq |\mathbf{f}(\mathbf{y})|\,|\mathbf{g}(\mathbf{y})-\mathbf{K}|+|\mathbf{K}|\,|\mathbf{f}(\mathbf{y})-\mathbf{L}|.\end{aligned}$$

There exists δ_1 such that if $0<|\mathbf{y}-\mathbf{x}|<\delta_1$ and $\mathbf{y}\in D(\mathbf{f})$, then

$$|\mathbf{f}(\mathbf{y})-\mathbf{L}|<1,$$

and so for such $\mathbf{y}$, the triangle inequality implies, $|\mathbf{f}(\mathbf{y})|<1+|\mathbf{L}|$. Therefore, for $0<|\mathbf{y}-\mathbf{x}|<\delta_1$,

$$|\mathbf{f}\cdot\mathbf{g}(\mathbf{y})-\mathbf{L}\cdot\mathbf{K}|\leq(1+|\mathbf{K}|+|\mathbf{L}|)\left[|\mathbf{g}(\mathbf{y})-\mathbf{K}|+|\mathbf{f}(\mathbf{y})-\mathbf{L}|\right]. \tag{7.5}$$

Now let $0<\delta_2$ be such that if $\mathbf{y}\in D(\mathbf{f})$ and $0<|\mathbf{x}-\mathbf{y}|<\delta_2$,

$$|\mathbf{f}(\mathbf{y})-\mathbf{L}|<\frac{\varepsilon}{2(1+|\mathbf{K}|+|\mathbf{L}|)},\ |\mathbf{g}(\mathbf{y})-\mathbf{K}|<\frac{\varepsilon}{2(1+|\mathbf{K}|+|\mathbf{L}|)}.$$

Then letting $0<\delta\leq\min(\delta_1,\delta_2)$, it follows from 7.5 that

$$|\mathbf{f}\cdot\mathbf{g}(\mathbf{y})-\mathbf{L}\cdot\mathbf{K}|<\varepsilon$$

and this proves 7.2.

The proof of 7.3 is left to you.

Consider 7.4. Since $\mathbf{h}$ is continuous near $\mathbf{L}$, it follows that for $\varepsilon>0$ given, there exists $\eta>0$ such that if $|\mathbf{y}-\mathbf{L}|<\eta$, then

$$|\mathbf{h}(\mathbf{y})-\mathbf{h}(\mathbf{L})|<\varepsilon$$

Now since $\lim_{\mathbf{y}\to\mathbf{x}}\mathbf{f}(\mathbf{y})=\mathbf{L}$, there exists $\delta>0$ such that if $0<|\mathbf{y}-\mathbf{x}|<\delta$, then

$$|\mathbf{f}(\mathbf{y})-\mathbf{L}|<\eta.$$

Therefore, if $0<|\mathbf{y}-\mathbf{x}|<\delta$,

$$|\mathbf{h}(\mathbf{f}(\mathbf{y}))-\mathbf{h}(\mathbf{L})|<\varepsilon.$$

It only remains to verify the last assertion. Assume $|\mathbf{f}(\mathbf{y})-\mathbf{b}|\leq r$. It is required to show that $|\mathbf{L}-\mathbf{b}|\leq r$. If this is not true, then $|\mathbf{L}-\mathbf{b}|>r$. Consider $B(\mathbf{L},|\mathbf{L}-\mathbf{b}|-r)$. Since $\mathbf{L}$ is the limit of $\mathbf{f}$, it follows $\mathbf{f}(\mathbf{y})\in B(\mathbf{L},|\mathbf{L}-\mathbf{b}|-r)$ whenever $\mathbf{y}\in D(\mathbf{f})$ is close enough to $\mathbf{x}$. Thus, by the triangle inequality,

$$|\mathbf{f}(\mathbf{y})-\mathbf{L}|<|\mathbf{L}-\mathbf{b}|-r$$

and so

$$\begin{aligned}r &< |\mathbf{L}-\mathbf{b}|-|\mathbf{f}(\mathbf{y})-\mathbf{L}|\leq||\mathbf{b}-\mathbf{L}|-|\mathbf{f}(\mathbf{y})-\mathbf{L}||\\ &\leq |\mathbf{b}-\mathbf{f}(\mathbf{y})|,\end{aligned}$$

a contradiction to the assumption that $|\mathbf{b}-\mathbf{f}(\mathbf{y})|\leq r$.

Theorem 7.4.6 *For $\mathbf{f}:D(\mathbf{f})\to\mathbb{R}^q$ and $\mathbf{x}\in D(\mathbf{f})$ a limit point of $D(\mathbf{f})$, $\mathbf{f}$ is continuous at $\mathbf{x}$ if and only if*

$$\lim_{\mathbf{y}\to\mathbf{x}}\mathbf{f}(\mathbf{y})=\mathbf{f}(\mathbf{x}).$$

Proof: First suppose $\mathbf{f}$ is continuous at $\mathbf{x}$ a limit point of $D(\mathbf{f})$. Then for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $|\mathbf{y}-\mathbf{x}| < \delta$ and $\mathbf{y} \in D(\mathbf{f})$, then $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$. In particular, this holds if $0 < |\mathbf{x}-\mathbf{y}| < \delta$ and this is just the definition of the limit. Hence $\mathbf{f}(\mathbf{x}) = \lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y})$.

Next suppose $\mathbf{x}$ is a limit point of $D(\mathbf{f})$ and $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x})$. This means that if $\varepsilon > 0$ there exists $\delta > 0$ such that for $0 < |\mathbf{x}-\mathbf{y}| < \delta$ and $\mathbf{y} \in D(\mathbf{f})$, it follows $|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})| < \varepsilon$. However, if $\mathbf{y} = \mathbf{x}$, then $|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})| = |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x})| = 0$ and so whenever $\mathbf{y} \in D(\mathbf{f})$ and $|\mathbf{x}-\mathbf{y}| < \delta$, it follows $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$, showing $\mathbf{f}$ is continuous at $\mathbf{x}$.

The following theorem is important.

Theorem 7.4.7 *Suppose* $\mathbf{f} : D(\mathbf{f}) \to \mathbb{R}^q$. *Then for* $\mathbf{x}$ *a limit point of* $D(\mathbf{f})$,

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L} \tag{7.6}$$

if and only if

$$\lim_{\mathbf{y}\to\mathbf{x}} f_k(\mathbf{y}) = L_k \tag{7.7}$$

where $\mathbf{f}(\mathbf{y}) \equiv (f_1(\mathbf{y}), \cdots, f_p(\mathbf{y}))$ *and* $\mathbf{L} \equiv (L_1, \cdots, L_p)$.

In the case where $q = 3$ *and* $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$ *and* $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{g}(\mathbf{y}) = \mathbf{K}$, *then*

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) \times \mathbf{g}(\mathbf{y}) = \mathbf{L} \times \mathbf{K}. \tag{7.8}$$

Proof: Suppose 7.6. Then letting $\varepsilon > 0$ be given there exists $\delta > 0$ such that if $0 < |\mathbf{y}-\mathbf{x}| < \delta$, it follows

$$|f_k(\mathbf{y}) - L_k| \leq |\mathbf{f}(\mathbf{y}) - \mathbf{L}| < \varepsilon$$

which verifies 7.7.

Now suppose 7.7 holds. Then letting $\varepsilon > 0$ be given, there exists δ_k such that if $0 < |\mathbf{y}-\mathbf{x}| < \delta_k$, then

$$|f_k(\mathbf{y}) - L_k| < \frac{\varepsilon}{\sqrt{p}}.$$

Let $0 < \delta < \min(\delta_1, \cdots, \delta_p)$. Then if $0 < |\mathbf{y}-\mathbf{x}| < \delta$, it follows

$$\begin{aligned} |\mathbf{f}(\mathbf{y}) - \mathbf{L}| &= \left(\sum_{k=1}^{p} |f_k(\mathbf{y}) - L_k|^2 \right)^{1/2} \\ &< \left(\sum_{k=1}^{p} \frac{\varepsilon^2}{p} \right)^{1/2} = \varepsilon. \end{aligned}$$

It remains to verify 7.8. But from the first part of this theorem and the description of the cross product presented earlier in terms of the permutation symbol,

$$\begin{aligned} \lim_{\mathbf{y}\to\mathbf{x}} (\mathbf{f}(\mathbf{y}) \times \mathbf{g}(\mathbf{y}))_i &= \lim_{\mathbf{y}\to\mathbf{x}} \varepsilon_{ijk} f_j(\mathbf{y}) g_k(\mathbf{y}) \\ &= \varepsilon_{ijk} L_j K_k = (\mathbf{L} \times \mathbf{K})_i. \end{aligned}$$

Therefore, from the first part of this theorem, this establishes 11.5. This completes the proof.

Example 7.4.8 *Find* $\lim_{(x,y)\to(3,1)} \left(\frac{x^2-9}{x-3}, y \right)$.

It is clear that $\lim_{(x,y)\to(3,1)} \frac{x^2-9}{x-3} = 6$ and $\lim_{(x,y)\to(3,1)} y = 1$. Therefore, this limit equals $(6,1)$.

Example 7.4.9 *Find* $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$.

First of all observe the domain of the function is $\mathbb{R}^2 \setminus \{(0,0)\}$, every point in $\mathbb{R}^2$ except the origin. Therefore, $(0,0)$ is a limit point of the domain of the function so it might make sense to take a limit. However, just as in the case of a function of one variable, the limit may not exist. In fact, this is the case here. To see this, take points on the line $y = 0$. At these points, the value of the function equals 0. Now consider points on the line $y = x$ where the value of the function equals $1/2$. Since arbitrarily close to $(0,0)$ there are points where the function equals $1/2$ and points where the function has the value 0, it follows there can be no limit. Just take $\varepsilon = 1/10$ for example. You can't be within $1/10$ of $1/2$ and also within $1/10$ of 0 at the same time.

Note it is necessary to rely on the definition of the limit much more than in the case of a function of one variable and there are no easy ways to do limit problems for functions of more than one variable. It is what it is and you will not deal with these concepts without suffering and anguish.

7.5 Properties Of Continuous Functions

Functions of p variables have many of the same properties as functions of one variable. First there is a version of the extreme value theorem generalizing the one dimensional case.

Theorem 7.5.1 *Let C be closed and bounded and let $f : C \to \mathbb{R}$ be continuous. Then f achieves its maximum and its minimum on C. This means there exist, $\mathbf{x}_1, \mathbf{x}_2 \in C$ such that for all $\mathbf{x} \in C$,*

$$f(\mathbf{x}_1) \leq f(\mathbf{x}) \leq f(\mathbf{x}_2).$$

There is also the long technical theorem about sums and products of continuous functions. These theorems are proved in the next section.

Theorem 7.5.2 *The following assertions are valid*

1. *The function, $a\mathbf{f} + b\mathbf{g}$ is continuous at $\mathbf{x}$ when $\mathbf{f}$, $\mathbf{g}$ are continuous at $\mathbf{x} \in D(\mathbf{f}) \cap D(\mathbf{g})$ and $a, b \in \mathbb{R}$.*
2. *If and f and g are each real valued functions continuous at $\mathbf{x}$, then fg is continuous at $\mathbf{x}$. If, in addition to this, $g(\mathbf{x}) \neq 0$, then f/g is continuous at $\mathbf{x}$.*
3. *If $\mathbf{f}$ is continuous at $\mathbf{x}$, $\mathbf{f}(\mathbf{x}) \in D(\mathbf{g}) \subseteq \mathbb{R}^p$, and $\mathbf{g}$ is continuous at $\mathbf{f}(\mathbf{x})$, then $\mathbf{g} \circ \mathbf{f}$ is continuous at $\mathbf{x}$.*
4. *If $\mathbf{f} = (f_1, \cdots, f_q) : D(\mathbf{f}) \to \mathbb{R}^q$, then $\mathbf{f}$ is continuous if and only if each f_k is a continuous real valued function.*
5. *The function $f : \mathbb{R}^p \to \mathbb{R}$, given by $f(\mathbf{x}) = |\mathbf{x}|$ is continuous.*

7.6 Exercises

1. Let $\mathbf{f}(t) = \left(t, t^2+1, \frac{t}{t+1}\right)$ and let $\mathbf{g}(t) = \left(t+1, 1, \frac{t}{t^2+1}\right)$. Find $\mathbf{f} \cdot \mathbf{g}$.
2. Let $\mathbf{f}, \mathbf{g}$ be given in the previous problem. Find $\mathbf{f} \times \mathbf{g}$.

3. Find $D(\mathbf{f})$ if $\mathbf{f}(x,y,z,w) = \left(\frac{xy}{zw}, \sqrt{6-x^2y^2}\right)$.

4. Let $\mathbf{f}(t) = \left(t, t^2, t^3\right)$, $\mathbf{g}(t) = \left(1, t, t^2\right)$, and $\mathbf{h}(t) = (\sin t, t, 1)$. Find the time rate of change of the volume of the parallelepiped spanned by the vectors $\mathbf{f}, \mathbf{g}$, and $\mathbf{h}$.

5. Let $\mathbf{f}(t) = (t, \sin t)$. Show f is continuous at every point t.

6. Suppose $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|$ where K is a constant. Show that $\mathbf{f}$ is everywhere continuous. Functions satisfying such an inequality are called Lipschitz functions.

7. Suppose $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|^{\alpha}$ where K is a constant and $\alpha \in (0,1)$. Show that $\mathbf{f}$ is everywhere continuous.

8. Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is given by $f(\mathbf{x}) = 3x_1x_2 + 2x_3^2$. Use Theorem 7.3.2 to verify that f is continuous. **Hint:** You should first verify that the function, $\pi_k : \mathbb{R}^3 \to \mathbb{R}$ given by $\pi_k(\mathbf{x}) = x_k$ is a continuous function.

9. Show that if $f : \mathbb{R}^q \to \mathbb{R}$ is a polynomial then it is continuous.

10. State and prove a theorem which involves quotients of functions encountered in the previous problem.

11. Let
$$f(x,y) \equiv \begin{cases} \frac{xy}{x^2+y^2} \text{ if } (x,y) \neq (0,0) \\ 0 \text{ if } (x,y) = (0,0) \end{cases}.$$
Find $\lim_{(x,y)\to(0,0)} f(x,y)$ if it exists. If it does not exist, tell why it does not exist. **Hint:** Consider along the line $y = x$ and along the line $y = 0$.

12. Find the following limits if possible

    (a) $\lim_{(x,y)\to(0,0)} \frac{x^2-y^2}{x^2+y^2}$

    (b) $\lim_{(x,y)\to(0,0)} \frac{x\left(x^2-y^2\right)}{(x^2+y^2)}$

    (c) $\lim_{(x,y)\to(0,0)} \frac{\left(x^2-y^4\right)^2}{(x^2+y^4)^2}$ **Hint:** Consider along $y = 0$ and along $x = y^2$.

    (d) $\lim_{(x,y)\to(0,0)} x \sin\left(\frac{1}{x^2+y^2}\right)$

    (e) $\lim_{(x,y)\to(1,2)} \frac{-2yx^2+8yx+34y+3y^3-18y^2+6x^2-13x-20-xy^2-x^3}{-y^2+4y-5-x^2+2x}$. **Hint:** It might help to write this in terms of the variables $(s,t) = (x-1, y-2)$.

13. In the definition of limit, why must $\mathbf{x}$ be a limit point of $D(\mathbf{f})$? **Hint:** If $\mathbf{x}$ were not a limit point of $D(\mathbf{f})$, show there exists $\delta > 0$ such that $B(\mathbf{x}, \delta)$ contains no points of $D(\mathbf{f})$ other than possibly $\mathbf{x}$ itself. Argue that 33.3 is a limit and that so is 22 and 7 and 11. In other words the concept is totally worthless.

14. Suppose $\lim_{x\to 0} f(x,0) = 0 = \lim_{y\to 0} f(0,y)$. Does it follow that
$$\lim_{(x,y)\to(0,0)} f(x,y) = 0?$$
Prove or give counter example.

15. $\mathbf{f} : D \subseteq \mathbb{R}^p \to \mathbb{R}^q$ is Lipschitz continuous or just Lipschitz for short if there exists a constant, K such that

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K |\mathbf{x} - \mathbf{y}|$$

for all $\mathbf{x}, \mathbf{y} \in D$. Show every Lipschitz function is uniformly continuous which means that given $\varepsilon > 0$ there exists $\delta > 0$ independent of $\mathbf{x}$ such that if $|\mathbf{x} - \mathbf{y}| < \delta$, then $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$.

16. If $\mathbf{f}$ is uniformly continuous, does it follow that $|\mathbf{f}|$ is also uniformly continuous? If $|\mathbf{f}|$ is uniformly continuous does it follow that $\mathbf{f}$ is uniformly continuous? Answer the same questions with "uniformly continuous" replaced with "continuous". Explain why.

17. Let f be defined on the positive integers. Thus $D(f) = \mathbb{N}$. Show that f is automatically continuous at every point of $D(f)$. Is it also uniformly continuous? What does this mean about the concept of continuous functions being those which can be graphed without taking the pencil off the paper?

18. In Problem 12c show $\lim_{t \to 0} f(tx, ty) = 1$ for any choice of (x, y). Using Problem 12c what does this tell you about limits existing just because the limit along any line exists.

19. Let $f(x, y, z) = x^2 y + \sin(xyz)$. Does f achieve a maximum on the set

$$\left\{(x, y, z) : x^2 + y^2 + 2z^2 \leq 8\right\}?$$

Explain why.

20. Suppose $\mathbf{x}$ is defined to be a limit point of a set, A if and only if for all $r > 0$, $B(\mathbf{x}, r)$ contains a point of A different than $\mathbf{x}$. Show this is equivalent to the above definition of limit point.

21. Give an example of a set of points in $\mathbb{R}^3$ which has no limit points. Show that if $D(\mathbf{f})$ equals this set, then $\mathbf{f}$ is continuous. Show that more generally, if $\mathbf{f}$ is any function for which $D(\mathbf{f})$ has no limit points, then $\mathbf{f}$ is continuous.

22. Let $\{\mathbf{x}_k\}_{k=1}^n$ be any finite set of points in $\mathbb{R}^p$. Show this set has no limit points.

23. Suppose S is any set of points such that every pair of points is at least as far apart as 1. Show S has no limit points.

24. Find $\lim_{\mathbf{x} \to \mathbf{0}} \frac{\sin(|\mathbf{x}|)}{|\mathbf{x}|}$ and prove your answer from the definition of limit.

25. Suppose $\mathbf{g}$ is a continuous vector valued function of one variable defined on $[0, \infty)$. Prove

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \mathbf{g}(|\mathbf{x}|) = \mathbf{g}(|\mathbf{x}_0|).$$

26. Give some examples of limit problems for functions of many variables which have limits and prove your assertions.

7.7 Some Fundamentals

This section contains the proofs of the theorems which were stated without proof along with some other significant topics which will be useful later. These topics are of fundamental significance but are difficult.

Theorem 7.7.1 *The following assertions are valid*

1. *The function,* $a\mathbf{f}+b\mathbf{g}$ *is continuous at* $\mathbf{x}$ *when* $\mathbf{f}$, $\mathbf{g}$ *are continuous at* $\mathbf{x}\in D(\mathbf{f})\cap D(\mathbf{g})$ *and* $a,b\in\mathbb{R}$.
2. *If and* f *and* g *are each real valued functions continuous at* $\mathbf{x}$, *then* fg *is continuous at* $\mathbf{x}$. *If, in addition to this,* $g(\mathbf{x})\neq 0$, *then* f/g *is continuous at* $\mathbf{x}$.
3. *If* $\mathbf{f}$ *is continuous at* $\mathbf{x}$, $\mathbf{f}(\mathbf{x})\in D(\mathbf{g})\subseteq\mathbb{R}^p$, *and* $\mathbf{g}$ *is continuous at* $\mathbf{f}(\mathbf{x})$,*then* $\mathbf{g}\circ\mathbf{f}$ *is continuous at* $\mathbf{x}$.
4. *If* $\mathbf{f}=(f_1,\cdots,f_q):D(\mathbf{f})\to\mathbb{R}^q$, *then* $\mathbf{f}$ *is continuous if and only if each* f_k *is a continuous real valued function.*
5. *The function* $f:\mathbb{R}^p\to\mathbb{R}$, *given by* $f(\mathbf{x})=|\mathbf{x}|$ *is continuous.*

Proof: Begin with 1.) Let $\varepsilon>0$ be given. By assumption, there exist $\delta_1>0$ such that whenever $|\mathbf{x}-\mathbf{y}|<\delta_1$, it follows $|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y})|<\frac{\varepsilon}{2(|a|+|b|+1)}$ and there exists $\delta_2>0$ such that whenever $|\mathbf{x}-\mathbf{y}|<\delta_2$, it follows that $|\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{y})|<\frac{\varepsilon}{2(|a|+|b|+1)}$. Then let $0<\delta\leq\min(\delta_1,\delta_2)$. If $|\mathbf{x}-\mathbf{y}|<\delta$, then everything happens at once. Therefore, using the triangle inequality

$$|a\mathbf{f}(\mathbf{x})+b\mathbf{f}(\mathbf{x})-(a\mathbf{g}(\mathbf{y})+b\mathbf{g}(\mathbf{y}))|$$

$$\begin{aligned}&\leq|a|\,|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y})|+|b|\,|\mathbf{g}(\mathbf{x})-\mathbf{g}(\mathbf{y})|\\&<|a|\left(\frac{\varepsilon}{2(|a|+|b|+1)}\right)+|b|\left(\frac{\varepsilon}{2(|a|+|b|+1)}\right)<\varepsilon.\end{aligned}$$

Now begin on 2.) There exists $\delta_1>0$ such that if $|\mathbf{y}-\mathbf{x}|<\delta_1$, then $|f(\mathbf{x})-f(\mathbf{y})|<1$. Therefore, for such $\mathbf{y}$,

$$|f(\mathbf{y})|<1+|f(\mathbf{x})|.$$

It follows that for such $\mathbf{y}$,

$$|fg(\mathbf{x})-fg(\mathbf{y})|\leq|f(\mathbf{x})g(\mathbf{x})-g(\mathbf{x})f(\mathbf{y})|+|g(\mathbf{x})f(\mathbf{y})-f(\mathbf{y})g(\mathbf{y})|$$

$$\begin{aligned}&\leq|g(\mathbf{x})|\,|f(\mathbf{x})-f(\mathbf{y})|+|f(\mathbf{y})|\,|g(\mathbf{x})-g(\mathbf{y})|\\&\leq(1+|g(\mathbf{x})|+|f(\mathbf{y})|)\,[|g(\mathbf{x})-g(\mathbf{y})|+|f(\mathbf{x})-f(\mathbf{y})|]\\&\leq(2+|g(\mathbf{x})|+|f(\mathbf{x})|)\,[|g(\mathbf{x})-g(\mathbf{y})|+|f(\mathbf{x})-f(\mathbf{y})|]\end{aligned}$$

Now let $\varepsilon > 0$ be given. There exists δ_2 such that if $|\mathbf{x}-\mathbf{y}| < \delta_2$, then

$$|g(\mathbf{x}) - g(\mathbf{y})| < \frac{\varepsilon}{2(2+|g(\mathbf{x})|+|f(\mathbf{x})|)},$$

and there exists δ_3 such that if $|\mathbf{x}-\mathbf{y}| < \delta_3$, then

$$|f(\mathbf{x}) - f(\mathbf{y})| < \frac{\varepsilon}{2(2+|g(\mathbf{x})|+|f(\mathbf{x})|)}$$

Now let $0 < \delta \leq \min(\delta_1, \delta_2, \delta_3)$. Then if $|\mathbf{x}-\mathbf{y}| < \delta$, all the above hold at once and

$$|fg(\mathbf{x}) - fg(\mathbf{y})| \leq$$

$$\begin{aligned}
&(2+|g(\mathbf{x})|+|f(\mathbf{x})|)\left[|g(\mathbf{x})-g(\mathbf{y})|+|f(\mathbf{x})-f(\mathbf{y})|\right] \\
&< (2+|g(\mathbf{x})|+|f(\mathbf{x})|)\left(\frac{\varepsilon}{2(2+|g(\mathbf{x})|+|f(\mathbf{x})|)}+\frac{\varepsilon}{2(2+|g(\mathbf{x})|+|f(\mathbf{x})|)}\right)=\varepsilon.
\end{aligned}$$

This proves the first part of 2.) To obtain the second part, let δ_1 be as described above and let $\delta_0 > 0$ be such that for $|\mathbf{x}-\mathbf{y}| < \delta_0$,

$$|g(\mathbf{x}) - g(\mathbf{y})| < |g(\mathbf{x})|/2$$

and so by the triangle inequality,

$$-|g(\mathbf{x})|/2 \leq |g(\mathbf{y})| - |g(\mathbf{x})| \leq |g(\mathbf{x})|/2$$

which implies $|g(\mathbf{y})| \geq |g(\mathbf{x})|/2$, and $|g(\mathbf{y})| < 3|g(\mathbf{x})|/2$.

Then if $|\mathbf{x}-\mathbf{y}| < \min(\delta_0, \delta_1)$,

$$\begin{aligned}
\left|\frac{f(\mathbf{x})}{g(\mathbf{x})} - \frac{f(\mathbf{y})}{g(\mathbf{y})}\right| &= \left|\frac{f(\mathbf{x})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{x})}{g(\mathbf{x})g(\mathbf{y})}\right| \\
&\leq \frac{|f(\mathbf{x})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{x})|}{\left(\frac{|g(\mathbf{x})|^2}{2}\right)} \\
&= \frac{2|f(\mathbf{x})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{x})|}{|g(\mathbf{x})|^2}
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{2}{|g(\mathbf{x})|^2}\left[|f(\mathbf{x})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{y}) + f(\mathbf{y})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{x})|\right] \\
&\leq \frac{2}{|g(\mathbf{x})|^2}\left[|g(\mathbf{y})||f(\mathbf{x})-f(\mathbf{y})| + |f(\mathbf{y})||g(\mathbf{y})-g(\mathbf{x})|\right] \\
&\leq \frac{2}{|g(\mathbf{x})|^2}\left[\frac{3}{2}|\mathbf{g}(\mathbf{x})||f(\mathbf{x})-f(\mathbf{y})| + (1+|f(\mathbf{x})|)|g(\mathbf{y})-g(\mathbf{x})|\right] \\
&\leq \frac{2}{|g(\mathbf{x})|^2}(1+2|f(\mathbf{x})|+2|g(\mathbf{x})|)\left[|f(\mathbf{x})-f(\mathbf{y})|+|g(\mathbf{y})-g(\mathbf{x})|\right] \\
&\equiv M\left[|f(\mathbf{x})-f(\mathbf{y})|+|g(\mathbf{y})-g(\mathbf{x})|\right]
\end{aligned}$$

where

$$M \equiv \frac{2}{|g(\mathbf{x})|^2}(1+2|f(\mathbf{x})|+2|g(\mathbf{x})|)$$

Now let δ_2 be such that if $|\mathbf{x}-\mathbf{y}| < \delta_2$, then

$$|f(\mathbf{x}) - f(\mathbf{y})| < \frac{\varepsilon}{2}M^{-1}$$

and let δ_3 be such that if $|\mathbf{x}-\mathbf{y}|<\delta_3$, then

$$|g(\mathbf{y})-g(\mathbf{x})|<\frac{\varepsilon}{2}M^{-1}.$$

Then if $0<\delta\leq\min(\delta_0,\delta_1,\delta_2,\delta_3)$, and $|\mathbf{x}-\mathbf{y}|<\delta$, everything holds and

$$\begin{aligned}\left|\frac{f(\mathbf{x})}{g(\mathbf{x})}-\frac{f(\mathbf{y})}{g(\mathbf{y})}\right| &\leq M\left[|f(\mathbf{x})-f(\mathbf{y})|+|g(\mathbf{y})-g(\mathbf{x})|\right]\\ &< M\left[\frac{\varepsilon}{2}M^{-1}+\frac{\varepsilon}{2}M^{-1}\right]=\varepsilon.\end{aligned}$$

This completes the proof of the second part of 2.) Note that in these proofs no effort is made to find some sort of "best" δ. The problem is one which has a yes or a no answer. Either is it or it is not continuous.

Now begin on 3.). If $\mathbf{f}$ is continuous at $\mathbf{x}$, $\mathbf{f}(\mathbf{x})\in D(\mathbf{g})\subseteq\mathbb{R}^p$, and $\mathbf{g}$ is continuous at $\mathbf{f}(\mathbf{x})$,then $\mathbf{g}\circ\mathbf{f}$ is continuous at $\mathbf{x}$. Let $\varepsilon>0$ be given. Then there exists $\eta>0$ such that if $|\mathbf{y}-\mathbf{f}(\mathbf{x})|<\eta$ and $\mathbf{y}\in D(\mathbf{g})$, it follows that $|\mathbf{g}(\mathbf{y})-\mathbf{g}(\mathbf{f}(\mathbf{x}))|<\varepsilon$. It follows from continuity of $\mathbf{f}$ at $\mathbf{x}$ that there exists $\delta>0$ such that if $|\mathbf{x}-\mathbf{z}|<\delta$ and $\mathbf{z}\in D(\mathbf{f})$, then $|\mathbf{f}(\mathbf{z})-\mathbf{f}(\mathbf{x})|<\eta$. Then if $|\mathbf{x}-\mathbf{z}|<\delta$ and $\mathbf{z}\in D(\mathbf{g}\circ\mathbf{f})\subseteq D(\mathbf{f})$, all the above hold and so

$$|\mathbf{g}(\mathbf{f}(\mathbf{z}))-\mathbf{g}(\mathbf{f}(\mathbf{x}))|<\varepsilon.$$

This proves part 3.)

Part 4.) says: If $\mathbf{f}=(f_1,\cdots,f_q):D(\mathbf{f})\to\mathbb{R}^q$, then $\mathbf{f}$ is continuous if and only if each f_k is a continuous real valued function. Then

$$\begin{aligned}|f_k(\mathbf{x})-f_k(\mathbf{y})| &\leq |\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y})|\\ &\equiv\left(\sum_{i=1}^{q}|f_i(\mathbf{x})-f_i(\mathbf{y})|^2\right)^{1/2}\\ &\leq\sum_{i=1}^{q}|f_i(\mathbf{x})-f_i(\mathbf{y})|.\end{aligned}\tag{7.9}$$

Suppose first that $\mathbf{f}$ is continuous at $\mathbf{x}$. Then there exists $\delta>0$ such that if $|\mathbf{x}-\mathbf{y}|<\delta$, then $|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y})|<\varepsilon$. The first part of the above inequality then shows that for each $k=1,\cdots,q$, $|f_k(\mathbf{x})-f_k(\mathbf{y})|<\varepsilon$. This shows the only if part. Now suppose each function, f_k is continuous. Then if $\varepsilon>0$ is given, there exists $\delta_k>0$ such that whenever $|\mathbf{x}-\mathbf{y}|<\delta_k$

$$|f_k(\mathbf{x})-f_k(\mathbf{y})|<\varepsilon/q.$$

Now let $0<\delta\leq\min(\delta_1,\cdots,\delta_q)$. For $|\mathbf{x}-\mathbf{y}|<\delta$, the above inequality holds for all k and so the last part of 7.9 implies

$$\begin{aligned}|\mathbf{f}(\mathbf{x})-\mathbf{f}(\mathbf{y})| &\leq\sum_{i=1}^{q}|f_i(\mathbf{x})-f_i(\mathbf{y})|\\ &<\sum_{i=1}^{q}\frac{\varepsilon}{q}=\varepsilon.\end{aligned}$$

This proves part 4.)

To verify part 5.), let $\varepsilon>0$ be given and let $\delta=\varepsilon$. Then if $|\mathbf{x}-\mathbf{y}|<\delta$, the triangle inequality implies

$$\begin{aligned}|f(\mathbf{x})-f(\mathbf{y})| &=||\mathbf{x}|-|\mathbf{y}||\\ &\leq|\mathbf{x}-\mathbf{y}|<\delta=\varepsilon.\end{aligned}$$

This proves part 5.) and completes the proof of the theorem.

7.7.1 The Nested Interval Lemma

Here is a multidimensional version of the nested interval lemma.

Lemma 7.7.2 *Let* $I_k = \prod_{i=1}^p \left[a_i^k, b_i^k\right] \equiv \left\{\mathbf{x} \in \mathbb{R}^p : x_i \in \left[a_i^k, b_i^k\right]\right\}$ *and suppose that for all* $k = 1, 2, \cdots$,

$$I_k \supseteq I_{k+1}.$$

Then there exists a point, $\mathbf{c} \in \mathbb{R}^p$ *which is an element of every* I_k.

Proof: Since $I_k \supseteq I_{k+1}$, it follows that for each $i = 1, \cdots, p$, $\left[a_i^k, b_i^k\right] \supseteq \left[a_i^{k+1}, b_i^{k+1}\right]$. This implies that for each i,

$$a_i^k \leq a_i^{k+1},\ b_i^k \geq b_i^{k+1}. \tag{7.10}$$

Consequently, if $k \leq l$,

$$a_i^l \leq b_i^l \leq b_i^k. \tag{7.11}$$

Now define

$$c_i \equiv \sup\left\{a_i^l : l = 1, 2, \cdots\right\}$$

By the first inequality in 7.10,

$$c_i = \sup\left\{a_i^l : l = k, k+1, \cdots\right\} \tag{7.12}$$

for each $k = 1, 2 \cdots$. Therefore, picking any k,7.11 shows that b_i^k is an upper bound for the set, $\left\{a_i^l : l = k, k+1, \cdots\right\}$ and so it is at least as large as the least upper bound of this set which is the definition of c_i given in 7.12. Thus, for each i and each k,

$$a_i^k \leq c_i \leq b_i^k.$$

Defining $\mathbf{c} \equiv (c_1, \cdots, c_p)$, $\mathbf{c} \in I_k$ for all k. This proves the lemma.

If you don't like the proof,you could prove the lemma for the one variable case first and then do the following.

Lemma 7.7.3 *Let* $I_k = \prod_{i=1}^p \left[a_i^k, b_i^k\right] \equiv \left\{\mathbf{x} \in \mathbb{R}^p : x_i \in \left[a_i^k, b_i^k\right]\right\}$ *and suppose that for all* $k = 1, 2, \cdots$,

$$I_k \supseteq I_{k+1}.$$

Then there exists a point, $\mathbf{c} \in \mathbb{R}^p$ *which is an element of every* I_k.

Proof: For each $i = 1, \cdots, p$, $\left[a_i^k, b_i^k\right] \supseteq \left[a_i^{k+1}, b_i^{k+1}\right]$ and so by the nested interval theorem for one dimensional problems, there exists a point $c_i \in \left[a_i^k, b_i^k\right]$ for all k. Then letting $\mathbf{c} \equiv (c_1, \cdots, c_p)$ it follows $\mathbf{c} \in I_k$ for all k. This proves the lemma.

7.7.2 The Extreme Value Theorem

Definition 7.7.4 *A set,* $C \subseteq \mathbb{R}^p$ *is said to be* ***bounded*** *if* $C \subseteq \prod_{i=1}^p [a_i, b_i]$ *for some choice of intervals,* $[a_i, b_i]$ *where* $-\infty < a_i < b_i < \infty$. *The* ***diameter*** *of a set,* S, *is defined as*

$$\operatorname{diam}(S) \equiv \sup\{|\mathbf{x} - \mathbf{y}| : \mathbf{x}, \mathbf{y} \in S\}.$$

A function, $\mathbf{f}$ *having values in* $\mathbb{R}^p$ *is said to be bounded if the set of values of* $\mathbf{f}$ *is a bounded set.*

Thus $\operatorname{diam}(S)$ is just a careful description of what you would think of as the diameter. It measures how stretched out the set is.

Lemma 7.7.5 *Let $C \subseteq \mathbb{R}^p$ be closed and bounded and let $f : C \to \mathbb{R}$ be continuous. Then f is bounded.*

Proof: Suppose not. Since C is bounded, it follows $C \subseteq \prod_{i=1}^p [a_i, b_i] \equiv I_0$ for some closed intervals, $[a_i, b_i]$. Consider all sets of the form $\prod_{i=1}^p [c_i, d_i]$ where $[c_i, d_i]$ equals either $\left[a_i, \frac{a_i+b_i}{2}\right]$ or $[c_i, d_i] = \left[\frac{a_i+b_i}{2}, b_i\right]$. Thus there are 2^p of these sets because there are two choices for the i^{th} slot for $i = 1, \cdots, p$. Also, if $\mathbf{x}$ and $\mathbf{y}$ are two points in one of these sets,

$$|x_i - y_i| \leq 2^{-1} |b_i - a_i| .$$

Observe that $\operatorname{diam}(I_0) = \left(\sum_{i=1}^p |b_i - a_i|^2\right)^{1/2}$ because for $\mathbf{x}, \mathbf{y} \in I_0, |x_i - y_i| \leq |a_i - b_i|$ for each $i = 1, \cdots, p$,

$$\begin{aligned} |\mathbf{x} - \mathbf{y}| &= \left(\sum_{i=1}^p |x_i - y_i|^2\right)^{1/2} \\ &\leq 2^{-1} \left(\sum_{i=1}^p |b_i - a_i|^2\right)^{1/2} \equiv 2^{-1} \operatorname{diam}(I_0) . \end{aligned}$$

Denote by $\{J_1, \cdots, J_{2^p}\}$ these sets determined above. It follows the diameter of each set is no larger than $2^{-1} \operatorname{diam}(I_0)$. In particular, since $\mathbf{d} \equiv (d_1, \cdots, d_p)$ and $\mathbf{c} \equiv (c_1, \cdots, c_p)$ are two such points, for each J_k,

$$\operatorname{diam}(J_k) \equiv \left(\sum_{i=1}^p |d_i - c_i|^2\right)^{1/2} \leq 2^{-1} \operatorname{diam}(I_0)$$

Since the union of these sets equals all of I_0, it follows

$$C = \cup_{k=1}^{2^p} J_k \cap C.$$

If f is not bounded on C, it follows that for some k, f is not bounded on $J_k \cap C$. Let $I_1 \equiv J_k$ and let $C_1 = C \cap I_1$. Now do to I_1 and C_1 what was done to I_0 and C to obtain $I_2 \subseteq I_1$, and for $\mathbf{x}, \mathbf{y} \in I_2$,

$$|\mathbf{x} - \mathbf{y}| \leq 2^{-1} \operatorname{diam}(I_1) \leq 2^{-2} \operatorname{diam}(I_2) ,$$

and f is unbounded on $I_2 \cap C_1 \equiv C_2$. Continue in this way obtaining sets, I_k such that $I_k \supseteq I_{k+1}$ and $\operatorname{diam}(I_k) \leq 2^{-k} \operatorname{diam}(I_0)$ and f is unbounded on $I_k \cap C$. By the nested interval lemma, there exists a point, $\mathbf{c}$ which is contained in each I_k.

Claim: $\mathbf{c} \in C$.

Proof of claim: Suppose $\mathbf{c} \notin C$. Since C is a closed set, there exists $r > 0$ such that $B(\mathbf{c}, r)$ is contained completely in $\mathbb{R}^p \setminus C$. In other words, $B(\mathbf{c}, r)$ contains no points of C. Let k be so large that $\operatorname{diam}(I_0) 2^{-k} < r$. Then since $\mathbf{c} \in I_k$, and any two points of I_k are closer than $\operatorname{diam}(I_0) 2^{-k}$, I_k must be contained in $B(\mathbf{c}, r)$ and so has no points of C in it, contrary to the manner in which the I_k are defined in which f is unbounded on $I_k \cap C$. Therefore, $\mathbf{c} \in C$ as claimed.

Now for k large enough, and $\mathbf{x} \in C \cap I_k$, the continuity of f implies $|f(\mathbf{c}) - f(\mathbf{x})| < 1$ contradicting the manner in which I_k was chosen since this inequality implies f is bounded on $I_k \cap C$. This proves the theorem.

Here is a proof of the extreme value theorem.

Theorem 7.7.6 *Let C be closed and bounded and let $f : C \to \mathbb{R}$ be continuous. Then f achieves its maximum and its minimum on C. This means there exist, $\mathbf{x}_1, \mathbf{x}_2 \in C$ such that for all $\mathbf{x} \in C$,*

$$f(\mathbf{x}_1) \leq f(\mathbf{x}) \leq f(\mathbf{x}_2) .$$

Proof: Let $M = \sup\{f(\mathbf{x}) : \mathbf{x} \in C\}$. Then by Lemma 7.7.5, M is a finite number. Is $f(\mathbf{x}_2) = M$ for some x_2? if not, you could consider the function,

$$g(\mathbf{x}) \equiv \frac{1}{M - f(\mathbf{x})}$$

and g would be a continuous and unbounded function defined on C, contrary to Lemma 7.7.5. Therefore, there exists $\mathbf{x}_2 \in C$ such that $f(\mathbf{x}_2) = M$. A similar argument applies to show the existence of $\mathbf{x}_1 \in C$ such that

$$f(\mathbf{x}_1) = \inf\{f(\mathbf{x}) : \mathbf{x} \in C\}.$$

This proves the theorem.

7.7.3 Sequences And Completeness

Definition 7.7.7 *A function whose domain is defined as a set of the form*

$$\{k, k+1, k+2, \cdots\}$$

for k an integer is known as a sequence. Thus you can consider $f(k), f(k+1), f(k+2)$, etc. Usually the domain of the sequence is either $\mathbb{N}$, the natural numbers consisting of $\{1, 2, 3, \cdots\}$ or the nonnegative integers, $\{0, 1, 2, 3, \cdots\}$. Also, it is traditional to write f_1, f_2, etc. instead of $f(1), f(2), f(3)$ etc. when referring to sequences. In the above context, f_k is called the first term, f_{k+1} the second and so forth. It is also common to write the sequence, not as f but as $\{f_i\}_{i=k}^{\infty}$ or just $\{f_i\}$ for short. The letter used for the name of the sequence is not important. Thus it is all right to let a be the name of a sequence or to refer to it as $\{a_i\}$. When the sequence has values in $\mathbb{R}^p$, it is customary to write it in bold face. Thus $\{\mathbf{a}_i\}$ would refer to a sequence having values in $\mathbb{R}^p$ for some $p > 1$.

Example 7.7.8 *Let $\{a_k\}_{k=1}^{\infty}$ be defined by $a_k \equiv k^2 + 1$.*

This gives a sequence. In fact, $a_7 = a(7) = 7^2 + 1 = 50$ just from using the formula for the k^{th} term of the sequence.

It is nice when sequences come to us in this way from a formula for the k^{th} term. However, this is often not the case. Sometimes sequences are defined recursively. This happens, when the first several terms of the sequence are given and then a rule is specified which determines a_{n+1} from knowledge of $a_1, \cdots, a_n$. This rule which specifies a_{n+1} from knowledge of a_k for $k \leq n$ is known as a recurrence relation.

Example 7.7.9 *Let $a_1 = 1$ and $a_2 = 1$. Assuming $a_1, \cdots, a_{n+1}$ are known, $a_{n+2} \equiv a_n + a_{n+1}$.*

Thus the first several terms of this sequence, listed in order, are 1, 1, 2, 3, 5, 8,$\cdots$. This particular sequence is called the Fibonacci sequence and is important in the study of reproducing rabbits.

Example 7.7.10 *Let $\mathbf{a}_k = (k, \sin(k))$. Thus this sequence has values in $\mathbb{R}^2$.*

Definition 7.7.11 *Let $\{\mathbf{a}_n\}$ be a sequence and let $n_1 < n_2 < n_3, \cdots$ be any strictly increasing list of integers such that n_1 is at least as large as the first index used to define the sequence $\{\mathbf{a}_n\}$. Then if $\mathbf{b}_k \equiv \mathbf{a}_{n_k}$, $\{\mathbf{b}_k\}$ is called a subsequence of $\{\mathbf{a}_n\}$.*

For example, suppose $a_n = \left(n^2+1\right)$. Thus $a_1 = 2$, $a_3 = 10$, etc. If

$$n_1 = 1, n_2 = 3, n_3 = 5, \cdots, n_k = 2k-1,$$

then letting $b_k = a_{n_k}$, it follows

$$b_k = \left(\left(2k-1\right)^2+1\right) = 4k^2 - 4k + 2.$$

Definition 7.7.12 *A sequence, $\{\mathbf{a}_k\}$ is said to* ***converge*** *to $\mathbf{a}$ if for every $\varepsilon > 0$ there exists n_ε such that if $n > n_\varepsilon$, then $|\mathbf{a} - \mathbf{a}_\varepsilon| < \varepsilon$. The usual notation for this is $\lim_{n\to\infty} \mathbf{a}_n = \mathbf{a}$ although it is often written as $\mathbf{a}_n \to \mathbf{a}$.*

The following theorem says the limit, if it exists, is unique.

Theorem 7.7.13 *If a sequence, $\{\mathbf{a}_n\}$ converges to $\mathbf{a}$ and to $\mathbf{b}$ then $\mathbf{a} = \mathbf{b}$.*

Proof: There exists n_ε such that if $n > n_\varepsilon$ then $|\mathbf{a}_n - \mathbf{a}| < \frac{\varepsilon}{2}$ and if $n > n_\varepsilon$, then $|\mathbf{a}_n - \mathbf{b}| < \frac{\varepsilon}{2}$. Then pick such an n.

$$|\mathbf{a} - \mathbf{b}| < |\mathbf{a} - \mathbf{a}_n| + |\mathbf{a}_n - \mathbf{b}| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since ε is arbitrary, this proves the theorem.

The following is the definition of a Cauchy sequencein $\mathbb{R}^p$.

Definition 7.7.14 *$\{\mathbf{a}_n\}$ is a Cauchy sequence if for all $\varepsilon > 0$, there exists n_ε such that whenever $n, m \geq n_\varepsilon$,*

$$|\mathbf{a}_n - \mathbf{a}_m| < \varepsilon.$$

A sequence is Cauchy means the terms are "bunching up to each other" as m, n get large.

Theorem 7.7.15 *The set of terms in a Cauchy sequence in $\mathbb{R}^p$ is bounded in the sense that for all n, $|\mathbf{a}_n| < M$ for some $M < \infty$.*

Proof: Let $\varepsilon = 1$ in the definition of a Cauchy sequence and let $n > n_1$. Then from the definition,

$$|\mathbf{a}_n - \mathbf{a}_{n_1}| < 1.$$

It follows that for all $n > n_1$,

$$|\mathbf{a}_n| < 1 + |\mathbf{a}_{n_1}|.$$

Therefore, for all n,

$$|\mathbf{a}_n| \leq 1 + |\mathbf{a}_{n_1}| + \sum_{k=1}^{n_1} |\mathbf{a}_k|.$$

This proves the theorem.

Theorem 7.7.16 *If a sequence $\{\mathbf{a}_n\}$ in $\mathbb{R}^p$ converges, then the sequence is a Cauchy sequence. Also, if some subsequence of a Cauchy sequence converges, then the original sequence converges.*

Proof: Let $\varepsilon > 0$ be given and suppose $\mathbf{a}_n \to \mathbf{a}$. Then from the definition of convergence, there exists n_ε such that if $n > n_\varepsilon$, it follows that

$$|\mathbf{a}_n - \mathbf{a}| < \frac{\varepsilon}{2}$$

Therefore, if $m, n \geq n_\varepsilon + 1$, it follows that

$$|\mathbf{a}_n - \mathbf{a}_m| \leq |\mathbf{a}_n - \mathbf{a}| + |\mathbf{a} - \mathbf{a}_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

showing that, since $\varepsilon > 0$ is arbitrary, $\{\mathbf{a}_n\}$ is a Cauchy sequence. It remains to show the last claim. Suppose then that $\{\mathbf{a}_n\}$ is a Cauchy sequence and $\mathbf{a} = \lim_{k\to\infty} \mathbf{a}_{n_k}$ where $\{\mathbf{a}_{n_k}\}_{k=1}^{\infty}$ is a subsequence. Let $\varepsilon > 0$ be given. Then there exists K such that if $k, l \geq K$, then $|\mathbf{a}_k - \mathbf{a}_l| < \frac{\varepsilon}{2}$. Then if $k > K$, it follows $n_k > K$ because $n_1, n_2, n_3, \cdots$ is strictly increasing as the subscript increases. Also, there exists K_1 such that if $k > K_1, |\mathbf{a}_{n_k} - \mathbf{a}| < \frac{\varepsilon}{2}$. Then letting $n > \max(K, K_1)$, pick $k > \max(K, K_1)$. Then

$$|\mathbf{a} - \mathbf{a}_n| \leq |\mathbf{a} - \mathbf{a}_{n_k}| + |\mathbf{a}_{n_k} - \mathbf{a}_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This proves the theorem.

Definition 7.7.17 *A set, K in $\mathbb{R}^p$ is said to be* ***sequentially compact*** *if every sequence in K has a subsequence which converges to a point of K.*

Theorem 7.7.18 *If $I_0 = \prod_{i=1}^{p} [a_i, b_i]$, $p \geq 1$, where $a_i \leq b_i$, then I_0 is sequentially compact.*

Proof: Let $\{\mathbf{a}_i\}_{i=1}^{\infty} \subseteq I_0$ and consider all sets of the form $\prod_{i=1}^{p} [c_i, d_i]$ where $[c_i, d_i]$ equals either $\left[a_i, \frac{a_i+b_i}{2}\right]$ or $[c_i, d_i] = \left[\frac{a_i+b_i}{2}, b_i\right]$. Thus there are 2^p of these sets because there are two choices for the i^{th} slot for $i = 1, \cdots, p$. Also, if $\mathbf{x}$ and $\mathbf{y}$ are two points in one of these sets,

$$|x_i - y_i| \leq 2^{-1} |b_i - a_i|.$$

$\operatorname{diam}(I_0) = \left(\sum_{i=1}^{p} |b_i - a_i|^2\right)^{1/2}$,

$$\begin{aligned} |\mathbf{x} - \mathbf{y}| &= \left(\sum_{i=1}^{p} |x_i - y_i|^2\right)^{1/2} \\ &\leq 2^{-1} \left(\sum_{i=1}^{p} |b_i - a_i|^2\right)^{1/2} \equiv 2^{-1} \operatorname{diam}(I_0). \end{aligned}$$

In particular, since $\mathbf{d} \equiv (d_1, \cdots, d_p)$ and $\mathbf{c} \equiv (c_1, \cdots, c_p)$ are two such points,

$$D_1 \equiv \left(\sum_{i=1}^{p} |d_i - c_i|^2\right)^{1/2} \leq 2^{-1} \operatorname{diam}(I_0)$$

Denote by $\{J_1, \cdots, J_{2^p}\}$ these sets determined above. Since the union of these sets equals all of $I_0 \equiv I$, it follows that for some J_k, the sequence, $\{\mathbf{a}_i\}$ is contained in J_k for infinitely many k. Let that one be called I_1. Next do for I_1 what was done for I_0 to get $I_2 \subseteq I_1$ such that the diameter is half that of I_1 and I_2 contains $\{\mathbf{a}_k\}$ for infinitely many values of k. Continue in this way obtaining a nested sequence of intervals, $\{I_k\}$ such that $I_k \supseteq I_{k+1}$, and if $\mathbf{x}, \mathbf{y} \in I_k$, then $|\mathbf{x} - \mathbf{y}| \leq 2^{-k} \operatorname{diam}(I_0)$, and I_n contains $\{\mathbf{a}_k\}$ for infinitely many values of k for each n. Then by the nested interval lemma, there exists $\mathbf{c}$ such that $\mathbf{c}$ is contained in each I_k. Pick $\mathbf{a}_{n_1} \in I_1$. Next pick $n_2 > n_1$ such that $\mathbf{a}_{n_2} \in I_2$. If $\mathbf{a}_{n_1}, \cdots, \mathbf{a}_{n_k}$ have been chosen, let $\mathbf{a}_{n_{k+1}} \in I_{k+1}$ and $n_{k+1} > n_k$. This can be done because in the construction, I_n contains $\{\mathbf{a}_k\}$ for infinitely many k. Thus the distance between $\mathbf{a}_{n_k}$ and $\mathbf{c}$ is no larger than $2^{-k} \operatorname{diam}(I_0)$ and so $\lim_{k\to\infty} \mathbf{a}_{n_k} = \mathbf{c} \in I_0$. This proves the theorem.

Theorem 7.7.19 *Every Cauchy sequence in $\mathbb{R}^p$ converges.*

Proof: Let $\{\mathbf{a}_k\}$ be a Cauchy sequence. By Theorem 7.7.15 there is some interval, $\prod_{i=1}^{p}[a_i,b_i]$ containing all the terms of $\{\mathbf{a}_k\}$. Therefore, by Theorem 7.7.18 a subsequence converges to a point of this interval. By Theorem 7.7.16 the original sequence converges. This proves the theorem.

7.7.4 Continuity And The Limit Of A Sequence

Just as in the case of a function of one variable, there is a very useful way of thinking of continuity in terms of limits of sequences found in the following theorem. In words, it says a function is continuous if it takes convergent sequences to convergent sequences whenever possible.

Theorem 7.7.20 *A function $\mathbf{f}: D(\mathbf{f}) \to \mathbb{R}^q$ is continuous at $\mathbf{x} \in D(\mathbf{f})$ if and only if, whenever $\mathbf{x}_n \to \mathbf{x}$ with $\mathbf{x}_n \in D(\mathbf{f})$, it follows $\mathbf{f}(\mathbf{x}_n) \to \mathbf{f}(\mathbf{x})$.*

Proof: Suppose first that $\mathbf{f}$ is continuous at $\mathbf{x}$ and let $\mathbf{x}_n \to \mathbf{x}$. Let $\varepsilon > 0$ be given. By continuity, there exists $\delta > 0$ such that if $|\mathbf{y}-\mathbf{x}| < \delta$, then $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$. However, there exists n_δ such that if $n \geq n_\delta$, then $|\mathbf{x}_n - \mathbf{x}| < \delta$ and so for all n this large,

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_n)| < \varepsilon$$

which shows $\mathbf{f}(\mathbf{x}_n) \to \mathbf{f}(\mathbf{x})$.

Now suppose the condition about taking convergent sequences to convergent sequences holds at $\mathbf{x}$. Suppose $\mathbf{f}$ fails to be continuous at $\mathbf{x}$. Then there exists $\varepsilon > 0$ and $\mathbf{x}_n \in D(f)$ such that $|\mathbf{x}-\mathbf{x}_n| < \frac{1}{n}$, yet

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_n)| \geq \varepsilon.$$

But this is clearly a contradiction because, although $\mathbf{x}_n \to \mathbf{x}$, $\mathbf{f}(\mathbf{x}_n)$ fails to converge to $\mathbf{f}(\mathbf{x})$. It follows $\mathbf{f}$ must be continuous after all. This proves the theorem.

7.8 Exercises

1. Suppose $\{\mathbf{x}_n\}$ is a sequence contained in a closed set, C which converges to $\mathbf{x}$. Show that $\mathbf{x} \in C$. **Hint:** Recall that a set is closed if and only if the complement of the set is open. That is if and only if $\mathbb{R}^n \setminus C$ is open.

2. Show using Problem 1 and Theorem 7.7.18 that every closed and bounded set is sequentially compact. **Hint:** If C is such a set, then $C \subseteq I_0 \equiv \prod_{i=1}^{n}[a_i,b_i]$. Now if $\{\mathbf{x}_n\}$ is a sequence in C, it must also be a sequence in I_0. Apply Problem 1 and Theorem 7.7.18.

3. Prove the extreme value theorem, a continuous function achieves its maximum and minimum on any closed and bounded set, C, using the result of Problem 2. **Hint:** Suppose $\lambda = \sup\{f(\mathbf{x}) : \mathbf{x} \in C\}$. Then there exists $\{\mathbf{x}_n\} \subseteq C$ such that $f(\mathbf{x}_n) \to \lambda$. Now select a convergent subsequence using Problem 2. Do the same for the minimum.

4. Let C be a closed and bounded set and suppose $\mathbf{f}: C \to \mathbb{R}^m$ is continuous. Show that $\mathbf{f}$ must also be **uniformly continuous.** This means: For every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $\mathbf{x}, \mathbf{y} \in C$ and $|\mathbf{x}-\mathbf{y}| < \delta$, it follows $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$. This is a good time to review the definition of continuity so you will see the difference. **Hint:** Suppose it is not so. Then there exists $\varepsilon > 0$ and $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ such that $|\mathbf{x}_k - \mathbf{y}_k| < \frac{1}{k}$ but $|\mathbf{f}(\mathbf{x}_k) - \mathbf{f}(\mathbf{y}_k)| \geq \varepsilon$. Now use Problem 2 to obtain a convergent subsequence.

5. Suppose every Cauchy sequence converges in $\mathbb{R}$. Show this implies the least upper bound axiom which is the usual way to state completeness for $\mathbb{R}$. Explain why the convergence of Cauchy sequences is equivalent to every nonempty set which is bounded above has a least upper bound in $\mathbb{R}$.

6. From Problem 2 every closed and bounded set is sequentially compact. Are these the only sets which are sequentially compact? Explain.

7. A set whose elements are open sets, $\mathcal{C}$ is called an **open cover** of H if $\cup\mathcal{C} \supseteq H$. In other words, $\mathcal{C}$ is an open cover of H if every point of H is in at least one set of $\mathcal{C}$. Show that if $\mathcal{C}$ is an open cover of a closed and bounded set H then there exists $\delta > 0$ such that whenever $\mathbf{x} \in H$, $B(\mathbf{x}, \delta)$ is contained in some set of $\mathcal{C}$. This number, δ is called a **Lebesgue number**. **Hint:** If there is no Lebesgue number for H, let $H \subseteq I = \prod_{i=1}^{n} [a_i, b_i]$. Use the process of chopping the intervals in half to get a sequence of nested intervals, I_k contained in I where $\operatorname{diam}(I_k) \leq 2^{-k} \operatorname{diam}(I)$ and there is no Lebesgue number for the open cover on $H_k \equiv H \cap I_k$. Now use the nested interval theorem to get $\mathbf{c}$ in all these H_k. For some $r > 0$ it follows $B(\mathbf{c}, r)$ is contained in some open set of U. But for large k, it must be that $H_k \subseteq B(\mathbf{c}, r)$ which contradicts the construction. You fill in the details.

8. A set is **compact** if for every open cover of the set, there exists a finite subset of the open cover which also covers the set. Show every closed and bounded set in $\mathbb{R}^p$ is compact. Next show that if a set in $\mathbb{R}^p$ is compact, then it must be closed and bounded. This is called the Heine Borel theorem.

9. Suppose S is a nonempty set in $\mathbb{R}^p$. Define

$$\operatorname{dist}(\mathbf{x},S) \equiv \inf\{|\mathbf{x}-\mathbf{y}| : \mathbf{y} \in S\}.$$

   Show that

$$|\operatorname{dist}(\mathbf{x},S) - \operatorname{dist}(\mathbf{y},S)| \leq |\mathbf{x}-\mathbf{y}|.$$

   Hint: Suppose $\operatorname{dist}(\mathbf{x}, S) < \operatorname{dist}(\mathbf{y}, S)$. If these are equal there is nothing to show. Explain why there exists $\mathbf{z} \in S$ such that $|\mathbf{x}-\mathbf{z}| < \operatorname{dist}(\mathbf{x},S) + \varepsilon$. Now explain why

$$|\operatorname{dist}(\mathbf{x},S) - \operatorname{dist}(\mathbf{y},S)| = \operatorname{dist}(\mathbf{y},S) - \operatorname{dist}(\mathbf{x},S) \leq |\mathbf{y}-\mathbf{z}| - (|\mathbf{x}-\mathbf{z}| - \varepsilon)$$

   Now use the triangle inequality and observe that ε is arbitrary.

10. Suppose H is a closed set and $H \subseteq U \subseteq \mathbb{R}^p$, an open set. Show there exists a continuous function defined on $\mathbb{R}^p$, f such that $f(\mathbb{R}^p) \subseteq [0,1]$, $f(\mathbf{x}) = 0$ if $\mathbf{x} \notin U$ and $f(\mathbf{x}) = 1$ if $\mathbf{x} \in H$. **Hint:** Try something like

$$\frac{\operatorname{dist}\left(\mathbf{x}, U^C\right)}{\operatorname{dist}(\mathbf{x}, U^C) + \operatorname{dist}(\mathbf{x}, H)},$$

   where $U^C \equiv \mathbb{R}^p \setminus U$, a closed set. You need to explain why the denominator is never equal to zero. The rest is supplied by Problem 9. This is a special case of a major theorem called Urysohn's lemma.

Vector Valued Functions Of One Variable

8.0.1 Outcomes

1. Identify a curve given its parameterization.
2. Determine combinations of vector functions such as sums, vector products, and scalar products.
3. Define limit, derivative, and integral for vector functions.
4. Evaluate limits, derivatives and integrals of vector functions.
5. Find the line tangent to a curve at a given point.
6. Recall, derive and apply rules to combinations of vector functions for the following:
   (a) limits
   (b) differentiation
   (c) integration
7. Describe what is meant by arc length.
8. Evaluate the arc length of a curve.
9. Evaluate the work done by a varying force over a curved path.

8.1 Limits Of A Vector Valued Function Of One Variable

Limits of vector valued functions have been considered earlier. Here it is desired to consider

$$\lim_{h\to 0}\frac{\mathbf{f}(t_0+h)-\mathbf{f}(t_0)}{h}$$

Specializing to functions of one variable, one can give a meaning to

$$\lim_{s\to t+}\mathbf{f}(s),\ \lim_{s\to t-}\mathbf{f}(s),\ \lim_{s\to\infty}\mathbf{f}(s),$$

and

$$\lim_{s\to-\infty}\mathbf{f}(s).$$

Definition 8.1.1 *In the case where $D(\mathbf{f})$ is only assumed to satisfy $D(\mathbf{f}) \supseteq (t, t+r)$,*

$$\lim_{s \to t+} \mathbf{f}(s) = \mathbf{L}$$

if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < s - t < \delta,$$

then

$$|\mathbf{f}(s) - \mathbf{L}| < \varepsilon.$$

In the case where $D(\mathbf{f})$ is only assumed to satisfy $D(\mathbf{f}) \supseteq (t-r, t)$,

$$\lim_{s \to t-} \mathbf{f}(s) = \mathbf{L}$$

if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < t - s < \delta,$$

then

$$|\mathbf{f}(s) - \mathbf{L}| < \varepsilon.$$

One can also consider limits as a variable "approaches" infinity. Of course nothing is "close" to infinity and so this requires a slightly different definition.

$$\lim_{t \to \infty} \mathbf{f}(t) = \mathbf{L}$$

if for every $\varepsilon > 0$ there exists l such that whenever $t > l$,

$$|\mathbf{f}(t) - \mathbf{L}| < \varepsilon \tag{8.1}$$

and

$$\lim_{t \to -\infty} \mathbf{f}(t) = \mathbf{L}$$

if for every $\varepsilon > 0$ there exists l such that whenever $t < l$, 8.1 holds.

Note that in all of this the definitions are identical to the case of scalar valued functions. The only difference is that here $|\cdot|$ refers to the norm or length in $\mathbb{R}^p$ where maybe $p > 1$.

Example 8.1.2 *Let $\mathbf{f}(t) = \left(\cos t, \sin t, t^2 + 1, \ln(t)\right)$. Find $\lim_{t \to \pi/2} \mathbf{f}(t)$.*

Use Theorem 7.4.7 on Page 139 and the continuity of the functions to write this limit equals

$$\begin{aligned} & \left(\lim_{t \to \pi/2} \cos t, \lim_{t \to \pi/2} \sin t, \lim_{t \to \pi/2} \left(t^2 + 1\right), \lim_{t \to \pi/2} \ln(t) \right) \\ = & \left(0, 1, \ln\left(\frac{\pi^2}{4} + 1\right), \ln\left(\frac{\pi}{2}\right) \right). \end{aligned}$$

Example 8.1.3 *Let $\mathbf{f}(t) = \left(\frac{\sin t}{t}, t^2, t+1\right)$. Find $\lim_{t \to 0} \mathbf{f}(t)$.*

Recall that $\lim_{t \to 0} \frac{\sin t}{t} = 1$. Then from Theorem 7.4.7 on Page 139, $\lim_{t \to 0} \mathbf{f}(t) = (1, 0, 1)$.

8.2 The Derivative And Integral

The following definition is on the derivative and integral of a vector valued function of one variable.

Definition 8.2.1 *The derivative of a function,* $\mathbf{f}'(t)$, *is defined as the following limit whenever the limit exists. If the limit does not exist, then neither does* $\mathbf{f}'(t)$.

$$\lim_{h\to 0}\frac{\mathbf{f}(t+h)-\mathbf{f}(x)}{h}\equiv \mathbf{f}'(t)$$

The function of h *on the left is called the difference quotient just as it was for a scalar valued function. If* $\mathbf{f}(t)=(f_1(t),\cdots,f_p(t))$ *and* $\int_a^b f_i(t)\,dt$ *exists for each* $i=1,\cdots,p$, *then* $\int_a^b \mathbf{f}(t)\,dt$ *is defined as the vector,*

$$\left(\int_a^b f_1(t)\,dt,\cdots,\int_a^b f_p(t)\,dt\right).$$

This is what is meant by saying $\mathbf{f}\in R([a,b])$.

This is exactly like the definition for a scalar valued function. As before,

$$\mathbf{f}'(x)=\lim_{y\to x}\frac{\mathbf{f}(y)-\mathbf{f}(x)}{y-x}.$$

As in the case of a scalar valued function, differentiability implies continuity but not the other way around.

Theorem 8.2.2 *If* $\mathbf{f}'(t)$ *exists, then* $\mathbf{f}$ *is continuous at* t.

Proof: Suppose $\varepsilon>0$ is given and choose $\delta_1>0$ such that if $|h|<\delta_1$,

$$\left|\frac{\mathbf{f}(t+h)-\mathbf{f}(t)}{h}-\mathbf{f}'(t)\right|<1.$$

then for such h, the triangle inequality implies

$$|\mathbf{f}(t+h)-\mathbf{f}(t)|<|h|+|\mathbf{f}'(t)|\,|h|.$$

Now letting $\delta<\min\left(\delta_1,\frac{\varepsilon}{1+|\mathbf{f}'(x)|}\right)$ it follows if $|h|<\delta$, then

$$|\mathbf{f}(t+h)-\mathbf{f}(t)|<\varepsilon.$$

Letting $y=h+t$, this shows that if $|y-t|<\delta$,

$$|\mathbf{f}(y)-\mathbf{f}(t)|<\varepsilon$$

which proves $\mathbf{f}$ is continuous at t. This proves the theorem.

As in the scalar case, there is a fundamental theorem of calculus.

Theorem 8.2.3 *If* $\mathbf{f}\in R([a,b])$ *and if* $\mathbf{f}$ *is continuous at* $t\in(a,b)$, *then*

$$\frac{d}{dt}\left(\int_a^t \mathbf{f}(s)\,ds\right)=\mathbf{f}(t).$$

Proof: Say $\mathbf{f}(t) = (f_1(t), \cdots, f_p(t))$. Then it follows

$$\frac{1}{h}\int_a^{t+h} \mathbf{f}(s)\, ds - \frac{1}{h}\int_a^t \mathbf{f}(s)\, ds = \left(\frac{1}{h}\int_t^{t+h} f_1(s)\, ds, \cdots, \frac{1}{h}\int_t^{t+h} f_p(s)\, ds\right)$$

and $\lim_{h\to 0} \frac{1}{h}\int_t^{t+h} f_i(s)\, ds = f_i(t)$ for each $i = 1, \cdots, p$ from the fundamental theorem of calculus for scalar valued functions. Therefore,

$$\lim_{h\to 0} \frac{1}{h}\int_a^{t+h} \mathbf{f}(s)\, ds - \frac{1}{h}\int_a^t \mathbf{f}(s)\, ds = (f_1(t), \cdots, f_p(t)) = \mathbf{f}(t)$$

and this proves the claim.

Example 8.2.4 *Let* $\mathbf{f}(x) = \mathbf{c}$ *where* $\mathbf{c}$ *is a constant. Find* $\mathbf{f}'(x)$.

The difference quotient,

$$\frac{\mathbf{f}(x+h) - \mathbf{f}(x)}{h} = \frac{\mathbf{c} - \mathbf{c}}{h} = \mathbf{0}$$

Therefore,

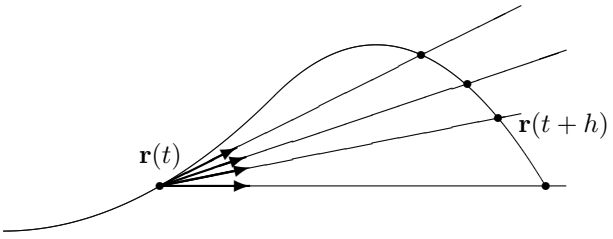
$$\lim_{h\to 0} \frac{\mathbf{f}(x+h) - \mathbf{f}(x)}{h} = \lim_{h\to 0} \mathbf{0} = \mathbf{0}$$

Example 8.2.5 *Let* $\mathbf{f}(t) = (at, bt)$ *where* a, b *are constants. Find* $\mathbf{f}'(t)$.

From the above discussion this derivative is just the vector valued functions whose components consist of the derivatives of the components of $\mathbf{f}$. Thus $\mathbf{f}'(t) = (a, b)$.

8.2.1 Geometric And Physical Significance Of The Derivative

Suppose $\mathbf{r}$ is a vector valued function of a parameter, t not necessarily time and consider the following picture of the points traced out by $\mathbf{r}$.



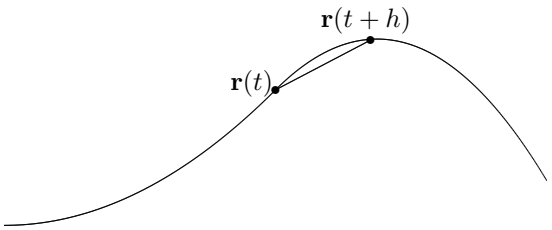
In this picture there are unit vectors in the direction of the vector from $\mathbf{r}(t)$ to $\mathbf{r}(t+h)$. You can see that it is reasonable to suppose these unit vectors, if they converge, converge to a unit vector, $\mathbf{T}$ which is tangent to the curve at the point $\mathbf{r}(t)$. Now each of these unit vectors is of the form

$$\frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{|\mathbf{r}(t+h) - \mathbf{r}(t)|} \equiv \mathbf{T}_h.$$

Thus $\mathbf{T}_h \to \mathbf{T}$, a unit tangent vector to the curve at the point $\mathbf{r}(t)$. Therefore,

$$\begin{aligned} \mathbf{r}'(t) &\equiv \lim_{h\to 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h} = \lim_{h\to 0} \frac{|\mathbf{r}(t+h) - \mathbf{r}(t)|}{h} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{|\mathbf{r}(t+h) - \mathbf{r}(t)|} \\ &= \lim_{h\to 0} \frac{|\mathbf{r}(t+h) - \mathbf{r}(t)|}{h} \mathbf{T}_h = |\mathbf{r}'(t)|\, \mathbf{T}. \end{aligned}$$

In the case that t is time, the expression $|\mathbf{r}(t+h)-\mathbf{r}(t)|$ is a good approximation for the distance traveled by the object on the time interval $[t,t+h]$. The real distance would be the length of the curve joining the two points but if h is very small, this is essentially equal to $|\mathbf{r}(t+h)-\mathbf{r}(t)|$ as suggested by the picture below.



Therefore,

$$\frac{|\mathbf{r}(t+h)-\mathbf{r}(t)|}{h}$$

gives for small h, the approximate distance travelled on the time interval, $[t,t+h]$ divided by the length of time, h. Therefore, this expression is really the average speed of the object on this small time interval and so the limit as $h\to 0$, deserves to be called the instantaneous speed of the object. Thus $|\mathbf{r}'(t)|\,\mathbf{T}$ represents the speed times a unit direction vector, $\mathbf{T}$ which defines the direction in which the object is moving. Thus $\mathbf{r}'(t)$ is the velocity of the object. This is the physical significance of the derivative when t is time.

How do you go about computing $\mathbf{r}'(t)$? Letting $\mathbf{r}(t)=(r_1(t),\cdots,r_q(t))$, the expression

$$\frac{\mathbf{r}(t_0+h)-\mathbf{r}(t_0)}{h} \tag{8.2}$$

is equal to

$$\left(\frac{r_1(t_0+h)-r_1(t_0)}{h},\cdots,\frac{r_q(t_0+h)-r_q(t_0)}{h}\right).$$

Then as h converges to 0, 8.2 converges to

$$\mathbf{v}\equiv(v_1,\cdots,v_q)$$

where $v_k=r_k'(t)$. This by Theorem 7.4.7 on Page 139, which says that the term in 8.2 gets close to a vector, $\mathbf{v}$ if and only if all the coordinate functions of the term in 8.2 get close to the corresponding coordinate functions of $\mathbf{v}$.

In the case where t is time, this simply says the velocity vector equals the vector whose components are the derivatives of the components of the displacement vector, $\mathbf{r}(t)$.

In any case, the vector, $\mathbf{T}$ determines a direction vector which is tangent to the curve at the point, $\mathbf{r}(t)$ and so it is possible to find parametric equations for the line tangent to the curve at various points.

Example 8.2.6 *Let* $\mathbf{r}(t)=\left(\sin t,t^2,t+1\right)$ *for* $t\in[0,5]$. *Find a tangent line to the curve parameterized by* $\mathbf{r}$ *at the point* $\mathbf{r}(2)$.

From the above discussion, a direction vector has the same direction as $\mathbf{r}'(2)$. Therefore, it suffices to simply use $\mathbf{r}'(2)$ as a direction vector for the line. $\mathbf{r}'(2)=(\cos 2,4,1)$. Therefore, a parametric equation for the tangent line is

$$(\sin 2,4,3)+t(\cos 2,4,1)=(x,y,z).$$

Example 8.2.7 *Let* $\mathbf{r}(t) = \left(\sin t, t^2, t+1\right)$ *for* $t \in [0,5]$. *Find the velocity vector when* $t = 1$.

From the above discussion, this is simply $\mathbf{r}'(1) = (\cos 1, 2, 1)$.

8.2.2 Differentiation Rules

There are rules which relate the derivative to the various operations done with vectors such as the dot product, the cross product, and vector addition and scalar multiplication.

Theorem 8.2.8 *Let* $a, b \in \mathbb{R}$ *and suppose* $\mathbf{f}'(t)$ *and* $\mathbf{g}'(t)$ *exist. Then the following formulas are obtained.*

$$(a\mathbf{f} + b\mathbf{g})'(t) = a\mathbf{f}'(t) + b\mathbf{g}'(t). \tag{8.3}$$

$$(\mathbf{f} \cdot \mathbf{g})'(t) = \mathbf{f}'(t) \cdot \mathbf{g}(t) + \mathbf{f}(t) \cdot \mathbf{g}'(t) \tag{8.4}$$

If $\mathbf{f}, \mathbf{g}$ *have values in* $\mathbb{R}^3$, *then*

$$(\mathbf{f} \times \mathbf{g})'(t) = \mathbf{f}(t) \times \mathbf{g}'(t) + \mathbf{f}'(t) \times \mathbf{g}(t) \tag{8.5}$$

The formulas, 8.4, and 8.5 are referred to as the product rule.

Proof: The first formula is left for you to prove. Consider the second, 8.4.

$$\lim_{h\to 0} \frac{\mathbf{f} \cdot \mathbf{g}(t+h) - \mathbf{fg}(t)}{h}$$

$$\begin{aligned}
&= \lim_{h\to 0} \frac{\mathbf{f}(t+h) \cdot \mathbf{g}(t+h) - \mathbf{f}(t+h) \cdot \mathbf{g}(t)}{h} + \frac{\mathbf{f}(t+h) \cdot \mathbf{g}(t) - \mathbf{f}(t) \cdot \mathbf{g}(t)}{h} \\
&= \lim_{h\to 0} \left(\mathbf{f}(t+h) \cdot \frac{(\mathbf{g}(t+h) - \mathbf{g}(t))}{h} + \frac{(\mathbf{f}(t+h) - \mathbf{f}(t))}{h} \cdot \mathbf{g}(t) \right) \\
&= \lim_{h\to 0} \sum_{k=1}^{n} f_k(t+h) \frac{(g_k(t+h) - g_k(t))}{h} + \sum_{k=1}^{n} \frac{(f_k(t+h) - f_k(t))}{h} g_k(t) \\
&= \sum_{k=1}^{n} f_k(t)\, g_k'(t) + \sum_{k=1}^{n} f_k'(t)\, g_k(t) \\
&= \mathbf{f}'(t) \cdot \mathbf{g}(t) + \mathbf{f}(t) \cdot \mathbf{g}'(t).
\end{aligned}$$

Formula 8.5 is left as an exercise which follows from the product rule and the definition of the cross product in terms of components given on Page 107.

Example 8.2.9 *Let*

$$\mathbf{r}(t) = \left(t^2, \sin t, \cos t\right)$$

and let $\mathbf{p}(t) = (t, \ln(t+1), 2t)$. *Find* $(\mathbf{r}(t) \times \mathbf{p}(t))'$.

From 8.5 this equals$(2t, \cos t, -\sin t) \times (t, \ln(t+1), 2t) + \left(t^2, \sin t, \cos t\right) \times \left(1, \frac{1}{t+1}, 2\right)$.

Example 8.2.10 *Let* $\mathbf{r}(t) = \left(t^2, \sin t, \cos t\right)$ *Find* $\int_0^\pi \mathbf{r}(t)\, dt$.

This equals $\left(\int_0^\pi t^2\, dt, \int_0^\pi \sin t\ dt, \int_0^\pi \cos t\, dt\right) = \left(\frac{1}{3}\pi^3, 2, 0\right)$.

Example 8.2.11 *An object has position* $\mathbf{r}(t) = \left(t^3, \frac{t}{1+1}, \sqrt{t^2+2}\right)$ *kilometers where* t *is given in hours. Find the velocity of the object in kilometers per hour when* $t = 1$.

Recall the velocity at time t was $\mathbf{r}'(t)$. Therefore, find $\mathbf{r}'(t)$ and plug in $t = 1$ to find the velocity.

$$\begin{aligned}\mathbf{r}'(t) &= \left(3t^2, \frac{1(1+t)-t}{(1+t)^2}, \frac{1}{2}\left(t^2+2\right)^{-1/2}2t\right)\\ &= \left(3t^2, \frac{1}{(1+t)^2}, \frac{1}{\sqrt{(t^2+2)}}t\right)\end{aligned}$$

When $t = 1$, the velocity is

$$\mathbf{r}'(1) = \left(3, \frac{1}{4}, \frac{1}{\sqrt{3}}\right) \text{ kilometers per hour.}$$

Obviously, this can be continued. That is, you can consider the possibility of taking the derivative of the derivative and then the derivative of that and so forth. The main thing to consider about this is the notation and it is exactly like it was in the case of a scalar valued function presented earlier. Thus $\mathbf{r}''(t)$ denotes the second derivative.

When you are given a vector valued function of one variable, sometimes it is possible to give a simple description of the curve which results. Usually it is not possible to do this!

Example 8.2.12 *Describe the curve which results from the vector valued function,* $\mathbf{r}(t) = (\cos 2t, \sin 2t, t)$ *where* $t \in \mathbb{R}$.

The first two components indicate that for $\mathbf{r}(t) = (x(t), y(t), z(t))$, the pair, $(x(t), y(t))$ traces out a circle. While it is doing so, $z(t)$ is moving at a steady rate in the positive direction. Therefore, the curve which results is a cork skrew shaped thing called a helix.

As an application of the theorems for differentiating curves, here is an interesting application. It is also a situation where the curve can be identified as something familiar.

Example 8.2.13 *Sound waves have the angle of incidence equal to the angle of reflection. Suppose you are in a large room and you make a sound. The sound waves spread out and you would expect your sound to be inaudible very far away. But what if the room were shaped so that the sound is reflected off the wall toward a single point, possibly far away from you? Then you might have the interesting phenomenon of someone far away hearing what you said quite clearly. How should the room be designed?*

Suppose you are located at the point $\mathbf{P}_0$ and the point where your sound is to be reflected is $\mathbf{P}_1$. Consider a plane which contains the two points and let $\mathbf{r}(t)$ denote a parameterization of the intersection of this plane with the walls of the room. Then the condition that the angle of reflection equals the angle of incidence reduces to saying the angle between $\mathbf{P}_0 - \mathbf{r}(t)$ and $-\mathbf{r}'(t)$ equals the angle between $\mathbf{P}_1 - \mathbf{r}(t)$ and $\mathbf{r}'(t)$. Draw a picture to see this. Therefore,

$$\frac{(\mathbf{P}_0 - \mathbf{r}(t)) \cdot (-\mathbf{r}'(t))}{|\mathbf{P}_0 - \mathbf{r}(t)|\,|\mathbf{r}'(t)|} = \frac{(\mathbf{P}_1 - \mathbf{r}(t)) \cdot (\mathbf{r}'(t))}{|\mathbf{P}_1 - \mathbf{r}(t)|\,|\mathbf{r}'(t)|}.$$

This reduces to

$$\frac{(\mathbf{r}(t) - \mathbf{P}_0) \cdot (-\mathbf{r}'(t))}{|\mathbf{r}(t) - \mathbf{P}_0|} = \frac{(\mathbf{r}(t) - \mathbf{P}_1) \cdot (\mathbf{r}'(t))}{|\mathbf{r}(t) - \mathbf{P}_1|} \tag{8.6}$$

Now

$$\frac{(\mathbf{r}(t) - \mathbf{P}_1) \cdot (\mathbf{r}'(t))}{|\mathbf{r}(t) - \mathbf{P}_1|} = \frac{d}{dt}|\mathbf{r}(t) - \mathbf{P}_1|$$

and a similar formula holds for $\mathbf{P}_1$ replaced with $\mathbf{P}_0$. This is because

$$|\mathbf{r}(t)-\mathbf{P}_1| = \sqrt{(\mathbf{r}(t)-\mathbf{P}_1)\cdot(\mathbf{r}(t)-\mathbf{P}_1)}$$

and so using the chain rule and product rule,

$$\begin{aligned}\frac{d}{dt}|\mathbf{r}(t)-\mathbf{P}_1| &= \frac{1}{2}((\mathbf{r}(t)-\mathbf{P}_1)\cdot(\mathbf{r}(t)-\mathbf{P}_1))^{-1/2}\,2((\mathbf{r}(t)-\mathbf{P}_1)\cdot\mathbf{r}'(t))\\ &= \frac{(\mathbf{r}(t)-\mathbf{P}_1)\cdot(\mathbf{r}'(t))}{|\mathbf{r}(t)-\mathbf{P}_1|}.\end{aligned}$$

Therefore, from 8.6,

$$\frac{d}{dt}(|\mathbf{r}(t)-\mathbf{P}_1|)+\frac{d}{dt}(|\mathbf{r}(t)-\mathbf{P}_0|)=0$$

showing that $|\mathbf{r}(t)-\mathbf{P}_1|+|\mathbf{r}(t)-\mathbf{P}_0| = C$ for some constant, C.This implies the curve of intersection of the plane with the room is an ellipse having $\mathbf{P}_0$ and $\mathbf{P}_1$ as the foci.

8.2.3 Leibniz's Notation

Leibniz's notation also generalizes routinely. For example, $\frac{d\mathbf{y}}{dt} = \mathbf{y}'(t)$ with other similar notations holding.

8.3 Product Rule For Matrices*

Here is the concept of the product rule extended to matrix multiplication.

Definition 8.3.1 *Let $A(t)$ be an $m\times n$ matrix. Say $A(t)=(A_{ij}(t))$. Suppose also that $A_{ij}(t)$ is a differentiable function for all i,j. Then define $A'(t)\equiv\left(A'_{ij}(t)\right)$. That is, $A'(t)$ is the matrix which consists of replacing each entry by its derivative. Such an $m\times n$ matrix in which the entries are differentiable functions is called a differentiable matrix.*

The next lemma is just a version of the product rule.

Lemma 8.3.2 *Let $A(t)$ be an $m\times n$ matrix and let $B(t)$ be an $n\times p$ matrix with the property that all the entries of these matrices are differentiable functions. Then*

$$(A(t)B(t))' = A'(t)B(t)+A(t)B'(t).$$

Proof: $(A(t)B(t))' = \left(C'_{ij}(t)\right)$ where $C_{ij}(t) = A_{ik}(t)B_{kj}(t)$ and the repeated index summation convention is being used. Therefore,

$$\begin{aligned}C'_{ij}(t) &= A'_{ik}(t)B_{kj}(t)+A_{ik}(t)B'_{kj}(t)\\ &= (A'(t)B(t))_{ij}+(A(t)B'(t))_{ij}\\ &= (A'(t)B(t)+A(t)B'(t))_{ij}\end{aligned}$$

Therefore, the ij^{th} entry of $A(t)B(t)$ equals the ij^{th} entry of $A'(t)B(t)+A(t)B'(t)$ and this proves the lemma.

8.4 Moving Coordinate Systems*

Let $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ be a right handed[1] orthonormal basis of vectors for each t. It is assumed these vectors are C^1 functions of t. Letting the positive x axis extend in the direction of $\mathbf{i}(t)$, the positive y axis extend in the direction of $\mathbf{j}(t)$, and the positive z axis extend in the direction of $\mathbf{k}(t)$, yields a moving coordinate system. Now let $\mathbf{u} = (u_1, u_2, u_3) \in \mathbb{R}^3$ and let t_0 be some reference time. For example you could let $t_0 = 0$. Then define the components of $\mathbf{u}$ with respect to these vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ at time t_0 as

$$\mathbf{u} \equiv u_1\mathbf{i}(t_0) + u_2\mathbf{j}(t_0) + u_3\mathbf{k}(t_0).$$

Let $\mathbf{u}(t)$ be defined as the vector which has the same components with respect to $\mathbf{i}, \mathbf{j}, \mathbf{k}$ but at time t. Thus

$$\mathbf{u}(t) \equiv u_1\mathbf{i}(t) + u_2\mathbf{j}(t) + u_3\mathbf{k}(t).$$

and the vector has changed although the components have not.

For example, this is exactly the situation in the case of apparently fixed basis vectors on the earth if $\mathbf{u}$ is a position vector from the given spot on the earth's surface to a point regarded as fixed with the earth due to its keeping the same coordinates relative to coordinate axes which are fixed with the earth.

Now define a linear transformation $Q(t)$ mapping $\mathbb{R}^3$ to $\mathbb{R}^3$ by

$$Q(t)\mathbf{u} \equiv u_1\mathbf{i}(t) + u_2\mathbf{j}(t) + u_3\mathbf{k}(t)$$

where

$$\mathbf{u} \equiv u_1\mathbf{i}(t_0) + u_2\mathbf{j}(t_0) + u_3\mathbf{k}(t_0)$$

Thus letting $\mathbf{v}, \mathbf{u} \in \mathbb{R}^3$ be vectors and α, β, scalars,

$$Q(t)(\alpha\mathbf{u} + \beta\mathbf{v}) \equiv (\alpha u_1 + \beta v_1)\mathbf{i}(t) + (\alpha u_2 + \beta v_2)\mathbf{j}(t) + (\alpha u_3 + \beta v_3)\mathbf{k}(t)$$

$$\begin{aligned} &= (\alpha u_1\mathbf{i}(t) + \alpha u_2\mathbf{j}(t) + \alpha u_3\mathbf{k}(t)) + (\beta v_1\mathbf{i}(t) + \beta v_2\mathbf{j}(t) + \beta v_3\mathbf{k}(t)) \\ &= \alpha(u_1\mathbf{i}(t) + u_2\mathbf{j}(t) + u_3\mathbf{k}(t)) + \beta(v_1\mathbf{i}(t) + v_2\mathbf{j}(t) + v_3\mathbf{k}(t)) \\ &\equiv \alpha Q(t)\mathbf{u} + \beta Q(t)\mathbf{v} \end{aligned}$$

showing that $Q(t)$ is a linear transformation. Also, $Q(t)$ preserves all distances because, since the vectors, $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ form an orthonormal set,

$$|Q(t)\mathbf{u}| = \left(\sum_{i=1}^{3} \left(u^i\right)^2\right)^{1/2} = |\mathbf{u}|.$$

For simplicity, let $\mathbf{i}(t) = \mathbf{e}_1(t), \mathbf{j}(t) = \mathbf{e}_2(t), \mathbf{k}(t) = \mathbf{e}_3(t)$ and

$$\mathbf{i}(t_0) = \mathbf{e}_1(t_0), \mathbf{j}(t_0) = \mathbf{e}_2(t_0), \mathbf{k}(t_0) = \mathbf{e}_3(t_0).$$

Then using the repeated index summation convention,

$$\mathbf{u}(t) = u_j\mathbf{e}_j(t) = u_j\mathbf{e}_j(t) \cdot \mathbf{e}_i(t_0)\mathbf{e}_i(t_0)$$

and so with respect to the basis, $\mathbf{i}(t_0) = \mathbf{e}_1(t_0), \mathbf{j}(t_0) = \mathbf{e}_2(t_0), \mathbf{k}(t_0) = \mathbf{e}_3(t_0)$, the matrix of $Q(t)$ is

$$Q_{ij}(t) = \mathbf{e}_i(t_0) \cdot \mathbf{e}_j(t)$$

Recall this means you take a vector, $\mathbf{u} \in \mathbb{R}^3$ which is a list of the components of $\mathbf{u}$ with respect to $\mathbf{i}(t_0), \mathbf{j}(t_0), \mathbf{k}(t_0)$ and when you multiply by $Q(t)$ you get the components of $\mathbf{u}(t)$ with respect to $\mathbf{i}(t_0), \mathbf{j}(t_0), \mathbf{k}(t_0)$. I will refer to this matrix as $Q(t)$ to save notation.

[1] Recall that right handed implies $\mathbf{i} \times \mathbf{j} = \mathbf{k}$.

Lemma 8.4.1 *Suppose $Q(t)$ is a real, differentiable $n \times n$ matrix which preserves distances. Then $Q(t) Q(t)^T = Q(t)^T Q(t) = I$. Also, if $\mathbf{u}(t) \equiv Q(t)\mathbf{u}$, then there exists a vector, $\mathbf{\Omega}(t)$ such that*

$$\mathbf{u}'(t) = \mathbf{\Omega}(t) \times \mathbf{u}(t).$$

Proof: Recall that $(\mathbf{z} \cdot \mathbf{w}) = \frac{1}{4}\left(|\mathbf{z}+\mathbf{w}|^2 - |\mathbf{z}-\mathbf{w}|^2\right)$. Therefore,

$$\begin{aligned}(Q(t)\mathbf{u} \cdot Q(t)\mathbf{w}) &= \frac{1}{4}\left(|Q(t)(\mathbf{u}+\mathbf{w})|^2 - |Q(t)(\mathbf{u}-\mathbf{w})|^2\right) \\ &= \frac{1}{4}\left(|\mathbf{u}+\mathbf{w}|^2 - |\mathbf{u}-\mathbf{w}|^2\right) \\ &= (\mathbf{u} \cdot \mathbf{w}).\end{aligned}$$

This implies

$$\left(Q(t)^T Q(t)\mathbf{u} \cdot \mathbf{w}\right) = (\mathbf{u} \cdot \mathbf{w})$$

for all $\mathbf{u}, \mathbf{w}$. Therefore, $Q(t)^T Q(t)\mathbf{u} = \mathbf{u}$ and so $Q(t)^T Q(t) = Q(t) Q(t)^T = I$. This proves the first part of the lemma.

It follows from the product rule, Lemma 8.3.2 that

$$Q'(t) Q(t)^T + Q(t) Q'(t)^T = 0$$

and so

$$Q'(t) Q(t)^T = -\left(Q'(t) Q(t)^T\right)^T. \tag{8.7}$$

From the definition, $Q(t)\mathbf{u} = \mathbf{u}(t)$,

$$\mathbf{u}'(t) = Q'(t)\mathbf{u} = Q'(t)\overbrace{Q(t)^T \mathbf{u}(t)}^{=\mathbf{u}}.$$

Then writing the matrix of $Q'(t) Q(t)^T$ with respect to $\mathbf{i}(t_0), \mathbf{j}(t_0), \mathbf{k}(t_0)$, it follows from 8.7 that the matrix of $Q'(t) Q(t)^T$ is of the form

$$\begin{pmatrix} 0 & -\omega_3(t) & \omega_2(t) \\ \omega_3(t) & 0 & -\omega_1(t) \\ -\omega_2(t) & \omega_1(t) & 0 \end{pmatrix}$$

for some time dependent scalars, ω_i. Therefore,

$$\begin{aligned}\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}'(t) &= \begin{pmatrix} 0 & -\omega_3(t) & \omega_2(t) \\ \omega_3(t) & 0 & -\omega_1(t) \\ -\omega_2(t) & \omega_1(t) & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}(t) \\ &= \begin{pmatrix} w_2(t) u_3(t) - w_3(t) u_2(t) \\ w_3(t) u_1(t) - w_1(t) u_3(t) \\ w_1(t) u_2(t) - w_2(t) u_1(t) \end{pmatrix}\end{aligned}$$

where the u_i are the components of the vector $\mathbf{u}(t)$ in terms of the fixed vectors

$$\mathbf{i}(t_0), \mathbf{j}(t_0), \mathbf{k}(t_0).$$

Therefore,

$$\mathbf{u}'(t) = \mathbf{\Omega}(t) \times \mathbf{u}(t) = Q'(t) Q(t)^T \mathbf{u}(t) \tag{8.8}$$

where

$$\mathbf{\Omega}(t) = \omega_1(t)\mathbf{i}(t_0) + \omega_2(t)\mathbf{j}(t_0) + \omega_3(t)\mathbf{k}(t_0).$$

because

$$\mathbf{\Omega}(t)\times\mathbf{u}(t)\equiv\begin{vmatrix}\mathbf{i}(t_0) & \mathbf{j}(t_0) & \mathbf{k}(t_0)\\ w_1 & w_2 & w_3\\ u_1 & u_2 & u_3\end{vmatrix}\equiv$$

$$\mathbf{i}(t_0)(w_2u_3-w_3u_2)+\mathbf{j}(t_0)(w_3u_1-w_1u_3)+\mathbf{k}(t_0)(w_1u_2-w_2u_1).$$

This proves the lemma and yields the existence part of the following theorem.

Theorem 8.4.2 *Let $\mathbf{i}(t),\mathbf{j}(t),\mathbf{k}(t)$ be as described. Then there exists a unique vector $\mathbf{\Omega}(t)$ such that if $\mathbf{u}(t)$ is a vector whose components are constant with respect to $\mathbf{i}(t),\mathbf{j}(t),\mathbf{k}(t)$, then*

$$\mathbf{u}'(t)=\mathbf{\Omega}(t)\times\mathbf{u}(t).$$

Proof: It only remains to prove uniqueness. Suppose $\mathbf{\Omega}_1$ also works. Then $\mathbf{u}(t)=Q(t)\mathbf{u}$ and so $\mathbf{u}'(t)=Q'(t)\mathbf{u}$ and

$$Q'(t)\mathbf{u}=\mathbf{\Omega}\times Q(t)\mathbf{u}=\mathbf{\Omega}_1\times Q(t)\mathbf{u}$$

for all $\mathbf{u}$. Therefore,

$$(\mathbf{\Omega}-\mathbf{\Omega}_1)\times Q(t)\mathbf{u}=\mathbf{0}$$

for all $\mathbf{u}$ and since $Q(t)$ is one to one and onto, this implies $(\mathbf{\Omega}-\mathbf{\Omega}_1)\times\mathbf{w}=\mathbf{0}$ for all $\mathbf{w}$ and thus $\mathbf{\Omega}-\mathbf{\Omega}_1=\mathbf{0}$. This proves the theorem.

Definition 8.4.3 *A **rigid body** in $\mathbb{R}^3$ has a moving coordinate system with the property that for an observer on the rigid body, the vectors, $\mathbf{i}(t),\mathbf{j}(t),\mathbf{k}(t)$ are constant. More generally, a vector $\mathbf{u}(t)$ is said to be fixed with the body if to a person on the body, the vector appears to have the same magnitude and same direction independent of t. Thus $\mathbf{u}(t)$ is fixed with the body if $\mathbf{u}(t)=u_1\mathbf{i}(t)+u_2\mathbf{j}(t)+u_3\mathbf{k}(t)$.*

The following comes from the above discussion.

Theorem 8.4.4 *Let $B(t)$ be the set of points in three dimensions occupied by a rigid body. Then there exists a vector $\mathbf{\Omega}(t)$ such that whenever $\mathbf{u}(t)$ is fixed with the rigid body,*

$$\mathbf{u}'(t)=\mathbf{\Omega}(t)\times\mathbf{u}(t).$$

8.5 Exercises

1. Find the following limits if possible
   (a) $\lim_{x\to 0+}\left(\frac{|x|}{x},\sin x/x,\cos x\right)$
   (b) $\lim_{x\to 0+}\left(\frac{x}{|x|},\sec x,e^x\right)$
   (c) $\lim_{x\to 4}\left(\frac{x^2-16}{x+4},x+7,\frac{\tan 4x}{5x}\right)$
   (d) $\lim_{x\to\infty}\left(\frac{x}{1+x^2},\frac{x^2}{1+x^2},\frac{\sin x^2}{x}\right)$
2. Find $\lim_{x\to 2}\left(\frac{x^2-4}{x+2},x^2+2x-1,\frac{x^2-4}{x-2}\right)$.
3. Prove from the definition that $\lim_{x\to a}(\sqrt[3]{x},x+1)=(\sqrt[3]{a},a+1)$ for all $a\in\mathbb{R}$. **Hint:** You might want to use the formula for the difference of two cubes,

$$a^3-b^3=(a-b)\left(a^2+ab+b^2\right).$$

4. Let $\mathbf{r}(t) = \left(4 + (t-1)^2, \sqrt{t^2+1}(t-1)^3, \frac{(t-1)^3}{t^5}\right)$ describe the position of an object in $\mathbb{R}^3$ as a function of t where t is measured in seconds and $\mathbf{r}(t)$ is measured in meters. Is the velocity of this object ever equal to zero? If so, find the value of t at which this occurs and the point in $\mathbb{R}^3$ at which the velocity is zero.

5. Let $\mathbf{r}(t) = \left(\sin 2t, t^2, 2t+1\right)$ for $t \in [0,4]$. Find a tangent line to the curve parameterized by $\mathbf{r}$ at the point $\mathbf{r}(2)$.

6. Let $\mathbf{r}(t) = \left(t, \sin t^2, t+1\right)$ for $t \in [0,5]$. Find a tangent line to the curve parameterized by $\mathbf{r}$ at the point $\mathbf{r}(2)$.

7. Let $\mathbf{r}(t) = \left(\sin t, t^2, \cos\left(t^2\right)\right)$ for $t \in [0,5]$. Find a tangent line to the curve parameterized by $\mathbf{r}$ at the point $\mathbf{r}(2)$.

8. Let $\mathbf{r}(t) = \left(\sin t, \cos\left(t^2\right), t+1\right)$ for $t \in [0,5]$. Find the velocity when $t = 3$.

9. Let $\mathbf{r}(t) = \left(\sin t, t^2, t+1\right)$ for $t \in [0,5]$. Find the velocity when $t = 3$.

10. Let $\mathbf{r}(t) = \left(t, \ln\left(t^2+1\right), t+1\right)$ for $t \in [0,5]$. Find the velocity when $t = 3$.

11. Suppose an object has position $\mathbf{r}(t) \in \mathbb{R}^3$ where $\mathbf{r}$ is differentiable and suppose also that $|\mathbf{r}(t)| = c$ where c is a constant.

    (a) Show first that this condition does not require $\mathbf{r}(t)$ to be a constant. **Hint:** You can do this either mathematically or by giving a physical example.

    (b) Show that you can conclude that $\mathbf{r}'(t) \cdot \mathbf{r}(t) = 0$. That is, the velocity is always perpendicular to the displacement.

12. Prove 8.5 from the component description of the cross product.

13. Prove 8.5 from the formula $(\mathbf{f} \times \mathbf{g})_i = \varepsilon_{ijk} f_j g_k$.

14. Prove 8.5 directly from the definition of the derivative without considering components.

15. A bezier curve in $\mathbb{R}^n$ is a vector valued function of the form

$$\mathbf{y}(t) = \sum_{k=0}^{n} \binom{n}{k} \mathbf{x}_k (1-t)^{n-k} t^k$$

    where here the $\binom{n}{k}$ are the binomial coefficients and $\mathbf{x}_k$ are $n+1$ points in $\mathbb{R}^n$. Show that $\mathbf{y}(0) = \mathbf{x}_0$, $\mathbf{y}(1) = \mathbf{x}_n$, and find $\mathbf{y}'(0)$ and $\mathbf{y}'(1)$. Recall that $\binom{n}{0} = \binom{n}{n} = 1$ and $\binom{n}{n-1} = \binom{n}{1} = n$. Curves of this sort are important in various computer programs.

16. Suppose $\mathbf{r}(t)$, $\mathbf{s}(t)$, and $\mathbf{p}(t)$ are three differentiable functions of t which have values in $\mathbb{R}^3$. Find a formula for $(\mathbf{r}(t) \times \mathbf{s}(t) \cdot \mathbf{p}(t))'$.

17. If $\mathbf{r}'(t) = \mathbf{0}$ for all $t \in (a,b)$, show there exists a constant vector, $\mathbf{c}$ such that $\mathbf{r}(t) = \mathbf{c}$ for all $t \in (a,b)$.

18. If $\mathbf{F}'(t) = \mathbf{f}(t)$ for all $t \in (a,b)$ and $\mathbf{F}$ is continuous on $[a,b]$, show $\int_a^b \mathbf{f}(t)\, dt = \mathbf{F}(b) - \mathbf{F}(a)$.

19. Verify that if $\mathbf{\Omega} \times \mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$, then $\mathbf{\Omega} = \mathbf{0}$.

20. Verify that if $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \cdot \mathbf{u} = 0$ and both $\mathbf{\Omega}$ and $\mathbf{\Omega}_1$ satisfy $\mathbf{\Omega} \times \mathbf{u} = \mathbf{v}$, then $\mathbf{\Omega}_1 = \mathbf{\Omega}$.

8.6 Exercises With Answers

1. Find the following limits if possible

   (a) $\lim_{x\to 0+}\left(\frac{|x|}{x}, \sin 2x/x, \frac{\tan x}{x}\right) = (1,2,1)$

   (b) $\lim_{x\to 0+}\left(\frac{x}{|x|}, \cos x, e^{2x}\right) = (1,1,1)$

   (c) $\lim_{x\to 4}\left(\frac{x^2-16}{x+4}, x-7, \frac{\tan 7x}{5x}\right) = \left(0,-3,\frac{7}{5}\right)$

2. Let $\mathbf{r}(t) = \left(4+(t-1)^2, \sqrt{t^2+1}(t-1)^3, \frac{(t-1)^3}{t^5}\right)$ describe the position of an object in $\mathbb{R}^3$ as a function of t where t is measured in seconds and $\mathbf{r}(t)$ is measured in meters. Is the velocity of this object ever equal to zero? If so, find the value of t at which this occurs and the point in $\mathbb{R}^3$ at which the velocity is zero.

   You need to differentiate this. $\mathbf{r}'(t) = \left(2(t-1), (t-1)^2\frac{4t^2-t+3}{\sqrt{(t^2+1)}}, -(t-1)^2\frac{2t-5}{t^6}\right)$. Now you need to find the value(s) of t where $\mathbf{r}'(t) = \mathbf{0}$.

3. Let $\mathbf{r}(t) = \left(\sin t, t^2, 2t+1\right)$ for $t \in [0,4]$. Find a tangent line to the curve parameterized by $\mathbf{r}$ at the point $\mathbf{r}(2)$.

   $\mathbf{r}'(t) = (\cos t, 2t, 2)$. When $t=2$, the point on the curve is $(\sin 2, 4, 5)$. A direction vector is $\mathbf{r}'(2)$ and so a tangent line is $\mathbf{r}(t) = (\sin 2, 4, 5) + t(\cos 2, 4, 2)$.

4. Let $\mathbf{r}(t) = \left(\sin t, \cos\left(t^2\right), t+1\right)$ for $t \in [0,5]$. Find the velocity when $t=3$.

   $\mathbf{r}'(t) = \left(\cos t, -2t\sin\left(t^2\right), 1\right)$. The velocity when $t=3$ is just

   $$\mathbf{r}'(3) = (\cos 3, -6\sin(9), 1).$$

5. Prove 8.5 directly from the definition of the derivative without considering components.

   The formula for the derivative of a cross product can be obtained in the usual way using rules of the cross product.

   $$\begin{aligned}\frac{\mathbf{u}(t+h)\times\mathbf{v}(t+h)-\mathbf{u}(t)\times\mathbf{v}(t)}{h} &= \frac{\mathbf{u}(t+h)\times\mathbf{v}(t+h)-\mathbf{u}(t+h)\times\mathbf{v}(t)}{h} \\ &\quad + \frac{\mathbf{u}(t+h)\times\mathbf{v}(t)-\mathbf{u}(t)\times\mathbf{v}(t)}{h} \\ &= \mathbf{u}(t+h)\times\left(\frac{\mathbf{v}(t+h)-\mathbf{v}(t)}{h}\right) + \left(\frac{\mathbf{u}(t+h)-\mathbf{u}(t)}{h}\right)\times\mathbf{v}(t)\end{aligned}$$

   Doesn't this remind you of the proof of the product rule? Now procede in the usual way. If you want to really understand this, you should consider why $\mathbf{u},\mathbf{v}\to\mathbf{u}\times\mathbf{v}$ is a continuous map. This follows from the geometric description of the cross product or more easily from the coordinate description.

6. Suppose $\mathbf{r}(t)$, $\mathbf{s}(t)$, and $\mathbf{p}(t)$ are three differentiable functions of t which have values in $\mathbb{R}^3$. Find a formula for $(\mathbf{r}(t)\times\mathbf{s}(t)\cdot\mathbf{p}(t))'$.

   From the product rules for the cross and dot product, this equals

   $$\begin{aligned}&(\mathbf{r}(t)\times\mathbf{s}(t))'\cdot\mathbf{p}(t) + \mathbf{r}(t)\times\mathbf{s}(t)\cdot\mathbf{p}'(t) \\ = \;& \mathbf{r}'(t)\times\mathbf{s}(t)\cdot\mathbf{p}(t) + \mathbf{r}(t)\times\mathbf{s}'(t)\cdot\mathbf{p}(t) + \mathbf{r}(t)\times\mathbf{s}(t)\cdot\mathbf{p}'(t)\end{aligned}$$

7. If $\mathbf{r}'(t) = \mathbf{0}$ for all $t \in (a,b)$, show there exists a constant vector, $\mathbf{c}$ such that $\mathbf{r}(t) = \mathbf{c}$ for all $t \in (a,b)$.

   Do this by considering standard one variable calculus and on the components of $\mathbf{r}(t)$.

8. If $\mathbf{F}'(t) = \mathbf{f}(t)$ for all $t \in (a,b)$ and $\mathbf{F}$ is continuous on $[a,b]$, show $\int_a^b \mathbf{f}(t)\,dt = \mathbf{F}(b) - \mathbf{F}(a)$.

   Do this by considering standard one variable calculus and on the components of $\mathbf{r}(t)$.

9. Verify that if $\mathbf{\Omega} \times \mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$, then $\mathbf{\Omega} = \mathbf{0}$.

   Geometrically this says that if $\mathbf{\Omega}$ is not equal to zero then it is parallel to every vector. Why does this make it obvious that $\mathbf{\Omega}$ must equal zero?

8.7 Newton's Laws Of Motion

Definition 8.7.1 *Let $\mathbf{r}(t)$ denote the position of an object. Then the acceleration of the object is defined to be $\mathbf{r}''(t)$.*

Newton's[2] first law is: "Every body persists in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by forces impressed on it."

Newton's second law is:

$$\mathbf{F} = m\mathbf{a} \tag{8.9}$$

where $\mathbf{a}$ is the acceleration and m is the mass of the object.

Newton's third law states: "To every action there is always opposed an equal reaction; or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts."

Of these laws, only the second two are independent of each other, the first law being implied by the second. The third law says roughly that if you apply a force to something, the thing applies the same force back.

The second law is the one of most interest. Note that the statement of this law depends on the concept of the derivative because the acceleration is defined as a derivative. Newton used calculus and these laws to solve profound problems involving the motion of the planets and other problems in mechanics. The next example involves the concept that if you know the force along with the initial velocity and initial position, then you can determine the position.

Example 8.7.2 *Let $\mathbf{r}(t)$ denote the position of an object of mass 2 kilogram at time t and suppose the force acting on the object is given by $\mathbf{F}(t) = \left(t, 1-t^2, 2e^{-t}\right)$. Suppose $\mathbf{r}(0) = (1,0,1)$ meters, and $\mathbf{r}'(0) = (0,1,1)$ meters/sec. Find $\mathbf{r}(t)$.*

By Newton's second law, $2\mathbf{r}''(t) = \mathbf{F}(t) = \left(t, 1-t^2, 2e^{-t}\right)$ and so

$$\mathbf{r}''(t) = \left(t/2, \left(1-t^2\right)/2, e^{-t}\right).$$

[2]Isaac Newton 1642-1727 is often credited with inventing calculus although this is not correct since most of the ideas were in existence earlier. However, he made major contributions to the subject partly in order to study physics and astronomy. He formulated the laws of gravity, made major contributions to optics, and stated the fundamental laws of mechanics listed here. He invented a version of the binomial theorem when he was only 23 years old and built a reflecting telescope. He showed that Kepler's laws for the motion of the planets came from calculus and his laws of gravitation. In 1686 he published an important book, Principia, in which many of his ideas are found. Newton was also very interested in theology and had strong views on the nature of God which were based on his study of the Bible and early Christian writings. He finished his life as Master of the Mint.

Therefore the velocity is given by

$$\mathbf{r}'(t) = \left(\frac{t^2}{4}, \frac{t - t^3/3}{2}, -e^{-t}\right) + \mathbf{c}$$

where $\mathbf{c}$ is a constant vector which must be determined from the initial condition given for the velocity. Thus letting $\mathbf{c} = (c_1, c_2, c_3)$,

$$(0, 1, 1) = (0, 0, -1) + (c_1, c_2, c_3)$$

which requires $c_1 = 0, c_2 = 1$, and $c_3 = 2$. Therefore, the velocity is found.

$$\mathbf{r}'(t) = \left(\frac{t^2}{4}, \frac{t - t^3/3}{2} + 1, -e^{-t} + 2\right).$$

Now from this, the displacement must equal

$$\mathbf{r}(t) = \left(\frac{t^3}{12}, \frac{t^2/2 - t^4/12}{2} + t, e^{-t} + 2t\right) + (C_1, C_2, C_3)$$

where the constant vector, (C_1, C_2, C_3) must be determined from the initial condition for the displacement. Thus

$$\mathbf{r}(0) = (1, 0, 1) = (0, 0, 1) + (C_1, C_2, C_3)$$

which means $C_1 = 1, C_2 = 0$, and $C_3 = 0$. Therefore, the displacement has also been found.

$$\mathbf{r}(t) = \left(\frac{t^3}{12} + 1, \frac{t^2/2 - t^4/12}{2} + t, e^{-t} + 2t\right) \text{ meters.}$$

Actually, in applications of this sort of thing acceleration does not usually come to you as a nice given function written in terms of simple functions you understand. Rather, it comes as measurements taken by instruments and the position is continuously being updated based on this information. Another situation which often occurs is the case when the forces on the object depend not just on time but also on the position or velocity of the object.

Example 8.7.3 *An artillery piece is fired at ground level on a level plain. The angle of elevation is $\pi/6$ radians and the speed of the shell is 400 meters per second. How far does the shell fly before hitting the ground?*

Neglect air resistance in this problem. Also let the direction of flight be along the positive x axis. Thus the initial velocity is the vector, $400\cos(\pi/6)\,\mathbf{i} + 400\sin(\pi/6)\,\mathbf{j}$ while the only force experienced by the shell after leaving the artillery piece is the force of gravity, $-mg\mathbf{j}$ where m is the mass of the shell. The acceleration of gravity equals 9.8 meters per sec^2 and so the following needs to be solved.

$$m\mathbf{r}''(t) = -mg\mathbf{j},\ \mathbf{r}(0) = (0, 0), \mathbf{r}'(0) = 400\cos(\pi/6)\,\mathbf{i} + 400\sin(\pi/6)\,\mathbf{j}.$$

Denoting $\mathbf{r}(t)$ as $(x(t), y(t))$,

$$x''(t) = 0,\ y''(t) = -g.$$

Therefore, $y'(t) = -gt + C$ and from the information on the initial velocity,

$$C = 400\sin(\pi/6) = 200.$$

Thus

$$y(t) = -4.9t^2 + 200t + D.$$

$D = 0$ because the artillery piece is fired at ground level which requires both x and y to equal zero at this time. Similarly, $x'(t) = 400\cos(\pi/6)$ so $x(t) = 400\cos(\pi/6)\,t = 200\sqrt{3}t$. The shell hits the ground when $y = 0$ and this occurs when $-4.9t^2 + 200t = 0$. Thus $t = 40.816\,326\,530\,6$ seconds and so at this time,

$$x = 200\sqrt{3}\,(40.816\,326\,530\,6) = 14139.190\,265\,9 \text{ meters.}$$

The next example is more complicated because it also takes in to account air resistance. We do not live in a vacuum.

Example 8.7.4 *A lump of "blue ice" escapes the lavatory of a jet flying at 600 miles per hour at an altitude of 30,000 feet. This blue ice weighs 64 pounds near the earth and experiences a force of air resistance equal to* $(-.1)\,\mathbf{r}'(t)$ *pounds. Find the position and velocity of the blue ice as a function of time measured in seconds. Also find the velocity when the lump hits the ground. Such lumps have been known to surprise people on the ground.*

The first thing needed is to obtain information which involves consistent units. The blue ice weighs 32 pounds near the earth. Thus 32 pounds is the force exerted by gravity on the lump and so its mass must be given by Newton's second law as follows.

$$64 = m \times 32.$$

Thus $m = 2$ slugs. The slug is the unit of mass in the system involving feet and pounds. The jet is flying at 600 miles per hour. I want to change this to feet per second. Thus it flies at

$$\frac{600 \times 5280}{60 \times 60} = 880 \text{ feet per second.}$$

The explanation for this is that there are 5280 feet in a mile and so it goes 600×5280 feet in one hour. There are 60×60 seconds in an hour. The position of the lump of blue ice will be computed from a point on the ground directly beneath the airplane at the instant the blue ice escapes and regard the airplane as moving in the direction of the positive x axis. Thus the initial displacement is

$$\mathbf{r}(0) = (0, 30000) \text{ feet}$$

and the initial velocity is

$$\mathbf{r}'(0) = (880, 0) \text{ feet/sec.}$$

The force of gravity is

$$(0, -64) \text{ pounds}$$

and the force due to air resistance is

$$(-.1)\,\mathbf{r}'(t) \text{ pounds.}$$

Newtons second law yields the following initial value problem for $\mathbf{r}(t) = (r_1(t), r_2(t))$.

$$2\,(r_1''(t), r_2''(t)) = (-.1)\,(r_1'(t), r_2'(t)) + (0, -64),\ (r_1(0), r_2(0)) = (0, 30000),$$
$$(r_1'(0), r_2'(0)) = (880, 0)$$

Therefore,

$$\begin{array}{c} 2r_1''(t) + (.1)\, r_1'(t) = 0 \\ 2r_2''(t) + (.1)\, r_2'(t) = -64 \\ r_1(0) = 0 \\ r_2(0) = 30000 \\ r_1'(0) = 880 \\ r_2'(0) = 0 \end{array} \tag{8.10}$$

To save on repetition solve

$$mr'' + kr' = c, r(0) = u, r'(0) = v.$$

Divide the differential equation by m and get

$$r'' + (k/m)\, r' = c/m.$$

Now multiply both sides by $e^{(k/m)t}$. You should check this gives

$$\frac{d}{dt}\left(e^{(k/m)t} r'\right) = (c/m)\, e^{(k/m)t}$$

Therefore,

$$e^{(k/m)t} r' = \frac{1}{k} e^{\frac{k}{m}t} c + C$$

and using the initial condition, $v = c/k + C$ and so

$$r'(t) = (c/k) + (v - (c/k))\, e^{-\frac{k}{m}t}$$

Now this implies

$$r(t) = (c/k)\, t - \frac{1}{k} m e^{-\frac{k}{m}t}\left(v - \frac{c}{k}\right) + D \tag{8.11}$$

where D is a constant to be determined from the initial conditions. Thus

$$u = -\frac{m}{k}\left(v - \frac{c}{k}\right) + D$$

and so

$$r(t) = (c/k)\, t - \frac{1}{k} m e^{-\frac{k}{m}t}\left(v - \frac{c}{k}\right) + \left(u + \frac{m}{k}\left(v - \frac{c}{k}\right)\right).$$

Now apply this to the system 8.10 to find

$$\begin{aligned} r_1(t) &= -\frac{1}{(.1)} 2\left(\exp\left(\frac{-(.1)}{2} t\right)\right)(880) + \left(\frac{2}{(.1)}(880)\right) \\ &= -17600.0 \exp(-.0\,5t) + 17600.0 \end{aligned}$$

and

$$\begin{aligned} r_2(t) &= (-64/(.1))\, t - \frac{1}{(.1)} 2\left(\exp\left(-\frac{(.1)}{2} t\right)\right)\left(\frac{64}{(.1)}\right) + \left(30000 + \frac{2}{(.1)}\left(\frac{64}{(.1)}\right)\right) \\ &= -640.0t - 12800.0 \exp(-.0\,5t) + 42800.0 \end{aligned}$$

This gives the coordinates of the position. What of the velocity? Using 8.11 in the same way to obtain the velocity,

$$\begin{aligned} r_1'(t) &= 880.0 \exp(-.0\,5t), \\ r_2'(t) &= -640.0 + 640.0 \exp(-.0\,5t). \end{aligned} \tag{8.12}$$

To determine the velocity when the blue ice hits the ground, it is necessary to find the value of t when this event takes place and then to use 8.12 to determine the velocity. It hits ground when $r_2(t) = 0$. Thus it suffices to solve the equation,

$$0 = -640.0t - 12800.0\exp(-.0\,5t) + 42800.0.$$

This is a fairly hard equation to solve using the methods of algebra. In fact, I do not have a good way to find this value of t using algebra. However if plugging in various values of t using a calculator you eventually find that when $t = 66.14$,

$$-640.0(66.14) - 12800.0\exp(-.0\,5(66.14)) + 42800.0 = 1.588 \text{ feet.}$$

This is close enough to hitting the ground and so plugging in this value for t yields the approximate velocity,

$$(880.0\exp(-.0\,5(66.14)), -640.0 + 640.0\exp(-.0\,5(66.14))) = (32.23, -616.56).$$

Notice how because of air resistance the component of velocity in the horizontal direction is only about 32 feet per second even though this component started out at 880 feet per second while the component in the vertical direction is -616 feet per second even though this component started off at 0 feet per second. You see that air resistance can be very important so it is not enough to pretend, as is often done in beginning physics courses that everything takes place in a vacuum. Actually, this problem used several physical simplifications. It was assumed the force acting on the lump of blue ice by gravity was constant. This is not really true because it actually depends on the distance between the center of mass of the earth and the center of mass of the lump. It was also assumed the air resistance is proportional to the velocity. This is an over simplification when high speeds are involved. However, increasingly correct models can be studied in a systematic way as above.

8.7.1 Kinetic Energy

Newton's second law is also the basis for the notion of **kinetic energy**. When a force is exerted on an object which causes the object to move, it follows that the force is doing work which manifests itself in a change of velocity of the object. How is the total work done on the object by the force related to the final velocity of the object? By Newton's second law, and letting $\mathbf{v}$ be the velocity,

$$\mathbf{F}(t) = m\mathbf{v}'(t).$$

Now in a small increment of time, $(t, t+dt)$, the work done on the object would be approximately equal to

$$dW = \mathbf{F}(t)\cdot\mathbf{v}(t)\,dt. \tag{8.13}$$

If no work has been done at time $t = 0$,then 8.13 implies

$$\frac{dW}{dt} = \mathbf{F}\cdot\mathbf{v},\ W(0) = 0.$$

Hence,

$$\frac{dW}{dt} = m\mathbf{v}'(t)\cdot\mathbf{v}(t) = \frac{m}{2}\frac{d}{dt}|\mathbf{v}(t)|^2.$$

Therefore, the total work done up to time t would be $W(t) = \frac{m}{2}|\mathbf{v}(t)|^2 - \frac{m}{2}|\mathbf{v}_0|^2$ where $|\mathbf{v}_0|$ denotes the initial speed of the object. This difference represents the change in the kinetic energy.

8.7.2 Impulse And Momentum

Work and energy involve a force acting on an object for some distance. Impulse involves a force which acts on an object for an interval of time.

Definition 8.7.5 *Let* $\mathbf{F}$ *be a force which acts on an object during the time interval,* $[a,b]$. *The* ***impulse*** *of this force is*

$$\int_a^b \mathbf{F}(t)\,dt.$$

This is defined as

$$\left(\int_a^b F_1(t)\,dt, \int_a^b F_2(t)\,dt, \int_a^b F_3(t)\,dt\right).$$

The ***linear momentum*** *of an object of mass* m *and velocity* $\mathbf{v}$ *is defined as*

$$\text{Linear momentum } = m\mathbf{v}.$$

The notion of impulse and momentum are related in the following theorem.

Theorem 8.7.6 *Let* $\mathbf{F}$ *be a force acting on an object of mass* m. *Then the impulse equals the change in momentum. More precisely,*

$$\int_a^b \mathbf{F}(t)\,dt = m\mathbf{v}(b) - m\mathbf{v}(a).$$

Proof: This is really just the fundamental theorem of calculus and Newton's second law applied to the components of $\mathbf{F}$.

$$\int_a^b \mathbf{F}(t)\,dt = \int_a^b m\frac{d\mathbf{v}}{dt}dt = m\mathbf{v}(b) - m\mathbf{v}(a) \tag{8.14}$$

Now suppose two point masses, A and B collide. Newton's third law says the force exerted by mass A on mass B is equal in magnitude but opposite in direction to the force exerted by mass B on mass A. Letting the collision take place in the time interval, $[a,b]$ and denoting the two masses by m_A and m_B and their velocities by $\mathbf{v}_A$ and $\mathbf{v}_B$ it follows that

$$m_A\mathbf{v}_A(b) - m_A\mathbf{v}_A(a) = \int_a^b (\text{Force of } B \text{ on } A)\,dt$$

and

$$\begin{aligned} m_B\mathbf{v}_B(b) - m_B\mathbf{v}_B(a) &= \int_a^b (\text{Force of } A \text{ on } B)\,dt \\ &= -\int_a^b (\text{Force of } B \text{ on } A)\,dt \\ &= -(m_A\mathbf{v}_A(b) - m_A\mathbf{v}_A(a)) \end{aligned}$$

and this shows

$$m_B\mathbf{v}_B(b) + m_A\mathbf{v}_A(b) = m_B\mathbf{v}_B(a) + m_A\mathbf{v}_A(a).$$

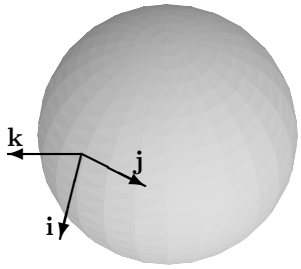
In other words, in a collision between two masses the total linear momentum before the collision equals the total linear momentum after the collision. This is known as the conservation of linear momentum.

8.8 Acceleration With Respect To Moving Coordinate Systems*

The idea is you have a coordinate system which is moving and this results in strange forces experienced relative to these moving coordinates systems. A good example is what we experience every day living on a rotating ball. Relative to our supposedly fixed coordinate system, we experience forces which account for many phenomena which are observed.

8.8.1 The Coriolis Acceleration

Imagine a point on the surface of the earth. Now consider unit vectors, one pointing South, one pointing East and one pointing directly away from the center of the earth.



Denote the first as $\mathbf{i}$, the second as $\mathbf{j}$ and the third as $\mathbf{k}$. If you are standing on the earth you will consider these vectors as fixed, but of course they are not. As the earth turns, they change direction and so each is in reality a function of t. Nevertheless, it is with respect to these apparently fixed vectors that you wish to understand acceleration, velocities, and displacements.

In general, let $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ be an orthonormal basis of vectors for each t, like the vectors described in the first paragraph. It is assumed these vectors are C^1 functions of t. Letting the positive x axis extend in the direction of $\mathbf{i}(t)$, the positive y axis extend in the direction of $\mathbf{j}(t)$, and the positive z axis extend in the direction of $\mathbf{k}(t)$, yields a moving coordinate system. By Theorem 8.4.2 on Page 163, there exists an angular velocity vector, $\mathbf{\Omega}(t)$ such that if $\mathbf{u}(t)$ is any vector which has constant components with respect to $\mathbf{i}(t), \mathbf{j}(t)$, and $\mathbf{k}(t)$, then

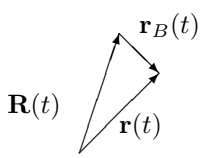
$$\mathbf{\Omega} \times \mathbf{u} = \mathbf{u}'. \tag{8.15}$$

Now let $\mathbf{R}(t)$ be a position vector of the moving coordinate system and let

$$\mathbf{r}(t) = \mathbf{R}(t) + \mathbf{r}_B(t)$$

where

$$\mathbf{r}_B(t) \equiv x(t)\,\mathbf{i}(t) + y(t)\,\mathbf{j}(t) + z(t)\,\mathbf{k}(t).$$



In the example of the earth, $\mathbf{R}(t)$ is the position vector of a point $\mathbf{p}(t)$ on the earth's surface and $\mathbf{r}_B(t)$ is the position vector of another point from $\mathbf{p}(t)$, thus regarding $\mathbf{p}(t)$ as the origin. $\mathbf{r}_B(t)$ is the position vector of a point as perceived by the observer on the earth with respect to the vectors he thinks of as fixed. Similarly, $\mathbf{v}_B(t)$ and $\mathbf{a}_B(t)$ will be the velocity and acceleration relative to $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$, and so $\mathbf{v}_B = x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}$ and $\mathbf{a}_B = x''\mathbf{i} + y''\mathbf{j} + z''\mathbf{k}$. Then

$$\mathbf{v} \equiv \mathbf{r}' = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}'.$$

By , 8.15, if $\mathbf{e} \in \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, $\mathbf{e}' = \boldsymbol{\Omega} \times \mathbf{e}$ because the components of these vectors with respect to $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are constant. Therefore,

$$\begin{aligned} x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}' &= x\boldsymbol{\Omega} \times \mathbf{i} + y\boldsymbol{\Omega} \times \mathbf{j} + z\boldsymbol{\Omega} \times \mathbf{k} \\ &= \boldsymbol{\Omega} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) \end{aligned}$$

and consequently,

$$\mathbf{v} = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + \boldsymbol{\Omega} \times \mathbf{r}_B = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + \boldsymbol{\Omega} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}).$$

Now consider the acceleration. Quantities which are relative to the moving coordinate system are distinguished by using the subscript, B.

$$\mathbf{a} = \mathbf{v}' = \mathbf{R}'' + x''\mathbf{i} + y''\mathbf{j} + z''\mathbf{k} + \overbrace{x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}'}^{\boldsymbol{\Omega} \times \mathbf{v}_B} + \boldsymbol{\Omega}' \times \mathbf{r}_B$$

$$+ \boldsymbol{\Omega} \times \left(\overbrace{x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}}^{\mathbf{v}_B} + \overbrace{x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}'}^{\boldsymbol{\Omega} \times \mathbf{r}_B(t)} \right)$$

$$= \mathbf{R}'' + \mathbf{a}_B + \boldsymbol{\Omega}' \times \mathbf{r}_B + 2\boldsymbol{\Omega} \times \mathbf{v}_B + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B).$$

The acceleration $\mathbf{a}_B$ is that perceived by an observer for whom the moving coordinate system is fixed. The term $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B)$ is called the centripetal acceleration. Solving for $\mathbf{a}_B$,

$$\mathbf{a}_B = \mathbf{a} - \mathbf{R}'' - \boldsymbol{\Omega}' \times \mathbf{r}_B - 2\boldsymbol{\Omega} \times \mathbf{v}_B - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B). \tag{8.16}$$

Here the term $-(\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B))$ is called the centrifugal acceleration, it being an acceleration felt by the observer relative to the moving coordinate system which he regards as fixed, and the term $-2\boldsymbol{\Omega} \times \mathbf{v}_B$ is called the Coriolis acceleration, an acceleration experienced by the observer as he moves relative to the moving coordinate system. The mass multiplied by the Coriolis acceleration defines the Coriolis force.

There is a ride found in some amusement parks in which the victims stand next to a circular wall covered with a carpet or some rough material. Then the whole circular room begins to revolve faster and faster. At some point, the bottom drops out and the victims are held in place by friction. The force they feel which keeps them stuck to the wall is called centrifugal force and it causes centrifugal acceleration. It is not necessary to move relative to coordinates fixed with the revolving wall in order to feel this force and it is pretty predictable. However, if the nauseated victim moves relative to the rotating wall, he will feel the effects of the Coriolis force and this force is really strange. The difference between these forces is that the Coriolis force is caused by movement relative to the moving coordinate system and the centrifugal force is not.

8.8.2 The Coriolis Acceleration On The Rotating Earth

Now consider the earth. Let $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$, be the usual basis vectors attached to the rotating earth. Thus $\mathbf{k}^*$ is fixed in space with $\mathbf{k}^*$ pointing in the direction of the north pole from the center of the earth while $\mathbf{i}^*$ and $\mathbf{j}^*$ point to fixed points on the surface of the earth. Thus $\mathbf{i}^*$ and $\mathbf{j}^*$ depend on t while $\mathbf{k}^*$ does not. Let $\mathbf{i}, \mathbf{j}, \mathbf{k}$ be the unit vectors described earlier with $\mathbf{i}$ pointing South, $\mathbf{j}$ pointing East, and $\mathbf{k}$ pointing away from the center of the earth at some point of the rotating earth's surface, $\mathbf{p}$. Letting $\mathbf{R}(t)$ be the position vector of the point $\mathbf{p}$, from the center of the earth, observe the coordinates of $\mathbf{R}(t)$ are constant with respect to $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$. Also, since the earth rotates from West to East and the speed of a point on the surface of the earth relative to an observer fixed in space is $\omega |\mathbf{R}| \sin\phi$ where ω is the angular speed of the earth about an axis through the poles, it follows from the geometric definition of the cross product that

$$\mathbf{R}' = \omega \mathbf{k}^* \times \mathbf{R}$$

Therefore, $\mathbf{\Omega} = \omega \mathbf{k}^*$ and so

$$\mathbf{R}'' = \overbrace{\mathbf{\Omega}' \times \mathbf{R}}^{=\mathbf{0}} + \mathbf{\Omega} \times \mathbf{R}' = \mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{R})$$

since $\mathbf{\Omega}$ does not depend on t. Formula 8.16 implies

$$\mathbf{a}_B = \mathbf{a} - \mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{R}) - 2\mathbf{\Omega} \times \mathbf{v}_B - \mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{r}_B). \tag{8.17}$$

In this formula, you can totally ignore the term $\mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{r}_B)$ because it is so small whenever you are considering motion near some point on the earth's surface. To see this, note $\omega \overbrace{(24)(3600)}^{\text{seconds in a day}} = 2\pi$, and so $\omega = 7.2722 \times 10^{-5}$ in radians per second. If you are using seconds to measure time and feet to measure distance, this term is therefore, no larger than

$$\left(7.2722 \times 10^{-5}\right)^2 |\mathbf{r}_B|.$$

Clearly this is not worth considering in the presence of the acceleration due to gravity which is approximately 32 feet per second squared near the surface of the earth.

If the acceleration $\mathbf{a}$, is due to gravity, then

$$\mathbf{a}_B = \mathbf{a} - \mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{R}) - 2\mathbf{\Omega} \times \mathbf{v}_B =$$

$$\overbrace{-\frac{GM(\mathbf{R}+\mathbf{r}_B)}{|\mathbf{R}+\mathbf{r}_B|^3} - \mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{R})}^{\equiv \mathbf{g}} - 2\mathbf{\Omega} \times \mathbf{v}_B \equiv \mathbf{g} - 2\mathbf{\Omega} \times \mathbf{v_B}.$$

Note that

$$\mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{R}) = (\mathbf{\Omega} \cdot \mathbf{R})\mathbf{\Omega} - |\mathbf{\Omega}|^2 \mathbf{R}$$

and so $\mathbf{g}$, the acceleration relative to the moving coordinate system on the earth is not directed exactly toward the center of the earth except at the poles and at the equator, although the components of acceleration which are in other directions are very small when compared with the acceleration due to the force of gravity and are often neglected. Therefore, if the only force acting on an object is due to gravity, the following formula describes the acceleration relative to a coordinate system moving with the earth's surface.

$$\mathbf{a}_B = \mathbf{g} - 2(\mathbf{\Omega} \times \mathbf{v}_B)$$

While the vector, $\mathbf{\Omega}$ is quite small, if the relative velocity, $\mathbf{v}_B$ is large, the Coriolis acceleration could be significant. This is described in terms of the vectors $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ next.

Letting (ρ, θ, ϕ) be the usual spherical coordinates of the point $\mathbf{p}(t)$ on the surface taken with respect to $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$ the usual way with ϕ the polar angle, it follows the $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$ coordinates of this point are

$$\begin{pmatrix} \rho\sin(\phi)\cos(\theta) \\ \rho\sin(\phi)\sin(\theta) \\ \rho\cos(\phi) \end{pmatrix}.$$

It follows,

$$\mathbf{i} = \cos(\phi)\cos(\theta)\,\mathbf{i}^* + \cos(\phi)\sin(\theta)\,\mathbf{j}^* - \sin(\phi)\,\mathbf{k}^*$$

$$\mathbf{j} = -\sin(\theta)\,\mathbf{i}^* + \cos(\theta)\,\mathbf{j}^* + 0\mathbf{k}^*$$

and

$$\mathbf{k} = \sin(\phi)\cos(\theta)\,\mathbf{i}^* + \sin(\phi)\sin(\theta)\,\mathbf{j}^* + \cos(\phi)\,\mathbf{k}^*.$$

It is necessary to obtain $\mathbf{k}^*$ in terms of the vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Thus the following equation needs to be solved for a, b, c to find $\mathbf{k}^* = a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$

$$\overbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}^{\mathbf{k}^*} = \begin{pmatrix} \cos(\phi)\cos(\theta) & -\sin(\theta) & \sin(\phi)\cos(\theta) \\ \cos(\phi)\sin(\theta) & \cos(\theta) & \sin(\phi)\sin(\theta) \\ -\sin(\phi) & 0 & \cos(\phi) \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{8.18}$$

The first column is $\mathbf{i}$, the second is $\mathbf{j}$ and the third is $\mathbf{k}$ in the above matrix. The solution is $a = -\sin(\phi), b = 0$, and $c = \cos(\phi)$.

Now the Coriolis acceleration on the earth equals

$$2(\mathbf{\Omega} \times \mathbf{v}_B) = 2\omega \left(\overbrace{-\sin(\phi)\,\mathbf{i}+0\mathbf{j}+\cos(\phi)\,\mathbf{k}}^{\mathbf{k}^*} \right) \times (x'\mathbf{i}+y'\mathbf{j}+z'\mathbf{k}).$$

This equals

$$2\omega\left[(-y'\cos\phi)\,\mathbf{i} + (x'\cos\phi + z'\sin\phi)\,\mathbf{j} - (y'\sin\phi)\,\mathbf{k}\right]. \tag{8.19}$$

Remember ϕ is fixed and pertains to the fixed point, $\mathbf{p}(t)$ on the earth's surface. Therefore, if the acceleration, $\mathbf{a}$ is due to gravity,

$$\mathbf{a}_B = \mathbf{g} - 2\omega\left[(-y'\cos\phi)\,\mathbf{i} + (x'\cos\phi + z'\sin\phi)\,\mathbf{j} - (y'\sin\phi)\,\mathbf{k}\right]$$

where $\mathbf{g} = -\frac{GM(\mathbf{R}+\mathbf{r}_B)}{|\mathbf{R}+\mathbf{r}_B|^3} - \mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})$ as explained above. The term $\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})$ is pretty small and so it will be neglected. However, the Coriolis force will not be neglected.

Example 8.8.1 *Suppose a rock is dropped from a tall building. Where will it strike?*

Assume $\mathbf{a} = -g\mathbf{k}$ and the $\mathbf{j}$ component of $\mathbf{a}_B$ is approximately

$$-2\omega(x'\cos\phi + z'\sin\phi).$$

The dominant term in this expression is clearly the second one because x' will be small. Also, the $\mathbf{i}$ and $\mathbf{k}$ contributions will be very small. Therefore, the following equation is descriptive of the situation.

$$\mathbf{a}_B = -g\mathbf{k} - 2z'\omega\sin\phi\mathbf{j}.$$

$z' = -gt$ approximately. Therefore, considering the $\mathbf{j}$ component, this is

$$2gt\omega \sin\phi.$$

Two integrations give $\left(\omega gt^3/3\right)\sin\phi$ for the $\mathbf{j}$ component of the relative displacement at time t.

This shows the rock does not fall directly towards the center of the earth as expected but slightly to the east.

Example 8.8.2 *In 1851 Foucault set a pendulum vibrating and observed the earth rotate out from under it. It was a very long pendulum with a heavy weight at the end so that it would vibrate for a long time without stopping*[3]*. This is what allowed him to observe the earth rotate out from under it. Clearly such a pendulum will take 24 hours for the plane of vibration to appear to make one complete revolution at the north pole. It is also reasonable to expect that no such observed rotation would take place on the equator. Is it possible to predict what will take place at various latitudes?*

Using 8.19, in 8.17,

$$\mathbf{a}_B = \mathbf{a} - \mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})$$
$$-2\omega\left[(-y'\cos\phi)\,\mathbf{i} + (x'\cos\phi + z'\sin\phi)\,\mathbf{j} - (y'\sin\phi)\,\mathbf{k}\right].$$

Neglecting the small term, $\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{R})$, this becomes

$$= -g\mathbf{k} + \mathbf{T}/m - 2\omega\left[(-y'\cos\phi)\,\mathbf{i} + (x'\cos\phi + z'\sin\phi)\,\mathbf{j} - (y'\sin\phi)\,\mathbf{k}\right]$$

where $\mathbf{T}$, the tension in the string of the pendulum, is directed towards the point at which the pendulum is supported, and m is the mass of the pendulum bob. The pendulum can be thought of as the position vector from $(0,0,l)$ to the surface of the sphere $x^2 + y^2 + (z-l)^2 = l^2$. Therefore,

$$\mathbf{T} = -T\frac{x}{l}\mathbf{i} - T\frac{y}{l}\mathbf{j} + T\frac{l-z}{l}\mathbf{k}$$

and consequently, the differential equations of relative motion are

$$x'' = -T\frac{x}{ml} + 2\omega y'\cos\phi$$

$$y'' = -T\frac{y}{ml} - 2\omega\left(x'\cos\phi + z'\sin\phi\right)$$

and

$$z'' = T\frac{l-z}{ml} - g + 2\omega y'\sin\phi.$$

If the vibrations of the pendulum are small so that for practical purposes, $z'' = z = 0$, the last equation may be solved for T to get

$$gm - 2\omega y'\sin(\phi)\,m = T.$$

Therefore, the first two equations become

$$x'' = -\left(gm - 2\omega m y'\sin\phi\right)\frac{x}{ml} + 2\omega y'\cos\phi$$

and

$$y'' = -\left(gm - 2\omega m y'\sin\phi\right)\frac{y}{ml} - 2\omega\left(x'\cos\phi + z'\sin\phi\right).$$

[3]There is such a pendulum in the Eyring building at BYU and to keep people from touching it, there is a little sign which says Warning! 1000 ohms.

All terms of the form xy' or $y'y$ can be neglected because it is assumed x and y remain small. Also, the pendulum is assumed to be long with a heavy weight so that x' and y' are also small. With these simplifying assumptions, the equations of motion become

$$x'' + g\frac{x}{l} = 2\omega y' \cos\phi$$

and

$$y'' + g\frac{y}{l} = -2\omega x' \cos\phi.$$

These equations are of the form

$$x'' + a^2 x = by',\ y'' + a^2 y = -bx' \tag{8.20}$$

where $a^2 = \frac{g}{l}$ and $b = 2\omega\cos\phi$. Then it is fairly tedious but routine to verify that for each constant, c,

$$x = c\sin\left(\frac{bt}{2}\right)\sin\left(\frac{\sqrt{b^2+4a^2}}{2}t\right),\ y = c\cos\left(\frac{bt}{2}\right)\sin\left(\frac{\sqrt{b^2+4a^2}}{2}t\right) \tag{8.21}$$

yields a solution to 8.20 along with the initial conditions,

$$x(0) = 0, y(0) = 0, x'(0) = 0, y'(0) = \frac{c\sqrt{b^2+4a^2}}{2}. \tag{8.22}$$

It is clear from experiments with the pendulum that the earth does indeed rotate out from under it causing the plane of vibration of the pendulum to appear to rotate. The purpose of this discussion is not to establish these self evident facts but to predict how long it takes for the plane of vibration to make one revolution. Therefore, there will be some instant in time at which the pendulum will be vibrating in a plane determined by **k** and **j**. (Recall **k** points away from the center of the earth and **j** points East.) At this instant in time, defined as $t = 0$, the conditions of 8.22 will hold for some value of c and so the solution to 8.20 having these initial conditions will be those of 8.21 by uniqueness of the initial value problem. Writing these solutions differently,

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = c\begin{pmatrix} \sin\left(\frac{bt}{2}\right) \\ \cos\left(\frac{bt}{2}\right) \end{pmatrix}\sin\left(\frac{\sqrt{b^2+4a^2}}{2}t\right)$$

This is very interesting! The vector, $c\begin{pmatrix} \sin\left(\frac{bt}{2}\right) \\ \cos\left(\frac{bt}{2}\right) \end{pmatrix}$ always has magnitude equal to $|c|$ but its direction changes very slowly because b is very small. The plane of vibration is determined by this vector and the vector **k**. The term $\sin\left(\frac{\sqrt{b^2+4a^2}}{2}t\right)$ changes relatively fast and takes values between -1 and 1. This is what describes the actual observed vibrations of the pendulum. Thus the plane of vibration will have made one complete revolution when $t = P$ for

$$\frac{bP}{2} \equiv 2\pi.$$

Therefore, the time it takes for the earth to turn out from under the pendulum is

$$P = \frac{4\pi}{2\omega\cos\phi} = \frac{2\pi}{\omega}\sec\phi.$$

Since ω is the angular speed of the rotating earth, it follows $\omega = \frac{2\pi}{24} = \frac{\pi}{12}$ in radians per hour. Therefore, the above formula implies

$$P = 24\sec\phi.$$

I think this is really amazing. You could actually determine latitude, not by taking readings with instruments using the North Star but by doing an experiment with a big pendulum. You would set it vibrating, observe P in hours, and then solve the above equation for ϕ. Also note the pendulum would not appear to change its plane of vibration at the equator because $\lim_{\phi\to\pi/2}\sec\phi=\infty$.

The Coriolis acceleration is also responsible for the phenomenon of the next example.

Example 8.8.3 *It is known that low pressure areas rotate counterclockwise as seen from above in the Northern hemisphere but clockwise in the Southern hemisphere. Why?*

Neglect accelerations other than the Coriolis acceleration and the following acceleration which comes from an assumption that the point $\mathbf{p}(t)$ is the location of the lowest pressure.

$$\mathbf{a}=-a(r_B)\,\mathbf{r}_B$$

where $r_B=r$ will denote the distance from the fixed point $\mathbf{p}(t)$ on the earth's surface which is also the lowest pressure point. Of course the situation could be more complicated but this will suffice to explain the above question. Then the acceleration observed by a person on the earth relative to the apparently fixed vectors, $\mathbf{i},\mathbf{k},\mathbf{j}$, is

$$\mathbf{a}_B=-a(r_B)\,(x\mathbf{i}+y\mathbf{j}+z\mathbf{k})-2\omega\left[-y'\cos(\phi)\,\mathbf{i}+\left(x'\cos(\phi)+z'\sin(\phi)\right)\mathbf{j}-\left(y'\sin(\phi)\,\mathbf{k}\right)\right]$$

Therefore, one obtains some differential equations from $\mathbf{a}_B=x''\mathbf{i}+y''\mathbf{j}+z''\mathbf{k}$ by matching the components. These are

$$\begin{aligned}x''+a(r_B)\,x &= 2\omega y'\cos\phi\\ y''+a(r_B)\,y &= -2\omega x'\cos\phi-2\omega z'\sin(\phi)\\ z''+a(r_B)\,z &= 2\omega y'\sin\phi\end{aligned}$$

Now remember, the vectors, $\mathbf{i},\mathbf{j},\mathbf{k}$ are fixed relative to the earth and so are constant vectors. Therefore, from the properties of the determinant and the above differential equations,

$$\left(\mathbf{r}_B'\times\mathbf{r}_B\right)'=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\ x'&y'&z'\\ x&y&z\end{vmatrix}'=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\ x''&y''&z''\\ x&y&z\end{vmatrix}$$

$$=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\ -a(r_B)\,x+2\omega y'\cos\phi & -a(r_B)\,y-2\omega x'\cos\phi-2\omega z'\sin(\phi) & -a(r_B)\,z+2\omega y'\sin\phi\\ x&y&z\end{vmatrix}$$

Then the $\mathbf{k}^{th}$ component of this cross product equals

$$\omega\cos(\phi)\left(y^2+x^2\right)'+2\omega xz'\sin(\phi).$$

The first term will be negative because it is assumed $\mathbf{p}(t)$ is the location of low pressure causing y^2+x^2 to be a decreasing function. If it is assumed there is not a substantial motion in the $\mathbf{k}$ direction, so that z is fairly constant and the last term can be neglected, then the $\mathbf{k}^{th}$ component of $\left(\mathbf{r}_B'\times\mathbf{r}_B\right)'$ is negative provided $\phi\in\left(0,\frac{\pi}{2}\right)$ and positive if $\phi\in\left(\frac{\pi}{2},\pi\right)$. Beginning with a point at rest, this implies $\mathbf{r}_B'\times\mathbf{r}_B=\mathbf{0}$ initially and then the above implies its $\mathbf{k}^{th}$ component is negative in the upper hemisphere when $\phi<\pi/2$ and positive in the lower hemisphere when $\phi>\pi/2$. Using the right hand and the geometric definition of the cross product, this shows clockwise rotation in the lower hemisphere and counter clockwise rotation in the upper hemisphere.

Note also that as ϕ gets close to $\pi/2$ near the equator, the above reasoning tends to break down because $\cos(\phi)$ becomes close to zero. Therefore, the motion towards the low pressure has to be more pronounced in comparison with the motion in the $\mathbf{k}$ direction in order to draw this conclusion.

8.9 Exercises

1. Show the solution to $\mathbf{v}' + r\mathbf{v} = \mathbf{c}$ with the initial condition, $\mathbf{v}(0) = \mathbf{v}_0$ is $\mathbf{v}(t) = \left(\mathbf{v}_0 - \frac{\mathbf{c}}{r}\right) e^{-rt} + (\mathbf{c}/r)$. If $\mathbf{v}$ is velocity and $r = k/m$ where k is a constant for air resistance and m is the mass, and $\mathbf{c} = \mathbf{f}/m$, argue from Newton's second law that this is the equation for finding the velocity, $\mathbf{v}$ of an object acted on by air resistance proportional to the velocity and a constant force, $\mathbf{f}$, possibly from gravity. Does there exist a terminal velocity? What is it? **Hint:** To find the solution to the equation, multiply both sides by e^{rt}. Verify that then $\frac{d}{dt}\left(e^{rt}\mathbf{v}\right) = \mathbf{c}e^{rt}$. Then integrating both sides, $e^{rt}\mathbf{v}(t) = \frac{1}{r}\mathbf{c}e^{rt} + \mathbf{C}$. Now you need to find $\mathbf{C}$ from using the initial condition which states $\mathbf{v}(0) = \mathbf{v}_0$.

2. Verify Formula 8.14 carefully by considering the components.

3. Suppose that the air resistance is proportional to the velocity but it is desired to find the constant of proportionality. Describe how you could find this constant.

4. Suppose an object having mass equal to 5 kilograms experiences a time dependent force, $\mathbf{F}(t) = e^{-t}\mathbf{i} + \cos(t)\,\mathbf{j} + t^2\mathbf{k}$ meters per sec^2. Suppose also that the object is at the point $(0,1,1)$ meters at time $t = 0$ and that its initial velocity at this time is $\mathbf{v} = \mathbf{i} + \mathbf{j} - \mathbf{k}$ meters per sec. Find the position of the object as a function of t.

5. Fill in the details for the derivation of kinetic energy. In particular verify that $m\mathbf{v}'(t) \cdot \mathbf{v}(t) = \frac{m}{2}\frac{d}{dt}|\mathbf{v}(t)|^2$. Also, why would $dW = \mathbf{F}(t) \cdot \mathbf{v}(t)\,dt$?

6. Suppose the force acting on an object, $\mathbf{F}$ is always perpendicular to the velocity of the object. Thus $\mathbf{F} \cdot \mathbf{v} = 0$. Show the Kinetic energy of the object is constant. Such forces are sometimes called forces of constraint because they do not contribute to the speed of the object, only its direction.

7. A cannon is fired at an angle, θ from ground level on a vast plain. The speed of the ball as it leaves the mouth of the cannon is known to be s meters per second. Neglecting air resistance, find a formula for how far the cannon ball goes before hitting the ground. Show the maximum range for the cannon ball is achieved when $\theta = \pi/4$.

8. Suppose in the context of Problem 7 that the cannon ball has mass 10 kilograms and it experiences a force of air resistance which is $.01\mathbf{v}$ Newtons where $\mathbf{v}$ is the velocity in meters per second. The acceleration of gravity is 9.8 meters per sec^2. Also suppose that the initial speed is 100 meters per second. Find a formula for the displacement, $\mathbf{r}(t)$ of the cannon ball. If the angle of elevation equals $\pi/4$, use a calculator or other means to estimate the time before the cannon ball hits the ground.

9. Show that Newton's first law can be obtained from the second law.

10. Show that if $\mathbf{v}'(t) = \mathbf{0}$, for all $t \in (a,b)$, then there exists a constant vector, $\mathbf{z}$ independent of t such that $\mathbf{v}(t) = \mathbf{z}$ for all t.

11. Suppose an object moves in three dimensional space in such a way that the only force acting on the object is directed toward a single fixed point in three dimensional space. Verify that the motion of the object takes place in a plane. **Hint:** Let $\mathbf{r}(t)$ denote the position vector of the object from the fixed point. Then the force acting on the object must be of the form $g(\mathbf{r}(t))\,\mathbf{r}(t)$ and by Newton's second law, this equals $m\mathbf{r}''(t)$. Therefore,

$$m\mathbf{r}'' \times \mathbf{r} = g(\mathbf{r})\,\mathbf{r} \times \mathbf{r} = \mathbf{0}.$$

Now argue that $\mathbf{r}'' \times \mathbf{r} = (\mathbf{r}' \times \mathbf{r})'$, showing that $(\mathbf{r}' \times \mathbf{r})$ must equal a constant vector, $\mathbf{z}$. Therefore, what can be said about $\mathbf{z}$ and $\mathbf{r}$?

12. Suppose the only forces acting on an object are the force of gravity, $-mg\mathbf{k}$ and a force, $\mathbf{F}$ which is perpendicular to the motion of the object. Thus $\mathbf{F} \cdot \mathbf{v} = \mathbf{0}$. Show the total energy of the object,

$$E \equiv \frac{1}{2} m |\mathbf{v}|^2 + mgz$$

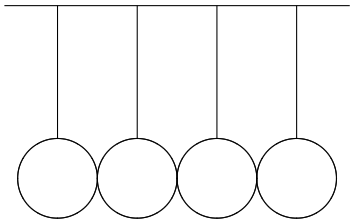
is constant. Here $\mathbf{v}$ is the velocity and the first term is the kinetic energy while the second is the potential energy. **Hint:** Use Newton's second law to show the time derivative of the above expression equals zero.

13. Using Problem 12, suppose an object slides down a frictionless inclined plane from a height of 100 feet. When it reaches the bottom, how fast will it be going? Assume it starts from rest.

14. The ballistic pendulum is an interesting device which is used to determine the speed of a bullet. It is a large massive block of wood hanging from a long string. A rifle is fired into the block of wood which then moves. The speed of the bullet can be determined from measuring how high the block of wood rises. Explain how this can be done and why. **Hint:** Let v be the speed of the bullet which has mass m and let the block of wood have mass M. By conservation of momentum $mv = (m + M) V$ where V is the speed of the block of wood immediately after the collision. Thus the energy is $\frac{1}{2} (m + M) V^2$ and this block of wood rises to a height of h. Now use Problem 12.

15. In the experiment of Problem 14, show the kinetic energy before the collision is greater than the kinetic energy after the collision. Thus linear momentum is conserved but energy is not. Such a collision is called inelastic.

16. There is a popular toy consisting of identical steel balls suspended from strings of equal length as illustrated in the following picture.



The ball at the right is lifted and allowed to swing. When it collides with the other balls, the ball on the left is observed to swing away from the others with the same speed the ball on the right had when it collided. Why does this happen? Why don't two or more of the stationary balls start to move, perhaps at a slower speed? This is an example of an elastic collision because energy is conserved. Of course this could change if you fixed things so the balls would stick to each other.

17. An illustration used in many beginning physics books is that of firing a rifle horizontally and dropping an identical bullet from the same height above the perfectly flat ground followed by an assertion that the two bullets will hit the ground at

exactly the same time. Is this true on the rotating earth assuming the experiment takes place over a large perfectly flat field so the curvature of the earth is not an issue? Explain. What other irregularities will occur? Recall the Coriolis force is $2\omega\left[\left(-y'\cos\phi\right)\mathbf{i}+\left(x'\cos\phi+z'\sin\phi\right)\mathbf{j}-\left(y'\sin\phi\right)\mathbf{k}\right]$ where $\mathbf{k}$ points away from the center of the earth, $\mathbf{j}$ points East, and $\mathbf{i}$ points South.

18. Suppose you have n masses, $m_1,\cdots,m_n$. Let the position vector of the i^{th} mass be $\mathbf{r}_i\left(t\right)$. The center of mass of these is defined to be

$$\mathbf{R}\left(t\right)\equiv\frac{\sum_{i=1}^{n}\mathbf{r}_i m_i}{\sum_{i=1}^{n}m_i}\equiv\frac{\sum_{i=1}^{n}\mathbf{r}_i\left(t\right)m_i}{M}.$$

Let $\mathbf{r}_{Bi}\left(t\right)=\mathbf{r}_i\left(t\right)-\mathbf{R}\left(t\right)$. Show that $\sum_{i=1}^{n}m_i\mathbf{r}_i\left(t\right)-\sum_i m_i\mathbf{R}\left(t\right)=\mathbf{0}$.

19. Suppose you have n masses, $m_1,\cdots,m_n$ which make up a moving rigid body. Let $\mathbf{R}\left(t\right)$ denote the position vector of the center of mass of these n masses. Find a formula for the total kinetic energy in terms of this position vector, the angular velocity vector, and the position vector of each mass from the center of mass. **Hint:** Use Problem 18.

8.10 Exercises With Answers

1. Show the solution to $\mathbf{v}'+r\mathbf{v}=\mathbf{c}$ with the initial condition, $\mathbf{v}\left(0\right)=\mathbf{v}_0$ is $\mathbf{v}\left(t\right)=\left(\mathbf{v}_0-\frac{\mathbf{c}}{r}\right)e^{-rt}+\left(\mathbf{c}/r\right)$. If $\mathbf{v}$ is velocity and $r=k/m$ where k is a constant for air resistance and m is the mass, and $\mathbf{c}=\mathbf{f}/m$, argue from Newton's second law that this is the equation for finding the velocity, $\mathbf{v}$ of an object acted on by air resistance proportional to the velocity and a constant force, $\mathbf{f}$, possibly from gravity. Does there exist a terminal velocity? What is it?

   Multiply both sides of the differential equation by e^{rt}. Then the left side becomes $\frac{d}{dt}\left(e^{rt}\mathbf{v}\right)=e^{rt}\mathbf{c}$. Now integrate both sides. This gives $e^{rt}\mathbf{v}\left(t\right)=\mathbf{C}+\frac{e^{rt}}{r}\mathbf{c}$. You finish the rest.

2. Suppose an object having mass equal to 5 kilograms experiences a time dependent force, $\mathbf{F}\left(t\right)=e^{-t}\mathbf{i}+\cos\left(t\right)\mathbf{j}+t^2\mathbf{k}$ meters per sec^2. Suppose also that the object is at the point $(0,1,1)$ meters at time $t=0$ and that its initial velocity at this time is $\mathbf{v}=\mathbf{i}+\mathbf{j}-\mathbf{k}$ meters per sec. Find the position of the object as a function of t.

   This is done by using Newton's law. Thus $5\frac{d^2\mathbf{r}}{dt^2}=e^{-t}\mathbf{i}+\cos\left(t\right)\mathbf{j}+t^2\mathbf{k}$ and so $5\frac{d\mathbf{r}}{dt}=-e^{-t}\mathbf{i}+\sin\left(t\right)\mathbf{j}+\left(t^3/3\right)\mathbf{k}+\mathbf{C}$. Find the constant, $\mathbf{C}$ by using the given initial velocity. Next do another integration obtaining another constant vector which will be determined by using the given initial position of the object.

3. Fill in the details for the derivation of kinetic energy. In particular verify that $m\mathbf{v}'\left(t\right)\cdot\mathbf{v}\left(t\right)=\frac{m}{2}\frac{d}{dt}\left|\mathbf{v}\left(t\right)\right|^2$. Also, why would $dW=\mathbf{F}\left(t\right)\cdot\mathbf{v}\left(t\right)dt$?

   Remember $\left|\mathbf{v}\right|^2=\mathbf{v}\cdot\mathbf{v}$. Now use the product rule.

4. Suppose the force acting on an object, $\mathbf{F}$ is always perpendicular to the velocity of the object. Thus $\mathbf{F}\cdot\mathbf{v}=0$. Show the Kinetic energy of the object is constant. Such forces are sometimes called forces of constraint because they do not contribute to the speed of the object, only its direction.

   $0=\mathbf{F}\cdot\mathbf{v}=m\mathbf{v}'\cdot\mathbf{v}$. Explain why this is $\frac{d}{dt}\left(m\frac{1}{2}\left|\mathbf{v}\right|^2\right)$, the derivative of the kinetic energy.

8.11 Line Integrals

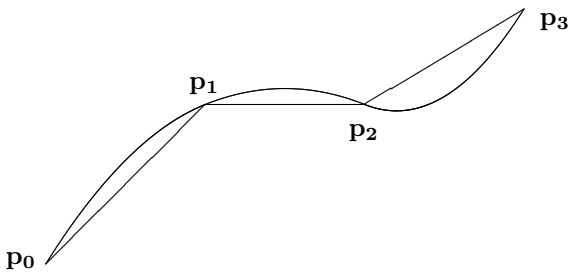
The concept of the integral can be extended to functions which are not defined on an interval of the real line but on some curve in $\mathbb{R}^n$. This is done by defining things in such a way that the more general concept reduces to the earlier notion. First it is necessary to consider what is meant by arc length.

8.11.1 Arc Length And Orientations

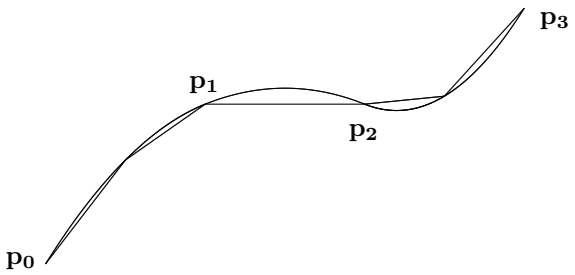
The application of the integral considered here is the concept of the **length of a curve**. C is a **smooth curve** in $\mathbb{R}^n$ if there exists an interval, $[a,b] \subseteq \mathbb{R}$ and functions $x_i : [a,b] \to \mathbb{R}$ such that the following conditions hold

1. x_i is continuous on $[a,b]$.
2. x_i' exists and is continuous and bounded on $[a,b]$, with $x_i'(a)$ defined as the derivative from the right,
$$\lim_{h \to 0+} \frac{x_i(a+h) - x_i(a)}{h},$$
and $x_i'(b)$ defined similarly as the derivative from the left.
3. For $\mathbf{p}(t) \equiv (x_1(t), \cdots, x_n(t))$, $t \to \mathbf{p}(t)$ is one to one on (a,b).
4. $|\mathbf{p}'(t)| \equiv \left(\sum_{i=1}^n |x_i'(t)|^2\right)^{1/2} \neq 0$ for all $t \in [a,b]$.
5. $C = \cup \{(x_1(t), \cdots, x_n(t)) : t \in [a,b]\}$.

The functions, $x_i(t)$, defined above are giving the coordinates of a point in $\mathbb{R}^n$ and the list of these functions is called a **parameterization** for the smooth curve. Note the natural direction of the interval also gives a direction for moving along the curve. Such a direction is called an orientation. The integral is used to define what is meant by the length of such a smooth curve. Consider such a smooth curve having parameterization $(x_1, \cdots, x_n)$. Forming a partition of $[a,b]$, $a = t_0 < \cdots < t_n = b$ and letting $\mathbf{p}_i = ($ $x_1(t_i)$, $\cdots$, $x_n(t_i)$ $)$, you could consider the polygon formed by lines from $\mathbf{p}_0$ to $\mathbf{p}_1$ and from $\mathbf{p}_1$ to $\mathbf{p}_2$ and from $\mathbf{p}_3$ to $\mathbf{p}_4$ etc. to be an approximation to the curve, C. The following picture illustrates what is meant by this.



Now consider what happens when the partition is refined by including more points. You can see from the following picture that the polygonal approximation would appear to be even better and that as more points are added in the partition, the sum of the lengths of the line segments seems to get close to something which deserves to be defined as the length of the curve, C.



Thus the length of the curve is approximated by

$$\sum_{k=1}^{n} |\mathbf{p}(t_k) - \mathbf{p}(t_{k-1})|.$$

Since the functions in the parameterization are differentiable, it is reasonable to expect this to be close to

$$\sum_{k=1}^{n} |\mathbf{p}'(t_{k-1})| (t_k - t_{k-1})$$

which is seen to be a Riemann sum for the integral

$$\int_a^b |\mathbf{p}'(t)| \, dt$$

and it is this integral which is defined as the length of the curve.

Would the same length be obtained if another parameterization were used? This is a very important question because the length of the curve should depend only on the curve itself and not on the method used to trace out the curve. The answer to this question is that the length of the curve does not depend on parameterization. The proof is somewhat technical so is given in the last section of this chapter.

Does the definition of length given above correspond to the usual definition of length in the case when the curve is a line segment? It is easy to see that it does so by considering two points in $\mathbb{R}^n$, $\mathbf{p}$ and $\mathbf{q}$. A parameterization for the line segment joining these two points is

$$f_i(t) \equiv tp_i + (1-t) q_i, \ t \in [0,1].$$

Using the definition of length of a smooth curve just given, the length according to this definition is

$$\int_0^1 \left(\sum_{i=1}^{n} (p_i - q_i)^2 \right)^{1/2} dt = |\mathbf{p} - \mathbf{q}|.$$

Thus this new definition which is valid for smooth curves which may not be straight line segments gives the usual length for straight line segments.

The proof that curve length is well defined for a smooth curve contains a result which deserves to be stated as a corollary. It is proved in Lemma 8.14.13 on Page 194 but the proof is mathematically fairly advanced so it is presented later.

Corollary 8.11.1 *Let C be a smooth curve and let $\mathbf{f} : [a,b] \to C$ and $\mathbf{g} : [c,d] \to C$ be two parameterizations satisfying 1 - 5. Then $\mathbf{g}^{-1} \circ \mathbf{f}$ is either strictly increasing or strictly decreasing.*

Definition 8.11.2 *If $\mathbf{g}^{-1} \circ \mathbf{f}$ is increasing, then $\mathbf{f}$ and $\mathbf{g}$ are said to be equivalent parameterizations and this is written as $\mathbf{f} \sim \mathbf{g}$. It is also said that the two parameterizations give the same orientation for the curve when $\mathbf{f} \sim \mathbf{g}$.*

When the parameterizations are equivalent, they preserve the direction, of motion along the curve and this also shows there are exactly two orientations of the curve since either $\mathbf{g}^{-1} \circ \mathbf{f}$ is increasing or it is decreasing. This is not hard to believe. In simple language, the message is that there are exactly two directions of motion along a curve. The difficulty is in proving this is actually the case.

Lemma 8.11.3 *The following hold for* $\sim$.

$$\mathbf{f} \sim \mathbf{f}, \tag{8.23}$$

$$\text{If } \mathbf{f} \sim \mathbf{g} \text{ then } \mathbf{g} \sim \mathbf{f}, \tag{8.24}$$

$$\text{If } \mathbf{f} \sim \mathbf{g} \text{ and } \mathbf{g} \sim \mathbf{h}, \text{ then } \mathbf{f} \sim \mathbf{h}. \tag{8.25}$$

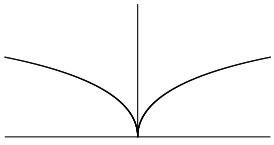
Proof: Formula 8.23 is obvious because $\mathbf{f}^{-1} \circ \mathbf{f}(t) = t$ so it is clearly an increasing function. If $\mathbf{f} \sim \mathbf{g}$ then $\mathbf{f}^{-1} \circ \mathbf{g}$ is increasing. Now $\mathbf{g}^{-1} \circ \mathbf{f}$ must also be increasing because it is the inverse of $\mathbf{f}^{-1} \circ \mathbf{g}$. This verifies 8.24. To see 8.25, $\mathbf{f}^{-1} \circ \mathbf{h} = \left(\mathbf{f}^{-1} \circ \mathbf{g}\right) \circ \left(\mathbf{g}^{-1} \circ \mathbf{h}\right)$ and so since both of these functions are increasing, it follows $\mathbf{f}^{-1} \circ \mathbf{h}$ is also increasing. This proves the lemma.

The symbol, $\sim$ is called an equivalence relation. If C is such a smooth curve just described, and if $\mathbf{f} : [a, b] \to C$ is a parameterization of C, consider $\mathbf{g}(t) \equiv \mathbf{f}((a + b) - t)$, also a parameterization of C. Now by Corollary 8.11.1, if $\mathbf{h}$ is a parameterization, then if $\mathbf{f}^{-1} \circ \mathbf{h}$ is not increasing, it must be the case that $\mathbf{g}^{-1} \circ \mathbf{h}$ is increasing. Consequently, either $\mathbf{h} \sim \mathbf{g}$ or $\mathbf{h} \sim \mathbf{f}$. These parameterizations, $\mathbf{h}$, which satisfy $\mathbf{h} \sim \mathbf{f}$ are called the equivalence class determined by $\mathbf{f}$ and those $\mathbf{h} \sim \mathbf{g}$ are called the equivalence class determined by $\mathbf{g}$. These two classes are called **orientations** of C. They give the direction of motion on C. You see that going from $\mathbf{f}$ to $\mathbf{g}$ corresponds to tracing out the curve in the opposite direction.

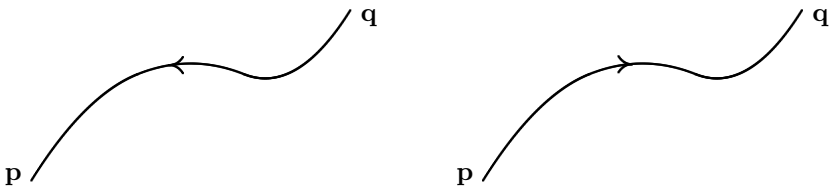
Sometimes people wonder why it is required, in the definition of a smooth curve that $\mathbf{p}'(t) \neq \mathbf{0}$. Imagine t is time and $\mathbf{p}(t)$ gives the location of a point in space. If $\mathbf{p}'(t)$ is allowed to equal zero, the point can stop and change directions abruptly, producing a pointy place in C. Here is an example.

Example 8.11.4 *Graph the curve* $\left(t^3, t^2\right)$ *for* $t \in [-1, 1]$.

In this case, $t = x^{1/3}$ and so $y = x^{2/3}$. Thus the graph of this curve looks like the picture below. Note the pointy place. Such a curve should not be considered smooth! If it were a banister and you were sliding down it, it would be clear at a certain point that the curve is not smooth. I think you may even get the point of this from the picture below.



So what is the thing to remember from all this? First, there are certain conditions which must be satisfied for a curve to be smooth. These are listed in 1 - 5. Next, if you have any curve, there are two directions you can move over this curve, each called an orientation. This is illustrated in the following picture.



Either you move from **p** to **q** or you move from **q** to **p**.

Definition 8.11.5 *A curve C is piecewise smooth if there exist points on this curve, $\mathbf{p}_0, \mathbf{p}_1, \cdots, \mathbf{p}_n$ such that, denoting $C_{\mathbf{p}_{k-1}\mathbf{p}_k}$ the part of the curve joining $\mathbf{p}_{k-1}$ and $\mathbf{p}_k$, it follows $C_{\mathbf{p}_{k-1}\mathbf{p}_k}$ is a smooth curve and $\cup_{k=1}^{n} C_{\mathbf{p}_{k-1}\mathbf{p}_k} = C$. In other words, it is piecewise smooth if it is made from a finite number of smooth curves linked together.*

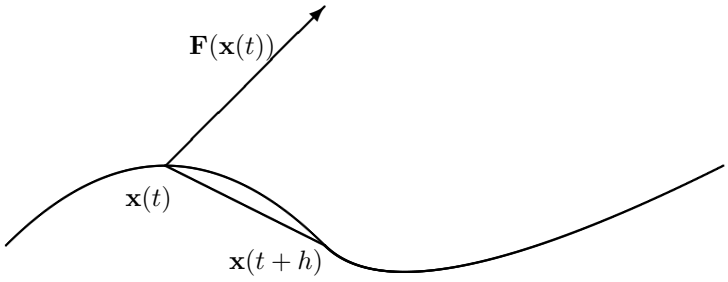
Note that Example 8.11.4 is an example of a piecewise smooth curve although it is not smooth.

8.11.2 Line Integrals And Work

Let C be a smooth curve contained in $\mathbb{R}^p$. A curve, C is an "**oriented curve**" if the only parameterizations considered are those which lie in exactly one of the two equivalence classes, each of which is called an "**orientation**". In simple language, orientation specifies a direction over which motion along the curve is to take place. Thus, it specifies the order in which the points of C are encountered. The pair of concepts consisting of the set of points making up the curve along with a direction of motion along the curve is called an **oriented curve**.

Definition 8.11.6 *Suppose $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^p$ is given for each $\mathbf{x} \in C$ where C is a smooth oriented curve and suppose $\mathbf{x} \to \mathbf{F}(\mathbf{x})$ is continuous. The mapping $\mathbf{x} \to \mathbf{F}(\mathbf{x})$ is called a **vector field**. In the case that $\mathbf{F}(\mathbf{x})$ is a force, it is called a **force field.***

Next the concept of work done by a force field, $\mathbf{F}$ on an object as it moves along the curve, C, in the direction determined by the given orientation of the curve will be defined. This is new. Earlier the work done by a force which acts on an object moving in a straight line was discussed but here the object moves over a curve. In order to define what is meant by the work, consider the following picture.



In this picture, the work done by $\mathbf{F}$ on an object which moves from the point $\mathbf{x}(t)$ to the point $\mathbf{x}(t+h)$ along the straight line shown would equal $\mathbf{F} \cdot (\mathbf{x}(t+h) - \mathbf{x}(t))$. It is reasonable to assume this would be a good approximation to the work done in moving along the curve joining $\mathbf{x}(t)$ and $\mathbf{x}(t+h)$ provided h is small enough. Also, provided h is small,

$$\mathbf{x}(t+h) - \mathbf{x}(t) \approx \mathbf{x}'(t)\, h$$

where the wriggly equal sign indicates the two quantities are close. In the notation of Leibniz, one writes dt for h and

$$dW = \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\, dt$$

or in other words,

$$\frac{dW}{dt} = \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t).$$

Defining the total work done by the force at $t = 0$, corresponding to the first endpoint of the curve, to equal zero, the work would satisfy the following initial value problem.

$$\frac{dW}{dt} = \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t),\ W(a) = 0.$$

This motivates the following definition of work.

Definition 8.11.7 *Let $\mathbf{F}(\mathbf{x})$ be given above. Then the work done by this force field on an object moving over the curve C in the direction determined by the specified orientation is defined as*

$$\int_C \mathbf{F} \cdot d\mathbf{R} \equiv \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t)\ dt$$

where the function, $\mathbf{x}$ is one of the allowed parameterizations of C in the given orientation of C. In other words, there is an interval, $[a,b]$ and as t goes from a to b, $\mathbf{x}(t)$ moves in the direction determined from the given orientation of the curve.

Theorem 8.11.8 *The symbol, $\int_C \mathbf{F} \cdot d\mathbf{R}$, is well defined in the sense that every parameterization in the given orientation of C gives the same value for $\int_C \mathbf{F} \cdot d\mathbf{R}$.*

Proof: Suppose $\mathbf{g} : [c,d] \to C$ is another allowed parameterization. Thus $\mathbf{g}^{-1} \circ \mathbf{f}$ is an increasing function, ϕ. Then since ϕ is increasing,

$$\int_c^d \mathbf{F}(\mathbf{g}(s)) \cdot \mathbf{g}'(s)\ ds = \int_a^b \mathbf{F}(\mathbf{g}(\phi(t))) \cdot \mathbf{g}'(\phi(t))\, \phi'(t)\ dt$$

$$= \int_a^b \mathbf{F}(\mathbf{f}(t)) \cdot \frac{d}{dt}\left(\mathbf{g}\left(\mathbf{g}^{-1} \circ \mathbf{f}(t)\right)\right)\ dt = \int_a^b \mathbf{F}(\mathbf{f}(t)) \cdot \mathbf{f}'(t)\ dt.$$

This proves the theorem.

Regardless the physical interpretation of $\mathbf{F}$, this is called the **line integral.** When $\mathbf{F}$ is interpreted as a force, the line integral measures the extent to which the motion over the curve in the indicated direction is aided by the force. If the net effect of the force on the object is to impede rather than to aid the motion, this will show up as the work being negative.

Does the concept of work as defined here coincide with the earlier concept of work when the object moves over a straight line when acted on by a constant force?

Let $\mathbf{p}$ and $\mathbf{q}$ be two points in $\mathbb{R}^n$ and suppose $\mathbf{F}$ is a constant force acting on an object which moves from $\mathbf{p}$ to $\mathbf{q}$ along the straight line joining these points. Then the work done is $\mathbf{F} \cdot (\mathbf{q} - \mathbf{p})$. Is the same thing obtained from the above definition? Let $\mathbf{x}(t) \equiv \mathbf{p} + t(\mathbf{q} - \mathbf{p}), t \in [0,1]$ be a parameterization for this oriented curve, the straight line in the direction from $\mathbf{p}$ to $\mathbf{q}$. Then $\mathbf{x}'(t) = \mathbf{q} - \mathbf{p}$ and $\mathbf{F}(\mathbf{x}(t)) = \mathbf{F}$. Therefore, the above definition yields

$$\int_0^1 \mathbf{F} \cdot (\mathbf{q} - \mathbf{p})\ dt = \mathbf{F} \cdot (\mathbf{q} - \mathbf{p}).$$

Therefore, the new definition adds to but does not contradict the old one.

Example 8.11.9 *Suppose for $t \in [0,\pi]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + \cos(2t)\mathbf{j} + \sin(2t)\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$, $\mathbf{F}(x,y,z) \equiv 2xy\mathbf{i} + x^2\mathbf{j} + \mathbf{k}$. Find*

$$\int_C \mathbf{F}\cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t.

To find this line integral use the above definition and write

$$\int_C \mathbf{F}\cdot d\mathbf{R} = \int_0^\pi \left(2t\left(\cos(2t)\right), t^2, 1\right)\cdot \left(1, -2\sin(2t), 2\cos(2t)\right)\, dt$$

In evaluating this replace the x in the formula for $\mathbf{F}$ with t, the y in the formula for $\mathbf{F}$ with $\cos(2t)$ and the z in the formula for $\mathbf{F}$ with $\sin(2t)$ because these are the values of these variables which correspond to the value of t. Taking the dot product, this equals the following integral.

$$\int_0^\pi \left(2t\cos 2t - 2(\sin 2t)\, t^2 + 2\cos 2t\right)\, dt = \pi^2$$

Example 8.11.10 *Let C denote the oriented curve obtained by $\mathbf{r}(t) = \left(t, \sin t, t^3\right)$ where the orientation is determined by increasing t for $t \in [0,2]$. Also let $\mathbf{F} = (x, y, xz + z)$. Find $\int_C \mathbf{F}\cdot d\mathbf{R}$.*

You use the definition.

$$\begin{aligned}\int_C \mathbf{F}\cdot d\mathbf{R} &= \int_0^2 \left(t, \sin(t), (t+1)t^3\right)\cdot\left(1, \cos(t), 3t^2\right) dt \\ &= \int_0^2 \left(t + \sin(t)\cos(t) + 3(t+1)t^5\right) dt \\ &= \frac{1251}{14} - \frac{1}{2}\cos^2(2).\end{aligned}$$

Suppose you have a curve specified by $\mathbf{r}(s) = (x(s), y(s), z(s))$ and it has the property that $|\mathbf{r}'(s)| = 1$ for all $s \in [0,b]$. Then the length of this curve for s between 0 and s_1 is

$$\int_0^{s_1} |\mathbf{r}'(s)|\, ds = \int_0^{s_1} 1 ds = s_1.$$

This parameter is therefore called arc length because the length of the curve up to s equals s. Now you can always change the parameter to be arc length.

Proposition 8.11.11 *Suppose C is an oriented smooth curve parameterized by $\mathbf{r}(t)$ for $t \in [a,b]$. Then letting l denote the total length of C, there exists $\mathbf{R}(s)$, $s \in [0,l]$ another parameterization for this curve which preserves the orientation and such that $|\mathbf{R}'(s)| = 1$ so that s is arc length.*

Prove: Let $\phi(t) \equiv \int_a^t |\mathbf{r}'(\tau)|\, d\tau \equiv s$. Then s is an increasing function of t because

$$\frac{ds}{dt} = \phi'(t) = |\mathbf{r}'(t)| > 0.$$

Now define $\mathbf{R}(s) \equiv \mathbf{r}\left(\phi^{-1}(s)\right)$. Then

$$\begin{aligned}\mathbf{R}'(s) &= \mathbf{r}'\left(\phi^{-1}(s)\right)\left(\phi^{-1}\right)'(s) \\ &= \frac{\mathbf{r}'\left(\phi^{-1}(s)\right)}{\left|\mathbf{r}'\left(\phi^{-1}(s)\right)\right|}\end{aligned}$$

and so $|\mathbf{R}'(s)| = 1$ as claimed. $\mathbf{R}(l) = \mathbf{r}\left(\phi^{-1}(l)\right) = \mathbf{r}\left(\phi^{-1}\left(\int_a^b |\mathbf{r}'(\tau)|\, d\tau\right)\right) = \mathbf{r}(b)$ and $\mathbf{R}(0) = \mathbf{r}\left(\phi^{-1}(0)\right) = \mathbf{r}(a)$ and $\mathbf{R}$ delivers the same set of points in the same order as $\mathbf{r}$ because $\frac{ds}{dt} > 0$.

The arc length parameter is just like any other parameter in so far as considerations of line integrals are concerned because it was shown above that line integrals are independent of parameterization. However, when things are defined in terms of the arc length parameterization, it is clear they depend only on geometric properties of the curve itself and for this reason, the arc length parameterization is important in differential geometry.

8.11.3 Another Notation For Line Integrals

Definition 8.11.12 *Let $\mathbf{F}(x,y,z) = (P(x,y,z), Q(x,y,z), R(x,y,z))$ and let C be an oriented curve. Then another way to write $\int_C \mathbf{F}\cdot d\mathbf{R}$ is*

$$\int_C P dx + Q dy + R dz$$

This last is referred to as the integral of a **differential form**, $Pdx + Qdy + Rdz$. The study of differential forms is important. Formally, $d\mathbf{R} = (dx, dy, dz)$ and so the integrand in the above is formally $\mathbf{F}\cdot d\mathbf{R}$. Other occurances of this notation are handled similarly in 2 or higher dimensions.

8.12 Exercises

1. Suppose for $t \in [0, 2\pi]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + \cos(2t)\,\mathbf{j} + \sin(2t)\,\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$,

   $$\mathbf{F}(x,y,z) \equiv 2xy\mathbf{i} + \left(x^2 + 2zy\right)\mathbf{j} + y^2\mathbf{k}.$$

   Find the work,

   $$\int_C \mathbf{F}\cdot d\mathbf{R}$$

   where C is the curve traced out by this object which has the orientation determined by the direction of increasing t.

2. Here is a vector field, $\left(y, x + z^2, 2yz\right)$ and here is the parameterization of a curve, C. $\mathbf{R}(t) = (\cos 2t, 2\sin 2t, t)$ where t goes from 0 to $\pi/4$. Find $\int_C \mathbf{F}\cdot d\mathbf{R}$.

3. If f and g are both increasing functions, show $f \circ g$ is an increasing function also. Assume anything you like about the domains of the functions.

4. Suppose for $t \in [0, 3]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + t\mathbf{j} + t\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$, $\mathbf{F}(x,y,z) \equiv yz\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}$. Find

   $$\int_C \mathbf{F}\cdot d\mathbf{R}$$

   where C is the curve traced out by this object which has the orientation determined by the direction of increasing t. Repeat the problem for $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t\mathbf{k}$.

5. Suppose for $t \in [0,1]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + t\mathbf{j} + t\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$, $\mathbf{F}(x,y,z) \equiv z\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}$. Find

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t. Repeat the problem for $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t\mathbf{k}$.

6. Let $\mathbf{F}(x,y,z)$ be a given force field and suppose it acts on an object having mass, m on a curve with parameterization, $(x(t), y(t), z(t))$ for $t \in [a,b]$. Show directly that the work done equals the difference in the kinetic energy. **Hint:**

$$\int_a^b \mathbf{F}(x(t), y(t), z(t)) \cdot (x'(t), y'(t), z'(t))\, dt =$$

$$\int_a^b m(x''(t), y''(t), z''(t)) \cdot (x'(t), y'(t), z'(t))\, dt,$$

etc.

8.13 Exercises With Answers

1. Suppose for $t \in [0, 2\pi]$ the position of an object is given by $\mathbf{r}(t) = 2t\mathbf{i} + \cos(t)\,\mathbf{j} + \sin(t)\,\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$,

$$\mathbf{F}(x,y,z) \equiv 2xy\mathbf{i} + \left(x^2 + 2zy\right)\mathbf{j} + y^2\mathbf{k}.$$

Find the work,

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t.

You might think of $d\mathbf{R} = \mathbf{r}'(t)\,dt$ to help remember what to do. Then from the definition,

$$\int_C \mathbf{F} \cdot d\mathbf{R} =$$

$$\int_0^{2\pi} \left(2(2t)(\sin t), 4t^2 + 2\sin(t)\cos(t), \sin^2(t)\right) \cdot (2, -\sin(t), \cos(t))\, dt$$

$$= \int_0^{2\pi} \left(8t\sin t - \left(2\sin t\cos t + 4t^2\right)\sin t + \sin^2 t\cos t\right) dt = 16\pi^2 - 16\pi$$

2. Here is a vector field, $\left(y, x^2 + z, 2yz\right)$ and here is the parameterization of a curve, C. $\mathbf{R}(t) = (\cos 2t, 2\sin 2t, t)$ where t goes from 0 to $\pi/4$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$.

$$d\mathbf{R} = (-2\sin(2t), 4\cos(2t), 1)\, dt.$$

Then by the definition,

$$\int_C \mathbf{F} \cdot d\mathbf{R} =$$

$$\int_0^{\pi/4} \left(2\sin(2t), \cos^2(2t) + t, 4t\sin(2t)\right) \cdot (-2\sin(2t), 4\cos(2t), 1)\, dt$$

$$= \int_0^{\pi/4} \left(-4\sin^2 2t + 4\left(\cos^2 2t + t\right)\cos 2t + 4t\sin 2t\right) dt = \frac{4}{3}$$

3. Suppose for $t \in [0,1]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + t\mathbf{j} + t\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$,

$$\mathbf{F}(x,y,z) \equiv yz\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}.$$

Find

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t. Repeat the problem for $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t\mathbf{k}$.

You should get the same answer in this case. This is because the vector field happens to be conservative. (More on this later.)

8.14 Independence Of Parameterization*

Recall that if $\mathbf{p}(t) : t \in [a,b]$ was a parameterization of a smooth curve, C, the length of C is defined as

$$\int_a^b |\mathbf{p}'(t)|\, dt$$

If some other parameterization were used to trace out C, would the same answer be obtained? To answer this question in a satisfactory manner requires some hard calculus.

8.14.1 Hard Calculus

Definition 8.14.1 *A sequence $\{a_n\}_{n=1}^{\infty}$ converges to a,*

$$\lim_{n\to\infty} a_n = a \text{ or } a_n \to a$$

if and only if for every $\varepsilon > 0$ there exists n_ε such that whenever $n \geq n_\varepsilon$,

$$|a_n - a| < \varepsilon.$$

In words the definition says that given any measure of closeness, ε, the terms of the sequence are eventually all this close to a. Note the similarity with the concept of limit. Here, the word "eventually" refers to n being sufficiently large. Earlier, it referred to y being sufficiently close to x on one side or another or else x being sufficiently large in either the positive or negative directions. The limit of a sequence, if it exists, is unique.

Theorem 8.14.2 *If $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} a_n = a_1$ then $a_1 = a$.*

Proof: Suppose $a_1 \neq a$. Then let $0 < \varepsilon < |a_1 - a|/2$ in the definition of the limit. It follows there exists n_ε such that if $n \geq n_\varepsilon$, then $|a_n - a| < \varepsilon$ and $|a_n - a_1| < \varepsilon$. Therefore, for such n,

$$\begin{aligned} |a_1 - a| &\leq |a_1 - a_n| + |a_n - a| \\ &< \varepsilon + \varepsilon < |a_1 - a|/2 + |a_1 - a|/2 = |a_1 - a|, \end{aligned}$$

a contradiction.

Definition 8.14.3 *Let $\{a_n\}$ be a sequence and let $n_1 < n_2 < n_3, \cdots$ be any strictly increasing list of integers such that n_1 is at least as large as the first index used to define the sequence $\{a_n\}$. Then if $b_k \equiv a_{n_k}$, $\{b_k\}$ is called a subsequence of $\{a_n\}$.*

Theorem 8.14.4 *Let $\{x_n\}$ be a sequence with $\lim_{n\to\infty} x_n = x$ and let $\{x_{n_k}\}$ be a subsequence. Then $\lim_{k\to\infty} x_{n_k} = x$.*

Proof: Let $\varepsilon > 0$ be given. Then there exists n_ε such that if $n > n_\varepsilon$, then $|x_n - x| < \varepsilon$. Suppose $k > n_\varepsilon$. Then $n_k \geq k > n_\varepsilon$ and so

$$|x_{n_k} - x| < \varepsilon$$

showing $\lim_{k\to\infty} x_{n_k} = x$ as claimed.

There is a very useful way of thinking of continuity in terms of limits of sequences found in the following theorem. In words, it says a function is continuous if it takes convergent sequences to convergent sequences whenever possible.

Theorem 8.14.5 *A function $f : D(f) \to \mathbb{R}$ is continuous at $x \in D(f)$ if and only if, whenever $x_n \to x$ with $x_n \in D(f)$, it follows $f(x_n) \to f(x)$.*

Proof: Suppose first that f is continuous at x and let $x_n \to x$. Let $\varepsilon > 0$ be given. By continuity, there exists $\delta > 0$ such that if $|y - x| < \delta$, then $|f(x) - f(y)| < \varepsilon$. However, there exists n_δ such that if $n \geq n_\delta$, then $|x_n - x| < \delta$ and so for all n this large,

$$|f(x) - f(x_n)| < \varepsilon$$

which shows $f(x_n) \to f(x)$.

Now suppose the condition about taking convergent sequences to convergent sequences holds at x. Suppose f fails to be continuous at x. Then there exists $\varepsilon > 0$ and $x_n \in D(f)$ such that $|x - x_n| < \frac{1}{n}$, yet

$$|f(x) - f(x_n)| \geq \varepsilon.$$

But this is clearly a contradiction because, although $x_n \to x$, $f(x_n)$ fails to converge to $f(x)$. It follows f must be continuous after all. This proves the theorem.

Definition 8.14.6 *A set, $K \subseteq \mathbb{R}$ is sequentially compact if whenever $\{a_n\} \subseteq K$ is a sequence, there exists a subsequence, $\{a_{n_k}\}$ such that this subsequence converges to a point of K.*

The following theorem is part of a major advanced calculus theorem known as the Heine Borel theorem.

Theorem 8.14.7 *Every closed interval, $[a, b]$ is sequentially compact.*

Proof: Let $\{x_n\} \subseteq [a,b] \equiv I_0$. Consider the two intervals $\left[a, \frac{a+b}{2}\right]$ and $\left[\frac{a+b}{2}, b\right]$ each of which has length $(b-a)/2$. At least one of these intervals contains x_n for infinitely many values of n. Call this interval I_1. Now do for I_1 what was done for I_0. Split it in half and let I_2 be the interval which contains x_n for infinitely many values of n. Continue this way obtaining a sequence of nested intervals $I_0 \supseteq I_1 \supseteq I_2 \supseteq I_3 \cdots$ where the length of I_n is $(b-a)/2^n$. Now pick n_1 such that $x_{n_1} \in I_1$, n_2 such that $n_2 > n_1$ and $x_{n_2} \in I_2, n_3$ such that $n_3 > n_2$ and $x_{n_3} \in I_3$, etc. (This can be done because in each case the intervals contained x_n for infinitely many values of n.) By the nested interval lemma there exists a point, c contained in all these intervals. Furthermore,

$$|x_{n_k} - c| < (b-a)\, 2^{-k}$$

and so $\lim_{k\to\infty} x_{n_k} = c \in [a,b]$. This proves the theorem.

Lemma 8.14.8 *Let $\phi : [a,b] \to \mathbb{R}$ be a continuous function and suppose ϕ is $1-1$ on (a,b). Then ϕ is either strictly increasing or strictly decreasing on $[a,b]$. Furthermore, ϕ^{-1} is continuous.*

Proof: First it is shown that ϕ is either strictly increasing or strictly decreasing on (a,b).

If ϕ is not strictly decreasing on (a,b), then there exists $x_1 < y_1$, $x_1, y_1 \in (a,b)$ such that

$$(\phi(y_1) - \phi(x_1))(y_1 - x_1) > 0.$$

If for some other pair of points, $x_2 < y_2$ with $x_2, y_2 \in (a,b)$, the above inequality does not hold, then since ϕ is $1-1$,

$$(\phi(y_2) - \phi(x_2))(y_2 - x_2) < 0.$$

Let $x_t \equiv t x_1 + (1-t)\, x_2$ and $y_t \equiv t y_1 + (1-t)\, y_2$. Then $x_t < y_t$ for all $t \in [0,1]$ because

$$t x_1 \le t y_1 \text{ and } (1-t)\, x_2 \le (1-t)\, y_2$$

with strict inequality holding for at least one of these inequalities since not both t and $(1-t)$ can equal zero. Now define

$$h(t) \equiv (\phi(y_t) - \phi(x_t))(y_t - x_t).$$

Since h is continuous and $h(0) < 0$, while $h(1) > 0$, there exists $t \in (0,1)$ such that $h(t) = 0$. Therefore, both x_t and y_t are points of (a,b) and $\phi(y_t) - \phi(x_t) = 0$ contradicting the assumption that ϕ is one to one. It follows ϕ is either strictly increasing or strictly decreasing on (a,b).

This property of being either strictly increasing or strictly decreasing on (a,b) carries over to $[a,b]$ by the continuity of ϕ. Suppose ϕ is strictly increasing on (a,b), a similar argument holding for ϕ strictly decreasing on (a,b). If $x > a$, then pick $y \in (a,x)$ and from the above, $\phi(y) < \phi(x)$. Now by continuity of ϕ at a,

$$\phi(a) = \lim_{x\to a+} \phi(z) \le \phi(y) < \phi(x).$$

Therefore, $\phi(a) < \phi(x)$ whenever $x \in (a,b)$. Similarly $\phi(b) > \phi(x)$ for all $x \in (a,b)$.

It only remains to verify ϕ^{-1} is continuous. Suppose then that $s_n \to s$ where s_n and s are points of $\phi([a,b])$. It is desired to verify that $\phi^{-1}(s_n) \to \phi^{-1}(s)$. If this does not happen, there exists $\varepsilon > 0$ and a subsequence, still denoted by s_n such that $\left|\phi^{-1}(s_n) - \phi^{-1}(s)\right| \ge \varepsilon$. Using the sequential compactness of $[a,b]$ (Theorem 7.7.18 on Page 150) there exists a further subsequence, still denoted by n, such that $\phi^{-1}(s_n) \to t_1 \in [a,b], t_1 \ne \phi^{-1}(s)$. Then by continuity of ϕ, it follows $s_n \to \phi(t_1)$ and so $s = \phi(t_1)$. Therefore, $t_1 = \phi^{-1}(s)$ after all. This proves the lemma.

Corollary 8.14.9 *Let $f:(a,b)\to\mathbb{R}$ be one to one and continuous. Then $f(a,b)$ is an open interval, (c,d) and $f^{-1}:(c,d)\to(a,b)$ is continuous.*

Proof: Since f is either strictly increasing or strictly decreasing, it follows that $f(a,b)$ is an open interval, (c,d). Assume f is decreasing. Now let $x\in(a,b)$. Why is f^{-1} is continuous at $f(x)$? Since f is decreasing, if $f(x)<f(y)$, then $y\equiv f^{-1}(f(y))<x\equiv f^{-1}(f(x))$ and so f^{-1} is also decreasing. Let $\varepsilon>0$ be given. Let $\varepsilon>\eta>0$ and $(x-\eta,x+\eta)\subseteq(a,b)$. Then $f(x)\in(f(x+\eta),f(x-\eta))$. Let

$$\delta=\min\left(f(x)-f(x+\eta),f(x-\eta)-f(x)\right).$$

Then if

$$|f(z)-f(x)|<\delta,$$

it follows

$$z\equiv f^{-1}(f(z))\in(x-\eta,x+\eta)\subseteq(x-\varepsilon,x+\varepsilon)$$

so

$$\left|f^{-1}(f(z))-x\right|=\left|f^{-1}(f(z))-f^{-1}(f(x))\right|<\varepsilon.$$

This proves the theorem in the case where f is strictly decreasing. The case where f is increasing is similar.

Theorem 8.14.10 *Let $f:[a,b]\to\mathbb{R}$ be continuous and one to one. Suppose $f'(x_1)$ exists for some $x_1\in[a,b]$ and $f'(x_1)\neq0$. Then $\left(f^{-1}\right)'(f(x_1))$ exists and is given by the formula, $\left(f^{-1}\right)'(f(x_1))=\frac{1}{f'(x_1)}$.*

Proof: By Lemma 8.14.8 f is either strictly increasing or strictly decreasing and f^{-1} is continuous on $[a,b]$. Therefore there exists $\eta>0$ such that if $0<|f(x_1)-f(x)|<\eta$, then

$$0<|x_1-x|=\left|f^{-1}(f(x_1))-f^{-1}(f(x))\right|<\delta$$

where δ is small enough that for $0<|x_1-x|<\delta$,

$$\left|\frac{x-x_1}{f(x)-f(x_1)}-\frac{1}{f'(x_1)}\right|<\varepsilon.$$

It follows that if $0<|f(x_1)-f(x)|<\eta$,

$$\left|\frac{f^{-1}(f(x))-f^{-1}(f(x_1))}{f(x)-f(x_1)}-\frac{1}{f'(x_1)}\right|=\left|\frac{x-x_1}{f(x)-f(x_1)}-\frac{1}{f'(x_1)}\right|<\varepsilon$$

Therefore, since $\varepsilon>0$ is arbitrary,

$$\lim_{y\to f(x_1)}\frac{f^{-1}(y)-f^{-1}(f(x_1))}{y-f(x_1)}=\frac{1}{f'(x_1)}$$

and this proves the theorem.

The following obvious corollary comes from the above by not bothering with end points.

Corollary 8.14.11 *Let $f:(a,b)\to\mathbb{R}$ be continuous and one to one. Suppose $f'(x_1)$ exists for some $x_1\in(a,b)$ and $f'(x_1)\neq0$. Then $\left(f^{-1}\right)'(f(x_1))$ exists and is given by the formula, $\left(f^{-1}\right)'(f(x_1))=\frac{1}{f'(x_1)}$.*

This is one of those theorems which is very easy to remember if you neglect the difficult questions and simply focus on formal manipulations. Consider the following.

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

Now use the chain rule on both sides to write

$$\left(f^{-1}\right)'(f(x))\, f'(x) = 1,$$

and then divide both sides by $f'(x)$ to obtain

$$\left(f^{-1}\right)'(f(x)) = \frac{1}{f'(x)}.$$

Of course this gives the conclusion of the above theorem rather effortlessly and it is formal manipulations like this which aid many of us in remembering formulas such as the one given in the theorem.

8.14.2 Independence Of Parameterization

Theorem 8.14.12 *Let $\phi : [a,b] \to [c,d]$ be one to one and suppose ϕ' exists and is continuous on $[a,b]$. Then if f is a continuous function defined on $[a,b]$ which is Riemann integrable*[4],

$$\int_c^d f(s)\, ds = \int_a^b f(\phi(t)) \left|\phi'(t)\right| dt$$

Proof: Let $F'(s) = f(s)$. (For example, let $F(s) = \int_a^s f(r)\, dr$.) Then the first integral equals $F(d) - F(c)$ by the fundamental theorem of calculus. By Lemma 8.14.8, ϕ is either strictly increasing or strictly decreasing. Suppose ϕ is strictly decreasing. Then $\phi(a) = d$ and $\phi(b) = c$. Therefore, $\phi' \leq 0$ and the second integral equals

$$-\int_a^b f(\phi(t))\, \phi'(t)\, dt = \int_b^a \frac{d}{dt}(F(\phi(t)))\, dt$$

$$= F(\phi(a)) - F(\phi(b)) = F(d) - F(c).$$

The case when ϕ is increasing is similar. This proves the theorem.

Lemma 8.14.13 *Let $\mathbf{f} : [a,b] \to C$, $\mathbf{g} : [c,d] \to C$ be parameterizations of a smooth curve which satisfy conditions 1 - 5. Then $\boldsymbol{\phi}(t) \equiv \mathbf{g}^{-1} \circ \mathbf{f}(t)$ is $1-1$ on (a,b), continuous on $[a,b]$, and either strictly increasing or strictly decreasing on $[a,b]$.*

Proof: It is obvious ϕ is $1-1$ on (a,b) from the conditions $\mathbf{f}$ and $\mathbf{g}$ satisfy. It only remains to verify continuity on $[a,b]$ because then the final claim follows from Lemma 8.14.8. If ϕ is not continuous on $[a,b]$, then there exists a sequence, $\{t_n\} \subseteq [a,b]$ such that $t_n \to t$ but $\phi(t_n)$ fails to converge to $\phi(t)$. Therefore, for some $\varepsilon > 0$ there exists a subsequence, still denoted by n such that $|\phi(t_n) - \phi(t)| \geq \varepsilon$. Using the sequential compactness of $[c,d]$, (See Theorem 7.7.18 on Page 150.) there is a further subsequence, still denoted by n such that $\{\phi(t_n)\}$ converges to a point, s, of $[c,d]$ which is not equal to $\phi(t)$. Thus $\mathbf{g}^{-1} \circ \mathbf{f}(t_n) \to s$ and still $t_n \to t$. Therefore, the continuity of $\mathbf{f}$ and $\mathbf{g}$ imply $\mathbf{f}(t_n) \to \mathbf{g}(s)$ and $\mathbf{f}(t_n) \to \mathbf{f}(t)$. Therefore, $\mathbf{g}(s) = \mathbf{f}(t)$ and so $s = \mathbf{g}^{-1} \circ \mathbf{f}(t) = \phi(t)$, a contradiction. Therefore, ϕ is continuous as claimed.

Theorem 8.14.14 *The length of a smooth curve is not dependent on parameterization.*

[4]Recall that all continuous functions of this sort are Riemann integrable.

Proof: Let C be the curve and suppose $\mathbf{f}:[a,b]\to C$ and $\mathbf{g}:[c,d]\to C$ both satisfy conditions 1 - 5. Is it true that $\int_a^b |\mathbf{f}'(t)|\,dt = \int_c^d |\mathbf{g}'(s)|\,ds$?

Let $\phi(t) \equiv \mathbf{g}^{-1}\circ\mathbf{f}(t)$ for $t\in[a,b]$. Then by the above lemma ϕ is either strictly increasing or strictly decreasing on $[a,b]$. Suppose for the sake of simplicity that it is strictly increasing. The decreasing case is handled similarly.

Let $s_0\in\phi([a+\delta,b-\delta])\subset(c,d)$. Then by assumption 4, $g_i'(s_0)\neq 0$ for some i. By continuity of g_i', it follows $g_i'(s)\neq 0$ for all $s\in I$ where I is an open interval contained in $[c,d]$ which contains s_0. It follows that on this interval, g_i is either strictly increasing or strictly decreasing. Therefore, $J\equiv g_i(I)$ is also an open interval and you can define a differentiable function, $h_i: J\to I$ by

$$h_i(g_i(s)) = s.$$

This implies that for $s\in I$,

$$h_i'(g_i(s)) = \frac{1}{g_i'(s)}. \tag{8.26}$$

Now letting $s=\phi(t)$ for $s\in I$, it follows $t\in J_1$, an open interval. Also, for s and t related this way, $\mathbf{f}(t)=\mathbf{g}(s)$ and so in particular, for $s\in I$,

$$g_i(s) = f_i(t).$$

Consequently,

$$s = h_i(f_i(t)) = \phi(t)$$

and so, for $t\in J_1$,

$$\phi'(t) = h_i'(f_i(t))\,f_i'(t) = h_i'(g_i(s))\,f_i'(t) = \frac{f_i'(t)}{g_i'(\phi(t))} \tag{8.27}$$

which shows that ϕ' exists and is continuous on J_1, an open interval containing $\phi^{-1}(s_0)$. Since s_0 is arbitrary, this shows ϕ' exists on $[a+\delta,b-\delta]$ and is continuous there.

Now $\mathbf{f}(t) = \mathbf{g}\circ\left(\mathbf{g}^{-1}\circ\mathbf{f}\right)(t) = \mathbf{g}(\phi(t))$ and it was just shown that ϕ' is a continuous function on $[a-\delta,b+\delta]$. It follows

$$\mathbf{f}'(t) = \mathbf{g}'(\phi(t))\,\phi'(t)$$

and so, by Theorem 8.14.12,

$$\begin{aligned}\int_{\phi(a+\delta)}^{\phi(b-\delta)} |\mathbf{g}'(s)|\,ds &= \int_{a+\delta}^{b-\delta} |\mathbf{g}'(\phi(t))|\,\left|\phi'(t)\right|\,dt \\ &= \int_{a+\delta}^{b-\delta} |\mathbf{f}'(t)|\,dt.\end{aligned}$$

Now using the continuity of $\phi, \mathbf{g}'$, and $\mathbf{f}'$ on $[a,b]$ and letting $\delta\to 0+$ in the above, yields

$$\int_c^d |\mathbf{g}'(s)|\,ds = \int_a^b |\mathbf{f}'(t)|\,dt$$

and this proves the theorem.

Motion On A Space Curve

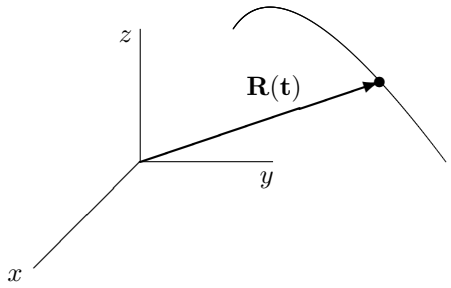
9.0.3 Outcomes

1. Recall the definitions of unit tangent, unit normal, and osculating plane.
2. Calculate the curvature for a space curve.
3. Given the position vector function of a moving object, calculate the velocity, speed, and acceleration of the object and write the acceleration in terms of its tangential and normal components.
4. Derive formulas for the curvature of a parameterized curve and the curvature of a plane curve given as a function.

9.1 Space Curves

A fly buzzing around the room, a person riding a roller coaster, and a satellite orbiting the earth all have something in common. They are moving over some sort of curve in three dimensions.

Denote by $\mathbf{R}(t)$ the position vector of the point on the curve which occurs at time t. Assume that $\mathbf{R}', \mathbf{R}''$ exist and is continuous. Thus $\mathbf{R}' = \mathbf{v}$, the velocity and $\mathbf{R}'' = \mathbf{a}$ is the acceleration.



Lemma 9.1.1 *Define* $\mathbf{T}(t) \equiv \mathbf{R}'(t) / |\mathbf{R}'(t)|$. *Then* $|\mathbf{T}(t)| = 1$ *and if* $\mathbf{T}'(t) \neq 0$, *then there exists a unit vector,* $\mathbf{N}(t)$ *perpendicular to* $\mathbf{T}(t)$ *and a scalar valued function,* $\kappa(t)$, *with* $\mathbf{T}'(t) = \kappa(t) |\mathbf{v}| \mathbf{N}(t)$.

Proof: It follows from the definition that $|\mathbf{T}| = 1$. Therefore, $\mathbf{T} \cdot \mathbf{T} = 1$ and so, upon differentiating both sides,

$$\mathbf{T}' \cdot \mathbf{T} + \mathbf{T} \cdot \mathbf{T}' = 2\mathbf{T}' \cdot \mathbf{T} = 0.$$

Therefore, $\mathbf{T}'$ is perpendicular to $\mathbf{T}$. Let

$$\mathbf{N}(t) \equiv \frac{\mathbf{T}'}{|\mathbf{T}'|}.$$

Then letting $|\mathbf{T}'| \equiv \kappa(t)|\mathbf{v}(t)|$, it follows

$$\mathbf{T}'(t) = \kappa(t)|\mathbf{v}(t)|\mathbf{N}(t).$$

This proves the lemma.

The plane determined by the two vectors, $\mathbf{T}$ and $\mathbf{N}$ is called the **osculating**[1] **plane**. It identifies a particular plane which is in a sense tangent to this space curve. In the case where $|\mathbf{T}'(t)| = 0$ near the point of interest, $\mathbf{T}(t)$ equals a constant and so the space curve is a straight line which it would be supposed has no curvature. Also, the principal normal is undefined in this case. This makes sense because if there is no curving going on, there is no special direction normal to the curve at such points which could be distinguished from any other direction normal to the curve. In the case where $|\mathbf{T}'(t)| = 0$, $\kappa(t) = 0$ and the radius of curvature would be considered infinite.

Definition 9.1.2 *The vector, $\mathbf{T}(t)$ is called the **unit tangent vector** and the vector, $\mathbf{N}(t)$ is called the **principal normal.** The function, $\kappa(t)$ in the above lemma is called the **curvature**. The **radius of curvature** is defined as $\rho = 1/\kappa$.*

The important thing about this is that it is possible to write the acceleration as the sum of two vectors, one perpendicular to the direction of motion and the other in the direction of motion.

Theorem 9.1.3 *For $\mathbf{R}(t)$ the position vector of a space curve, the acceleration is given by the formula*

$$\begin{aligned} \mathbf{a} &= \frac{d|\mathbf{v}|}{dt}\mathbf{T} + \kappa|\mathbf{v}|^2\mathbf{N} \\ &\equiv a_T\mathbf{T} + a_N\mathbf{N}. \end{aligned} \tag{9.1}$$

Furthermore, $a_T^2 + a_N^2 = |\mathbf{a}|^2$.

Proof:

$$\begin{aligned} \mathbf{a} &= \frac{d\mathbf{v}}{dt} = \frac{d}{dt}(\mathbf{R}') = \frac{d}{dt}(|\mathbf{v}|\mathbf{T}) \\ &= \frac{d|\mathbf{v}|}{dt}\mathbf{T} + |\mathbf{v}|\mathbf{T}' \\ &= \frac{d|\mathbf{v}|}{dt}\mathbf{T} + |\mathbf{v}|^2\kappa\mathbf{N}. \end{aligned}$$

This proves the first part.

For the second part,

$$\begin{aligned} |\mathbf{a}|^2 &= (a_T\mathbf{T} + a_N\mathbf{N}) \cdot (a_T\mathbf{T} + a_N\mathbf{N}) \\ &= a_T^2\mathbf{T}\cdot\mathbf{T} + 2a_Na_T\mathbf{T}\cdot\mathbf{N} + a_N^2\mathbf{N}\cdot\mathbf{N} \\ &= a_T^2 + a_N^2 \end{aligned}$$

because $\mathbf{T}\cdot\mathbf{N} = 0$. This proves the theorem.

[1] To osculate means to kiss. Thus this plane could be called the kissing plane. However, that does not sound formal enough so it is called the osculating plane.

Finally, it is well to point out that the curvature is a property of the curve itself, and does not depend on the paramterization of the curve. If the curve is given by two different vector valued functions, $\mathbf{R}(t)$ and $\mathbf{R}(\tau)$, then from the formula above for the curvature,

$$\kappa(t) = \frac{|\mathbf{T}'(t)|}{|\mathbf{v}(t)|} = \frac{\left|\frac{d\mathbf{T}}{d\tau}\frac{d\tau}{dt}\right|}{\left|\frac{d\mathbf{R}}{d\tau}\frac{d\tau}{dt}\right|} = \frac{\left|\frac{d\mathbf{T}}{d\tau}\right|}{\left|\frac{d\mathbf{R}}{d\tau}\right|} \equiv \kappa(\tau).$$

From this, it is possible to give an important formula from physics. Suppose an object orbits a point at constant speed, v. In the above notation, $|\mathbf{v}| = v$. What is the centripetal acceleration of this object? You may know from a physics class that the answer is v^2/r where r is the radius. This follows from the above quite easily. The parameterization of the object which is as described is

$$\mathbf{R}(t) = \left(r\cos\left(\frac{v}{r}t\right), r\sin\left(\frac{v}{r}t\right)\right).$$

Therefore, $\mathbf{T} = \left(-\sin\left(\frac{v}{r}t\right), \cos\left(\frac{v}{r}t\right)\right)$ and

$$\mathbf{T}' = \left(-\frac{v}{r}\cos\left(\frac{v}{r}t\right), -\frac{v}{r}\sin\left(\frac{v}{r}t\right)\right).$$

Thus, $\kappa = |\mathbf{T}'(t)|/v = \frac{1}{r}$. It follows

$$\mathbf{a} = \frac{dv}{dt}\mathbf{T} + v^2\kappa\mathbf{N} = \frac{v^2}{r}\mathbf{N}.$$

The vector, $\mathbf{N}$ points from the object toward the center of the circle because it is a positive multiple of the vector, $\left(-\frac{v}{r}\cos\left(\frac{v}{r}t\right), -\frac{v}{r}\sin\left(\frac{v}{r}t\right)\right)$.

Formula 9.1 also yields an easy way to find the curvature. Take the cross product of both sides with $\mathbf{v}$, the velocity. Then

$$\begin{aligned}\mathbf{a}\times\mathbf{v} &= \frac{d|\mathbf{v}|}{dt}\mathbf{T}\times\mathbf{v} + |\mathbf{v}|^2\kappa\mathbf{N}\times\mathbf{v}\\ &= \frac{d|\mathbf{v}|}{dt}\mathbf{T}\times\mathbf{v} + |\mathbf{v}|^3\kappa\mathbf{N}\times\mathbf{T}\end{aligned}$$

Now $\mathbf{T}$ and $\mathbf{v}$ have the same direction so the first term on the right equals zero. Taking the magnitude of both sides, and using the fact that $\mathbf{N}$ and $\mathbf{T}$ are two perpendicular unit vectors,

$$|\mathbf{a}\times\mathbf{v}| = |\mathbf{v}|^3\kappa$$

and so

$$\kappa = \frac{|\mathbf{a}\times\mathbf{v}|}{|\mathbf{v}|^3}. \tag{9.2}$$

Example 9.1.4 *Let $\mathbf{R}(t) = \left(\cos(t), t, t^2\right)$ for $t \in [0,3]$. Find the speed, velocity, curvature, and write the acceleration in terms of normal and tangential components.*

First of all $\mathbf{v}(t) = (-\sin t, 1, 2t)$ and so the speed is given by

$$|\mathbf{v}| = \sqrt{\sin^2(t) + 1 + 4t^2}.$$

Therefore,

$$a_T = \frac{d}{dt}\left(\sqrt{\sin^2(t) + 1 + 4t^2}\right) = \frac{\sin(t)\cos(t) + 4t}{\sqrt{(2 + 4t^2 - \cos^2 t)}}.$$

It remains to find a_N. To do this, you can find the curvature first if you like.

$$\mathbf{a}(t) = \mathbf{R}''(t) = (-\cos t, 0, 2).$$

Then

$$\begin{aligned}\kappa &= \frac{|(-\cos t, 0, 2) \times (-\sin t, 1, 2t)|}{\left(\sqrt{\sin^2(t) + 1 + 4t^2}\right)^3} \\ &= \frac{\sqrt{4 + (-2\sin(t) + 2(\cos(t))\, t)^2 + \cos^2(t)}}{\left(\sqrt{\sin^2(t) + 1 + 4t^2}\right)^3}\end{aligned}$$

Then

$$a_N = \kappa |\mathbf{v}|^2$$

$$\begin{aligned}&= \frac{\sqrt{4 + (-2\sin(t) + 2(\cos(t))\, t)^2 + \cos^2(t)}}{\left(\sqrt{\sin^2(t) + 1 + 4t^2}\right)^3} \left(\sin^2(t) + 1 + 4t^2\right) \\ &= \frac{\sqrt{4 + (-2\sin(t) + 2(\cos(t))\, t)^2 + \cos^2(t)}}{\sqrt{\sin^2(t) + 1 + 4t^2}}.\end{aligned}$$

You can observe the formula $a_N^2 + a_T^2 = |\mathbf{a}|^2$ holds. Indeed $a_N^2 + a_T^2 =$

$$\begin{aligned}&\left(\frac{\sqrt{4 + (-2\sin(t) + 2(\cos(t))\, t)^2 + \cos^2(t)}}{\sqrt{\sin^2(t) + 1 + 4t^2}}\right)^2 + \left(\frac{\sin(t)\cos(t) + 4t}{\sqrt{(2 + 4t^2 - \cos^2 t)}}\right)^2 \\ = \;& \frac{4 + (-2\sin t + 2(\cos t)\, t)^2 + \cos^2 t}{\sin^2 t + 1 + 4t^2} + \frac{(\sin t \cos t + 4t)^2}{2 + 4t^2 - \cos^2 t} = \cos^2 t + 4 = |\mathbf{a}|^2\end{aligned}$$

9.1.1 Some Simple Techniques

Recall the formula for acceleration is

$$\mathbf{a} = a_T \mathbf{T} + a_N \mathbf{N} \tag{9.3}$$

where $a_T = \frac{d|\mathbf{v}|}{dt}$ and $a_N = \kappa |\mathbf{v}|^2$. Of course one way to find a_T and a_N is to just find $|\mathbf{v}|$, $\frac{d|\mathbf{v}|}{dt}$ and κ and plug in. However, there is another way which might be easier. Take the dot product of both sides with $\mathbf{T}$. This gives,

$$\mathbf{a} \cdot \mathbf{T} = a_T \mathbf{T} \cdot \mathbf{T} + a_N \mathbf{N} \cdot \mathbf{T} = a_T.$$

Thus

$$\mathbf{a} = (\mathbf{a} \cdot \mathbf{T})\, \mathbf{T} + a_N \mathbf{N}$$

and so

$$\mathbf{a} - (\mathbf{a} \cdot \mathbf{T})\, \mathbf{T} = a_N \mathbf{N} \tag{9.4}$$

and taking norms of both sides,

$$|\mathbf{a} - (\mathbf{a} \cdot \mathbf{T})\, \mathbf{T}| = a_N.$$

Also from 9.4,

$$\frac{\mathbf{a} - (\mathbf{a} \cdot \mathbf{T})\, \mathbf{T}}{|\mathbf{a} - (\mathbf{a} \cdot \mathbf{T})\, \mathbf{T}|} = \frac{a_N \mathbf{N}}{a_N |\mathbf{N}|} = \mathbf{N}.$$

Also recall

$$\kappa = \frac{|\mathbf{a} \times \mathbf{v}|}{|\mathbf{v}|^3},\ a_T^2 + a_N^2 = |\mathbf{a}|^2$$

This is usually easier than computing $\mathbf{T}'/|\mathbf{T}'|$. To illustrate the use of these simple observations, consider the example worked above which was fairly messy. I will make it easier by selecting a value of t and by using the above simplifying techniques.

Example 9.1.5 *Let* $\mathbf{R}(t) = \left(\cos(t), t, t^2\right)$ *for* $t \in [0,3]$. *Find the speed, velocity, curvature, and write the acceleration in terms of normal and tangential components when* $t = 0$. *Also find* $\mathbf{N}$ *at the point where* $t = 0$.

First I need to find the velocity and acceleration. Thus

$$\mathbf{v} = (-\sin t, 1, 2t),\ \mathbf{a} = (-\cos t, 0, 2)$$

and consequently,

$$\mathbf{T} = \frac{(-\sin t, 1, 2t)}{\sqrt{\sin^2(t) + 1 + 4t^2}}.$$

When $t = 0$, this reduces to

$$\mathbf{v}(0) = (0,1,0),\ \mathbf{a} = (-1,0,2),\ |\mathbf{v}(0)| = 1,\ \mathbf{T} = (0,1,0),$$

and consequently,

$$\mathbf{T} = (0,1,0).$$

Then the tangential component of acceleration when $t = 0$ is

$$a_T = (-1,0,2) \cdot (0,1,0) = 0$$

Now $|\mathbf{a}|^2 = 5$ and so $a_N = \sqrt{5}$ because $a_T^2 + a_N^2 = |\mathbf{a}|^2$. Thus $\sqrt{5} = \kappa |\mathbf{v}(0)|^2 = \kappa \cdot 1 = \kappa$. Next lets find $\mathbf{N}$. From $\mathbf{a} = a_T \mathbf{T} + a_N \mathbf{N}$ it follows

$$(-1,0,2) = 0 \cdot \mathbf{T} + \sqrt{5}\mathbf{N}$$

and so

$$\mathbf{N} = \frac{1}{\sqrt{5}}(-1,0,2).$$

This was pretty easy.

Example 9.1.6 *Find a formula for the curvature of the curve given by the graph of* $y = f(x)$ *for* $x \in [a,b]$. *Assume whatever you like about smoothness of* f.

You need to write this as a parametric curve. This is most easily accomplished by letting $t = x$. Thus a parameterization is

$$(t, f(t), 0) : t \in [a,b].$$

Then you can use the formula given above. The acceleration is $(0, f''(t), 0)$ and the velocity is $(1, f'(t), 0)$. Therefore,

$$\mathbf{a} \times \mathbf{v} = (0, f''(t), 0) \times (1, f'(t), 0) = (0, 0, -f''(t)).$$

Therefore, the curvature is given by

$$\frac{|\mathbf{a} \times \mathbf{v}|}{|\mathbf{v}|^3} = \frac{|f''(t)|}{\left(1 + f'(t)^2\right)^{3/2}}.$$

Sometimes curves don't come to you parametrically. This is unfortunate when it occurs but you can sometimes find a parametric description of such curves. It should be emphasized that it is only sometimes when you can actually find a parameterization. General systems of nonlinear equations cannot be solved using algebra.

Example 9.1.7 *Find a parameterization for the intersection of the surfaces* $y+3z=2x^2+4$ *and* $y+2z=x+1$.

You need to solve for x and y in terms of x. This yields

$$z=2x^2-x+3,\ y=-4x^2+3x-5.$$

Therefore, letting $t=x$, the parameterization is $(x,y,z)=\left(t,-4t^2-5+3t,-t+3+2t^2\right)$.

Example 9.1.8 *Find a parametrization for the straight line joining* $(3,2,4)$ *and* $(1,10,5)$.

$(x,y,z)=(3,2,4)+t(-2,8,1)=(3-2t,2+8t,4+t)$ where $t\in[0,1]$. Note where this came from. The vector, $(-2,8,1)$ is obtained from $(1,10,5)-(3,2,4)$. Now you should check to see this works.

9.2 Geometry Of Space Curves*

If you are interested in more on space curves, you should read this section. Otherwise, procede to the exercises. Denote by $\mathbf{R}(s)$ the function which takes s to a point on this curve where s is arc length. Thus $\mathbf{R}(s)$ equals the point on the curve which occurs when you have traveled a distance of s along the curve from one end. This is known as the parameterization of the curve in terms of arc length. Note also that it incorporates an orientation on the curve because there are exactly two ends you could begin measuring length from. In this section, assume anything about smoothness and continuity to make the following manipulations valid. In particular, assume that $\mathbf{R}'$ exists and is continuous.

Lemma 9.2.1 *Define* $\mathbf{T}(s)\equiv\mathbf{R}'(s)$. *Then* $|\mathbf{T}(s)|=1$ *and if* $\mathbf{T}'(s)\neq 0$, *then there exists a unit vector,* $\mathbf{N}(s)$ *perpendicular to* $\mathbf{T}(s)$ *and a scalar valued function,* $\kappa(s)$ *with* $\mathbf{T}'(s)=\kappa(s)\mathbf{N}(s)$.

Proof: First, $s=\int_0^s|\mathbf{R}'(r)|\,dr$ because of the definition of arc length. Therefore, from the fundamental theorem of calculus, $1=|\mathbf{R}'(s)|=|\mathbf{T}(s)|$. Therefore, $\mathbf{T}\cdot\mathbf{T}=1$ and so upon differentiating this on both sides, yields $\mathbf{T}'\cdot\mathbf{T}+\mathbf{T}\cdot\mathbf{T}'=0$ which shows $\mathbf{T}\cdot\mathbf{T}'=0$. Therefore, the vector, $\mathbf{T}'$ is perpendicular to the vector, $\mathbf{T}$. In case $\mathbf{T}'(s)\neq\mathbf{0}$, let $\mathbf{N}(s)=\frac{\mathbf{T}'(s)}{|\mathbf{T}'(s)|}$ and so $\mathbf{T}'(s)=|\mathbf{T}'(s)|\,\mathbf{N}(s)$, showing the scalar valued function is $\kappa(s)=|\mathbf{T}'(s)|$. This proves the lemma.

The radius of curvature is defined as $\rho=\frac{1}{\kappa}$. Thus at points where there is a lot of curvature, the radius of curvature is small and at points where the curvature is small, the radius of curvature is large. The plane determined by the two vectors, $\mathbf{T}$ and $\mathbf{N}$ is called the osculating plane. It identifies a particular plane which is in a sense tangent to this space curve. In the case where $|\mathbf{T}'(s)|=0$ near the point of interest, $\mathbf{T}(s)$ equals a constant and so the space curve is a straight line which it would be supposed has no curvature. Also, the principal normal is undefined in this case. This makes sense because if there is no curving going on, there is no special direction normal to the curve at such points which could be distinguished from any other direction normal to the curve. In the case where $|\mathbf{T}'(s)|=0$, $\kappa(s)=0$ and the radius of curvature would be considered infinite.

Definition 9.2.2 *The vector,* $\mathbf{T}(s)$ *is called the unit tangent vector and the vector,* $\mathbf{N}(s)$ *is called the* ***principal normal.*** *The function,* $\kappa(s)$ *in the above lemma is called the* ***curvature.*** *When* $\mathbf{T}'(s) \neq 0$ *so the principal normal is defined, the vector,* $\mathbf{B}(s) \equiv \mathbf{T}(s) \times \mathbf{N}(s)$ *is called the* ***binormal.***

The binormal is normal to the osculating plane and $\mathbf{B}'$ tells how fast this vector changes. Thus it measures the rate at which the curve twists.

Lemma 9.2.3 *Let* $\mathbf{R}(s)$ *be a parameterization of a space curve with respect to arc length and let the vectors,* $\mathbf{T}, \mathbf{N}$, *and* $\mathbf{B}$ *be as defined above. Then* $\mathbf{B}' = \mathbf{T} \times \mathbf{N}'$ *and there exists a scalar function,* $\tau(s)$ *such that* $\mathbf{B}' = \tau \mathbf{N}$.

Proof: From the definition of $\mathbf{B} = \mathbf{T} \times \mathbf{N}$, and you can differentiate both sides and get $\mathbf{B}' = \mathbf{T}' \times \mathbf{N} + \mathbf{T} \times \mathbf{N}'$. Now recall that $\mathbf{T}'$ is a multiple called curvature multiplied by $\mathbf{N}$ so the vectors, $\mathbf{T}'$ and $\mathbf{N}$ have the same direction and $\mathbf{B}' = \mathbf{T} \times \mathbf{N}'$. Therefore, $\mathbf{B}'$ is either zero or is perpendicular to $\mathbf{T}$. But also, from the definition of $\mathbf{B}, \mathbf{B}$ is a unit vector and so $\mathbf{B}(s) \cdot \mathbf{B}(s) = 0$. Differentiating this,$\mathbf{B}'(s) \cdot \mathbf{B}(s) + \mathbf{B}(s) \cdot \mathbf{B}'(s) = 0$ showing that $\mathbf{B}'$ is perpendicular to $\mathbf{B}$ also. Therefore, $\mathbf{B}'$ is a vector which is perpendicular to both vectors, $\mathbf{T}$ and $\mathbf{B}$ and since this is in three dimensions, $\mathbf{B}'$ must be some scalar multiple of $\mathbf{N}$ and it is this multiple called τ. Thus $\mathbf{B}' = \tau \mathbf{N}$ as claimed.

Lets go over this last claim a little more. The following situation is obtained. There are two vectors, $\mathbf{T}$ and $\mathbf{B}$ which are perpendicular to each other and both $\mathbf{B}'$ and $\mathbf{N}$ are perpendicular to these two vectors, hence perpendicular to the plane determined by them. Therefore, $\mathbf{B}'$ must be a multiple of $\mathbf{N}$. Take a piece of paper, draw two unit vectors on it which are perpendicular. Then you can see that any two vectors which are perpendicular to this plane must be multiples of each other.

The scalar function, τ is called the torsion. In case $\mathbf{T}' = 0$, none of this is defined because in this case there is not a well defined osculating plane. The conclusion of the following theorem is called the Serret Frenet formulas.

Theorem 9.2.4 *(Serret Frenet) Let* $\mathbf{R}(s)$ *be the parameterization with respect to arc length of a space curve and* $\mathbf{T}(s) = \mathbf{R}'(s)$ *is the unit tangent vector. Suppose* $|\mathbf{T}'(s)| \neq 0$ *so the principal normal,* $\mathbf{N}(s) = \frac{\mathbf{T}'(s)}{|\mathbf{T}'(s)|}$ *is defined. The binormal is the vector* $\mathbf{B} \equiv \mathbf{T} \times \mathbf{N}$ *so* $\mathbf{T}, \mathbf{N}, \mathbf{B}$ *forms a right handed system of unit vectors each of which is perpendicular to every other. Then the following system of differential equations holds in* $\mathbb{R}^9$.

$$\mathbf{B}' = \tau \mathbf{N},\ \mathbf{T}' = \kappa \mathbf{N},\ \mathbf{N}' = -\kappa \mathbf{T} - \tau \mathbf{B}$$

where κ *is the curvature and is nonnegative and* τ *is the* ***torsion.***

Proof: $\kappa \geq 0$ because $\kappa = |\mathbf{T}'(s)|$. The first two equations are already established. To get the third, note that $\mathbf{B} \times \mathbf{T} = \mathbf{N}$ which follows because $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is given to form a right handed system of unit vectors each perpendicular to the others. (Use your right hand.) Now take the derivative of this expression. thus

$$\begin{aligned} \mathbf{N}' &= \mathbf{B}' \times \mathbf{T} + \mathbf{B} \times \mathbf{T}' \\ &= \tau \mathbf{N} \times \mathbf{T} + \kappa \mathbf{B} \times \mathbf{N}. \end{aligned}$$

Now recall again that $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is a right hand system. Thus $\mathbf{N} \times \mathbf{T} = -\mathbf{B}$ and $\mathbf{B} \times \mathbf{N} = -\mathbf{T}$. This establishes the Frenet Serret formulas.

This is an important example of a system of differential equations in $\mathbb{R}^9$. It is a remarkable result because it says that from knowledge of the two scalar functions, τ and κ, and initial values for $\mathbf{B}, \mathbf{T}$, and $\mathbf{N}$ when $s = 0$ you can obtain the binormal, unit tangent, and principal normal vectors. It is just the solution of an initial value

problem for a system of ordinary differential equations. Having done this, you can reconstruct the entire space curve starting at some point, $\mathbf{R}_0$ because $\mathbf{R}'(s) = \mathbf{T}(s)$ and so $\mathbf{R}(s) = \mathbf{R}_0 + \int_0^s \mathbf{T}'(r)\, dr$.

The vectors, $\mathbf{B}, \mathbf{T}$, and $\mathbf{N}$ are vectors which are functions of position on the space curve. Often, especially in applications, you deal with a space curve which is parameterized by a function of t where t is time. Thus a value of t would correspond to a point on this curve and you could let $\mathbf{B}(t), \mathbf{T}(t)$, and $\mathbf{N}(t)$ be the binormal, unit tangent, and principal normal at this point of the curve. The following example is typical.

Example 9.2.5 *Given the circular helix,* $\mathbf{R}(t) = (a\cos t)\,\mathbf{i} + (a\sin t)\,\mathbf{j} + (bt)\,\mathbf{k}$, *find the arc length,* $s(t)$, *the unit tangent vector,* $\mathbf{T}(t)$, *the principal normal,* $\mathbf{N}(t)$, *the binormal,* $\mathbf{B}(t)$, *the curvature,* $\kappa(t)$, *and the torsion,* $\tau(t)$. *Here* $t \in [0, T]$.

The arc length is $s(t) = \int_0^t \left(\sqrt{a^2+b^2}\right) dr = \left(\sqrt{a^2+b^2}\right) t$. Now the tangent vector is obtained using the chain rule as

$$\begin{aligned}\mathbf{T} &= \frac{d\mathbf{R}}{ds} = \frac{d\mathbf{R}}{dt}\frac{dt}{ds} = \frac{1}{\sqrt{a^2+b^2}}\mathbf{R}'(t) \\ &= \frac{1}{\sqrt{a^2+b^2}}\left((-a\sin t)\,\mathbf{i} + (a\cos t)\,\mathbf{j} + b\mathbf{k}\right)\end{aligned}$$

The principal normal:

$$\begin{aligned}\frac{d\mathbf{T}}{ds} &= \frac{d\mathbf{T}}{dt}\frac{dt}{ds} \\ &= \frac{1}{a^2+b^2}\left((-a\cos t)\,\mathbf{i} + (-a\sin t)\,\mathbf{j} + 0\mathbf{k}\right)\end{aligned}$$

and so

$$\mathbf{N} = \frac{d\mathbf{T}}{ds} \Big/ \left|\frac{d\mathbf{T}}{ds}\right| = -\left((\cos t)\,\mathbf{i} + (\sin t)\,\mathbf{j}\right)$$

The binormal:

$$\begin{aligned}\mathbf{B} &= \frac{1}{\sqrt{a^2+b^2}}\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -a\sin t & a\cos t & b \\ -\cos t & -\sin t & 0 \end{vmatrix} \\ &= \frac{1}{\sqrt{a^2+b^2}}\left((b\sin t)\,\mathbf{i} - b\cos t\mathbf{j} + a\mathbf{k}\right)\end{aligned}$$

Now the curvature, $\kappa(t) = \left|\frac{d\mathbf{T}}{ds}\right| = \sqrt{\left(\frac{a\cos t}{a^2+b^2}\right)^2 + \left(\frac{a\sin t}{a^2+b^2}\right)^2} = \frac{a}{a^2+b^2}$. Note the curvature is constant in this example. The final task is to find the torsion. Recall that $\mathbf{B}' = \tau\mathbf{N}$ where the derivative on $\mathbf{B}$ is taken with respect to arc length. Therefore, remembering that t is a function of s,

$$\begin{aligned}\mathbf{B}'(s) &= \frac{1}{\sqrt{a^2+b^2}}\left((b\cos t)\,\mathbf{i} + (b\sin t)\,\mathbf{j}\right)\frac{dt}{ds} \\ &= \frac{1}{a^2+b^2}\left((b\cos t)\,\mathbf{i} + (b\sin t)\,\mathbf{j}\right) \\ &= \tau\left(-(\cos t)\,\mathbf{i} - (\sin t)\,\mathbf{j}\right) = \tau\mathbf{N}\end{aligned}$$

and it follows $-b/\left(a^2+b^2\right) = \tau$.

An important application of the usefulness of these ideas involves the decomposition of the acceleration in terms of these vectors of an object moving over a space curve.

Corollary 9.2.6 *Let* $\mathbf{R}(t)$ *be a space curve and denote by* $\mathbf{v}(t)$ *the velocity,* $\mathbf{v}(t) = \mathbf{R}'(t)$ *and let* $v(t) \equiv |\mathbf{v}(t)|$ *denote the speed and let* $\mathbf{a}(t)$ *denote the acceleration. Then* $\mathbf{v} = v\mathbf{T}$ *and* $\mathbf{a} = \frac{dv}{dt}\mathbf{T} + \kappa v^2 \mathbf{N}$.

Proof: $\mathbf{T} = \frac{d\mathbf{R}}{ds} = \frac{d\mathbf{R}}{dt}\frac{dt}{ds} = \mathbf{v}\frac{dt}{ds}$. Also, $s = \int_0^t v(r)\,dr$ and so $\frac{ds}{dt} = v$ which implies $\frac{dt}{ds} = \frac{1}{v}$. Therefore, $\mathbf{T} = \mathbf{v}/v$ which implies $\mathbf{v} = v\mathbf{T}$ as claimed.

Now the acceleration is just the derivative of the velocity and so by the Serrat Frenet formulas,

$$\begin{aligned} \mathbf{a} &= \frac{dv}{dt}\mathbf{T} + v\frac{d\mathbf{T}}{dt} \\ &= \frac{dv}{dt}\mathbf{T} + v\frac{d\mathbf{T}}{ds}v = \frac{dv}{dt}\mathbf{T} + v^2\kappa\mathbf{N} \end{aligned}$$

Note how this decomposes the acceleration into a component tangent to the curve and one which is normal to it. Also note that from the above, $v|\mathbf{T}'|\frac{\mathbf{T}'(t)}{|\mathbf{T}'|} = v^2\kappa\mathbf{N}$ and so $\frac{|\mathbf{T}'|}{v} = \kappa$ and $\mathbf{N} = \frac{\mathbf{T}'(t)}{|\mathbf{T}'|}$

From this, it is possible to give an important formula from physics. Suppose an object orbits a point at constant speed, v. What is the centripetal acceleration of this object? You may know from a physics class that the answer is v^2/r where r is the radius. This follows from the above quite easily. The parameterization of the object which is as described is

$$\mathbf{R}(t) = \left(r\cos\left(\frac{v}{r}t\right), r\sin\left(\frac{v}{r}t\right)\right).$$

Therefore, $\mathbf{T} = \left(-\sin\left(\frac{v}{r}t\right), \cos\left(\frac{v}{r}t\right)\right)$ and $\mathbf{T}' = \left(-\frac{v}{r}\cos\left(\frac{v}{r}t\right), -\frac{v}{r}\sin\left(\frac{v}{r}t\right)\right)$. Thus,

$$\kappa = |\mathbf{T}'(t)|/v = \frac{1}{r}.$$

It follows $\mathbf{a} = \frac{dv}{dt}\mathbf{T} + v^2\kappa\mathbf{N} = \frac{v^2}{r}\mathbf{N}$. The vector, $\mathbf{N}$ points from the object toward the center of the circle because it is a positive multiple of the vector, $\left(-\frac{v}{r}\cos\left(\frac{v}{r}t\right), -\frac{v}{r}\sin\left(\frac{v}{r}t\right)\right)$.

9.3 Exercises

1. Find a parametrization for the intersection of the planes $2x + y + 3z = -2$ and $3x - 2y + z = -4$.
2. Find a parametrization for the intersection of the plane $3x + y + z = -3$ and the circular cylinder $x^2 + y^2 = 1$.
3. Find a parametrization for the intersection of the plane $4x + 2y + 3z = 2$ and the elliptic cylinder $x^2 + 4z^2 = 9$.
4. Find a parametrization for the straight line joining $(1, 2, 1)$ and $(-1, 4, 4)$.
5. Find a parametrization for the intersection of the surfaces $3y + 3z = 3x^2 + 2$ and $3y + 2z = 3$.
6. Find a formula for the curvature of the curve, $y = \sin x$ in the xy plane.
7. Find a formula for the curvature of the space curve in $\mathbb{R}^2$, $(x(t), y(t))$.
8. An object moves over the helix, $(\cos 3t, \sin 3t, 5t)$. Find the normal and tangential components of the acceleration of this object as a function of t and write the acceleration in the form $a_T\mathbf{T} + a_N\mathbf{N}$.

9. An object moves over the helix, $(\cos t, \sin t, t)$. Find the normal and tangential components of the acceleration of this object as a function of t and write the acceleration in the form $a_T\mathbf{T} + a_N\mathbf{N}$.

10. An object moves in $\mathbb{R}^3$ according to the formula, $\left(\cos 3t, \sin 3t, t^2\right)$. Find the normal and tangential components of the acceleration of this object as a function of t and write the acceleration in the form $a_T\mathbf{T} + a_N\mathbf{N}$.

11. An object moves over the helix, $(\cos t, \sin t, 2t)$. Find the osculating plane at the point of the curve corresponding to $t = \pi/4$.

12. An object moves over a circle of radius r according to the formula,

$$\mathbf{r}(t) = (r\cos(\omega t), r\sin(\omega t))$$

where $v = r\omega$. Show that the speed of the object is constant and equals to v. Tell why $a_T = 0$ and find a_N, $\mathbf{N}$. This yields the formula for centripetal acceleration from beginning physics classes.

13. Suppose $|\mathbf{R}(t)| = c$ where c is a constant and $\mathbf{R}(t)$ is the position vector of an object. Show the velocity, $\mathbf{R}'(t)$ is always perpendicular to $\mathbf{R}(t)$.

14. An object moves in three dimensions and the only force on the object is a central force. This means that if $\mathbf{r}(t)$ is the position of the object, $\mathbf{a}(t) = k(\mathbf{r}(t))\mathbf{r}(t)$ where k is some function. Show that if this happens, then the motion of the object must be in a plane. **Hint:** First argue that $\mathbf{a} \times \mathbf{r} = \mathbf{0}$. Next show that $(\mathbf{a} \times \mathbf{r}) = (\mathbf{v} \times \mathbf{r})'$. Therefore, $(\mathbf{v} \times \mathbf{r})' = \mathbf{0}$. Explain why this requires $\mathbf{v} \times \mathbf{r} = \mathbf{c}$ for some vector, $\mathbf{c}$ which does not depend on t. Then explain why $\mathbf{c} \cdot \mathbf{r} = 0$. This implies the motion is in a plane. Why? What are some examples of central forces?

15. Let $\mathbf{R}(t) = (\cos t)\,\mathbf{i} + (\cos t)\,\mathbf{j} + \left(\sqrt{2}\sin t\right)\mathbf{k}$. Find the arc length, s as a function of the parameter, t, if $t = 0$ is taken to correspond to $s = 0$.

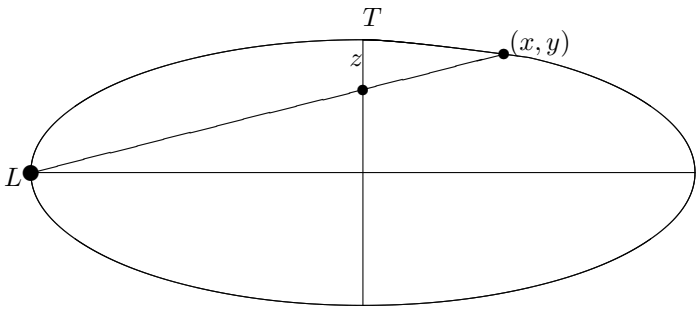
16. Let $\mathbf{R}(t) = 2\mathbf{i} + (4t+2)\,\mathbf{j} + 4t\mathbf{k}$. Find the arc length, s as a function of the parameter, t, if $t = 0$ is taken to correspond to $s = 0$.

17. Let $\mathbf{R}(t) = e^{5t}\mathbf{i} + e^{-5t}\mathbf{j} + 5\sqrt{2}t\mathbf{k}$. Find the arc length, s as a function of the parameter, t, if $t = 0$ is taken to correspond to $s = 0$.

18. An object moves along the x axis toward $(0,0)$ and then along the curve $y = x^2$ in the direction of increasing x at constant speed. Is the force acting on the object a continuous function? Explain. Is there any physically reasonable way to make this force continuous by relaxing the requirement that the object move at constant speed? If the curve were part of a railroad track, what would happen at the point where $x = 0$?

19. An object of mass m moving over a space curve is acted on by a force, $\mathbf{F}$. Show the work done by this force equals ma_T (length of the curve). In other words, it is only the tangential component of the force which does work.

20. The edge of an elliptical skating rink represented in the following picture has a light at its left end and satisfies the equation $\frac{x^2}{900} + \frac{y^2}{256} = 1$. (Distances measured in yards.)



A hockey puck slides from the point, T towards the center of the rink at the rate of 2 yards per second. What is the speed of its shadow along the wall when $z = 8$? **Hint:** You need to find $\sqrt{x'^2 + y'^2}$ at the instant described.

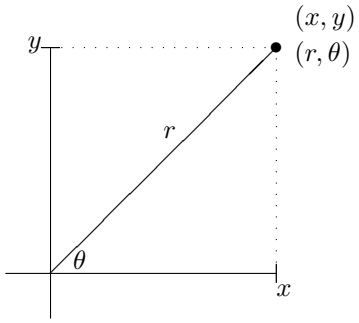
Some Curvilinear Coordinate Systems

10.0.1 Outcomes

1. Recall and use polar coordinates.
2. Graph relations involving polar coordinates.
3. Find the area of regions defined in terms of polar coordinates.
4. Recall and understand the derivation of Kepler's laws.
5. Recall and apply the concept of acceleration in polar coordinates.
6. Recall and use cylindrical and spherical coordinates.

10.1 Polar Coordinates

So far points have been identified in terms of Cartesian coordinates but there are other ways of specifying points in two and three dimensional space. These other ways involve using a list of two or three numbers which have a totally different meaning than Cartesian coordinates to specify a point in two or three dimensional space. In general these lists of numbers which have a different meaning than Cartesian coordinates are called Curvilinear coordinates. Probably the simplest curvilinear coordinate system is that of **polar coordinates**. The idea is suggested in the following picture.



You see in this picture, the number r identifies the distance of the point from the origin, $(0, 0)$ while θ is the angle shown between the positive x axis and the line from the origin to the point. This angle will always be given in radians and is in the interval $[0, 2\pi)$. Thus the given point, indicated by a small dot in the picture, can be described

in terms of the Cartesian coordinates, (x, y) or the polar coordinates, (r, θ). How are the two coordinates systems related? From the picture,

$$x = r\cos(\theta),\ y = r\sin(\theta). \tag{10.1}$$

Example 10.1.1 *The polar coordinates of a point in the plane are* $\left(5, \frac{\pi}{6}\right)$. *Find the Cartesian or rectangular coordinates of this point.*

From 10.1, $x = 5\cos\left(\frac{\pi}{6}\right) = \frac{5}{2}\sqrt{3}$ and $y = 5\sin\left(\frac{\pi}{6}\right) = \frac{5}{2}$. Thus the Cartesian coordinates are $\left(\frac{5}{2}\sqrt{3}, \frac{5}{2}\right)$.

Example 10.1.2 *Suppose the Cartesian coordinates of a point are* $(3, 4)$. *Find the polar coordinates.*

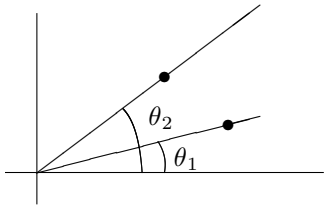
Recall that r is the distance form $(0, 0)$ and so $r = 5 = \sqrt{3^2 + 4^2}$. It remains to identify the angle. Note the point is in the first quadrant, (Both the x and y values are positive.) Therefore, the angle is something between 0 and $\pi/2$ and also $3 = 5\cos(\theta)$, and $4 = 5\sin(\theta)$. Therefore, dividing yields $\tan(\theta) = 4/3$. At this point, use a calculator or a table of trigonometric functions to find that at least approximately, $\theta = .927\,295$ radians.

10.1.1 Graphs In Polar Coordinates

Just as in the case of rectangular coordinates, it is possible to use relations between the polar coordinates to specify points in the plane. The process of sketching their graphs is very similar to that used to sketch graphs of functions in rectangular coordinates. I will only consider the case where the relation between the polar coordinates is of the form, $r = f(\theta)$. To graph such a relation, you can make a table of the form

θ	r
θ_1	$f(\theta_1)$
θ_2	$f(\theta_2)$
$\vdots$	$\vdots$

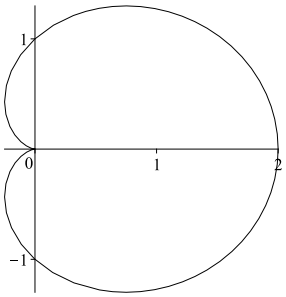
and then graph the resulting points and connect them up with a curve. The following picture illustrates how to begin this process.



To obtain the point in the plane which goes with the pair $(\theta, f(\theta))$, you draw the ray through the origin which makes an angle of θ with the positive x axis. Then you move along this ray a distance of $f(\theta)$ to obtain the point. As in the case with rectangular coordinates, this process is tedious and is best done by a computer algebra system.

Example 10.1.3 *Graph the polar equation,* $r = 1 + \cos\theta$.

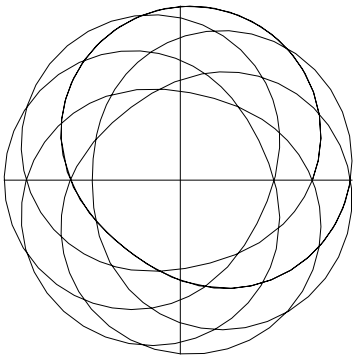
To do this, I will use Maple. The command which produces the polar graph of this is: > plot(1+cos(t),t=0..2*Pi,coords=polar); It tells Maple that r is given by $1 + \cos(t)$ and that $t \in [0, 2\pi]$. The variable t is playing the role of θ. It is easier to type t than θ in Maple.



You can also see just from your knowledge of the trig. functions that the graph should look something like this. When $\theta = 0, r = 2$ and then as θ increases to $\pi/2$, you see that $\cos\theta$ decreases to 0. Thus the line from the origin to the point on the curve should get shorter as θ goes from 0 to $\pi/2$. Then from $\pi/2$ to π, $\cos\theta$ gets negative eventually equaling -1 at $\theta = \pi$. Thus $r = 0$ at this point. Viewing the graph, you see this is exactly what happens. The above function is called a **cardioid**.

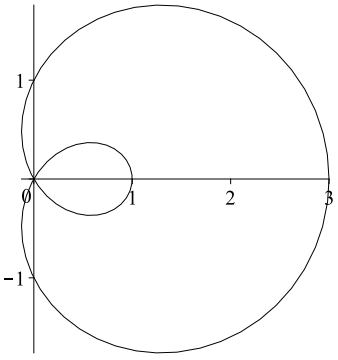
Here is another example. This is the graph obtained from $r = 3 + \sin\left(\frac{7\theta}{6}\right)$.

Example 10.1.4 *Graph* $r = 3 + \sin\left(\frac{7\theta}{6}\right)$ *for* $\theta \in [0, 14\pi]$.



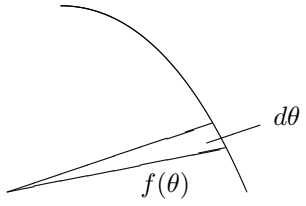
In polar coordinates people sometimes allow r to be negative. When this happens, it means that to obtain the point in the plane, you go in the opposite direction along the ray which starts at the origin and makes an angle of θ with the positive x axis. I do not believe the fussiness occasioned by this extra generality is justified by any sufficiently interesting application so no more will be said about this. It is mainly a fun way to obtain pretty pictures. Here is such an example.

Example 10.1.5 *Graph* $r = 1 + 2\cos\theta$ *for* $\theta \in [0, 2\pi]$.



10.2 The Area In Polar Coordinates

How can you find the area of the region determined by $0 \leq r \leq f(\theta)$ for $\theta \in [a,b]$, assuming this is a well defined set of points in the plane? See Example 10.1.5 with $\theta \in [0,2\pi]$ to see something which it would be better to avoid. I have in mind the situation where every ray through the origin having angle θ for $\theta \in [a,b]$ intersects the graph of $r = f(\theta)$ in exactly one point. To see how to find the area of such a region, consider the following picture.



This is a representation of a small triangle obtained from two rays whose angles differ by only $d\theta$. What is the area of this triangle, dA? It would be

$$\frac{1}{2}\sin(d\theta)\, f(\theta)^2 \approx \frac{1}{2} f(\theta)^2\, d\theta = dA$$

with the approximation getting better as the angle gets smaller. Thus the area should solve the initial value problem,

$$\frac{dA}{d\theta} = \frac{1}{2} f(\theta)^2,\ A(a) = 0.$$

Therefore, the total area would be given by the integral,

$$\frac{1}{2}\int_a^b f(\theta)^2\, d\theta. \tag{10.2}$$

Example 10.2.1 *Find the area of the cardioid, $r = 1 + \cos\theta$ for $\theta \in [0,2\pi]$.*

From the graph of the cardioid presented earlier, you can see the region of interest satisfies the conditions above that every ray intersects the graph in only one point. Therefore, from 10.2 this area is

$$\frac{1}{2}\int_0^{2\pi} (1+\cos(\theta))^2\, d\theta = \frac{3}{2}\pi.$$

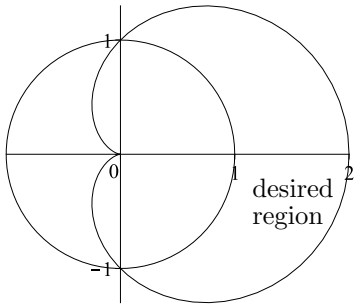
Example 10.2.2 *Verify the area of a circle of radius a is πa^2.*

The polar equation is just $r = a$ for $\theta \in [0,2\pi]$. Therefore, the area should be

$$\frac{1}{2}\int_0^{2\pi} a^2 d\theta = \pi a^2.$$

Example 10.2.3 *Find the area of the region inside the cardioid, $r = 1 + \cos\theta$ and outside the circle, $r = 1$ for $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.*

As is usual in such cases, it is a good idea to graph the curves involved to get an idea what is wanted.



The area of this region would be the area of the part of the cardioid corresponding to $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ minus the area of the part of the circle in the first quadrant. Thus the area is

$$\frac{1}{2}\int_{-\pi/2}^{\pi/2} (1+\cos(\theta))^2\, d\theta - \frac{1}{2}\int_{-\pi/2}^{\pi/2} 1 d\theta = \frac{1}{4}\pi + 2.$$

This example illustrates the following procedure for finding the area between the graphs of two curves given in polar coordinates.

Procedure 10.2.4 *Suppose that for all $\theta \in [a,b], 0 < g(\theta) < f(\theta)$. To find the area of the region defined in terms of polar coordinates by $g(\theta) < r < f(\theta)$, $\theta \in [a,b]$, you do the following.*

$$\frac{1}{2}\int_a^b \left(f(\theta)^2 - g(\theta)^2\right) d\theta.$$

10.3 Exercises

1. The following are the polar coordinates of points. Find the rectangular coordinates.

   (a) $\left(5, \frac{\pi}{6}\right)$
   (b) $\left(3, \frac{\pi}{3}\right)$
   (c) $\left(4, \frac{2\pi}{3}\right)$
   (d) $\left(2, \frac{3\pi}{4}\right)$
   (e) $\left(3, \frac{7\pi}{6}\right)$
   (f) $\left(8, \frac{11\pi}{6}\right)$

2. The following are the rectangular coordinates of points. Find the polar coordinates of these points.

   (a) $\left(\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2}\right)$
   (b) $\left(\frac{3}{2}, \frac{3}{2}\sqrt{3}\right)$
   (c) $\left(-\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2}\right)$
   (d) $\left(-\frac{5}{2}, \frac{5}{2}\sqrt{3}\right)$
   (e) $\left(-\sqrt{3}, -1\right)$
   (f) $\left(\frac{3}{2}, -\frac{3}{2}\sqrt{3}\right)$

3. In general it is a stupid idea to try to use algebra to invert and solve for a set of curvilinear coordinates such as polar or cylindrical coordinates in term of Cartesian coordinates. Not only is it often very difficult or even impossible to do it[1], but also it takes you in entirely the wrong direction because the whole point of introducing the new coordinates is to write everything in terms of these new coordinates and not in terms of Cartesian coordinates. However, sometimes this inversion can be done. Describe how to solve for r and θ in terms of x and y in polar coordinates.

4. Suppose $r = \frac{a}{1+e\sin\theta}$ where $e \in [0,1]$. By changing to rectangular coordinates, show this is either a parabola, an ellipse or a hyperbola. Determine the values of e which correspond to the various cases.

5. In Example 10.1.4 suppose you graphed it for $\theta \in [0, k\pi]$ where k is a positive integer. What is the smallest value of k such that the graph will look exactly like the one presented in the example?

6. Suppose you were to graph $r = 3 + \sin\left(\frac{m}{n}\theta\right)$ where m, n are integers. Can you give some description of what the graph will look like for $\theta \in [0, k\pi]$ for k a very large positive integer? How would things change if you did $r = 3 + \sin(\alpha\theta)$ where α is an irrational number?

7. Graph $r = 1 + \sin\theta$ for $\theta \in [0, 2\pi]$.

8. Graph $r = 2 + \sin\theta$ for $\theta \in [0, 2\pi]$.

9. Graph $r = 1 + 2\sin\theta$ for $\theta \in [0, 2\pi]$.

10. Graph $r = 2 + \sin(2\theta)$ for $\theta \in [0, 2\pi]$.

11. Graph $r = 1 + \sin(2\theta)$ for $\theta \in [0, 2\pi]$.

12. Graph $r = 1 + \sin(3\theta)$ for $\theta \in [0, 2\pi]$.

13. Find the area of the bounded region determined by $r = 1 + \sin(3\theta)$ for $\theta \in [0, 2\pi]$.

14. Find the area inside $r = 1 + \sin\theta$ and outside the circle $r = 1/2$.

15. Find the area inside the circle $r = 1/2$ and outside the region defined by $r = 1 + \sin\theta$.

10.4 Exercises With Answers

1. The following are the polar coordinates of points. Find the rectangular coordinates.

   (a) $\left(3, \frac{\pi}{6}\right)$ Rectangular coordinates: $\left(3\cos\left(\frac{\pi}{6}\right), 3\sin\left(\frac{\pi}{6}\right)\right) = \left(\frac{3}{2}\sqrt{3}, \frac{3}{2}\right)$

   (b) $\left(2, \frac{\pi}{3}\right)$Rectangular coordinates: $\left(2\cos\left(\frac{\pi}{3}\right), 2\sin\left(\frac{\pi}{3}\right)\right) = \left(1, \sqrt{3}\right)$

   (c) $\left(7, \frac{2\pi}{3}\right)$Rectangular coordinates: $\left(7\cos\left(\frac{2\pi}{3}\right), 7\sin\left(\frac{2\pi}{3}\right)\right) = \left(-\frac{7}{2}, \frac{7}{2}\sqrt{3}\right)$

   (d) $\left(6, \frac{3\pi}{4}\right)$Rectangular coordinates: $\left(6\cos\left(\frac{3\pi}{4}\right), 6\sin\left(\frac{3\pi}{4}\right)\right) = \left(-3\sqrt{2}, 3\sqrt{2}\right)$

2. The following are the rectangular coordinates of points. Find the polar coordinates of these points.

[1] It is no problem for these simple cases of curvilinear coordinates. However, it is a major difficulty in general. Algebra is simply not adequate to solve systems of nonlinear equations.

(a) $\left(5\sqrt{2}, 5\sqrt{2}\right)$Polar coordinates: $\theta = \pi/4$ because $\tan(\theta) = 1$.

$r = \sqrt{\left(5\sqrt{2}\right)^2 + \left(5\sqrt{2}\right)^2} = 10$

(b) $\left(3, 3\sqrt{3}\right)$Polar coordinates: $\theta = \pi/3$ because $\tan(\theta) = \sqrt{3}$.

$r = \sqrt{(3)^2 + \left(3\sqrt{3}\right)^2} = 6$

(c) $\left(-\sqrt{2}, \sqrt{2}\right)$Polar coordinates: $\theta = -\pi/4$ because $\tan(\theta) = -1$.

$r = \sqrt{\left(\sqrt{2}\right)^2 + \left(\sqrt{2}\right)^2} = 2$

(d) $\left(-3, 3\sqrt{3}\right)$Polar coordinates: $\theta = -\pi/3$ because $\tan(\theta) = -\sqrt{3}$.

$r = \sqrt{(3)^2 + \left(3\sqrt{3}\right)^2} = 6$

3. In general it is a stupid idea to try to use algebra to invert and solve for a set of curvilinear coordinates such as polar or cylindrical coordinates in term of Cartesian coordinates. Not only is it often very difficult or even impossible to do it[2], but also it takes you in entirely the wrong direction because the whole point of introducing the new coordinates is to write everything in terms of these new coordinates and not in terms of Cartesian coordinates. However, sometimes this inversion can be done. Describe how to solve for r and θ in terms of x and y in polar coordinates.

   This is what you were doing in the previous problem in special cases. If $x = r\cos\theta$ and $y = r\sin\theta$, then $\tan\theta = \frac{y}{x}$. This is how you can do it. You complete the solution. Tell how to find r. What do you do in case $x = 0$?

4. Suppose $r = \frac{a}{1+e\sin\theta}$ where $e \in [0,1]$. By changing to rectangular coordinates, show this is either a parabola, an ellipse or a hyperbola. Determine the values of e which correspond to the various cases.

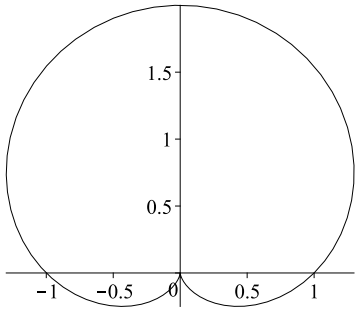
   Here is how you get started. $r + er\sin\theta = a$. Therefore, $\sqrt{x^2+y^2} + ey = a$ and so $\sqrt{x^2+y^2} = a - ey$.

5. Suppose you were to graph $r = 3 + \sin\left(\frac{m}{n}\theta\right)$ where m, n are integers. Can you give some description of what the graph will look like for $\theta \in [0, k\pi]$ for k a very large positive integer? How would things change if you did $r = 3 + \sin(\alpha\theta)$ where α is an irrational number?

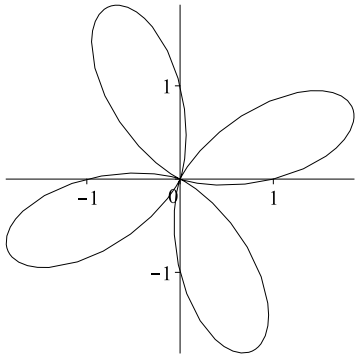
   The graph repeats when for two values of θ which differ by an integer multiple of 2π the corresponding values of r and r' also are equal. (Why?) Why isn't it enough to simply have the values of r equal? Thus you need $3+\sin(\theta\alpha) = 3+\sin((\theta + 2k\pi)\alpha)$ and $\cos(\theta\alpha) = \cos((\theta + 2k\pi)\alpha)$ for this to happen. The only way this can occur is for $(\theta + 2k\pi)\alpha - \theta\alpha$ to be a multiple of 2π. Why? However, this equals $2k\pi\alpha$ and if α is irrational, you can't have $k\alpha$ equal to an integer. Why? In the other case where $\alpha = \frac{m}{n}$ the graph will repeat.

6. Graph $r = 1 + \sin\theta$ for $\theta \in [0, 2\pi]$.

[2]It is no problem for these simple cases of curvilinear coordinates. However, it is a major difficulty in general. Algebra is simply not adequate to solve systems of nonlinear equations.



7. Find the area of the bounded region determined by $r = 1 + \sin(4\theta)$ for $\theta \in [0, 2\pi]$. First you should graph this thing to get an idea what is needed.



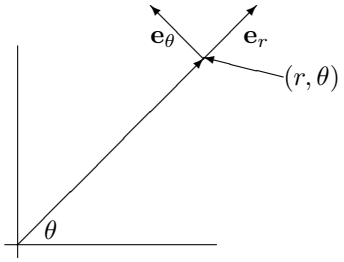
You see that you could simply take the area of one of the petals and then multiply by 4. To get the one which is mostly in the first quadrant, you should let θ go from $-\pi/8$ to $3\pi/8$. Thus the area of one petal is $\frac{1}{2}\int_{-\pi/8}^{3\pi/8}(1+\sin(4\theta))^2\,d\theta = \frac{3}{8}\pi$. Then you would need to multiply this by 4 to get the whole area. This gives $3\pi/2$. Alternatively, you could just do

$$\frac{1}{2}\int_0^{2\pi}(1+\sin(4\theta))^2\,d\theta = \frac{3}{2}\pi.$$

Be sure to always graph the polar function to be sure what you have in mind is appropriate. Sometimes, as indicated above, funny things happen with polar graphs.

10.5 The Acceleration In Polar Coordinates

Sometimes you have information about forces which act not in the direction of the coordinate axes but in some other direction. When this is the case, it is often useful to express things in terms of different coordinates which are consistent with these directions. A good example of this is the force exerted by the sun on a planet. This force is always directed toward the sun and so the force vector changes as the planet moves. To discuss this, consider the following simple diagram in which two unit vectors, $\mathbf{e}_r$ and $\mathbf{e}_\theta$ are shown.



The vector, $\mathbf{e}_r = (\cos\theta, \sin\theta)$ and the vector, $\mathbf{e}_\theta = (-\sin\theta, \cos\theta)$. You should convince yourself that the picture above corresponds to this definition of the two vectors. Note that $\mathbf{e}_r$ is a unit vector pointing away from $\mathbf{0}$ and

$$\mathbf{e}_\theta = \frac{d\mathbf{e}_r}{d\theta},\ \mathbf{e}_r = -\frac{d\mathbf{e}_\theta}{d\theta}. \tag{10.3}$$

Now consider the position vector from $\mathbf{0}$ of a point in the plane, $\mathbf{r}(t)$. Then

$$\mathbf{r}(t) = r(t)\,\mathbf{e}_r(\theta(t))$$

where $r(t) = |\mathbf{r}(t)|$. Thus $r(t)$ is just the distance from the origin, $\mathbf{0}$ to the point. What is the velocity and acceleration? Using the chain rule,

$$\frac{d\mathbf{e}_r}{dt} = \frac{d\mathbf{e}_r}{d\theta}\theta'(t),\ \frac{d\mathbf{e}_\theta}{dt} = \frac{d\mathbf{e}_\theta}{d\theta}\theta'(t)$$

and so from 10.3,

$$\frac{d\mathbf{e}_r}{dt} = \theta'(t)\,\mathbf{e}_\theta,\ \frac{d\mathbf{e}_\theta}{dt} = -\theta'(t)\,\mathbf{e}_r \tag{10.4}$$

Using 10.4 as needed along with the product rule and the chain rule,

$$\begin{aligned}\mathbf{r}'(t) &= r'(t)\,\mathbf{e}_r + r(t)\,\frac{d}{dt}(\mathbf{e}_r(\theta(t)))\\ &= r'(t)\,\mathbf{e}_r + r(t)\,\theta'(t)\,\mathbf{e}_\theta.\end{aligned}$$

Next consider the acceleration.

$$\begin{aligned}\mathbf{r}''(t) &= r''(t)\,\mathbf{e}_r + r'(t)\,\frac{d\mathbf{e}_r}{dt} + r'(t)\,\theta'(t)\,\mathbf{e}_\theta + r(t)\,\theta''(t)\,\mathbf{e}_\theta + r(t)\,\theta'(t)\,\frac{d}{dt}(\mathbf{e}_\theta)\\ &= r''(t)\,\mathbf{e}_r + 2r'(t)\,\theta'(t)\,\mathbf{e}_\theta + r(t)\,\theta''(t)\,\mathbf{e}_\theta + r(t)\,\theta'(t)\,(-\mathbf{e}_r)\,\theta'(t)\\ &= \left(r''(t) - r(t)\,\theta'(t)^2\right)\mathbf{e}_r + \left(2r'(t)\,\theta'(t) + r(t)\,\theta''(t)\right)\mathbf{e}_\theta.\end{aligned} \tag{10.5}$$

This is a very profound formula. Consider the following examples.

Example 10.5.1 *Suppose an object of mass m moves at a uniform speed, s, around a circle of radius R. Find the force acting on the object.*

By Newton's second law, the force acting on the object is $m\mathbf{r}''$. In this case, $r(t) = R$, a constant and since the speed is constant, $\theta'' = 0$. Therefore, the term in 10.5 corresponding to $\mathbf{e}_\theta$ equals zero and $m\mathbf{r}'' = -R\theta'(t)^2\,\mathbf{e}_r$. The speed of the object is s and so it moves s/R radians in unit time. Thus $\theta'(t) = s/R$ and so

$$m\mathbf{r}'' = -mR\left(\frac{s}{R}\right)^2\mathbf{e}_r = -m\frac{s^2}{R}\mathbf{e}_r.$$

This is the familiar formula for centripetal force from elementary physics, obtained as a very special case of 10.5.

Example 10.5.2 *A platform rotates at a constant speed in the counter clockwise direction and an object of mass m moves from the center of the platform toward the edge at constant speed. What forces act on this object?*

Let v denote the constant speed of the object moving toward the edge of the platform. Then

$$r'(t) = v, r''(t) = 0, \theta''(t) = 0,$$

while $\theta'(t) = \omega$, a positive constant. From 10.5

$$m\mathbf{r}''(t) = -mr(t)\omega^2\mathbf{e}_r + m2v\omega\mathbf{e}_\theta.$$

Thus the object experiences centripetal force from the first term and also a funny force from the second term which is in the direction of rotation of the platform. You can observe this by experiment if you like. Go to a playground and have someone spin one of those merry go rounds while you ride it and move from the center toward the edge. The term $2r'\theta'$ is called the Coriolis force.

Suppose at each point of space, $\mathbf{r}$ is associated a force, $\mathbf{F}(\mathbf{r})$ which a given object of mass m will experience if its position vector is $\mathbf{r}$. This is called a force field. a force field is a central force field if $\mathbf{F}(\mathbf{r}) = g(\mathbf{r})\mathbf{e}_r$. Thus in a central force field, the force an object experiences will always be directed toward or away from the origin, $\mathbf{0}$. The following simple lemma is very interesting because it says that in a central force field, objects must move in a plane.

Lemma 10.5.3 *Suppose an object moves in three dimensions in such a way that the only force acting on the object is a central force. Then the motion of the object is in a plane.*

Proof: Let $\mathbf{r}(t)$ denote the position vector of the object. Then from the definition of a central force and Newton's second law,

$$m\mathbf{r}'' = g(\mathbf{r})\mathbf{r}.$$

Therefore, $m\mathbf{r}'' \times \mathbf{r} = m(\mathbf{r}' \times \mathbf{r})' = g(\mathbf{r})\mathbf{r} \times \mathbf{r} = \mathbf{0}$. Therefore, $(\mathbf{r}' \times \mathbf{r}) = \mathbf{n}$, a constant vector and so $\mathbf{r} \cdot \mathbf{n} = \mathbf{r} \cdot (\mathbf{r}' \times \mathbf{r}) = 0$ showing that $\mathbf{n}$ is a normal vector to a plane which contains $\mathbf{r}(t)$ for all t. This proves the lemma.

10.6 Planetary Motion

Kepler's laws of planetary motion state that planets move around the sun along an ellipse, the equal area law described above holds, and there is a formula for the time it takes for the planet to move around the sun. These laws, discovered by Kepler, were shown by Newton to be consequences of his law of gravitation which states that the force acting on a mass, m by a mass, M is given by

$$\mathbf{F} = -GMm\left(\frac{1}{r^3}\right)\mathbf{r} = -GMm\left(\frac{1}{r^2}\right)\mathbf{e}_r$$

where r is the distance between centers of mass and $\mathbf{r}$ is the position vector from M to m. Here G is the gravitation constant. This is called an inverse square law. Gravity acts according to this law and so does electrostatic force. The constant, G, is very small when usual units are used and it has been computed using a very delicate experiment. It is now accepted to be

$$6.67 \times 10^{-11} \text{ Newton meter}^2/\text{kilogram}^2.$$

The experiment involved a light source shining on a mirror attached to a quartz fiber from which was suspended a long rod with two equal masses at the ends which were attracted by two larger masses. The gravitation force between the suspended masses and the two large masses caused the fibre to twist ever so slightly and this twisting was measured by observing the deflection of the light reflected from the mirror on a scale placed some distance from the fibre. The constant was first measured successfully by Lord Cavendish in 1798 and the present accepted value was obtained in 1942. Experiments like these are major accomplishments.

In the following argument, M is the mass of the sun and m is the mass of the planet. (It could also be a comet or an asteroid.)

10.6.1 The Equal Area Rule

An object moves in three dimensions in such a way that the only force acting on the object is a central force. Then the object moves in a plane and the radius vector from the origin to the object sweeps out area at a constant rate. This is the equal area rule. In the context of planetary motion it is called Kepler's second law.

Lemma 10.5.3 says the object moves in a plane. From the assumption that the force field is a central force field, it follows from 10.5 that

$$2r'(t)\,\theta'(t) + r(t)\,\theta''(t) = 0$$

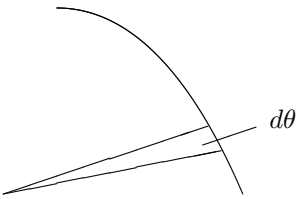
Multiply both sides of this equation by r. This yields

$$2rr'\theta' + r^2\theta'' = \left(r^2\theta'\right)' = 0. \tag{10.6}$$

Consequently,

$$r^2\theta' = c \tag{10.7}$$

for some constant, C. Now consider the following picture.



In this picture, $d\theta$ is the indicated angle and the two lines determining this angle are position vectors for the object at point t and point $t + dt$. The area of the sector, dA, is essentially $r^2 d\theta$ and so $dA = \frac{1}{2}r^2 d\theta$. Therefore,

$$\frac{dA}{dt} = \frac{1}{2}r^2\frac{d\theta}{dt} = \frac{c}{2}. \tag{10.8}$$

10.6.2 Inverse Square Law Motion, Kepler's First Law

Consider the first of Kepler's laws, the one which states that planets move along ellipses. From Lemma 10.5.3, the motion is in a plane. Now from 10.5 and Newton's second law,

$$\left(r''(t) - r(t)\,\theta'(t)^2\right)\mathbf{e}_r + \left(2r'(t)\,\theta'(t) + r(t)\,\theta''(t)\right)\mathbf{e}_\theta = -\frac{GMm}{m}\left(\frac{1}{r^2}\right)\mathbf{e}_r = -k\left(\frac{1}{r^2}\right)\mathbf{e}_r$$

Thus $k = GM$ and

$$r''(t) - r(t)\,\theta'(t)^2 = -k\left(\frac{1}{r^2}\right),\ 2r'(t)\,\theta'(t) + r(t)\,\theta''(t) = 0. \tag{10.9}$$

As in 10.6,$\left(r^2\theta'\right)' = 0$ and so there exists a constant, c, such that

$$r^2\theta' = c. \tag{10.10}$$

Now the other part of 10.9 and 10.10 implies

$$r''(t) - r(t)\,\theta'(t)^2 = r''(t) - r(t)\left(\frac{c^2}{r^4}\right) = -k\left(\frac{1}{r^2}\right). \tag{10.11}$$

It is only r as a function of θ which is of interest. Using the chain rule,

$$r' = \frac{dr}{d\theta}\frac{d\theta}{dt} = \frac{dr}{d\theta}\left(\frac{c}{r^2}\right) \tag{10.12}$$

and so also

$$\begin{aligned} r'' &= \frac{d^2r}{d\theta^2}\left(\frac{d\theta}{dt}\right)\left(\frac{c}{r^2}\right) + \frac{dr}{d\theta}(-2)(c)\left(r^{-3}\right)\frac{dr}{d\theta}\frac{d\theta}{dt} \\ &= \frac{d^2r}{d\theta^2}\left(\frac{c}{r^2}\right)^2 - 2\left(\frac{dr}{d\theta}\right)^2\left(\frac{c^2}{r^5}\right) \end{aligned} \tag{10.13}$$

Using 10.13 and 10.12 in 10.11 yields

$$\frac{d^2r}{d\theta^2}\left(\frac{c}{r^2}\right)^2 - 2\left(\frac{dr}{d\theta}\right)^2\left(\frac{c^2}{r^5}\right) - r(t)\left(\frac{c^2}{r^4}\right) = -k\left(\frac{1}{r^2}\right).$$

Now multiply both sides of this equation by r^4/c^2 to obtain

$$\frac{d^2r}{d\theta^2} - 2\left(\frac{dr}{d\theta}\right)^2\frac{1}{r} - r = \frac{-kr^2}{c^2}. \tag{10.14}$$

This is a nice differential equation for r as a function of θ but it is not clear what its solution is. It turns out to be convenient to define a new dependent variable, $\rho \equiv r^{-1}$ so $r = \rho^{-1}$. Then

$$\frac{dr}{d\theta} = (-1)\,\rho^{-2}\frac{d\rho}{d\theta},\ \frac{d^2r}{d\theta^2} = 2\rho^{-3}\left(\frac{d\rho}{d\theta}\right)^2 + (-1)\,\rho^{-2}\frac{d^2\rho}{d\theta^2}.$$

Substituting this in to 10.14 yields

$$2\rho^{-3}\left(\frac{d\rho}{d\theta}\right)^2 + (-1)\,\rho^{-2}\frac{d^2\rho}{d\theta^2} - 2\left(\rho^{-2}\frac{d\rho}{d\theta}\right)^2\rho - \rho^{-1} = \frac{-k\rho^{-2}}{c^2}.$$

which simplifies to

$$(-1)\,\rho^{-2}\frac{d^2\rho}{d\theta^2} - \rho^{-1} = \frac{-k\rho^{-2}}{c^2}$$

since those two terms which involve $\left(\frac{d\rho}{d\theta}\right)^2$ cancel. Now multiply both sides by $-\rho^2$ and this yields

$$\frac{d^2\rho}{d\theta^2} + \rho = \frac{k}{c^2}, \tag{10.15}$$

which is a much nicer differential equation. Let $R = \rho - \frac{k}{c^2}$. Then in terms of R, this differential equation is

$$\frac{d^2R}{d\theta^2} + R = 0.$$

Multiply both sides by $\frac{dR}{d\theta}$.

$$\frac{1}{2}\frac{d}{d\theta}\left(\left(\frac{dR}{d\theta}\right)^2 + R^2\right) = 0$$

and so

$$\left(\frac{dR}{d\theta}\right)^2 + R^2 = \delta^2 \tag{10.16}$$

for some $\delta > 0$. Therefore, there exists an angle, $\psi = \psi(\theta)$ such that

$$R = \delta\sin(\psi), \frac{dR}{d\theta} = \delta\cos(\psi)$$

because 10.16 says $\left(\frac{1}{\delta}\frac{dR}{d\theta}, \frac{1}{\delta}R\right)$ is a point on the unit circle. But differentiating, the first of the above equations,

$$\frac{dR}{d\theta} = \delta\cos(\psi)\frac{d\psi}{d\theta} = \delta\cos(\psi)$$

and so $\frac{d\psi}{d\theta} = 1$. Therefore, $\psi = \theta + \phi$. Choosing the coordinate system appropriately, you can assume $\phi = 0$. Therefore,

$$R = \rho - \frac{k}{c^2} = \frac{1}{r} - \frac{k}{c^2} = \delta\sin(\theta)$$

and so, solving for r,

$$\begin{aligned} r &= \frac{1}{\left(\frac{k}{c^2}\right) + \delta\sin\theta} = \frac{c^2/k}{1 + (c^2/k)\,\delta\sin\theta} \\ &= \frac{p\varepsilon}{1 + \varepsilon\sin\theta} \end{aligned}$$

where

$$\varepsilon = \left(c^2/k\right)\delta \text{ and } p = c^2/k\varepsilon. \tag{10.17}$$

Here all these constants are nonnegative.

Thus

$$r + \varepsilon r\sin\theta = \varepsilon p$$

and so $r = (\varepsilon p - \varepsilon y)$. Then squaring both sides,

$$x^2 + y^2 = (\varepsilon p - \varepsilon y)^2 = \varepsilon^2 p^2 - 2p\varepsilon^2 y + \varepsilon^2 y^2$$

And so

$$x^2 + \left(1 - \varepsilon^2\right)y^2 = \varepsilon^2 p^2 - 2p\varepsilon^2 y. \tag{10.18}$$

In case $\varepsilon = 1$, this reduces to the equation of a parabola. If $\varepsilon < 1$, this reduces to the equation of an ellipse and if $\varepsilon > 1$, this is called a hyperbola. This proves that objects which are acted on only by a force of the form given in the above example move along hyperbolas, ellipses or circles. The case where $\varepsilon = 0$ corresponds to a circle. The constant, ε is called the eccentricity. This is called Kepler's first law in the case of a planet.

10.6.3 Kepler's Third Law

Kepler's third law involves the time it takes for the planet to orbit the sun. From 10.18 you can complete the square and obtain

$$x^2 + \left(1-\varepsilon^2\right)\left(y + \frac{p\varepsilon^2}{1-\varepsilon^2}\right)^2 = \varepsilon^2 p^2 + \frac{p^2\varepsilon^4}{\left(1-\varepsilon^2\right)} = \frac{\varepsilon^2 p^2}{\left(1-\varepsilon^2\right)},$$

and this yields

$$x^2 / \left(\frac{\varepsilon^2 p^2}{1-\varepsilon^2}\right) + \left(y + \frac{p\varepsilon^2}{1-\varepsilon^2}\right)^2 / \left(\frac{\varepsilon^2 p^2}{\left(1-\varepsilon^2\right)^2}\right) = 1. \tag{10.19}$$

Now note this is the equation of an ellipse and that the diameter of this ellipse is

$$\frac{2\varepsilon p}{\left(1-\varepsilon^2\right)} \equiv 2a. \tag{10.20}$$

This follows because

$$\frac{\varepsilon^2 p^2}{\left(1-\varepsilon^2\right)^2} \geq \frac{\varepsilon^2 p^2}{1-\varepsilon^2}.$$

Now let T denote the time it takes for the planet to make one revolution about the sun. Using this formula, and 10.8 the following equation must hold.

$$\overbrace{\pi \frac{\varepsilon p}{\sqrt{1-\varepsilon^2}} \frac{\varepsilon p}{\left(1-\varepsilon^2\right)}}^{\text{area of ellipse}} = T\frac{c}{2}$$

Therefore,

$$T = \frac{2}{c}\frac{\pi\varepsilon^2 p^2}{\left(1-\varepsilon^2\right)^{3/2}}$$

and so

$$T^2 = \frac{4\pi^2\varepsilon^4 p^4}{c^2\left(1-\varepsilon^2\right)^3}$$

Now using 10.17, recalling that $k = GM$, and 10.20,

$$\begin{aligned} T^2 &= \frac{4\pi^2\varepsilon^4 p^4}{k\varepsilon p\left(1-\varepsilon^2\right)^3} = \frac{4\pi^2\left(\varepsilon p\right)^3}{k\left(1-\varepsilon^2\right)^3} \\ &= \frac{4\pi^2 a^3}{k} = \frac{4\pi^2 a^3}{GM}. \end{aligned}$$

Written more memorably, this has shown

$$T^2 = \frac{4\pi^2}{GM}\left(\frac{\text{diameter of ellipse}}{2}\right)^3. \tag{10.21}$$

This relationship is known as Kepler's third law.

10.7 Exercises

1. Suppose you know how the spherical coordinates of a moving point change as a function of t. Can you figure out the velocity of the point? Specifically, suppose $\phi(t) = t, \theta(t) = 1+t$, and $\rho(t) = t$. Find the speed and the velocity of the object in terms of Cartesian coordinates. **Hint:** You would need to find $x'(t), y'(t)$, and $z'(t)$. Then in terms of Cartesian coordinates, the velocity would be $x'(t)\,\mathbf{i} + y'(t)\,\mathbf{j} + z'(t)\,\mathbf{k}$.

2. Find the length of the cardioid, $r = 1+\cos\theta, \theta \in [0, 2\pi]$. **Hint:** A parameterization is $x(\theta) = (1+\cos\theta)\cos\theta, y(\theta) = (1+\cos\theta)\sin\theta$.

3. In general, show the length of the curve given in polar coordinates by $r = f(\theta), \theta \in [a, b]$ equals $\int_a^b \sqrt{f'(\theta)^2 + f(\theta)^2}d\theta$.

4. Suppose the curve given in polar coordinates by $r = f(\theta)$ for $\theta \in [a, b]$ is rotated about the y axis. Find a formula for the resulting surface of revolution.

5. Suppose an object moves in such a way that $r^2\theta'$ is a constant. Show the only force acting on the object is a central force.

6. Explain why low pressure areas rotate counter clockwise in the Northern hemisphere and clockwise in the Southern hemisphere. **Hint:** Note that from the point of view of an observer fixed in space above the North pole, the low pressure area already has a counter clockwise rotation because of the rotation of the earth and its spherical shape. Now consider 10.7. In the low pressure area stuff will move toward the center so r gets smaller. How are things different in the Southern hemisphere?

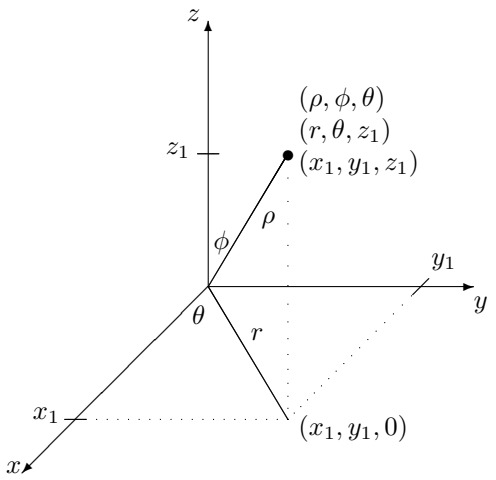
7. What are some physical assumptions which are made in the above derivation of Kepler's laws from Newton's laws of motion?

8. The orbit of the earth is pretty nearly circular and the distance from the sun to the earth is about 149×10^6 kilometers. Using 10.21 and the above value of the universal gravitation constant, determine the mass of the sun. The earth goes around it in 365 days. (Actually it is 365.256 days.)

9. It is desired to place a satellite above the equator of the earth which will rotate about the center of mass of the earth every 24 hours. Is it necessary that the orbit be circular? What if you want the satellite to stay above the same point on the earth at all times? If the orbit is to be circular and the satellite is to stay above the same point, at what distance from the center of mass of the earth should the satellite be? You may use that the mass of the earth is 5.98×10^{24} kilograms. Such a satellite is called geosynchronous.

10.8 Spherical And Cylindrical Coordinates

Now consider two three dimensional generalizations of polar coordinates. The following picture serves as motivation for the definition of these two other coordinate systems.



In this picture, ρ is the distance between the origin, the point whose Cartesian coordinates are $(0,0,0)$ and the point indicated by a dot and labelled as (x_1, y_1, z_1), (r, θ, z_1), and (ρ, ϕ, θ). The angle between the positive z axis and the line between the origin and the point indicated by a dot is denoted by ϕ, and θ, is the angle between the positive x axis and the line joining the origin to the point $(x_1, y_1, 0)$ as shown, while r is the length of this line. Thus r and θ determine a point in the plane determined by letting $z = 0$ and r and θ are the usual polar coordinates. Thus $r \geq 0$ and $\theta \in [0, 2\pi)$. Letting z_1 denote the usual z coordinate of a point in three dimensions, like the one shown as a dot, (r, θ, z_1) are the cylindrical coordinates of the dotted point. The spherical coordinates are determined by (ρ, ϕ, θ). When ρ is specified, this indicates that the point of interest is on some sphere of radius ρ which is centered at the origin. Then when ϕ is given, the location of the point is narrowed down to a circle and finally, θ determines which point is on this circle. Let $\phi \in [0, \pi], \theta \in [0, 2\pi)$, and $\rho \in [0, \infty)$. The picture shows how to relate these new coordinate systems to Cartesian coordinates. For Cylindrical coordinates,

$$\begin{aligned} x &= r\cos(\theta), \\ y &= r\sin(\theta), \\ z &= z \end{aligned}$$

and for spherical coordinates,

$$\begin{aligned} x &= \rho\sin(\phi)\cos(\theta), \\ y &= \rho\sin(\phi)\sin(\theta), \\ z &= \rho\cos(\phi). \end{aligned}$$

Spherical coordinates should be especially interesting to you because you live on the surface of a sphere. This has been known for several hundred years. You may also know that the standard way to determine position on the earth is to give the longitude and latitude. The latitude corresponds to ϕ and the longitude corresponds to θ.[3]

[3] Actually latitude is determined on maps and in navigation by measuring the angle from the equator rather than the pole but it is essentially the same idea.

Example 10.8.1 *Express the surface,* $z = \frac{1}{\sqrt{3}}\sqrt{x^2+y^2}$ *in spherical coordinates.*

This is

$$\rho\cos(\phi) = \frac{1}{\sqrt{3}}\sqrt{(\rho\sin(\phi)\cos(\theta))^2 + (\rho\sin(\phi)\sin(\theta))^2} = \frac{1}{3}\sqrt{3}\rho\sin\phi.$$

Therefore, this reduces to

$$\tan\phi = \sqrt{3}$$

and so this is just $\phi = \pi/3$.

Example 10.8.2 *Express the surface,* $y = x$ *in terms of spherical coordinates.*

This says $\rho\sin(\phi)\sin(\theta) = \rho\sin(\phi)\cos(\theta)$. Thus $\sin\theta = \cos\theta$. You could also write $\tan\theta = 1$.

Example 10.8.3 *Express the surface,* $x^2 + y^2 = 4$ *in cylindrical coordinates.*

This says $r^2\cos^2\theta + r^2\sin^2\theta = 4$. Thus $r = 2$.

10.9 Exercises

1. The following are the cylindrical coordinates of points. Find the rectangular and spherical coordinates.
   (a) $\left(5, \frac{5\pi}{6}, -3\right)$
   (b) $\left(3, \frac{\pi}{3}, 4\right)$
   (c) $\left(4, \frac{2\pi}{3}, 1\right)$
   (d) $\left(2, \frac{3\pi}{4}, -2\right)$
   (e) $\left(3, \frac{3\pi}{2}, -1\right)$
   (f) $\left(8, \frac{11\pi}{6}, -11\right)$

2. The following are the rectangular coordinates of points. Find the cylindrical and spherical coordinates of these points.
   (a) $\left(\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2}, -3\right)$
   (b) $\left(\frac{3}{2}, \frac{3}{2}\sqrt{3}, 2\right)$
   (c) $\left(-\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2}, 11\right)$
   (d) $\left(-\frac{5}{2}, \frac{5}{2}\sqrt{3}, 23\right)$
   (e) $\left(-\sqrt{3}, -1, -5\right)$
   (f) $\left(\frac{3}{2}, -\frac{3}{2}\sqrt{3}, -7\right)$

3. The following are spherical coordinates of points in the form (ρ, ϕ, θ). Find the rectangular and cylindrical coordinates.
   (a) $\left(4, \frac{\pi}{4}, \frac{5\pi}{6}\right)$
   (b) $\left(2, \frac{\pi}{3}, \frac{2\pi}{3}\right)$
   (c) $\left(3, \frac{5\pi}{6}, \frac{3\pi}{2}\right)$
   (d) $\left(4, \frac{\pi}{2}, \frac{7\pi}{4}\right)$

(e) $\left(4, \frac{2\pi}{3}, \frac{\pi}{6}\right)$
(f) $\left(4, \frac{3\pi}{4}, \frac{5\pi}{3}\right)$

4. The following are rectangular coordinates of points. Find the spherical and cylindrical coordinates.

(a) $\left(\sqrt{2}, \sqrt{6}, 2\sqrt{2}\right)$
(b) $\left(-\frac{1}{2}\sqrt{3}, \frac{3}{2}, 1\right)$
(c) $\left(-\frac{3}{4}\sqrt{2}, \frac{3}{4}\sqrt{2}, -\frac{3}{2}\sqrt{3}\right)$
(d) $\left(-\sqrt{3}, 1, 2\sqrt{3}\right)$
(e) $\left(-\frac{1}{4}\sqrt{2}, \frac{1}{4}\sqrt{6}, -\frac{1}{2}\sqrt{2}\right)$
(f) $\left(-\frac{9}{4}\sqrt{3}, \frac{27}{4}, -\frac{9}{2}\right)$

5. Describe how to solve the problem of finding spherical coordinates given rectangular coordinates.

6. A point has Cartesian coordinates, $(1, 2, 3)$. Find its spherical and cylindrical coordinates using a calculator or other electronic gadget.

7. Describe the following surface in rectangular coordinates. $\phi = \pi/4$ where ϕ is the polar angle in spherical coordinates.

8. Describe the following surface in rectangular coordinates. $\theta = \pi/4$ where θ is the angle measured from the postive x axis spherical coordinates.

9. Describe the following surface in rectangular coordinates. $\theta = \pi/4$ where θ is the angle measured from the postive x axis cylindrical coordinates.

10. Describe the following surface in rectangular coordinates. $r = 5$ where r is one of the cylindrical coordinates.

11. Describe the following surface in rectangular coordinates. $\rho = 4$ where ρ is the distance to the origin.

12. Give the cone, $z = \sqrt{x^2 + y^2}$ in cylindrical coordinates and in spherical coordinates.

13. Write the following in spherical coordinates.

(a) $z = x^2 + y^2$.
(b) $x^2 - y^2 = 1$
(c) $z^2 + x^2 + y^2 = 6$
(d) $z = \sqrt{x^2 + y^2}$
(e) $y = x$
(f) $z = x$

14. Write the following in cylindrical coordinates.

(a) $z = x^2 + y^2$.
(b) $x^2 - y^2 = 1$
(c) $z^2 + x^2 + y^2 = 6$
(d) $z = \sqrt{x^2 + y^2}$
(e) $y = x$
(f) $z = x$

10.10 Exercises With Answers

1. The following are the cylindrical coordinates of points. Find the rectangular and spherical coordinates.

   (a) $\left(5, \frac{5\pi}{3}, -3\right)$ Rectangular coordinates:

   $$\left(5\cos\left(\frac{5\pi}{6}\right), 5\sin\left(\frac{5\pi}{3}\right), -3\right) = \left(-\frac{5}{2}\sqrt{3}, -\frac{5}{2}\sqrt{3}, -3\right)$$

   (b) $\left(3, \frac{\pi}{2}, 4\right)$Rectangular coordinates:

   $$\left(3\cos\left(\frac{\pi}{2}\right), 3\sin\left(\frac{\pi}{2}\right), 4\right) = (0, 3, 4)$$

   (c) $\left(4, \frac{3\pi}{4}, 1\right)$Rectangular coordinates:

   $$\left(4\cos\left(\frac{3\pi}{4}\right), 4\sin\left(\frac{3\pi}{4}\right), 1\right) = \left(-2\sqrt{2}, 2\sqrt{2}, 1\right)$$

2. The following are the rectangular coordinates of points. Find the cylindrical and spherical coordinates of these points.

   (a) $\left(\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2}, -3\right)$Cylindrical coordinates:

   $$\left(\sqrt{\left(\frac{5}{2}\sqrt{2}\right)^2 + \left(\frac{5}{2}\sqrt{2}\right)^2}, \frac{\pi}{4}, -3\right) = \left(5, \frac{1}{4}\pi, -3\right)$$

   Spherical coordinates: $\left(\sqrt{34}, \frac{\pi}{4}, \phi\right)$ where $\cos\phi = \frac{-3}{\sqrt{34}}$

   (b) $\left(1, \sqrt{3}, 2\right)$ Cylindrical coordinates:

   $$\left(\sqrt{(1)^2 + \left(\sqrt{3}\right)^2}, \frac{\pi}{3}, 2\right) = \left(2, \frac{1}{3}\pi, 2\right)$$

   Spherical coordinates: $\left(2\sqrt{2}, \frac{\pi}{4}, \phi\right)$ where $\cos\phi = \frac{2}{2\sqrt{2}}$ so $\phi = \frac{\pi}{4}$.

3. The following are spherical coordinates of points in the form (ρ, ϕ, θ). Find the rectangular and cylindrical coordinates.

   (a) $\left(4, \frac{\pi}{4}, \frac{5\pi}{6}\right)$ Rectangular coordinates:

   $$\left(4\sin\left(\frac{\pi}{4}\right)\cos\left(\frac{5\pi}{6}\right), 4\sin\left(\frac{\pi}{4}\right)\sin\left(\frac{5\pi}{6}\right), 4\cos\left(\frac{\pi}{4}\right)\right) = \left(-\sqrt{2}\sqrt{3}, \sqrt{2}, 2\sqrt{2}\right)$$

   Cylindrical coordinates: $\left(4\sin\left(\frac{\pi}{4}\right), \frac{5\pi}{6}, 4\cos\left(\frac{\pi}{4}\right)\right) = \left(2\sqrt{2}, \frac{5}{6}\pi, 2\sqrt{2}\right)$.

   (b) $\left(2, \frac{\pi}{3}, \frac{3\pi}{4}\right)$ Rectangular coordinates:

   $$\left(2\sin\left(\frac{\pi}{3}\right)\cos\left(\frac{5\pi}{6}\right), 2\sin\left(\frac{\pi}{3}\right)\sin\left(\frac{5\pi}{6}\right), 2\cos\left(\frac{\pi}{3}\right)\right) = \left(-\frac{3}{2}, \frac{1}{2}\sqrt{3}, 1\right)$$

   Cylindrical coordinates: $\left(2\sin\left(\frac{\pi}{3}\right), \frac{3\pi}{4}, 2\cos\left(\frac{\pi}{3}\right)\right) = \left(\sqrt{3}, \frac{3}{4}\pi, 1\right)$.

(c) $\left(2, \frac{\pi}{6}, \frac{3\pi}{2}\right)$ Rectangular coordinates:

$$\left(2\sin\left(\frac{\pi}{6}\right)\cos\left(\frac{3\pi}{2}\right), 2\sin\left(\frac{\pi}{6}\right)\sin\left(\frac{3\pi}{2}\right), 2\cos\left(\frac{\pi}{6}\right)\right) = \left(0, -1, \sqrt{3}\right)$$

Cylindrical coordinates: $\left(2\sin\left(\frac{\pi}{6}\right), \frac{3\pi}{2}, 2\cos\left(\frac{\pi}{6}\right)\right) = \left(1, \frac{3}{2}\pi, \sqrt{3}\right)$.

4. The following are rectangular coordinates of points. Find the spherical and cylindrical coordinates.

   (a) $\left(\sqrt{2}, \sqrt{6}, 2\sqrt{2}\right)$ To find θ, note that $\tan\theta = \frac{\sqrt{6}}{\sqrt{2}} = \sqrt{3}$ and so $\theta = \frac{\pi}{3}$. $\rho = \sqrt{2+6+8} = 4$. $\cos\phi = \frac{2\sqrt{2}}{4} = \frac{\sqrt{2}}{2}$ so $\phi = \frac{\pi}{4}$. The spherical coordinates are therefore, $\left(4, \frac{\pi}{4}, \frac{\pi}{3}\right)$. The cylindrical coordinates are $\left(4\sin\left(\frac{\pi}{4}\right), \frac{\pi}{3}, 4\cos\left(\frac{\pi}{4}\right)\right) = \left(2\sqrt{2}, \frac{1}{3}\pi, 2\sqrt{2}\right)$. I can't stand to do any more of these but you can do the others the same way.

5. Describe how to solve the problem of finding spherical coordinates given rectangular coordinates.

   This is not easy and is somewhat unpleasant but everyone should do this once in their life. If x, y, z are the rectangular coordinates, you can get ρ as $\sqrt{x^2+y^2+z^2}$. Now $\cos\phi = \frac{z}{\rho}$. Finally, you need θ. You know ϕ and ρ. $x = \rho\sin\phi\cos\theta$ and $y = \rho\sin\phi\sin\theta$. Therefore, you can find θ in the same way you did for polar coordinates. Here $r = \rho\sin\phi$.

6. A point has Cartesian coordinates, $(1, 2, 3)$. Find its spherical and cylindrical coordinates using a calculator or other electronic gadget.

   See how to do it using Problem 5.

7. Describe the following surface in rectangular coordinates. $\phi = \pi/3$ where ϕ is the polar angle in spherical coordinates.

   This is a cone such that the angle between the positive z axis and the side of the cone seen from the side equals $\pi/3$.

8. Give the cone, $z = 2\sqrt{x^2+y^2}$ in cylindrical coordinates and in spherical coordinates.

   Cylindrical: $z = 2r$ Spherical: $\rho\cos\phi = 2\rho\sin\phi$. So it is $\tan\phi = \frac{1}{2}$.

9. Write the following in spherical coordinates.

   (a) $z = 2\left(x^2+y^2\right)$.
   $\rho\cos\phi = 2\rho^2\sin^2\phi$ or in other words $\cos\phi = 2\rho\sin^2\phi$

   (b) $x^2 - y^2 = 1$ $\left(\rho\sin\phi\cos\theta\right)^2 - \left(\rho\sin\phi\sin\theta\right)^2 = \rho^2\sin^2\phi\cos 2\theta = 1$

10. Write the following in cylindrical coordinates.

    (a) $z = x^2 + y^2$. $z = r^2$

    (b) $x^2 - y^2 = 1$ $r^2\cos^2\theta - r^2\sin^2\theta = r^2\cos 2\theta = 1$

Part IV

Vector Calculus In Many Variables

Functions Of Many Variables

11.0.1 Outcomes

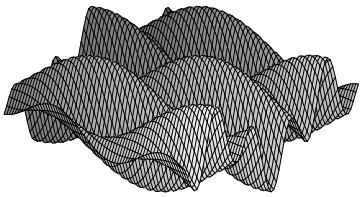
1. Represent a function of two variables by level curves.
2. Identify the characteristics of a function from a graph of its level curves.
3. Recall and use the concept of limit point.
4. Describe the geometrical significance of a directional derivative.
5. Give the relationship between partial derivatives and directional derivatives.
6. Compute partial derivatives and directional derivatives from their definitions.
7. Evaluate higher order partial derivatives.
8. State conditions under which mixed partial derivatives are equal.
9. Verify equations involving partial derivatives.
10. Describe the gradient of a scalar valued function and use to compute the directional derivative.
11. Explain why the directional derivative is maximized in the direction of the gradient and minimized in the direction of minus the gradient.

11.1 The Graph Of A Function Of Two Variables

With vector valued functions of many variables, it doesn't take long before it is impossible to draw meaningful pictures. This is because one needs more than three dimensions to accomplish the task and we can only visualize things in three dimensions. Ultimately, one of the main purposes of calculus is to free us from the tyranny of art. In calculus, we are permitted and even required to think in a meaningful way about things which cannot be drawn. However, it is certainly interesting to consider some things which can be visualized and this will help to formulate and understand more general notions which make sense in contexts which cannot be visualized. One of these is the concept of a scalar valued function of two variables.

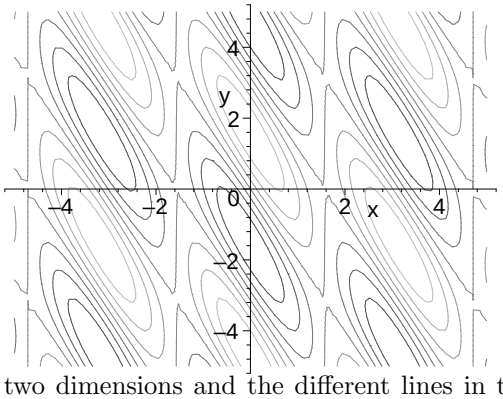
Let $f(x,y)$ denote a scalar valued function of two variables evaluated at the point (x,y). Its graph consists of the set of points, (x,y,z) such that $z = f(x,y)$. How does one go about depicting such a graph? The usual way is to fix one of the variables, say x and consider the function $z = f(x,y)$ where y is allowed to vary and x is fixed. Graphing this would give a curve which lies in the surface to be depicted. Then do the same thing for other values of x and the result would depict the graph desired graph. Computers

do this very well. The following is the graph of the function $z = \cos(x)\sin(2x+y)$ drawn using Maple, a computer algebra system.[1].



Notice how elaborate this picture is. The lines in the drawing correspond to taking one of the variables constant and graphing the curve which results. The computer did this drawing in seconds but you couldn't do it as well if you spent all day on it. I used a grid consisting of 70 choices for x and 70 choices for y.

Sometimes attempts are made to understand three dimensional objects like the above graph by looking at contour graphs in two dimensions. The contour graph of the above three dimensional graph is below and comes from using the computer algebra system again.



This is in two dimensions and the different lines in two dimensions correspond to points on the three dimensional graph which have the same z value. If you have looked at a weather map, these lines are called isotherms or isobars depending on whether the function involved is temperature or pressure. In a contour geographic map, the contour lines represent constant altitude. If many contour lines are close to each other, this indicates rapid change in the altitude, temperature, pressure, or whatever else may be measured.

A scalar function of three variables, cannot be visualized because four dimensions are required. However, some people like to try and visualize even these examples. This is done by looking at level surfaces in $\mathbb{R}^3$ which are defined as surfaces where the function assumes a constant value. They play the role of contour lines for a function of two variables. As a simple example, consider $f(x,y,z) = x^2+y^2+z^2$. The level surfaces of this function would be concentric spheres centered at $\mathbf{0}$. (Why?) Another way to visualize objects in higher dimensions involves the use of color and animation. However, there really are limits to what you can accomplish in this direction. So much for art.

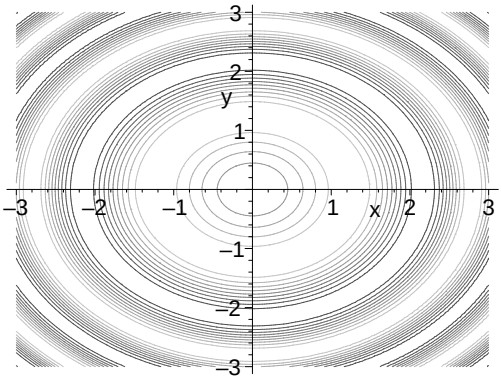
However, the concept of level curves is quite useful because these can be drawn.

Example 11.1.1 *Determine from a contour map where the function,*

$$f(x,y) = \sin\left(x^2+y^2\right)$$

[1]I used Maple and exported the graph as an eps. file which I then imported into this document.

is steepest.



In the picture, the steepest places are where the contour lines are close together because they correspond to various values of the function. You can look at the picture and see where they are close and where they are far. This is the advantage of a contour map.

11.2 Review Of Limits

Recall the concept of limit of a function of many variables. When $\mathbf{f} : D(\mathbf{f}) \to \mathbb{R}^q$ one can only consider in a meaningful way limits at limit points of the set, $D(\mathbf{f})$.

Definition 11.2.1 *Let A denote a nonempty subset of $\mathbb{R}^p$. A point, $\mathbf{x}$ is said to be a **limit point** of the set, A if for every $r > 0, B(\mathbf{x}, r)$ contains infinitely many points of A.*

Example 11.2.2 *Let S denote the set, $\{(x, y, z) \in \mathbb{R}^3 : x, y, z \text{ are all in } \mathbb{N}\}$. Which points are limit points?*

This set does not have any because any two of these points are at least as far apart as 1. Therefore, if $\mathbf{x}$ is any point of $\mathbb{R}^3, B(\mathbf{x}, 1/4)$ contains at most one point.

Example 11.2.3 *Let U be an open set in $\mathbb{R}^3$. Which points of U are limit points of U?*

They all are. From the definition of U being open, if $\mathbf{x} \in U$, There exists $B(\mathbf{x}, r) \subseteq U$ for some $r > 0$. Now consider the line segment $\mathbf{x} + tr\mathbf{e}_1$ where $t \in [0, 1/2]$. This describes infinitely many points and they are all in $B(\mathbf{x}, r)$ because

$$|\mathbf{x} + tr\mathbf{e}_1 - \mathbf{x}| = tr < r.$$

Therefore, every point of U is a limit point of U.

The case where U is open will be the one of most interest but many other sets have limit points.

Definition 11.2.4 *Let $\mathbf{f} : D(\mathbf{f}) \subseteq \mathbb{R}^p \to \mathbb{R}^q$ where $q, p \geq 1$ be a function and let $\mathbf{x}$ be a limit point of $D(\mathbf{f})$. Then*

$$\lim_{\mathbf{y} \to \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$$

if and only if the following condition holds. For all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < |\mathbf{y} - \mathbf{x}| < \delta \text{ and } \mathbf{y} \in D(\mathbf{f})$$

then,

$$|\mathbf{L} - \mathbf{f}(\mathbf{y})| < \varepsilon.$$

The condition that $\mathbf{x}$ must be a limit point of $D(\mathbf{f})$ if you are to take a limit at $\mathbf{x}$ is what makes the limit well defined.

Proposition 11.2.5 *Let $\mathbf{f} : D(\mathbf{f}) \subseteq \mathbb{R}^p \to \mathbb{R}^q$ where $q, p \geq 1$ be a function and let $\mathbf{x}$ be a limit point of $D(\mathbf{f})$. Then if $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y})$ exists, it must be unique.*

Proof: Suppose $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}_1$ and $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}_2$. Then for $\varepsilon > 0$ given, let $\delta_i > 0$ correspond to $\mathbf{L}_i$ in the definition of the limit and let $\delta = \min(\delta_1, \delta_2)$. Since $\mathbf{x}$ is a limit point, there exists $\mathbf{y} \in B(\mathbf{x}, \delta) \cap D(\mathbf{f})$. Therefore,

$$\begin{aligned} |\mathbf{L}_1 - \mathbf{L}_2| &\leq |\mathbf{L}_1 - \mathbf{f}(\mathbf{y})| + |\mathbf{f}(\mathbf{y}) - \mathbf{L}_2| \\ &< \varepsilon + \varepsilon = 2\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this shows $\mathbf{L}_1 = \mathbf{L}_2$. The following theorem summarized many important interactions involving continuity. Most of this theorem has been proved in Theorem 7.4.5 on Page 137 and Theorem 7.4.7 on Page 139.

Theorem 11.2.6 *Suppose $\mathbf{x}$ is a limit point of $D(\mathbf{f})$ and $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$, $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{g}(\mathbf{y}) = \mathbf{K}$ where $\mathbf{K}$ and $\mathbf{L}$ are vectors in $\mathbb{R}^p$ for $p \geq 1$. Then if $a, b \in \mathbb{R}$,*

$$\lim_{\mathbf{y}\to\mathbf{x}} a\mathbf{f}(\mathbf{y}) + b\mathbf{g}(\mathbf{y}) = a\mathbf{L} + b\mathbf{K}, \tag{11.1}$$

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f} \cdot \mathbf{g}(\mathbf{y}) = \mathbf{L} \cdot \mathbf{K} \tag{11.2}$$

Also, if $\mathbf{h}$ is a continuous function defined near $\mathbf{L}$, then

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{h} \circ \mathbf{f}(\mathbf{y}) = \mathbf{h}(\mathbf{L}). \tag{11.3}$$

For a vector valued function, $\mathbf{f}(\mathbf{y}) = (f_1(\mathbf{y}), \cdots, f_q(\mathbf{y}))$, $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L} = (L_1 \cdot \cdot, L_k)^T$ if and only if

$$\lim_{\mathbf{y}\to\mathbf{x}} f_k(\mathbf{y}) = L_k \tag{11.4}$$

for each $k = 1, \cdots, p$.

In the case where $\mathbf{f}$ and $\mathbf{g}$ have values in $\mathbb{R}^3$

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) \times \mathbf{g}(\mathbf{y}) = \mathbf{L} \times \mathbf{K}. \tag{11.5}$$

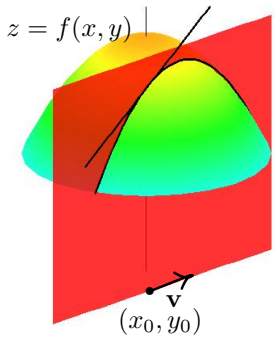
Also recall Theorem 7.4.6 on Page 138.

Theorem 11.2.7 *For $\mathbf{f} : D(\mathbf{f}) \to \mathbb{R}^q$ and $\mathbf{x} \in D(\mathbf{f})$ such that $\mathbf{x}$ is a limit point of $D(\mathbf{f})$, it follows $\mathbf{f}$ is continuous at $\mathbf{x}$ if and only if $\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x})$.*

11.3 The Directional Derivative And Partial Derivatives

11.3.1 The Directional Derivative

The directional derivative is just what its name suggests. It is the derivative of a function in a particular direction. The following picture illustrates the situation in the case of a function of two variables.



In this picture, $\mathbf{v} \equiv (v_1, v_2)$ is a unit vector in the xy plane and $\mathbf{x}_0 \equiv (x_0, y_0)$ is a point in the xy plane. When (x, y) moves in the direction of $\mathbf{v}$, this results in a change in $z = f(x, y)$ as shown in the picture. The directional derivative in this direction is defined as

$$\lim_{t \to 0} \frac{f(x_0 + tv_1, y_0 + tv_2) - f(x_0, y_0)}{t}.$$

It tells how fast z is changing in this direction. If you looked at it from the side, you would be getting the slope of the indicated tangent line. A simple example of this is a person climbing a mountain. He could go various directions, some steeper than others. The directional derivative is just a measure of the steepness in a given direction. This motivates the following general definition of the directional derivative.

Definition 11.3.1 *Let $f : U \to \mathbb{R}$ where U is an open set in $\mathbb{R}^n$ and let $\mathbf{v}$ be a unit vector. For $\mathbf{x} \in U$, define the* ***directional derivative*** *of f in the direction, $\mathbf{v}$, at the point $\mathbf{x}$ as*

$$D_{\mathbf{v}} f(\mathbf{x}) \equiv \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}.$$

Example 11.3.2 *Find the directional derivative of the function, $f(x, y) = x^2 y$ in the direction of $\mathbf{i} + \mathbf{j}$ at the point $(1, 2)$.*

First you need a unit vector which has the same direction as the given vector. This unit vector is $\mathbf{v} \equiv \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. Then to find the directional derivative from the definition, write the difference quotient described above. Thus $f(\mathbf{x} + t\mathbf{v}) = \left(1 + \frac{t}{\sqrt{2}}\right)^2 \left(2 + \frac{t}{\sqrt{2}}\right)$ and $f(\mathbf{x}) = 2$. Therefore,

$$\frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t} = \frac{\left(1 + \frac{t}{\sqrt{2}}\right)^2 \left(2 + \frac{t}{\sqrt{2}}\right) - 2}{t},$$

and to find the directional derivative, you take the limit of this as $t \to 0$. However, this difference quotient equals $\frac{1}{4}\sqrt{2}\left(10 + 4t\sqrt{2} + t^2\right)$ and so, letting $t \to 0$,

$$D_{\mathbf{v}} f(1, 2) = \left(\frac{5}{2}\sqrt{2}\right).$$

There is something you must keep in mind about this. The direction vector must always be a unit vector[2].

[2]Actually, there is a more general formulation of the notion of directional derivative known as the Gateaux derivative in which the length of $\mathbf{v}$ is not equal to one but it will not be considered.

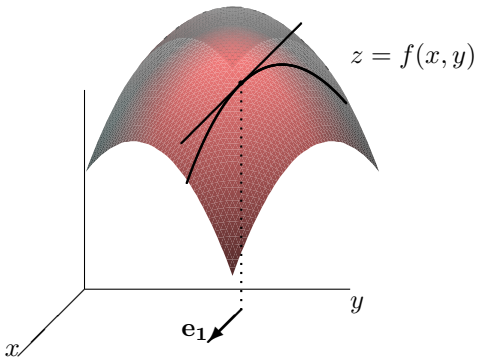
11.3.2 Partial Derivatives

There are some special unit vectors which come to mind immediately. These are the vectors, $\mathbf{e}_i$ where

$$\mathbf{e}_i = (0, \cdots, 0, 1, 0, \cdots 0)^T$$

and the 1 is in the i^{th} position.

Thus in case of a function of two variables, the directional derivative in the direction $\mathbf{i} = \mathbf{e_1}$ is the slope of the indicated straight line in the following picture.



As in the case of a general directional derivative, you fix y and take the derivative of the function, $x \to f(x, y)$. More generally, even in situations which cannot be drawn, the definition of a partial derivative is as follows.

Definition 11.3.3 *Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$. Then letting $\mathbf{x} = (x_1, \cdots, x_n)^T$ be a typical element of $\mathbb{R}^n$,*

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) \equiv D_{\mathbf{e}_i} f(\mathbf{x}).$$

*This is called the **partial derivative** of f. Thus,*

$$\begin{aligned} \frac{\partial f}{\partial x_i}(\mathbf{x}) &\equiv \lim_{t \to 0} \frac{f(\mathbf{x}+t\mathbf{e}_i) - f(\mathbf{x})}{t} \\ &= \lim_{t \to 0} \frac{f(x_1, \cdots, x_i + t, \cdots x_n) - f(x_1, \cdots, x_i, \cdots x_n)}{t}, \end{aligned}$$

and to find the partial derivative, differentiate with respect to the variable of interest and regard all the others as constants. Other notation for this partial derivative is $f_{x_i}, f_{,i}$, or $D_i f$. If $y = f(\mathbf{x})$, the partial derivative of f with respect to x_i may also be denoted by

$$\frac{\partial y}{\partial x_i} \text{ or } y_{x_i}.$$

Example 11.3.4 *Find $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$ if $f(x, y) = y \sin x + x^2 y + z$.*

From the definition above, $\frac{\partial f}{\partial x} = y \cos x + 2xy$, $\frac{\partial f}{\partial y} = \sin x + x^2$, and $\frac{\partial f}{\partial z} = 1$. Having taken one partial derivative, there is no reason to stop doing it. Thus, one could take the partial derivative with respect to y of the partial derivative with respect to x, denoted by $\frac{\partial^2 f}{\partial y \partial x}$ or f_{xy}. In the above example,

$$\frac{\partial^2 f}{\partial y \partial x} = f_{xy} = \cos x + 2x.$$

Also observe that

$$\frac{\partial^2 f}{\partial x \partial y} = f_{yx} = \cos x + 2x.$$

Higher order partial derivatives are defined by analogy to the above. Thus in the above example,

$$f_{yxx} = -\sin x + 2.$$

These partial derivatives, f_{xy} are called mixed partial derivatives.

There is an interesting relationship between the directional derivatives and the partial derivatives, provided the partial derivatives exist and are continuous.

Definition 11.3.5 *Suppose $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is an open set and the partial derivatives of f all exist and are continuous on U. Under these conditions, define the* ***gradient*** *of f denoted $\nabla f(\mathbf{x})$ to be the vector*

$$\nabla f(\mathbf{x}) = \left(f_{x_1}(\mathbf{x}), f_{x_2}(\mathbf{x}), \cdots, f_{x_n}(\mathbf{x})\right)^T.$$

Proposition 11.3.6 *In the situation of Definition 11.3.5 and for $\mathbf{v}$ a unit vector,*

$$D_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}.$$

This proposition will be proved in a more general setting later. For now, you can use it to compute directional derivatives.

Example 11.3.7 *Find the directional derivative of the function, $f(x,y) = \sin\left(2x^2 + y^3\right)$ at $(1,1)$ in the direction $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$.*

First find the gradient.

$$\nabla f(x,y) = \left(4x\cos\left(2x^2+y^3\right), 3y^2\cos\left(2x^2+y^3\right)\right)^T.$$

Therefore,

$$\nabla f(1,1) = \left(4\cos(3), 3\cos(3)\right)^T$$

The directional derivative is therefore,

$$\left(4\cos(3), 3\cos(3)\right)^T \cdot \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T = \frac{7}{2}(\cos 3)\sqrt{2}.$$

Another important observation is that the gradient gives the direction in which the function changes most rapidly.

Proposition 11.3.8 *In the situation of Definition 11.3.5, suppose $\nabla f(\mathbf{x}) \neq \mathbf{0}$. Then the direction in which f increases most rapidly, that is the direction in which the directional derivative is largest, is the direction of the gradient. Thus $\mathbf{v} = \nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$ is the unit vector which maximizes $D_{\mathbf{v}} f(\mathbf{x})$ and this maximum value is $|\nabla f(\mathbf{x})|$. Similarly, $\mathbf{v} = -\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$ is the unit vector which minimizes $D_{\mathbf{v}} f(\mathbf{x})$ and this minimum value is $-|\nabla f(\mathbf{x})|$.*

Proof: Let $\mathbf{v}$ be any unit vector. Then from Proposition 11.3.6,

$$D_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v} = |\nabla f(\mathbf{x})|\,|\mathbf{v}| \cos\theta = |\nabla f(\mathbf{x})| \cos\theta$$

where θ is the included angle between these two vectors, $\nabla f(\mathbf{x})$ and $\mathbf{v}$. Therefore, $D_{\mathbf{v}} f(\mathbf{x})$ is maximized when $\cos\theta = 1$ and minimized when $\cos\theta = -1$. The first case corresonds to the angle between the two vectors being 0 which requires they point in

the same direction in which case, it must be that $\mathbf{v} = \nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$ and $D_{\mathbf{v}} f(\mathbf{x}) = |\nabla f(\mathbf{x})|$. The second case occurs when θ is π and in this case the two vectors point in opposite directions and the directional derivative equals $-|\nabla f(\mathbf{x})|$.

The concept of a **directional derivative for a vector valued function** is also easy to define although the geometric significance expressed in pictures is not.

Definition 11.3.9 *Let $\mathbf{f} : U \to \mathbb{R}^p$ where U is an open set in $\mathbb{R}^n$ and let $\mathbf{v}$ be a unit vector. For $\mathbf{x} \in U$, define the directional derivative of $\mathbf{f}$ in the direction, $\mathbf{v}$, at the point $\mathbf{x}$ as*

$$D_{\mathbf{v}}\mathbf{f}(\mathbf{x}) \equiv \lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{v}) - \mathbf{f}(\mathbf{x})}{t}.$$

Example 11.3.10 *Let $\mathbf{f}(x,y) = \left(xy^2, yx\right)^T$. Find the directional derivative in the direction $(1,2)^T$ at the point (x,y).*

First, a unit vector in this direction is $\left(1/\sqrt{5}, 2/\sqrt{5}\right)^T$ and from the definition, the desired limit is

$$\begin{aligned} & \lim_{t \to 0} \frac{\left(\left(x + t\left(1/\sqrt{5}\right)\right)\left(y + t\left(2/\sqrt{5}\right)\right)^2 - xy^2, \left(x + t\left(1/\sqrt{5}\right)\right)\left(y + t\left(2/\sqrt{5}\right)\right) - xy\right)}{t} \\ = & \lim_{t \to 0} \left(\frac{4}{5}xy\sqrt{5} + \frac{4}{5}xt + \frac{1}{5}\sqrt{5}y^2 + \frac{4}{5}ty + \frac{4}{25}t^2\sqrt{5}, \frac{2}{5}x\sqrt{5} + \frac{1}{5}y\sqrt{5} + \frac{2}{5}t\right) \\ = & \left(\frac{4}{5}xy\sqrt{5} + \frac{1}{5}\sqrt{5}y^2, \frac{2}{5}x\sqrt{5} + \frac{1}{5}y\sqrt{5}\right). \end{aligned}$$

You see from this example and the above definition that all you have to do is to form the vector which is obtained by replacing each component of the vector with its directional derivative. In particular, you can take partial derivatives of vector valued functions and use the same notation.

Example 11.3.11 *Find the partial derivative with respect to x of the function $\mathbf{f}(x,y,z,w) = \left(xy^2, z\sin(xy), z^3x\right)^T$.*

From the above definition, $\mathbf{f}_x(x,y,z) = D_1\mathbf{f}(x,y,z) = \left(y^2, zy\cos(xy), z^3\right)^T$.

11.4 Mixed Partial Derivatives

Under certain conditions the **mixed partial derivatives** will always be equal. This astonishing fact is due to Euler in 1734.

Theorem 11.4.1 *Suppose $f : U \subseteq \mathbb{R}^2 \to \mathbb{R}$ where U is an open set on which f_x, f_y, f_{xy} and f_{yx} exist. Then if f_{xy} and f_{yx} are continuous at the point $(x,y) \in U$, it follows*

$$f_{xy}(x,y) = f_{yx}(x,y).$$

Proof: Since U is open, there exists $r > 0$ such that $B((x,y), r) \subseteq U$. Now let $|t|, |s| < r/2$ and consider

$$\Delta(s,t) \equiv \frac{1}{st}\{\overbrace{f(x+t, y+s) - f(x+t, y)}^{h(t)} - \overbrace{(f(x, y+s) - f(x,y))}^{h(0)}\}. \tag{11.6}$$

Note that $(x+t, y+s) \in U$ because

$$\begin{aligned} |(x+t, y+s) - (x,y)| &= |(t,s)| = \left(t^2+s^2\right)^{1/2} \\ &\leq \left(\frac{r^2}{4}+\frac{r^2}{4}\right)^{1/2} = \frac{r}{\sqrt{2}} < r. \end{aligned}$$

As implied above, $h(t) \equiv f(x+t, y+s) - f(x+t, y)$. Therefore, by the mean value theorem from calculus and the (one variable) chain rule,

$$\begin{aligned} \Delta(s,t) &= \frac{1}{st}(h(t) - h(0)) = \frac{1}{st} h'(\alpha t)\, t \\ &= \frac{1}{s}\left(f_x(x+\alpha t, y+s) - f_x(x+\alpha t, y)\right) \end{aligned}$$

for some $\alpha \in (0,1)$. Applying the mean value theorem again,

$$\Delta(s,t) = f_{xy}(x+\alpha t, y+\beta s)$$

where $\alpha, \beta \in (0,1)$.

If the terms $f(x+t, y)$ and $f(x, y+s)$ are interchanged in 11.6, $\Delta(s,t)$ is also unchanged and the above argument shows there exist $\gamma, \delta \in (0,1)$ such that

$$\Delta(s,t) = f_{yx}(x+\gamma t, y+\delta s).$$

Letting $(s,t) \to (0,0)$ and using the continuity of f_{xy} and f_{yx} at (x,y),

$$\lim_{(s,t)\to(0,0)} \Delta(s,t) = f_{xy}(x,y) = f_{yx}(x,y).$$

This proves the theorem.

The following is obtained from the above by simply fixing all the variables except for the two of interest.

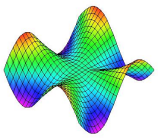
Corollary 11.4.2 *Suppose U is an open subset of $\mathbb{R}^n$ and $f : U \to \mathbb{R}$ has the property that for two indices, k, l, f_{x_k}, $f_{x_l}, f_{x_l x_k}$, and $f_{x_k x_l}$ exist on U and $f_{x_k x_l}$ and $f_{x_l x_k}$ are both continuous at $\mathbf{x} \in U$. Then $f_{x_k x_l}(\mathbf{x}) = f_{x_l x_k}(\mathbf{x})$.*

It is necessary to assume the mixed partial derivatives are continuous in order to assert they are equal. The following is a well known example [3].

Example 11.4.3 *Let*

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

Here is a picture of the graph of this function. It looks innocuous but isn't.



From the definition of partial derivatives it follows immediately that $f_x(0,0) = f_y(0,0) = 0$. Using the standard rules of differentiation, for $(x,y) \neq (0,0)$,

$$f_x = y\frac{x^4 - y^4 + 4x^2y^2}{\left(x^2+y^2\right)^2},\ f_y = x\frac{x^4 - y^4 - 4x^2y^2}{\left(x^2+y^2\right)^2}$$

Now

$$f_{xy}(0,0) \equiv \lim_{y\to 0} \frac{f_x(0,y) - f_x(0,0)}{y} = \lim_{y\to 0} \frac{-y^4}{(y^2)^2} = -1$$

while

$$f_{yx}(0,0) \equiv \lim_{x\to 0} \frac{f_y(x,0) - f_y(0,0)}{x} = \lim_{x\to 0} \frac{x^4}{(x^2)^2} = 1$$

showing that although the mixed partial derivatives do exist at $(0,0)$, they are not equal there.

11.5 Partial Differential Equations

Partial differential equations are equations which involve the partial derivatives of some function. The most famous partial differential equations involve the **Laplacian**, named after Laplace[3].

Definition 11.5.1 *Let u be a function of n variables. Then $\Delta u \equiv \sum_{k=1}^{n} u_{x_k x_k}$. This is also written as $\nabla^2 u$. The symbol, Δ or ∇^2 is called the Laplacian. When $\Delta u = 0$ the function, u is called **harmonic**.**Laplace's equation** is $\Delta u = 0$. The **heat equation** is $u_t - \Delta u = 0$ and the **wave equation** is $u_{tt} - \Delta u = 0$.*

Example 11.5.2 *Find the Laplacian of $u(x,y) = x^2 - y^2$.*

$u_{xx} = 2$ while $u_{yy} = -2$. Therefore, $\Delta u = u_{xx} + u_{yy} = 2 - 2 = 0$. Thus this function is harmonic, $\Delta u = 0$.

Example 11.5.3 *Find $u_t - \Delta u$ where $u(t,x,y) = e^{-t}\cos x$.*

In this case, $u_t = -e^{-t}\cos x$ while $u_{yy} = 0$ and $u_{xx} = -e^{-t}\cos x$ therefore, $u_t - \Delta u = 0$ and so u solves the heat equation, $u_t - \Delta u = 0$.

Example 11.5.4 *Let $u(t,x) = \sin t \cos x$. Find $u_{tt} - \Delta u$.*

In this case, $u_{tt} = -\sin t \cos x$ while $\Delta u = -\sin t \cos x$. Therefore, u is a solution of the wave equation, $u_{tt} - \Delta u = 0$.

11.6 Exercises

1. Find the directional derivative of $f(x,y,z) = x^2 y + z^4$ in the direction of the vector, $(1,3,-1)$ when $(x,y,z) = (1,1,1)$.

2. Find the directional derivative of $f(x,y,z) = \sin\left(x + y^2\right) + z$ in the direction of the vector, $(1,2,-1)$ when $(x,y,z) = (1,1,1)$.

3. Find the directional derivative of $f(x,y,z) = \ln\left(x + y^2\right) + z^2$ in the direction of the vector, $(1,1,-1)$ when $(x,y,z) = (1,1,1)$.

4. Find the largest value of the directional derivative of $f(x,y,z) = \ln\left(x + y^2\right) + z^2$ at the point $(1,1,1)$.

5. Find the smallest value of the directional derivative of $f(x,y,z) = x\sin\left(4xy^2\right) + z^2$ at the point $(1,1,1)$.

[3]Laplace was a great physicist of the 1700's. He made fundamental contributions to mechanics and astronomy.

6. An ant falls to the top of a stove having temperature $T(x,y) = x^2 \sin(x+y)$ at the point $(2,3)$. In what direction should the ant go to minimize the temperature? In what direction should he go to maximize the temperature?

7. Find the partial derivative with respect to y of the function

$$\mathbf{f}(x,y,z,w) = \left(y^2, z^2 \sin(xy), z^3 x\right)^T.$$

8. Find the partial derivative with respect to x of the function

$$\mathbf{f}(x,y,z,w) = \left(wx, zx \sin(xy), z^3 x\right)^T.$$

9. Find $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$ for $f =$

   (a) $x^2 y + \cos(xy) + z^3 y$

   (b) $e^{x^2+y^2} z \sin(x+y)$

   (c) $z^2 \sin^3\left(e^{x^2+y^3}\right)$

   (d) $x^2 \cos\left(\sin\left(\tan\left(z^2+y^2\right)\right)\right)$

   (e) x^{y^2+z}

10. Suppose

$$f(x,y) = \begin{cases} \frac{2xy+6x^3+12xy^2+18yx^2+36y^3+\sin\left(x^3\right)+\tan\left(3y^3\right)}{3x^2+6y^2} \text{ if } (x,y) \neq (0,0) \\ 0 \text{ if } (x,y) = (0,0). \end{cases}$$

    Find $\frac{\partial f}{\partial x}(0,0)$ and $\frac{\partial f}{\partial y}(0,0)$.

11. Why must the vector in the definition of the directional derivative be a unit vector? **Hint:** Suppose not. Would the directional derivative be a correct manifestation of steepness?

12. Find $f_x, f_y, f_z, f_{xy}, f_{yx}, f_{xz}, f_{zx}, f_{zy}, f_{yz}$ for the following. Verify the mixed partial derivatives are equal.

    (a) $x^2 y^3 z^4 + \sin(xyz)$

    (b) $\sin(xyz) + x^2 yz$

    (c) $z \ln\left|x^2+y^2+1\right|$

    (d) $e^{x^2+y^2+z^2}$

    (e) $\tan(xyz)$

13. Suppose $f : U \to \mathbb{R}$ where U is an open set and suppose that $\mathbf{x} \in U$ has the property that for all $\mathbf{y}$ near $\mathbf{x}$, $f(\mathbf{x}) \leq f(\mathbf{y})$. Prove that if f has all of its partial derivatives at $\mathbf{x}$, then $f_{x_i}(\mathbf{x}) = 0$ for each x_i. **Hint:** This is just a repeat of the usual one variable theorem seen in beginning calculus. You just do this one variable argument for each variable to get the conclusion.

14. As an important application of Problem 13 consider the following. Experiments are done at n times, $t_1, t_2, \cdots, t_n$ and at each time there results a collection of numerical outcomes. Denote by $\{(t_i, x_i)\}_{i=1}^p$ the set of all such pairs and try to find numbers a and b such that the line $x = at + b$ approximates these ordered pairs as well as possible in the sense that out of all choices of a and b, $\sum_{i=1}^p (at_i + b - x_i)^2$

is as small as possible. In other words, you want to minimize the function of two variables, $f(a,b) \equiv \sum_{i=1}^{p} (at_i + b - x_i)^2$. Find a formula for a and b in terms of the given ordered pairs. You will be finding the formula for the least squares regression line.

15. Show that if $v(x,y) = u(\alpha x, \beta y)$, then $v_x = \alpha u_x$ and $v_y = \beta u_y$. State and prove a generalization to any number of variables.

16. Let f be a function which has continuous derivatives. Show $u(t,x) = f(x - ct)$ solves the wave equation, $u_{tt} - c^2 \Delta u = 0$. What about $u(x,t) = f(x + ct)$?

17. D'Alembert found a formula for the solution to the wave equation, $u_{tt} = c^2 u_{xx}$ along with the initial conditions $u(x,0) = f(x), u_t(x,0) = g(x)$. Here is how he did it. He looked for a solution of the form $u(x,t) = h(x + ct) + k(x - ct)$ and then found h and k in terms of the given functions f and g. He ended up with something like

$$u(x,t) = \frac{1}{2c} \int_{x-ct}^{x+ct} g(r)\, dr + \frac{1}{2} (f(x + ct) + f(x - ct)).$$

Fill in the details.

18. Determine which of the following functions satisfy Laplace's equation.

(a) $x^3 - 3xy^2$
(b) $3x^2y - y^3$
(c) $x^3 - 3xy^2 + 2x^2 - 2y^2$
(d) $3x^2y - y^3 + 4xy$
(e) $3x^2 - y^3 + 4xy$
(f) $3x^2y - y^3 + 4y$
(g) $x^3 - 3x^2y^2 + 2x^2 - 2y^2$

19. Show that $z = \frac{xy}{y-x}$ is a solution to the partial differential equation, $x^2 \frac{\partial^2 z}{\partial x^2} + 2xy \frac{\partial^2 z}{\partial x \partial y} + y^2 \frac{\partial^2 z}{\partial y^2} = 0$.

20. Show that $z = \sqrt{x^2 + y^2}$ is a solution to $x \frac{\partial z}{\partial x} + y \frac{\partial z}{\partial y} = 0$.

21. Show that if $\Delta u = \lambda u$, then $e^{\lambda t} u$ solves the heat equation, $u_t - \Delta u = 0$.

22. Show that if a, b are scalars and u, v are functions which satisfy Laplace's equation then $au + bv$ also satisfies Laplace's equation. Verify a similar statement for the heat and wave equations.

23. Show that $u(x,t) = \frac{1}{\sqrt{t}} e^{-x^2/4c^2 t}$ solves the heat equation, $u_t = c^2 u_{xx}$.

The Derivative Of A Function Of Many Variables

12.0.1 Outcomes

1. Define differentiability and explain what the derivative is for a function of n variables.
2. Describe the relation between existence of partial derivatives, continuity, and differentiability.
3. Give examples of functions which have partial derivatives but are not continuous, examples of functions which are differentiable but not C^1, and examples of functions which are continuous without having partial derivatives.
4. Evaluate derivatives of composite functions using the chain rule.
5. Solve related rates problems using the chain rule.

12.1 The Derivative Of Functions Of One Variable

First recall the notion of the derivative of a function of one variable.

Observation 12.1.1 *Suppose a function, f of one variable has a derivative at x. Then*

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

This observation follows from the definition of the derivative of a function of one variable, namely

$$f'(x)\equiv\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}.$$

Definition 12.1.2 *A vector valued function of a vector, $\mathbf{v}$ is called $\mathbf{o}(\mathbf{v})$ if*

$$\lim_{|\mathbf{v}|\to 0}\frac{\mathbf{o}(\mathbf{v})}{|\mathbf{v}|}=\mathbf{0}. \tag{12.1}$$

Thus the function $f(x+h)-f(x)-f'(x)h$ is $o(h)$. The expression, $o(h)$, is used like an adjective. It is like saying the function is white or black or green or fat or thin. The term is used very imprecisely. Thus

$$\mathbf{o}(\mathbf{v})=\mathbf{o}(\mathbf{v})+\mathbf{o}(\mathbf{v}),\mathbf{o}(\mathbf{v})=45\mathbf{o}(\mathbf{v}),\mathbf{o}(\mathbf{v})=\mathbf{o}(\mathbf{v})-\mathbf{o}(\mathbf{v}),etc.$$

When you add two functions with the property of the above definition, you get another one having that same property. When you multiply by 45 the property is also retained as it is when you subtract two such functions. How could something so sloppy be useful? The notation is useful precisely because it prevents you from obsessing over things which are not relevant and should be ignored.

Theorem 12.1.3 *Let $f : (a,b) \to \mathbb{R}$ be a function of one variable. Then $f'(x)$ exists if and only if*

$$f(x+h) - f(x) = ph + o(h) \tag{12.2}$$

In this case, $p = f'(x)$.

Proof: From the above observation it follows that if $f'(x)$ does exist, then 12.2 holds.

Suppose then that 12.2 is true. Then

$$\frac{f(x+h) - f(x)}{h} - p = \frac{o(h)}{h}.$$

Taking a limit, you see that

$$p = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

and that in fact this limit exists which shows that $p = f'(x)$. This proves the theorem.

This theorem shows that one way to define $f'(x)$ is as the number, p, if there is one which has the property that

$$f(x+h) = f(x) + ph + o(h).$$

You should think of p as the linear transformation resulting from multiplication by the 1×1 matrix, (p).

Example 12.1.4 *Let $f(x) = x^3$. Find $f'(x)$.*

$f(x+h) = (x+h)^3 = x^3 + 3x^2h + 3xh^2 + h^3 = f(x) + 3x^2h + \left(3xh + h^2\right)h$. Since $\left(3xh + h^2\right)h = o(h)$, it follows $f'(x) = 3x^2$.

Example 12.1.5 *Let $f(x) = \sin(x)$. Find $f'(x)$.*

$$\begin{aligned} f(x+h) - f(x) &= \sin(x+h) - \sin(x) = \sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x) \\ &= \cos(x)\sin(h) + \sin(x)\frac{(\cos(h)-1)}{h}h \\ &= \cos(x)h + \cos(x)\frac{(\sin(h)-h)}{h}h + \sin x\frac{(\cos(h)-1)}{h}h. \end{aligned}$$

Now

$$\cos(x)\frac{(\sin(h)-h)}{h}h + \sin x\frac{(\cos(h)-1)}{h}h = o(h). \tag{12.3}$$

Remember the fundamental limits which allowed you to find the derivative of $\sin(x)$ were

$$\lim_{h \to 0} \frac{\sin(h)}{h} = 1, \ \lim_{h \to 0} \frac{\cos(h) - 1}{h} = 0. \tag{12.4}$$

These same limits are what is needed to verify 12.3.

12.2 The Derivative Of Functions Of Many Variables

This way of thinking about the derivative is exactly what is needed to define the derivative of a function of n variables. Recall the following definition.

Definition 12.2.1 *A function, T which maps $\mathbb{R}^n$ to $\mathbb{R}^p$ is called a linear transformation if for every pair of scalars, a, b and vectors, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, it follows that $T(a\mathbf{x} + b\mathbf{y}) = aT(\mathbf{x}) + bT(\mathbf{y})$.*

Recall that from the properties of matrix multiplication, it follows that if A is an $n \times p$ matrix, and if $\mathbf{x}, \mathbf{y}$ are vectors in $\mathbb{R}^n$, then $A(a\mathbf{x} + b\mathbf{y}) = aA(\mathbf{x}) + bA(\mathbf{y})$. Thus you can define a linear transformation by multiplying by a matrix. Of course the simplest example is that of a 1×1 matrix or number. You can think of the number 3 as a linear transformation, T mapping $\mathbb{R}$ to $\mathbb{R}$ according to the rule $Tx = 3x$. It satisfies the properties needed for a linear transformation because $3(ax + by) = a3x + b3y = aTx + bTy$. The case of the derivative of a scalar valued function of one variable is of this sort. You get a number for the derivative. However, you can think of this number as a linear transformation. Of course it is not worth the fuss to do so for a function of one variable but this is the way you must think of it for a function of n variables.

Definition 12.2.2 *Let $\mathbf{f} : U \to \mathbb{R}^p$ where U is an open set in $\mathbb{R}^n$ for $n, p \geq 1$ and let $\mathbf{x} \in U$ be given. Then $\mathbf{f}$ is defined to be **differentiable** at $\mathbf{x} \in U$ if and only if there exist column vectors, $\mathbf{v}_i$ such that for $\mathbf{h} = (h_1 \cdots, h_n)^T$,*

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + \sum_{i=1}^{n} \mathbf{v}_i h_i + \mathbf{o}(\mathbf{h}). \tag{12.5}$$

The derivative of the function, $\mathbf{f}$, denoted by $D\mathbf{f}(\mathbf{x})$, is the linear transformation defined by multiplying by the matrix whose columns are the $p \times 1$ vectors, $\mathbf{v}_i$. Thus if $\mathbf{w}$ is a vector in $\mathbb{R}^n$,

$$D\mathbf{f}(\mathbf{x})\mathbf{w} \equiv \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} \mathbf{w}.$$

It is common to think of this matrix as the derivative but strictly speaking, this is incorrect. The derivative is a "linear transformation" determined by multiplication by this matrix, called the **standard matrix** because it is based on the standard basis vectors for $\mathbb{R}^n$. The subtle issues involved in a thorough exploration of this issue will be avoided for now. It will be fine to think of the above matrix as the derivative. Other notations which are often used for this matrix or the linear transformation are $\mathbf{f}'(\mathbf{x})$, $J(\mathbf{x})$, and even $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ or $\frac{d\mathbf{f}}{d\mathbf{x}}$.

Theorem 12.2.3 *Suppose $\mathbf{f}$ is as given above in 12.5. Then*

$$\mathbf{v}_k = \lim_{h \to 0} \frac{\mathbf{f}(\mathbf{x} + h\mathbf{e}_k) - \mathbf{f}(\mathbf{x})}{h} \equiv \frac{\partial \mathbf{f}}{\partial x_k}(\mathbf{x}),$$

the k^{th} partial derivative.

Proof: Let $\mathbf{h} = (0, \cdots, h, 0, \cdots, 0)^T = h\mathbf{e}_k$ where the h is in the k^{th} slot. Then 12.5 reduces to

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + \mathbf{v}_k h + \mathbf{o}(h).$$

Therefore, dividing by h

$$\frac{\mathbf{f}(\mathbf{x} + h\mathbf{e}_k) - \mathbf{f}(\mathbf{x})}{h} = \mathbf{v}_k + \frac{\mathbf{o}(h)}{h}$$

and taking the limit,

$$\lim_{h\to 0}\frac{\mathbf{f}(\mathbf{x}+h\mathbf{e}_k)-\mathbf{f}(\mathbf{x})}{h}=\lim_{h\to 0}\left(\mathbf{v}_k+\frac{\mathbf{o}(h)}{h}\right)=\mathbf{v}_k$$

and so, the above limit exists. This proves the theorem.

Let $\mathbf{f}:U\to\mathbb{R}^q$ where U is an open subset of $\mathbb{R}^p$ and $\mathbf{f}$ is differentiable. It was just shown

$$\mathbf{f}(\mathbf{x}+\mathbf{v})=\mathbf{f}(\mathbf{x})+\sum_{j=1}^{p}\frac{\partial\mathbf{f}(\mathbf{x})}{\partial x_j}v_j+\mathbf{o}(\mathbf{v}).$$

Taking the i^{th} coordinate of the above equation yields

$$f_i(\mathbf{x}+\mathbf{v})=f_i(\mathbf{x})+\sum_{j=1}^{p}\frac{\partial f_i(\mathbf{x})}{\partial x_j}v_j+o(\mathbf{v})$$

and it follows that the term with a sum is nothing more than the i^{th} component of $J(\mathbf{x})\mathbf{v}$ where $J(\mathbf{x})$ is the $q\times p$ matrix,

$$\begin{pmatrix}\frac{\partial f_1}{\partial x_1}&\frac{\partial f_1}{\partial x_2}&\cdots&\frac{\partial f_1}{\partial x_p}\\ \frac{\partial f_2}{\partial x_1}&\frac{\partial f_2}{\partial x_2}&\cdots&\frac{\partial f_2}{\partial x_p}\\ \vdots&\vdots&\ddots&\vdots\\ \frac{\partial f_q}{\partial x_1}&\frac{\partial f_q}{\partial x_2}&\cdots&\frac{\partial f_q}{\partial x_p}\end{pmatrix}.$$

This gives the form of the matrix which defines the linear transformation, $D\mathbf{f}(\mathbf{x})$. Thus

$$\mathbf{f}(\mathbf{x}+\mathbf{v})=\mathbf{f}(\mathbf{x})+J(\mathbf{x})\mathbf{v}+\mathbf{o}(\mathbf{v}) \tag{12.6}$$

and to reiterate, the linear transformation which results by multiplication by this $q\times p$ matrix is known as the derivative.

Sometimes x,y,z is written instead of x_1,x_2, and x_3. This is to save on notation and is easier to write and to look at although it lacks generality. When this is done it is understood that $x=x_1,y=x_2$, and $z=x_3$. Thus the derivative is the linear transformation determined by

$$\begin{pmatrix}f_{1x}&f_{1y}&f_{1z}\\ f_{2x}&f_{2y}&f_{2z}\\ f_{3x}&f_{3y}&f_{3z}\end{pmatrix}.$$

Example 12.2.4 *Let A be a constant $m\times n$ matrix and consider $\mathbf{f}(\mathbf{x})=A\mathbf{x}$. Find $D\mathbf{f}(\mathbf{x})$ if it exists.*

$$\mathbf{f}(\mathbf{x}+\mathbf{h})-\mathbf{f}(\mathbf{x})=A(\mathbf{x}+\mathbf{h})-A(\mathbf{x})=A\mathbf{h}=A\mathbf{h}+\mathbf{o}(\mathbf{h}).$$

In fact in this case, $\mathbf{o}(\mathbf{h})=\mathbf{0}$. Therefore, $D\mathbf{f}(\mathbf{x})=A$. Note that this looks the same as the case in one variable, $f(x)=ax$.

12.3 C^1 Functions

Given a function of many variables, how can you tell if it is differentiable? Sometimes you have to go directly to the definition and verify it is differrentiable from the definition. For example, you may have seen the following important example in one variable calculus.

Example 12.3.1 *Let* $f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 \text{ if } x = 0 \end{cases}$. *Find* $Df(0)$.

$f(h) - f(0) = 0h + h^2 \sin\left(\frac{1}{h}\right) = o(h)$ and so $Df(0) = 0$. If you find the derivative for $x \neq 0$, it is totally useless information if what you want is $Df(0)$. This is because the derivative, turns out to be discontinuous. Try it. Find the derivative for $x \neq 0$ and try to obtain $Df(0)$ from it. You see, in this example you had to revert to the definition to find the derivative.

It isn't really too hard to use the definition even for more ordinary examples.

Example 12.3.2 *Let* $\mathbf{f}(x,y) = \begin{pmatrix} x^2y + y^2 \\ y^3x \end{pmatrix}$. *Find* $D\mathbf{f}(1,2)$.

First of all note that the thing you are after is a 2×2 matrix.

$$\mathbf{f}(1,2) = \begin{pmatrix} 6 \\ 8 \end{pmatrix}.$$

Then

$$\mathbf{f}(1+h_1, 2+h_2) - \mathbf{f}(1,2)$$

$$\begin{aligned}
&= \begin{pmatrix} (1+h_1)^2(2+h_2) + (2+h_2)^2 \\ (2+h_2)^3(1+h_1) \end{pmatrix} - \begin{pmatrix} 6 \\ 8 \end{pmatrix} \\
&= \begin{pmatrix} 5h_2 + 4h_1 + 2h_1h_2 + 2h_1^2 + h_1^2h_2 + h_2^2 \\ 8h_1 + 12h_2 + 12h_1h_2 + 6h_2^2 + 6h_2^2h_1 + h_2^3 + h_2^3h_1 \end{pmatrix} \\
&= \begin{pmatrix} 4 & 5 \\ 8 & 12 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} 2h_1h_2 + 2h_1^2 + h_1^2h_2 + h_2^2 \\ 12h_1h_2 + 6h_2^2 + 6h_2^2h_1 + h_2^3 + h_2^3h_1 \end{pmatrix} \\
&= \begin{pmatrix} 4 & 5 \\ 8 & 12 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \mathbf{o}(\mathbf{h}).
\end{aligned}$$

Therefore, the standard matrix of the derivative is $\begin{pmatrix} 4 & 5 \\ 8 & 12 \end{pmatrix}$.

Most of the time, there is an easier way to conclude a derivative exists and to find it. It involves the notion of a C^1 function.

Definition 12.3.3 *When* $\mathbf{f} : U \to \mathbb{R}^p$ *for* U *an open subset of* $\mathbb{R}^n$ *and the vector valued functions,* $\frac{\partial \mathbf{f}}{\partial x_i}$ *are all continuous, (equivalently each* $\frac{\partial f_i}{\partial x_j}$ *is continuous), the function is said to be* $C^1(U)$. *If all the partial derivatives up to order* k *exist and are continuous, then the function is said to be* C^k.

It turns out that for a C^1 function, all you have to do is write the matrix described in Theorem 12.2.3 and this will be the derivative. There is no question of existence for the derivative for such functions. This is the importance of the next few theorems.

Theorem 12.3.4 *Let* U *be an open subset of* $\mathbb{R}^2$ *and suppose* $f : U \to \mathbb{R}$ *has the property that the partial derivatives* f_x *and* f_y *exist for* $(x,y) \in U$ *and are continuous at the point* (x_0, y_0). *Then*

$$f((x_0,y_0) + (v_1,v_2)) = f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)\, v_1 + \frac{\partial f}{\partial x}(x_0,y_0)\, v_2 + o(\mathbf{v}).$$

That is, f *is differentiable.*

Proof:

$$f((x_0,y_0)+(v_1,v_2)) - \left(f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)\,v_1 + \frac{\partial f}{\partial y}(x_0,y_0)\,v_2\right) \tag{12.7}$$

$$= (f(x_0+v_1,y_0+v_2) - f(x_0,y_0)) - \left(\frac{\partial f}{\partial x}(x_0,y_0)\,v_1 + \frac{\partial f}{\partial y}(x_0,y_0)\,v_2\right)$$

$$= \left(\overbrace{f(x_0+v_1,y_0+v_2) - f(x_0,y_0+v_2)}^{\text{changes only in first component}} + \overbrace{f(x_0,y_0+v_2) - f(x_0,y_0)}^{\text{changes only in second component}}\right)$$

$$- \left(\frac{\partial f}{\partial x}(x_0,y_0)\,v_1 + \frac{\partial f}{\partial y}(x_0,y_0)\,v_2\right)$$

By the mean value theorem, there exist numbers s and t in $[0,1]$ such that this equals

$$= \left(\frac{\partial f}{\partial x}(x_0+tv_1,y_0+v_2)\,v_1 + \frac{\partial f}{\partial y}(x_0,y_0+sv_2)\,v_2\right)$$

$$- \left(\frac{\partial f}{\partial x}(x_0,y_0)\,v_1 + \frac{\partial f}{\partial y}(x_0,y_0)\,v_2\right)$$

$$= \left(\frac{\partial f}{\partial x}(x_0+tv_1,y_0+v_2) - \frac{\partial f}{\partial x}(x_0,y_0)\right)v_1 + \left(\frac{\partial f}{\partial y}(x_0,y_0+sv_2) - \frac{\partial f}{\partial y}(x_0,y_0)\right)v_2$$

Therefore, letting $o(\mathbf{v})$ denote the expression in 12.7, and noticing that $|v_1|$ and $|v_2|$ are both no larger than $|\mathbf{v}|$,

$$|o(\mathbf{v})| \leq \left(\left|\frac{\partial f}{\partial x}(x_0+tv_1,y_0+v_2) - \frac{\partial f}{\partial x}(x_0,y_0)\right| + \left|\frac{\partial f}{\partial y}(x_0,y_0+sv_2) - \frac{\partial f}{\partial y}(x_0,y_0)\right|\right)|\mathbf{v}|.$$

It follows

$$\frac{|o(\mathbf{v})|}{|\mathbf{v}|} \leq \left|\frac{\partial f}{\partial x}(x_0+tv_1,y_0+v_2) - \frac{\partial f}{\partial x}(x_0,y_0)\right| + \left|\frac{\partial f}{\partial y}(x_0,y_0+sv_2) - \frac{\partial f}{\partial y}(x_0,y_0)\right|$$

Therefore, $\lim_{\mathbf{v}\to\mathbf{0}} \frac{|o(\mathbf{v})|}{|\mathbf{v}|} = 0$ because of the assumption that f_x and f_y are continuous at the point (x_0,y_0) and this proves the theorem.

Having proved a theorem for scalar valued functions, one for vector valued functions follows immediately.

Theorem 12.3.5 *Let U be an open subset of $\mathbb{R}^p$ for $p \geq 1$ and suppose $\mathbf{f}: U \to \mathbb{R}^q$ has the property that each component function, f_i is differentiable at $\mathbf{x}_0$. Then $\mathbf{f}$ is differentiable at $\mathbf{x}_0$.*

Proof: Let $\mathbf{f}(\mathbf{x}) \equiv (f_1(\mathbf{x}),\cdots,f_q(\mathbf{x}))^T$. From the assumption each component function is differentiable, the following holds for each $k = 1,\cdots,q$.

$$f_k(\mathbf{x}_0+\mathbf{v}) = f_k(\mathbf{x}_0) + \sum_{i=1}^{p} \frac{\partial f_k}{\partial x_i}(\mathbf{x}_0)\,v_i + o_k(\mathbf{v}).$$

Define $\mathbf{o}(\mathbf{v}) \equiv (o_1(\mathbf{v}),\cdots,o_q(\mathbf{v}))^T$. Then 12.1 on Page 243 holds for $\mathbf{o}(\mathbf{v})$ because it holds for each of the components of $\mathbf{o}(\mathbf{v})$. The above equation is then equivalent to

$$\mathbf{f}(\mathbf{x}_0+\mathbf{v}) = \mathbf{f}(\mathbf{x}_0) + \sum_{i=1}^{p} \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0)\,v_i + \mathbf{o}(\mathbf{v})$$

and so $\mathbf{f}$ is differentiable at $\mathbf{x}_0$.

Here is an example to illustrate.

Example 12.3.6 *Let* $\mathbf{f}(x,y) = \begin{pmatrix} x^2y + y^2 \\ y^3x \end{pmatrix}$. *Find* $D\mathbf{f}(x,y)$.

From Theorem 12.3.4 this function is differentiable because all possible partial derivatives are continuous. Thus

$$D\mathbf{f}(x,y) = \begin{pmatrix} 2xy & x^2 + 2y \\ y^3 & 3y^2x \end{pmatrix}.$$

In particular,

$$D\mathbf{f}(1,2) = \begin{pmatrix} 4 & 5 \\ 8 & 12 \end{pmatrix}.$$

Not surprisingly, the above theorem has an extension to more variables. First this is illustrated with an example.

Example 12.3.7 *Let* $\mathbf{f}(x_1,x_2,x_3) = \begin{pmatrix} x_1^2x_2 + x_2^2 \\ x_2x_1 + x_3 \\ \sin(x_1x_2x_3) \end{pmatrix}$. *Find* $D\mathbf{f}(x_1,x_2,x_3)$.

All possible partial derivatives are continuous so the function is differentiable. The matrix for this derivative is therefore the following 3×3 matrix

$$\begin{pmatrix} 2x_1x_2 & x_1^2 + 2x_2 & 0 \\ x_2 & x_1 & 1 \\ x_2x_3\cos(x_1x_2x_3) & x_1x_3\cos(x_1x_2x_3) & x_1x_2\cos(x_1x_2x_3) \end{pmatrix}$$

The following theorem is the general result.

Theorem 12.3.8 *Let* U *be an open subset of* $\mathbb{R}^p$ *for* $p \geq 1$ *and suppose* $f : U \to \mathbb{R}$ *has the property that the partial derivatives* f_{x_i} *exist for all* $\mathbf{x} \in U$ *and are continuous at the point* $\mathbf{x}_0 \in U$. *Then*

$$f(\mathbf{x}_0 + \mathbf{v}) = f(\mathbf{x}_0) + \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, v_i + o(\mathbf{v}).$$

That is, f *is differentiable at* $\mathbf{x}_0$ *and the derivative of* f *equals the linear transformation obtained by multiplying by the* $1 \times p$ *matrix,*

$$\left(\frac{\partial f}{\partial x_1}(\mathbf{x}_0), \cdots, \frac{\partial f}{\partial x_p}(\mathbf{x}_0) \right).$$

Proof: The proof is similar to the case of two variables. Letting $\mathbf{v} = (v_1 \cdots, v_p)^T$, denote by $\theta_i\mathbf{v}$ the vector

$$(0, \cdots, 0, v_i, v_{i+1}, \cdots, v_p)^T$$

Thus $\theta_0\mathbf{v} = \mathbf{v}$, $\theta_{p-1}(\mathbf{v}) = (0, \cdots, 0, v_p)^T$, and $\theta_p\mathbf{v} = \mathbf{0}$. Now

$$f(\mathbf{x}_0 + \mathbf{v}) - \left(f(\mathbf{x}_0) + \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, v_i \right) \tag{12.8}$$

$$= \sum_{i=1}^{p} \left(\overbrace{f(\mathbf{x}_0 + \theta_{i-1}\mathbf{v}) - f(\mathbf{x}_0 + \theta_i\mathbf{v})}^{\text{changes only in the } i^{th} \text{ position}} \right) - \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, v_i$$

Now by the mean value theorem there exist numbers $s_i \in (0,1)$ such that the above expression equals

$$= \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + \theta_i \mathbf{v} + s_i v_i)\, v_i - \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, v_i$$

and so letting $o(\mathbf{v})$ equal the expression in 12.8,

$$|o(\mathbf{v})| \leq \sum_{i=1}^{p} \left| \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + \theta_i \mathbf{v} + s_i v_i) - \frac{\partial f}{\partial x_i}(\mathbf{x}_0) \right| |v_i|$$

$$\leq \sum_{i=1}^{p} \left| \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + \theta_i \mathbf{v} + s_i v_i) - \frac{\partial f}{\partial x_i}(\mathbf{x}_0) \right| |\mathbf{v}|$$

and so

$$\lim_{\mathbf{v}\to\mathbf{0}} \frac{|o(\mathbf{v})|}{|\mathbf{v}|} \leq \lim_{\mathbf{v}\to\mathbf{0}} \sum_{i=1}^{p} \left| \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + \theta_i \mathbf{v} + s_i v_i) - \frac{\partial f}{\partial x_i}(\mathbf{x}_0) \right| = 0$$

because of continuity of the f_{x_i} at $\mathbf{x}_0$. This proves the theorem.

Letting $\mathbf{x} - \mathbf{x}_0 = \mathbf{v}$,

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{x}_0) + \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)(x_i - x_{0i}) + o(\mathbf{v}) \\ &= f(\mathbf{x}_0) + \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, v_i + o(\mathbf{v}). \end{aligned}$$

Example 12.3.9 *Suppose* $f(x,y,z) = xy + z^2$. *Find* $Df(1,2,3)$.

Taking the partial derivatives of f, $f_x = y, f_y = x, f_z = 2z$. These are all continuous. Therefore, the function has a derivative and $f_x(1,2,3) = 1, f_y(1,2,3) = 2$, and $f_z(1,2,3) = 6$. Therefore, $Df(1,2,3)$ is given by

$$Df(1,2,3) = (1,2,6).$$

Also, for (x,y,z) close to $(1,2,3)$,

$$\begin{aligned} f(x,y,z) &\approx f(1,2,3) + 1(x-1) + 2(y-2) + 6(z-3) \\ &= 11 + 1(x-1) + 2(y-2) + 6(z-3) = -12 + x + 2y + 6z \end{aligned}$$

In the case where $\mathbf{f}$ has values in $\mathbb{R}^q$ rather than $\mathbb{R}$, is there a similar theorem about differentiability of a C^1 function?

Theorem 12.3.10 *Let* U *be an open subset of* $\mathbb{R}^p$ *for* $p \geq 1$ *and suppose* $\mathbf{f} : U \to \mathbb{R}^q$ *has the property that the partial derivatives* $\mathbf{f}_{x_i}$ *exist for all* $\mathbf{x} \in U$ *and are continuous at the point* $\mathbf{x}_0 \in U$, *then*

$$\mathbf{f}(\mathbf{x}_0 + \mathbf{v}) = \mathbf{f}(\mathbf{x}_0) + \sum_{i=1}^{p} \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0)\, v_i + \mathbf{o}(\mathbf{v}) \tag{12.9}$$

and so $\mathbf{f}$ *is differentiable at* $\mathbf{x}_0$.

Proof: This follows from Theorem 12.3.5.

When a function is differentiable at $\mathbf{x}_0$ it follows the function must be continuous there. This is the content of the following important lemma.

Lemma 12.3.11 *Let $\mathbf{f} : U \to \mathbb{R}^q$ where U is an open subset of $\mathbb{R}^p$. If $\mathbf{f}$ is differentiable, then $\mathbf{f}$ is continuous at $\mathbf{x}_0$. Furthermore, if $C \geq \max\left\{\left|\frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0)\right|, i = 1, \cdots, p\right\}$, then whenever $|\mathbf{x} - \mathbf{x}_0|$ is small enough,*

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)| \leq (Cp + 1)|\mathbf{x} - \mathbf{x}_0| \tag{12.10}$$

Proof: Suppose $\mathbf{f}$ is differentiable. Since $\mathbf{o}(\mathbf{v})$ satisfies 12.1, there exists $\delta_1 > 0$ such that if $|\mathbf{x} - \mathbf{x}_0| < \delta_1$, then $|\mathbf{o}(\mathbf{x} - \mathbf{x}_0)| < |\mathbf{x} - \mathbf{x}_0|$. But also, by the triangle inequality, Corollary 1.5.5 on Page 23,

$$\left|\sum_{i=1}^{p} \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0)(x_i - x_{0i})\right| \leq C\sum_{i=1}^{p} |x_i - x_{0i}| \leq Cp|\mathbf{x} - \mathbf{x}_0|$$

Therefore, if $|\mathbf{x} - \mathbf{x}_0| < \delta_1$,

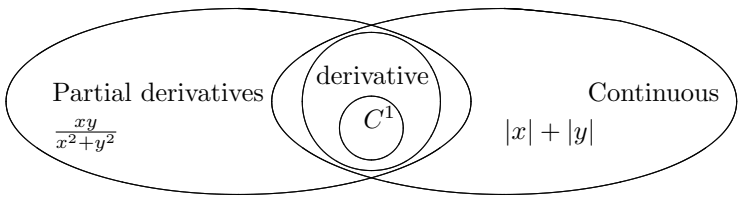
$$\begin{aligned} |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)| &\leq \left|\sum_{i=1}^{p} \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0)(x_i - x_{0i})\right| + |\mathbf{x} - \mathbf{x}_0| \\ &< (Cp + 1)|\mathbf{x} - \mathbf{x}_0| \end{aligned}$$

which verifies 12.10. Now letting $\varepsilon > 0$ be given, let $\delta = \min\left(\delta_1, \frac{\varepsilon}{Cp+1}\right)$. Then for $|\mathbf{x} - \mathbf{x}_0| < \delta$,

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)| < (Cp + 1)|\mathbf{x} - \mathbf{x}_0| < (Cp + 1)\frac{\varepsilon}{Cp + 1} = \varepsilon$$

showing $\mathbf{f}$ is continuous at $\mathbf{x}_0$.

There have been quite a few terms defined. First there was the concept of continuity. Next the concept of partial or directional derivative. Next there was the concept of differentiability and the derivative being a linear transformation determined by a certain matrix. Finally, it was shown that if a function is C^1, then it has a derivative. To give a rough idea of the relationships of these topics, here is a picture.



You might ask whether there are examples of functions which are differentiable but not C^1. Of course there are. In fact, Example 12.3.1 is just such an example as explained earlier. Then you should verify that $f'(x)$ exists for all $x \in \mathbb{R}$ but f' fails to be continuous at $x = 0$. Thus the function is differentiable at every point of $\mathbb{R}$ but fails to be C^1 at every point of $\mathbb{R}$.

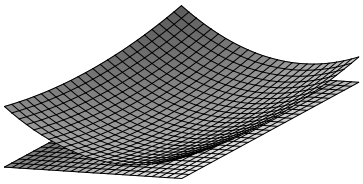
12.3.1 Approximation With A Tangent Plane

In the case where f is a scalar valued function of two variables, the geometric significance of the derivative can be exhibited in the following picture. Writing $\mathbf{v} \equiv (x - x_0, y - y_0)$,

the notion of differentiability at (x_0, y_0) reduces to

$$f(x,y) = f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial x}(x_0,y_0)(y-y_0) + o(\mathbf{v})$$

The right side of the above, $f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial x}(x_0,y_0)(y-y_0) = z$ is the equation of a plane approximating the graph of $z = f(x,y)$ for (x,y) near (x_0,y_0). Saying that the function is differentiable at (x_0,y_0) amounts to saying that the approximation delivered by this plane is very good if both $|x-x_0|$ and $|y-y_0|$ are small.



Example 12.3.12 *Suppose $f(x,y) = \sqrt{xy}$. Find the approximate change in f if x goes from 1 to 1.01 and y goes from 4 to 3.99.*

This can be done by noting that

$$\begin{aligned} f(1.01,3.99) - f(1,4) &\approx f_x(1,2)(.01) + f_y(1,2)(-.01) \\ &= 1(.01) + \frac{1}{4}(-.01) = 7.5 \times 10^{-3}. \end{aligned}$$

Of course the exact value would be

$$\sqrt{(1.01)(3.99)} - \sqrt{4} = 7.461\,083\,1 \times 10^{-3}.$$

12.4 The Chain Rule

12.4.1 The Chain Rule For Functions Of One Variable

First recall the chain rule for a function of one variable. Consider the following picture.

$$I \xrightarrow{g} J \xrightarrow{f} \mathbb{R}$$

Here I and J are open intervals and it is assumed that $g(I) \subseteq J$. The chain rule says that if $f'(g(x))$ exists and $g'(x)$ exists for $x \in I$, then the composition, $f \circ g$ also has a derivative at x and $(f \circ g)'(x) = f'(g(x))\, g'(x)$. Recall that $f \circ g$ is the name of the function defined by $f \circ g(x) \equiv f(g(x))$. In the notation of this chapter, the chain rule is written as

$$Df(g(x))\, Dg(x) = D(f \circ g)(x). \tag{12.11}$$

12.4.2 The Chain Rule For Functions Of Many Variables

Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^p$ be open sets and let $\mathbf{f}$ be a function defined on V having values in $\mathbb{R}^q$ while $\mathbf{g}$ is a function defined on U such that $\mathbf{g}(U) \subseteq V$ as in the following picture.

$$U \xrightarrow{\mathbf{g}} V \xrightarrow{\mathbf{f}} \mathbb{R}^q$$

The chain rule says that if the linear transformations (matrices) on the left in 12.11 both exist then the same formula holds in this more general case. Thus

$$D\mathbf{f}(\mathbf{g}(\mathbf{x}))\,D\mathbf{g}(\mathbf{x}) = D(\mathbf{f}\circ\mathbf{g})(\mathbf{x})$$

Note this all makes sense because $D\mathbf{f}(\mathbf{g}(\mathbf{x}))$ is a $q\times p$ matrix and $D\mathbf{g}(\mathbf{x})$ is a $p\times n$ matrix. Remember it is all right to do $(q\times p)(p\times n)$. The middle numbers match. More precisely,

Theorem 12.4.1 *(Chain rule) Let U be an open set in $\mathbb{R}^n$, let V be an open set in $\mathbb{R}^p$, let $\mathbf{g}: U\to\mathbb{R}^p$ be such that $\mathbf{g}(U)\subseteq V$, and let $\mathbf{f}: V\to\mathbb{R}^q$. Suppose $D\mathbf{g}(\mathbf{x})$ exists for some $\mathbf{x}\in U$ and that $D\mathbf{f}(\mathbf{g}(\mathbf{x}))$ exists. Then $D(\mathbf{f}\circ\mathbf{g})(\mathbf{x})$ exists and furthermore,*

$$D(\mathbf{f}\circ\mathbf{g})(\mathbf{x}) = D\mathbf{f}(\mathbf{g}(\mathbf{x}))\,D\mathbf{g}(\mathbf{x}). \tag{12.12}$$

In particular,

$$\frac{\partial(\mathbf{f}\circ\mathbf{g})(\mathbf{x})}{\partial x_j} = \sum_{i=1}^{p}\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i}\frac{\partial g_i(\mathbf{x})}{\partial x_j}. \tag{12.13}$$

There is an easy way to remember this in terms of the repeated index summation convention presented earlier. Let $\mathbf{y}=\mathbf{g}(\mathbf{x})$ and $\mathbf{z}=\mathbf{f}(\mathbf{y})$. Then the above says

$$\frac{\partial\mathbf{z}}{\partial y_i}\frac{\partial y_i}{\partial x_k} = \frac{\partial\mathbf{z}}{\partial x_k}. \tag{12.14}$$

Remember there is a sum on the repeated index. In particular, for each index, r,

$$\frac{\partial z_r}{\partial y_i}\frac{\partial y_i}{\partial x_k} = \frac{\partial z_r}{\partial x_k}.$$

The proof of this major theorem will be given at the end of this section. It will include the chain rule for functions of one variable as a special case. First here are some examples.

Example 12.4.2 *Let $f(u,v)=\sin(uv)$ and let $u(x,y,t)=t\sin x+\cos y$ and $v(x,y,t,s)=s\tan x+y^2+ts$. Letting $z=f(u,v)$ where u,v are as just described, find $\frac{\partial z}{\partial t}$ and $\frac{\partial z}{\partial x}$.*

From 12.14,

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial t}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial t} = v\cos(uv)\sin(x)+us\cos(uv).$$

Here $y_1=u, y_2=v, t=x_k$. Also,

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial x} = v\cos(uv)\,t\cos(x)+us\sec^2(x)\cos(uv).$$

Clearly you can continue in this way taking partial derivatives with respect to any of the other variables.

Example 12.4.3 *Let $w=f(u_1,u_2)=u_2\sin(u_1)$ and $u_1=x^2y+z, u_2=\sin(xy)$. Find $\frac{\partial w}{\partial x}$, $\frac{\partial w}{\partial y}$, and $\frac{\partial w}{\partial z}$.*

The derivative of f is of the form (w_x,w_y,w_z) and so it suffices to find the derivative of f using the chain rule. You need to find $Df(u_1,u_2)\,D\mathbf{g}(x,y,z)$ where $\mathbf{g}(x,y)=$

$\begin{pmatrix} x^2y+z \\ \sin(xy) \end{pmatrix}$. Then $D\mathbf{g}(x,y,z) = \begin{pmatrix} 2xy & x^2 & 1 \\ y\cos(xy) & x\cos(xy) & 0 \end{pmatrix}$. Also $Df(u_1,u_2) = (u_2\cos(u_1), \sin(u_1))$. Therefore, the derivative is

$$Df(u_1,u_2)\,D\mathbf{g}(x,y,z) = (u_2\cos(u_1),\sin(u_1)) \begin{pmatrix} 2xy & x^2 & 1 \\ y\cos(xy) & x\cos(xy) & 0 \end{pmatrix}$$

$$= \left(2u_2(\cos u_1)\,xy + (\sin u_1)\,y\cos xy,\, u_2(\cos u_1)\,x^2 + (\sin u_1)\,x\cos xy,\, u_2\cos u_1\right) = (w_x, w_y, w_z)$$

Thus $\frac{\partial w}{\partial x} = 2u_2(\cos u_1)\,xy + (\sin u_1)\,y\cos xy = 2(\sin(xy))\left(\cos\left(x^2y+z\right)\right)xy + \left(\sin\left(x^2y+z\right)\right)y\cos xy$. Similarly, you can find the other partial derivatives of w in terms of substituting in for u_1 and u_2 in the above. Note

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial u_1}\frac{\partial u_1}{\partial x} + \frac{\partial w}{\partial u_2}\frac{\partial u_2}{\partial x}.$$

In fact, in general if you have $w = f(u_1,u_2)$ and $\mathbf{g}(x,y,z) = \begin{pmatrix} u_1(x,y,z) \\ u_2(x,y,z) \end{pmatrix}$, then $D(f\circ\mathbf{g})(x,y,z)$ is of the form

$$\begin{aligned} & \begin{pmatrix} w_{u_1} & w_{u_2} \end{pmatrix} \begin{pmatrix} u_{1x} & u_{1y} & u_{1z} \\ u_{2x} & u_{2y} & u_{2z} \end{pmatrix} \\ = & \begin{pmatrix} w_{u_1}u_x + w_{u_2}u_{2x} & w_{u_1}u_y + w_{u_2}u_{2y} & w_{u_1}u_z + w_{u_2}u_{2z} \end{pmatrix}. \end{aligned}$$

Example 12.4.4 *Let* $w = f(u_1,u_2,u_3) = u_1^2 + u_3 + u_2$ *and* $\mathbf{g}(x,y,z) = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} x+2yz \\ x^2+y \\ z^2+x \end{pmatrix}$. *Find* $\frac{\partial w}{\partial x}$ *and* $\frac{\partial w}{\partial z}$.

By the chain rule,

$$(w_x, w_y, w_z) = \begin{pmatrix} w_{u_1} & w_{u_2} & w_{u_3} \end{pmatrix} \begin{pmatrix} u_{1x} & u_{1y} & u_{1z} \\ u_{2x} & u_{2y} & u_{2z} \\ u_{3x} & u_{3y} & u_{3z} \end{pmatrix}$$

$$= \begin{pmatrix} w_{u_1}u_{1x} + w_{u_2}u_{2x} + w_{u_3}u_{3x} & w_{u_1}u_{1y} + w_{u_2}u_{2y} + w_{u_3}u_{3y} & w_{u_1}u_{1z} + w_{u_2}u_{2z} + w_{u_3}u_{3z} \end{pmatrix}$$

Note the pattern.

$$\begin{aligned} w_x &= w_{u_1}u_{1x} + w_{u_2}u_{2x} + w_{u_3}u_{3x}, \\ w_y &= w_{u_1}u_{1y} + w_{u_2}u_{2y} + w_{u_3}u_{3y}, \\ w_z &= w_{u_1}u_{1z} + w_{u_2}u_{2z} + w_{u_3}u_{3z}. \end{aligned}$$

Therefore,

$$w_x = 2u_1(1) + 1(2x) + 1(1) = 2(x+2yz) + 2x + 1 = 4x + 4yz + 1$$

and

$$w_z = 2u_1(2y) + 1(0) + 1(2z) = 4(x+2yz)\,y + 2z = 4yx + 8y^2z + 2z.$$

Of course to find all the partial derivatives at once, you just use the chain rule. Thus you would get

$$\begin{aligned}\begin{pmatrix} w_x & w_y & w_z \end{pmatrix} &= \begin{pmatrix} 2u_1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2z & 2y \\ 2x & 1 & 0 \\ 1 & 0 & 2z \end{pmatrix} \\ &= \begin{pmatrix} 2u_1 + 2x + 1 & 4u_1 z + 1 & 4u_1 y + 2z \end{pmatrix} \\ &= \begin{pmatrix} 4x + 4yz + 1 & 4zx + 8yz^2 + 1 & 4yx + 8y^2 z + 2z \end{pmatrix}\end{aligned}$$

Example 12.4.5 *Let* $\mathbf{f}(u_1, u_2) = \begin{pmatrix} u_1^2 + u_2 \\ \sin(u_2) + u_1 \end{pmatrix}$ *and*

$$\mathbf{g}(x_1, x_2, x_3) = \begin{pmatrix} u_1(x_1, x_2, x_3) \\ u_2(x_1, x_2, x_3) \end{pmatrix} = \begin{pmatrix} x_1 x_2 + x_3 \\ x_2^2 + x_1 \end{pmatrix}.$$

Find $D(\mathbf{f} \circ \mathbf{g})(x_1, x_2, x_3)$.

To do this,

$$D\mathbf{f}(u_1, u_2) = \begin{pmatrix} 2u_1 & 1 \\ 1 & \cos u_2 \end{pmatrix}, D\mathbf{g}(x_1, x_2, x_3) = \begin{pmatrix} x_2 & x_1 & 1 \\ 1 & 2x_2 & 0 \end{pmatrix}.$$

Then

$$D\mathbf{f}(\mathbf{g}(x_1, x_2, x_3)) = \begin{pmatrix} 2(x_1 x_2 + x_3) & 1 \\ 1 & \cos\left(x_2^2 + x_1\right) \end{pmatrix}$$

and so by the chain rule,

$$D(\mathbf{f} \circ \mathbf{g})(x_1, x_2, x_3) =$$

$$\begin{aligned}&\overbrace{\begin{pmatrix} 2(x_1 x_2 + x_3) & 1 \\ 1 & \cos\left(x_2^2 + x_1\right) \end{pmatrix}}^{D\mathbf{f}(\mathbf{g}(\mathbf{x}))} \overbrace{\begin{pmatrix} x_2 & x_1 & 1 \\ 1 & 2x_2 & 0 \end{pmatrix}}^{D\mathbf{g}(\mathbf{x})} \\ = &\begin{pmatrix} (2x_1 x_2 + 2x_3) x_2 + 1 & (2x_1 x_2 + 2x_3) x_1 + 2x_2 & 2x_1 x_2 + 2x_3 \\ x_2 + \cos\left(x_2^2 + x_1\right) & x_1 + 2x_2 \left(\cos\left(x_2^2 + x_1\right)\right) & 1 \end{pmatrix}\end{aligned}$$

Therefore, in particular,

$$\frac{\partial f_1 \circ \mathbf{g}}{\partial x_1}(x_1, x_2, x_3) = (2x_1 x_2 + 2x_3) x_2 + 1,$$

$$\frac{\partial f_2 \circ \mathbf{g}}{\partial x_3}(x_1, x_2, x_3) = 1, \frac{\partial f_2 \circ \mathbf{g}}{\partial x_2}(x_1, x_2, x_3) = x_1 + 2x_2 \left(\cos\left(x_2^2 + x_1\right)\right).$$

etc.

In different notation, let $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \mathbf{f}(u_1, u_2) = \begin{pmatrix} u_1^2 + u_2 \\ \sin(u_2) + u_1 \end{pmatrix}$. Then

$$\frac{\partial z_1}{\partial x_1} = \frac{\partial z_1}{\partial u_1}\frac{\partial u_1}{\partial x_1} + \frac{\partial z_1}{\partial u_2}\frac{\partial u_2}{\partial x_1} = 2u_1 x_2 + 1 = 2(x_1 x_2 + x_3) x_2 + 1.$$

Example 12.4.6 *Let* $\mathbf{f}(u_1, u_2, u_3) = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} u_1^2 + u_2 u_3 \\ u_1^2 + u_2^3 \\ \ln\left(1 + u_3^2\right) \end{pmatrix}$ *and let*

$$\mathbf{g}(x_1, x_2, x_3, x_4) = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_2^2 + \sin(x_3) + \cos(x_4) \\ x_4^2 - x_1 \\ x_3^2 + x_4 \end{pmatrix}.$$

Find $(\mathbf{f} \circ \mathbf{g})'(\mathbf{x})$.

$$D\mathbf{f}(\mathbf{u}) = \begin{pmatrix} 2u_1 & u_3 & u_2 \\ 2u_1 & 3u_2^2 & 0 \\ 0 & 0 & \frac{2u_3}{(1+u_3^2)} \end{pmatrix}$$

Similarly,

$$D\mathbf{g}(\mathbf{x}) = \begin{pmatrix} 1 & 2x_2 & \cos(x_3) & -\sin(x_4) \\ -1 & 0 & 0 & 2x_4 \\ 0 & 0 & 2x_3 & 1 \end{pmatrix}.$$

Then by the chain rule, $D(\mathbf{f}\circ\mathbf{g})(\mathbf{x}) = D\mathbf{f}(\mathbf{u})\,D\mathbf{g}(\mathbf{x})$ where $\mathbf{u} = \mathbf{g}(\mathbf{x})$ as described above. Thus $D(\mathbf{f}\circ\mathbf{g})(\mathbf{x}) =$

$$\begin{pmatrix} 2u_1 & u_3 & u_2 \\ 2u_1 & 3u_2^2 & 0 \\ 0 & 0 & \frac{2u_3}{(1+u_3^2)} \end{pmatrix} \begin{pmatrix} 1 & 2x_2 & \cos(x_3) & -\sin(x_4) \\ -1 & 0 & 0 & 2x_4 \\ 0 & 0 & 2x_3 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2u_1 - u_3 & 4u_1x_2 & 2u_1\cos x_3 + 2u_2x_3 & -2u_1\sin x_4 + 2u_3x_4 + u_2 \\ 2u_1 - 3u_2^2 & 4u_1x_2 & 2u_1\cos x_3 & -2u_1\sin x_4 + 6u_2^2x_4 \\ 0 & 0 & 4\frac{u_3}{1+u_3^2}x_3 & 2\frac{u_3}{1+u_3^2} \end{pmatrix} \tag{12.15}$$

where each u_i is given by the above formulas. Thus $\frac{\partial z_1}{\partial x_1}$ equals

$$\begin{aligned} 2u_1 - u_3 &= 2\left(x_1 + x_2^2 + \sin(x_3) + \cos(x_4)\right) - \left(x_3^2 + x_4\right) \\ &= 2x_1 + 2x_2^2 + 2\sin x_3 + 2\cos x_4 - x_3^2 - x_4. \end{aligned}$$

while $\frac{\partial z_2}{\partial x_4}$ equals

$$-2u_1\sin x_4 + 6u_2^2x_4 = -2\left(x_1 + x_2^2 + \sin(x_3) + \cos(x_4)\right)\sin(x_4) + 6\left(x_4^2 - x_1\right)^2 x_4.$$

If you wanted $\frac{\partial \mathbf{z}}{\partial x_2}$ it would be the second column of the above matrix in 12.15. Thus $\frac{\partial \mathbf{z}}{\partial x_2}$ equals

$$\begin{pmatrix} \frac{\partial z_1}{\partial x_2} \\ \frac{\partial z_2}{\partial x_2} \\ \frac{\partial z_3}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 4u_1x_2 \\ 4u_1x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 4\left(x_1 + x_2^2 + \sin(x_3) + \cos(x_4)\right)x_2 \\ 4\left(x_1 + x_2^2 + \sin(x_3) + \cos(x_4)\right)x_2 \\ 0 \end{pmatrix}.$$

I hope that by now it is clear that all the information you could desire about various partial derivatives is available and it all reduces to matrix multiplication and the consideration of entries of the matrix obtained by multiplying the two derivatives.

12.4.3 Related Rates Problems

Sometimes several variables are related and given information about how one variable is changing, you want to find how the others are changing.The following law is discussed later in the book, on Page 387.

Example 12.4.7 *Bernoulli's law states that in an incompressible fluid,*

$$\frac{v^2}{2g} + z + \frac{P}{\gamma} = C$$

where C is a constant. Here v is the speed, P is the pressure, and z is the height above some reference point. The constants, g and γ are the acceleration of gravity and the weight density of the fluid. Suppose measurements indicate that $\frac{dv}{dt} = -3$, and $\frac{dz}{dt} = 2$. Find $\frac{dP}{dt}$ when $v = 7$ and $z = 8$ in terms of g and γ.

This is just an exercise in using the chain rule. Differentiate the two sides with respect to t.

$$\frac{1}{g}v\frac{dv}{dt}+\frac{dz}{dt}+\frac{1}{\gamma}\frac{dP}{dt}=0.$$

Then when $v=7$ and $z=8$, finding $\frac{dP}{dt}$ involves nothing more than solving the following for $\frac{dP}{dt}$.

$$\frac{7}{g}(-3)+2+\frac{1}{\gamma}\frac{dP}{dt}=0$$

Thus

$$\frac{dP}{dt}=\gamma\left(\frac{21}{g}-2\right)$$

at this instant in time.

Example 12.4.8 *In Bernoulli's law above, each of v, z, and P are functions of (x,y,z), the position of a point in the fluid. Find a formula for $\frac{\partial P}{\partial x}$ in terms of the partial derivatives of the other variables.*

This is an example of the chain rule. Differentiate both sides with respect to x.

$$\frac{v}{g}v_x+z_x+\frac{1}{\gamma}P_x=0$$

and so

$$P_x=-\left(\frac{vv_x+z_xg}{g}\right)\gamma$$

Example 12.4.9 *Suppose a level curve is of the form $f(x,y)=C$ and that near a point on this level curve, y is a differentiable function of x. Find $\frac{dy}{dx}$.*

This is an example of the chain rule. Differentiate both sides with respect to x. This gives

$$f_x+f_y\frac{dy}{dx}=0.$$

Solving for $\frac{dy}{dx}$ gives

$$\frac{dy}{dx}=\frac{-f_x(x,y)}{f_y(x,y)}.$$

Example 12.4.10 *Suppose a level surface is of the form $f(x,y,z)=C$. and that near a point, (x,y,z) on this level surface, z is a C^1 function of x and y. Find a formula for z_x.*

This is an exaple of the use of the chain rule. Differentiate both sides of the equation with respect to x. Since $y_x=0$, this yields

$$f_x+f_zz_x=0.$$

Then solving for z_x gives

$$z_x=\frac{-f_x(x,y,z)}{f_z(x,y,z)}$$

Example 12.4.11 *Polar coordinates are*

$$x=r\cos\theta,\ y=r\sin\theta.$$

Thus if f is a C^1 scalar valued function you could ask to express f_x in terms of the variables, r and θ. Do so.

This is an example of the chain rule. $f = f(r,\theta)$ and so

$$f_x = f_r r_x + f_\theta \theta_x.$$

This will be done if you can find r_x and θ_x. However you must find these in terms of r and θ, not in terms of x and y. Using the chain rule on the two equations for the transformation,

$$\begin{aligned} 1 &= r_x \cos\theta - (r\sin\theta)\,\theta_x \\ 0 &= r_x \sin\theta + (r\cos\theta)\,\theta_x \end{aligned}$$

Solving these using Cramer's rule yields

$$r_x = \cos(\theta),\ \theta_x = \frac{-\sin(\theta)}{r}$$

Hence f_x in polar coordinates is

$$f_x = f_r(r,\theta)\cos(\theta) - f_\theta(r,\theta)\left(\frac{\sin(\theta)}{r}\right)$$

12.4.4 The Derivative Of The Inverse Function

Example 12.4.12 *Let $\mathbf{f}: U \to V$ where U and V are open sets in $\mathbb{R}^n$ and $\mathbf{f}$ is one to one and onto. Suppose also that $\mathbf{f}$ and $\mathbf{f}^{-1}$ are both differentiable. How are $D\mathbf{f}^{-1}$ and $D\mathbf{f}$ related?*

This can be done as follows. From the assumptions, $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{f}(\mathbf{x}))$. Let $I\mathbf{x} = \mathbf{x}$. Then by Example 12.2.4 on Page 246 $DI = I$. By the chain rule,

$$I = DI = D\mathbf{f}^{-1}(\mathbf{f}(\mathbf{x}))\,(D\mathbf{f}(\mathbf{x})).$$

Therefore,

$$D\mathbf{f}(\mathbf{x})^{-1} = D\mathbf{f}^{-1}(\mathbf{f}(\mathbf{x})).$$

This is equivalent to

$$D\mathbf{f}\left(\mathbf{f}^{-1}(\mathbf{y})\right)^{-1} = D\mathbf{f}^{-1}(\mathbf{y})$$

or

$$D\mathbf{f}(\mathbf{x})^{-1} = D\mathbf{f}^{-1}(\mathbf{y}),\ \mathbf{y} = \mathbf{f}(\mathbf{x}).$$

This is just like a similar situation for functions of one variable. Remember

$$\left(f^{-1}\right)'(f(x)) = 1/f'(x).$$

In terms of the repeated index summation convention, suppose $\mathbf{y} = \mathbf{f}(\mathbf{x})$ so that $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y})$. Then the above can be written as

$$\delta_{ij} = \frac{\partial x_i}{\partial y_k}(\mathbf{f}(\mathbf{x}))\,\frac{\partial y_k}{\partial x_j}(\mathbf{x}).$$

12.4.5 Acceleration In Spherical Coordinates*

Example 12.4.13 *Recall spherical coordinates are given by*

$$x = \rho\sin\phi\cos\theta,\ y = \rho\sin\phi\sin\theta,\ z = \rho\cos\phi.$$

If an object moves in three dimensions, describe its acceleration in terms of spherical coordinates and the vectors,

$$\mathbf{e}_\rho = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)^T,$$

$$\mathbf{e}_\theta = (-\rho\sin\phi\sin\theta, \rho\sin\phi\cos\theta, 0)^T,$$

and

$$\mathbf{e}_\phi = (\rho\cos\phi\cos\theta, \rho\cos\phi\sin\theta, -\rho\sin\phi)^T.$$

Why these vectors? Note how they were obtained. Let

$$\mathbf{r}(\rho,\theta,\phi) = (\rho\sin\phi\cos\theta,\ \rho\sin\phi\sin\theta,\ \rho\cos\phi)^T$$

and fix ϕ and θ, letting only ρ change, this gives a curve in the direction of increasing ρ. Thus it is a vector which points away from the origin. Letting only ϕ change and fixing θ and ρ, this gives a vector which is tangent to the sphere of radius ρ and points South. Similarly, letting θ change and fixing the other two gives a vector which points East and is tangent to the sphere of radius ρ. It is thought by most people that we live on a large sphere. The model of a flat earth is not believed by anyone except perhaps beginning physics students. Given we live on a sphere, what directions would be most meaningful? Wouldn't it be the directions of the vectors just described?

Let $\mathbf{r}(t)$ denote the position vector of the object from the origin. Thus

$$\mathbf{r}(t) = \rho(t)\,\mathbf{e}_\rho(t) = \left((x(t), y(t), z(t))^T\right)$$

Now this implies the velocity is

$$\mathbf{r}'(t) = \rho'(t)\,\mathbf{e}_\rho(t) + \rho(t)\,(\mathbf{e}_\rho(t))'. \tag{12.16}$$

You see, $\mathbf{e}_\rho = \mathbf{e}_\rho(\rho,\theta,\phi)$ where each of these variables is a function of t.

$$\frac{\partial \mathbf{e}_\rho}{\partial\phi} = (\cos\phi\cos\theta, \cos\phi\sin\theta, -\sin\phi)^T = \frac{1}{\rho}\mathbf{e}_\phi,$$

$$\frac{\partial \mathbf{e}_\rho}{\partial\theta} = (-\sin\phi\sin\theta, \sin\phi\cos\theta, 0)^T = \frac{1}{\rho}\mathbf{e}_\theta,$$

and

$$\frac{\partial \mathbf{e}_\rho}{\partial\rho} = 0.$$

Therefore, by the chain rule,

$$\begin{aligned}\frac{d\mathbf{e}_\rho}{dt} &= \frac{\partial \mathbf{e}_\rho}{\partial\phi}\frac{d\phi}{dt} + \frac{\partial \mathbf{e}_\rho}{\partial\theta}\frac{d\theta}{dt}\\ &= \frac{1}{\rho}\frac{d\phi}{dt}\mathbf{e}_\phi + \frac{1}{\rho}\frac{d\theta}{dt}\mathbf{e}_\theta.\end{aligned}$$

By 12.16,

$$\mathbf{r}' = \rho'\mathbf{e}_\rho + \frac{d\phi}{dt}\mathbf{e}_\phi + \frac{d\theta}{dt}\mathbf{e}_\theta. \tag{12.17}$$

Now things get interesting. This must be differentiated with respect to t. To do so,

$$\frac{\partial \mathbf{e}_\theta}{\partial \theta} = (-\rho \sin\phi \cos\theta, -\rho \sin\phi \sin\theta, 0)^T = ?$$

where it is desired to find a, b, c such that $? = a\mathbf{e}_\theta + b\mathbf{e}_\phi + c\mathbf{e}_\rho$. Thus

$$\begin{pmatrix} -\rho \sin\phi \sin\theta & \rho \cos\phi \cos\theta & \sin\phi \cos\theta \\ \rho \sin\phi \cos\theta & \rho \cos\phi \sin\theta & \sin\phi \sin\theta \\ 0 & -\rho \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -\rho \sin\phi \cos\theta \\ -\rho \sin\phi \sin\theta \\ 0 \end{pmatrix}$$

Using Cramer's rule, the solution is $a = 0, b = -\cos\phi \sin\phi$, and $c = -\rho \sin^2\phi$. Thus

$$\begin{aligned} \frac{\partial \mathbf{e}_\theta}{\partial \theta} &= (-\rho \sin\phi \cos\theta, -\rho \sin\phi \sin\theta, 0)^T \\ &= (-\cos\phi \sin\phi)\, \mathbf{e}_\phi + \left(-\rho \sin^2\phi\right) \mathbf{e}_\rho. \end{aligned}$$

Also,

$$\frac{\partial \mathbf{e}_\theta}{\partial \phi} = (-\rho \cos\phi \sin\theta, \rho \cos\phi \cos\theta, 0)^T = (\cot\phi)\, \mathbf{e}_\theta$$

and

$$\frac{\partial \mathbf{e}_\theta}{\partial \rho} = (-\sin\phi \sin\theta, \sin\phi \cos\theta, 0)^T = \frac{1}{\rho}\mathbf{e}_\theta.$$

Now in 12.17 it is also necessary to consider $\mathbf{e}_\phi$.

$$\frac{\partial \mathbf{e}_\phi}{\partial \phi} = (-\rho \sin\phi \cos\theta, -\rho \sin\phi \sin\theta, -\rho \cos\phi)^T = -\rho \mathbf{e}_\rho$$

$$\begin{aligned} \frac{\partial \mathbf{e}_\phi}{\partial \theta} &= (-\rho \cos\phi \sin\theta, \rho \cos\phi \cos\theta, 0)^T \\ &= (\cot\phi)\, \mathbf{e}_\theta \end{aligned}$$

and finally,

$$\frac{\partial \mathbf{e}_\phi}{\partial \rho} = (\cos\phi \cos\theta, \cos\phi \sin\theta, -\sin\phi)^T = \frac{1}{\rho}\mathbf{e}_\phi.$$

With these formulas for various partial derivatives, the chain rule is used to obtain $\mathbf{r}''$ which will yield a formula for the acceleration in terms of the spherical coordinates and these special vectors. By the chain rule,

$$\begin{aligned} \frac{d}{dt}(\mathbf{e}_\rho) &= \frac{\partial \mathbf{e}_\rho}{\partial \theta}\theta' + \frac{\partial \mathbf{e}_\rho}{\partial \phi}\phi' + \frac{\partial \mathbf{e}_\rho}{\partial \rho}\rho' \\ &= \frac{\theta'}{\rho}\mathbf{e}_\theta + \frac{\phi'}{\rho}\mathbf{e}_\phi \end{aligned}$$

$$\begin{aligned} \frac{d}{dt}(\mathbf{e}_\theta) &= \frac{\partial \mathbf{e}_\theta}{\partial \theta}\theta' + \frac{\partial \mathbf{e}_\theta}{\partial \phi}\phi' + \frac{\partial \mathbf{e}_\theta}{\partial \rho}\rho' \\ &= \theta'\left((-\cos\phi \sin\phi)\, \mathbf{e}_\phi + \left(-\rho \sin^2\phi\right) \mathbf{e}_\rho\right) + \phi'(\cot\phi)\, \mathbf{e}_\theta + \frac{\rho'}{\rho}\mathbf{e}_\theta \end{aligned}$$

$$\begin{aligned} \frac{d}{dt}(\mathbf{e}_\phi) &= \frac{\partial \mathbf{e}_\phi}{\partial \theta}\theta' + \frac{\partial \mathbf{e}_\phi}{\partial \phi}\phi' + \frac{\partial \mathbf{e}_\phi}{\partial \rho}\rho' \\ &= \left(\theta' \cot\phi\right) \mathbf{e}_\theta + \phi'(-\rho \mathbf{e}_\rho) + \left(\frac{\rho'}{\rho}\mathbf{e}_\phi\right) \end{aligned}$$

By 12.17,

$$\mathbf{r}'' = \rho''\mathbf{e}_\rho + \phi''\mathbf{e}_\phi + \theta''\mathbf{e}_\theta + \rho'\left(\mathbf{e}_\rho\right)' + \phi'\left(\mathbf{e}_\phi\right)' + \theta'\left(\mathbf{e}_\theta\right)'$$

and from the above, this equals

$$\begin{aligned}
&\rho''\mathbf{e}_\rho + \phi''\mathbf{e}_\phi + \theta''\mathbf{e}_\theta + \rho'\left(\frac{\theta'}{\rho}\mathbf{e}_\theta + \frac{\phi'}{\rho}\mathbf{e}_\phi\right) + \\
&\phi'\left(\left(\theta'\cot\phi\right)\mathbf{e}_\theta + \phi'\left(-\rho\mathbf{e}_\rho\right) + \left(\frac{\rho'}{\rho}\mathbf{e}_\phi\right)\right) + \\
&\theta'\left(\theta'\left(\left(-\cos\phi\sin\phi\right)\mathbf{e}_\phi + \left(-\rho\sin^2\phi\right)\mathbf{e}_\rho\right) + \phi'\left(\cot\phi\right)\mathbf{e}_\theta + \frac{\rho'}{\rho}\mathbf{e}_\theta\right)
\end{aligned}$$

and now all that remains is to collect the terms. Thus $\mathbf{r}''$ equals

$$\begin{aligned}
\mathbf{r}'' &= \left(\rho'' - \rho\left(\phi'\right)^2 - \rho\left(\theta'\right)^2\sin^2\left(\phi\right)\right)\mathbf{e}_\rho + \left(\phi'' + \frac{2\rho'\phi'}{\rho} - \left(\theta'\right)^2\cos\phi\sin\phi\right)\mathbf{e}_\phi + \\
&\quad + \left(\theta'' + \frac{2\theta'\rho'}{\rho} + 2\phi'\theta'\cot\left(\phi\right)\right)\mathbf{e}_\theta.
\end{aligned}$$

and this gives the acceleration in spherical coordinates. Note the prominent role played by the chain rule. All of the above is done in books on mechanics for general curvilinear coordinate systems and in the more general context, special theorems are developed which make things go much faster but these theorems are all exercises in the chain rule.

As an example of how this could be used, consider a rocket. Suppose for simplicity that it experiences a force only in the direction of $\mathbf{e}_\rho$, directly away from the earth. Of course this force produces a corresponding acceleration which can be computed as a function of time. As the fuel is burned, the rocket becomes less massive and so the acceleration will be an increasing function of t. However, this would be a known function, say $a\left(t\right)$. Suppose you wanted to know the latitude and longitude of the rocket as a function of time. (There is no reason to think these will stay the same.) Then all that would be required would be to solve the system of differential equations[1],

$$\begin{aligned}
\rho'' - \rho\left(\phi'\right)^2 - \rho\left(\theta'\right)^2\sin^2\left(\phi\right) &= a\left(t\right), \\
\phi'' + \frac{2\rho'\phi'}{\rho} - \left(\theta'\right)^2\cos\phi\sin\phi &= 0, \\
\theta'' + \frac{2\theta'\rho'}{\rho} + 2\phi'\theta'\cot\left(\phi\right) &= 0
\end{aligned}$$

along with initial conditions, $\rho\left(0\right) = \rho_0$ (the distance from the launch site to the center of the earth.), $\rho'\left(0\right) = \rho_1$(the initial vertical component of velocity of the rocket, probably 0.) and then initial conditions for $\phi, \phi', \theta, \theta'$. The initial value problems could then be solved numerically and you would know the distance from the center of the earth as a function of t along with θ and ϕ. Thus you could predict where the booster shells would fall to earth so you would know where to look for them. Of course there are many variations of this. You might want to specify forces in the $\mathbf{e}_\theta$ and $\mathbf{e}_\phi$ direction as well and attempt to control the position of the rocket or rather its payload. The point is that if you are interested in doing all this in terms of ϕ, θ, and ρ, the above shows how to do it systematically and you see it is all an exercise in using the chain rule. More could be said here involving moving coordinate systems and the Coriolis force. You really might want to do everything with respect to a coordinate system which is fixed with respect to the moving earth.

[1]You won't be able to find the solution to equations like these in terms of simple functions. The existence of such functions is being assumed. The reason they exist often depends on the implicit function theorem, a big theorem in advanced calculus.

12.4.6 Proof Of The Chain Rule

As in the case of a function of one variable, it is important to consider the derivative of a composition of two functions. The proof of the chain rule depends on the following fundamental lemma.

Lemma 12.4.14 *Let $\mathbf{g} : U \to \mathbb{R}^p$ where U is an open set in $\mathbb{R}^n$ and suppose $\mathbf{g}$ has a derivative at $\mathbf{x} \in U$. Then $\mathbf{o}(\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x})) = \mathbf{o}(\mathbf{v})$.*

Proof: It is necessary to show

$$\lim_{\mathbf{v}\to\mathbf{0}} \frac{|\mathbf{o}(\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x}))|}{|\mathbf{v}|} = 0. \tag{12.18}$$

From Lemma 12.3.11, there exists $\delta > 0$ such that if $|\mathbf{v}| < \delta$, then

$$|\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x})| \leq (Cn+1)|\mathbf{v}|. \tag{12.19}$$

Now let $\varepsilon > 0$ be given. There exists $\eta > 0$ such that if $|\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x})| < \eta$, then

$$|\mathbf{o}(\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x}))| < \left(\frac{\varepsilon}{Cn+1}\right)|\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x})| \tag{12.20}$$

Let $|\mathbf{v}| < \min\left(\delta, \frac{\eta}{Cn+1}\right)$. For such $\mathbf{v}$, $|\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x})| \leq \eta$, which implies

$$\begin{aligned} |\mathbf{o}(\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x}))| &< \left(\frac{\varepsilon}{Cn+1}\right)|\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x})| \\ &< \left(\frac{\varepsilon}{Cn+1}\right)(Cn+1)|\mathbf{v}| \end{aligned}$$

and so

$$\frac{|\mathbf{o}(\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x}))|}{|\mathbf{v}|} < \varepsilon$$

which establishes 12.18. This proves the lemma.

Recall the notation $\mathbf{f} \circ \mathbf{g}(\mathbf{x}) \equiv \mathbf{f}(\mathbf{g}(\mathbf{x}))$. Thus $\mathbf{f} \circ \mathbf{g}$ is the name of a function and this function is defined by what was just written. The following theorem is known as the **chain rule**.

Theorem 12.4.15 *(Chain rule) Let U be an open set in $\mathbb{R}^n$, let V be an open set in $\mathbb{R}^p$, let $\mathbf{g} : U \to \mathbb{R}^p$ be such that $\mathbf{g}(U) \subseteq V$, and let $\mathbf{f} : V \to \mathbb{R}^q$. Suppose $D\mathbf{g}(\mathbf{x})$ exists for some $\mathbf{x} \in U$ and that $D\mathbf{f}(\mathbf{g}(\mathbf{x}))$ exists. Then $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$ exists and furthermore,*

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = D\mathbf{f}(\mathbf{g}(\mathbf{x}))\, D\mathbf{g}(\mathbf{x}). \tag{12.21}$$

In particular,

$$\frac{\partial(\mathbf{f} \circ \mathbf{g})(\mathbf{x})}{\partial x_j} = \sum_{i=1}^{p} \frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i} \frac{\partial g_i(\mathbf{x})}{\partial x_j}. \tag{12.22}$$

Proof: From the assumption that $D\mathbf{f}(\mathbf{g}(\mathbf{x}))$ exists,

$$\mathbf{f}(\mathbf{g}(\mathbf{x}+\mathbf{v})) = \mathbf{f}(\mathbf{g}(\mathbf{x})) + \sum_{i=1}^{p} \frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i}(g_i(\mathbf{x}+\mathbf{v}) - g_i(\mathbf{x})) + \mathbf{o}(\mathbf{g}(\mathbf{x}+\mathbf{v}) - \mathbf{g}(\mathbf{x}))$$

which by Lemma 12.4.14 equals

$$(\mathbf{f} \circ \mathbf{g})(\mathbf{x}+\mathbf{v}) = \mathbf{f}(\mathbf{g}(\mathbf{x}+\mathbf{v})) = \mathbf{f}(\mathbf{g}(\mathbf{x})) + \sum_{i=1}^{p} \frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i}(g_i(\mathbf{x}+\mathbf{v}) - g_i(\mathbf{x})) + \mathbf{o}(\mathbf{v}).$$

Now since $D\mathbf{g}(\mathbf{x})$ exists, the above becomes

$$\begin{aligned}(\mathbf{f}\circ\mathbf{g})(\mathbf{x}+\mathbf{v}) &= \mathbf{f}(\mathbf{g}(\mathbf{x}))+\sum_{i=1}^{p}\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i}\left(\sum_{j=1}^{n}\frac{\partial g_i(\mathbf{x})}{\partial x_j}v_j+\mathbf{o}(\mathbf{v})\right)+\mathbf{o}(\mathbf{v})\\ &= \mathbf{f}(\mathbf{g}(\mathbf{x}))+\sum_{i=1}^{p}\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i}\left(\sum_{j=1}^{n}\frac{\partial g_i(\mathbf{x})}{\partial x_j}v_j\right)+\sum_{i=1}^{p}\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i}\mathbf{o}(\mathbf{v})+\mathbf{o}(\mathbf{v})\\ &= (\mathbf{f}\circ\mathbf{g})(\mathbf{x})+\sum_{j=1}^{n}\left(\sum_{i=1}^{p}\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i}\frac{\partial g_i(\mathbf{x})}{\partial x_j}\right)v_j+\mathbf{o}(\mathbf{v})\end{aligned}$$

because $\sum_{i=1}^{p}\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i}\mathbf{o}(\mathbf{v})+\mathbf{o}(\mathbf{v})=\mathbf{o}(\mathbf{v})$. This establishes 12.22 because of Theorem 12.2.3 on Page 245. Thus

$$\begin{aligned}(D(\mathbf{f}\circ\mathbf{g})(\mathbf{x}))_{kj} &= \sum_{i=1}^{p}\frac{\partial f_k(\mathbf{g}(\mathbf{x}))}{\partial y_i}\frac{\partial g_i(\mathbf{x})}{\partial x_j}\\ &= \sum_{i=1}^{p}D\mathbf{f}(\mathbf{g}(\mathbf{x}))_{ki}(D\mathbf{g}(\mathbf{x}))_{ij}.\end{aligned}$$

Then 12.21 follows from the definition of matrix multiplication.

12.5 Lagrangian Mechanics*

A difficult and important problem is to come up with differential equations which model mechanical systems. Lagrange gave a way to do this. It will be presented here as a very interesting and important application of the chain rule. Lagrange developed this technique back in the 1700's. The presentation here follows [12]. Assume N point masses, located at the points $\mathbf{x}_1,\cdots,\mathbf{x}_N$ in $\mathbb{R}^3$ and let the mass of the α^{th} mass be m_α. Then according to Newton's second law,

$$m_\alpha\mathbf{x}_\alpha''=\mathbf{F}_\alpha(\mathbf{x}_\alpha,t). \tag{12.23}$$

The dependence of $\mathbf{F}_\alpha$ on the two indicated quantities is indicative of the situation where the force may change in time and position. Now define

$$\mathbf{x}\equiv(\mathbf{x}_1,\cdots,\mathbf{x}_N)\in\mathbb{R}^{3N}$$

and assume $\mathbf{x}\in M$ which is defined locally in the form $\mathbf{x}=\mathbf{G}(\mathbf{q},t)$. Here $\mathbf{q}\in\mathbb{R}^m$ where typically $m<3N$ and $\mathbf{G}(\cdot,t)$ is a smooth one to one mapping from V, an open subset of $\mathbb{R}^m$ onto a set of points near $\mathbf{x}$ which are on M. Also assume t is in an open subset of $\mathbb{R}$. In what follows a dot over a variable will indicate a derivative taken with respect to time. Two dots will indicate the second derivative with respect to time, etc. Then define $\mathbf{G}_\alpha$ by

$$\mathbf{x}_\alpha=\mathbf{G}_\alpha(\mathbf{q},t).$$

Using the summation convention and the chain rule,

$$\frac{d\mathbf{x}_\alpha}{dt}=\frac{\partial\mathbf{G}_\alpha}{\partial q^j}\frac{dq^j}{dt}+\frac{\partial\mathbf{G}_\alpha}{\partial t}.$$

Therefore, the kinetic energy is of the form

$$\begin{aligned} T &\equiv \sum_{\alpha=1}^{N} \frac{1}{2} m_\alpha \left(\frac{d\mathbf{x}_\alpha}{dt} \cdot \frac{d\mathbf{x}_\alpha}{dt} \right) \\ &= \sum_{\alpha=1}^{N} \frac{1}{2} m_\alpha \left(\sum_j \frac{\partial \mathbf{G}_\alpha}{\partial q^j} \frac{dq^j}{dt} + \frac{\partial \mathbf{G}_\alpha}{\partial t} \cdot \sum_r \frac{\partial \mathbf{G}_\alpha}{\partial q^r} \frac{dq^r}{dt} + \frac{\partial \mathbf{G}_\alpha}{\partial t} \right) \end{aligned}$$

$$= \sum_{j,r} \frac{1}{2} \left[\sum_\alpha m_\alpha \left(\frac{\partial \mathbf{G}_\alpha}{\partial q^j} \cdot \frac{\partial \mathbf{G}_\alpha}{\partial q^r} \right) \right] \dot{q}^r \dot{q}^j + \sum_\alpha \sum_j m_\alpha \left(\frac{\partial \mathbf{G}_\alpha}{\partial q^j} \cdot \frac{\partial \mathbf{G}_\alpha}{\partial t} \right) \dot{q}^j$$

$$+ \sum_\alpha \frac{1}{2} m_\alpha \left(\frac{\partial \mathbf{G}_\alpha}{\partial t} \cdot \frac{\partial \mathbf{G}_\alpha}{\partial t} \right) \tag{12.24}$$

where in the last equation $\dot{q}^k$ indicates $\frac{dq^k}{dt}$. Therefore,

$$\begin{aligned} \frac{\partial T}{\partial \dot{q}^k} &= \sum_{j=1}^{m} \left[\sum_{\alpha=1}^{N} m_\alpha \left(\frac{\partial \mathbf{G}_\alpha}{\partial q^j} \cdot \frac{\partial \mathbf{G}_\alpha}{\partial q^k} \right) \right] \dot{q}^j + \sum_\alpha m_\alpha \left(\frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot \frac{\partial \mathbf{G}_\alpha}{\partial t} \right) \\ &= \sum_{\alpha=1}^{N} \left(m_\alpha \frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot \sum_{j=1}^{m} \frac{\partial \mathbf{G}_\alpha}{\partial q^j} \dot{q}^j \right) + \sum_\alpha m_\alpha \left(\frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot \frac{\partial \mathbf{G}_\alpha}{\partial t} \right) \\ &= \left(\sum_{\alpha=1}^{N} m_\alpha \frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot \left(\mathbf{x}'_\alpha - \frac{\partial \mathbf{G}_\alpha}{\partial t} \right) \right) + \sum_\alpha m_\alpha \left(\frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot \frac{\partial \mathbf{G}_\alpha}{\partial t} \right) \\ &= \sum_{\alpha=1}^{N} \frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot m_\alpha \mathbf{x}'_\alpha \end{aligned}$$

Now using the chain rule and product rule again, along with Newton's second law,

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}^k} \right) &= \left(\sum_{\alpha=1}^{N} \left[\left(\sum_j \frac{\partial^2 \mathbf{G}_\alpha}{\partial q^k \partial q^j} \dot{q}^j \right) + \frac{\partial^2 \mathbf{G}_\alpha}{\partial t \partial q^k} \right] \cdot m_\alpha \mathbf{x}'_\alpha \right) \\ &\quad + \left(\sum_{\alpha=1}^{N} \frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot m_\alpha \mathbf{x}''_\alpha \right) \\ &= \left(\sum_{\alpha=1}^{N} \left[\left(\sum_j \frac{\partial^2 \mathbf{G}_\alpha}{\partial q^k \partial q^j} \dot{q}^j \right) + \frac{\partial^2 \mathbf{G}_\alpha}{\partial t \partial q^k} \right] \cdot m_\alpha \mathbf{x}'_\alpha \right) + \\ &\quad + \left(\sum_{\alpha=1}^{N} \frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot \mathbf{F}_\alpha \right) \\ &= \left(\sum_{\alpha=1}^{N} \left[\left(\sum_j \frac{\partial^2 \mathbf{G}_\alpha}{\partial q^k \partial q^j} \dot{q}^j \right) + \frac{\partial^2 \mathbf{G}_\alpha}{\partial t \partial q^k} \right] \cdot \right. \\ &\quad \left. m_\alpha \left(\sum_r \frac{\partial \mathbf{G}_\alpha}{\partial q^r} \dot{q}^r + \frac{\partial \mathbf{G}_\alpha}{\partial t} \right) \right) + \left(\sum_{\alpha=1}^{N} \frac{\partial \mathbf{G}_\alpha}{\partial q^k} \cdot \mathbf{F}_\alpha \right) \end{aligned} \tag{12.25}$$

$$= \sum_{rj}\left[\sum_{\alpha=1}^{N} m_\alpha\left(\frac{\partial^2 \mathbf{G}_\alpha}{\partial q^j \partial q^k}\cdot\frac{\partial \mathbf{G}_\alpha}{\partial q^r}\right)\right]\dot{q}^r\dot{q}^j + \sum_{\alpha=1}^{N}\sum_j \frac{\partial^2 \mathbf{G}_\alpha}{\partial q^k \partial q^j}\dot{q}^j \cdot m_\alpha \frac{\partial \mathbf{G}_\alpha}{\partial t}$$
$$+\left(\sum_\alpha\sum_r \frac{\partial^2 \mathbf{G}_\alpha}{\partial t \partial q^k}\cdot m_\alpha \frac{\partial \mathbf{G}_\alpha}{\partial q^r}\dot{q}^r\right) + \sum_\alpha \frac{\partial^2 \mathbf{G}_\alpha}{\partial t \partial q^k}\cdot m_\alpha \frac{\partial \mathbf{G}_\alpha}{\partial t} + \left(\sum_{\alpha=1}^{N}\frac{\partial \mathbf{G}_\alpha}{\partial q^k}\cdot \mathbf{F}_\alpha\right) \tag{12.26}$$

Next consider $\frac{\partial T}{\partial q^k}$. Recall 12.24,

$$T = \sum_{j,r}\frac{1}{2}\left[\sum_\alpha m_\alpha\left(\frac{\partial \mathbf{G}_\alpha}{\partial q^j}\cdot\frac{\partial \mathbf{G}_\alpha}{\partial q^r}\right)\right]\dot{q}^r\dot{q}^j + \sum_\alpha\sum_j m_\alpha\left(\frac{\partial \mathbf{G}_\alpha}{\partial q^j}\cdot\frac{\partial \mathbf{G}_\alpha}{\partial t}\right)\dot{q}^j$$
$$+\sum_\alpha \frac{1}{2}m_\alpha\left(\frac{\partial \mathbf{G}_\alpha}{\partial t}\cdot\frac{\partial \mathbf{G}_\alpha}{\partial t}\right) \tag{12.27}$$

From this formula,

$$\frac{\partial T}{\partial q^k} = \sum_{rj}\left[\sum_{\alpha=1}^{N} m_\alpha\left(\frac{\partial^2 \mathbf{G}_\alpha}{\partial q^j \partial q^k}\cdot\frac{\partial \mathbf{G}_\alpha}{\partial q^r}\right)\right]\dot{q}^r\dot{q}^j +$$
$$\sum_\alpha\sum_j m_\alpha\left(\frac{\partial^2 \mathbf{G}_\alpha}{\partial q^k \partial q^j}\cdot\frac{\partial \mathbf{G}_\alpha}{\partial t}\right)\dot{q}^j + \sum_\alpha\sum_j m_\alpha\left(\frac{\partial \mathbf{G}_\alpha}{\partial q^j}\cdot\frac{\partial^2 \mathbf{G}_\alpha}{\partial q^k \partial t}\right)\dot{q}^j$$
$$+\sum_\alpha m_\alpha\left(\frac{\partial^2 \mathbf{G}_\alpha}{\partial q^k \partial t}\cdot\frac{\partial \mathbf{G}_\alpha}{\partial t}\right). \tag{12.28}$$

Now upon comparing 12.28 and 12.26

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}^k}\right) - \frac{\partial T}{\partial q^k} = \sum_{\alpha=1}^{N}\frac{\partial \mathbf{G}_\alpha}{\partial q^k}\cdot\mathbf{F}_\alpha.$$

Resolve the force, $\mathbf{F}_\alpha$ into the sum of two forces, $\mathbf{F}_\alpha = \mathbf{F}_\alpha^a + \mathbf{F}_\alpha^c$ where $\mathbf{F}_\alpha^c$ is a force of constraint which is perpendcular to $\frac{\partial \mathbf{G}_\alpha}{\partial q^k}$ and the other force, $\mathbf{F}_\alpha^a$ which is left over is called the applied force. The applied force is allowed to have a component which is perpendicular to $\frac{\partial \mathbf{G}_\alpha}{\partial q^k}$. The only requirement of this sort is placed on $\mathbf{F}_\alpha^c$. Therefore,

$$\frac{\partial \mathbf{G}_\alpha}{\partial q^k}\cdot\mathbf{F}_\alpha = \frac{\partial \mathbf{G}_\alpha}{\partial q^k}\cdot\mathbf{F}_\alpha^a$$

and so in the end, you obtain the following interesting equation which is equivalent to Newton's second law.

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}^k}\right) - \frac{\partial T}{\partial q^k} = \sum_{\alpha=1}^{N}\frac{\partial \mathbf{G}_\alpha}{\partial q^k}\cdot\mathbf{F}_\alpha^a \tag{12.29}$$
$$= \frac{\partial \mathbf{G}}{\partial q^k}\cdot\mathbf{F}^a, \tag{12.30}$$

where $\mathbf{F}^a \equiv (\mathbf{F}_1^\alpha, \cdots, \mathbf{F}_N^\alpha)$ is referred to as the total applied force.

It is particularly agreeable when the total applied force comes as the gradient of a potential function. This means there exists a scalar function of $\mathbf{x}$, ϕ defined near $\mathbf{G}(V)$ such that

$$\mathbf{F}_\alpha^a(\mathbf{x},t) = -\nabla_\alpha\phi(\mathbf{x},t)$$

where the symbol ∇_α denotes the gradient with respect to $\mathbf{x}_\alpha$. More generally,

$$\mathbf{F}_\alpha^a(\mathbf{x},t) = -\nabla_\alpha \phi(\mathbf{x},t) + \mathbf{F}_\alpha^d$$

where $\mathbf{F}_\alpha^d$ is a force which is not a force of constraint or the gradient of a given function. For example, it could be a force of friction. Then

$$\mathbf{F}^a(\mathbf{x},t) = -\nabla \phi(\mathbf{x},t) + \mathbf{F}^d$$

where

$$\mathbf{F}^d = \left(\mathbf{F}_1^d, \cdots, \mathbf{F}_N^d\right)$$

Now let $T(\mathbf{q},\dot{\mathbf{q}}) - \phi(\mathbf{G}(\mathbf{q},t)) = L(\mathbf{q},\dot{\mathbf{q}})$. Then letting x^j denote the usual Cartesian coordinates of $\mathbf{x}$,

$$\begin{aligned}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^k}\right) - \frac{\partial L}{\partial q^k} &= \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}^k}\right) - \frac{\partial T}{\partial q^k} + \sum_j \frac{\partial \phi(\mathbf{x})}{\partial x^j}\frac{\partial x^j}{\partial q^k} \\ &= \frac{\partial \mathbf{G}}{\partial q^k}\cdot\left(-\nabla\phi(\mathbf{x}) + \mathbf{F}^d\right) + \frac{\partial \mathbf{G}}{\partial q^k}\cdot\nabla\phi = \frac{\partial \mathbf{G}}{\partial q^k}\cdot\mathbf{F}^d. \quad (12.31)\end{aligned}$$

These are called Lagrange's equations of motion and they are enormously significant because it is often possible to find the kinetic and potential energy in terms of variables q^k which are meaningful for a particular problem. The expression, $L(\mathbf{q},\dot{\mathbf{q}})$ is called the Lagrangian. This has proved part of the following theorem.

Theorem 12.5.1 *In the above context Newton's second law implies*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^k}\right) - \frac{\partial L}{\partial q^k} = \frac{\partial \mathbf{G}}{\partial q^k}\cdot\mathbf{F}^d. \quad (12.32)$$

*In particular, if the applied force is the gradient of $-\phi$, the right side reduces to 0*12.32. *If, in addition to this, the potential function is time independent then the total energy is conserved. That is,*

$$T(\mathbf{q},\dot{\mathbf{q}}) + \phi(\mathbf{G}(\mathbf{q},t)) = C \quad (12.33)$$

for some constant, C.

Proof: It remains to verify the assertion about the energy. In terms of the Cartesian coordinates,

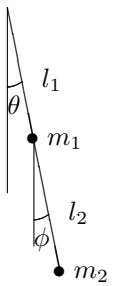
$$E = \sum_\alpha \frac{1}{2} m_\alpha \dot{\mathbf{x}}_\alpha \cdot \dot{\mathbf{x}}_\alpha + \phi(\mathbf{x},t).$$

Recall the applied force is given by $\mathbf{F}_\alpha^a = -\nabla_\alpha\phi(\mathbf{x},t) + \mathbf{F}_\alpha^d$. Differentiating with respect to time,

$$\begin{aligned}\frac{dE}{dt} &= \sum_\alpha m_\alpha \ddot{\mathbf{x}}_\alpha \cdot \dot{\mathbf{x}}_\alpha + \sum_j \frac{\partial \phi}{\partial x^j}\dot{x}^j + \frac{\partial \phi}{\partial t} \\ &= \sum_\alpha \mathbf{F}_\alpha \cdot \dot{\mathbf{x}}_\alpha + \sum_\alpha \nabla_\alpha \phi(\mathbf{x},t)\cdot\dot{\mathbf{x}}_\alpha + \frac{\partial \phi}{\partial t} \\ &= \sum_\alpha \mathbf{F}_\alpha^a \cdot \dot{\mathbf{x}}_\alpha + \sum_\alpha \nabla_\alpha \phi(\mathbf{x},t)\cdot\dot{\mathbf{x}}_\alpha + \frac{\partial \phi}{\partial t} \\ &= \sum_\alpha \left(-\nabla_\alpha\phi(\mathbf{x},t) + \mathbf{F}_\alpha^d\right)\cdot\dot{\mathbf{x}}_\alpha + \sum_\alpha \nabla_\alpha\phi(\mathbf{x},t)\cdot\dot{\mathbf{x}}_\alpha + \frac{\partial\phi}{\partial t} \\ &= \sum_\alpha \mathbf{F}_\alpha^d \cdot \dot{\mathbf{x}}_\alpha + \frac{\partial \phi}{\partial t}.\end{aligned}$$

Therefore, this shows 12.33 because in the case described, $\mathbf{F}_\alpha^d = 0$ and $\frac{\partial \phi}{\partial t} = 0$. In the case of friction, $\mathbf{F}_\alpha^d \cdot \dot{\mathbf{x}}_\alpha \leq 0$ and so in this case, if ϕ is time independent, the total energy is decreasing.

Example 12.5.2 *Consider the double pendulum.*



It is fairly easy to find the equations of motion in terms of the variables, ϕ and θ. These variables are the q^k mentioned above. Because the two rods joining the masses have fixed length, a constraint is introduced on the motion of the two masses. It is clear the position of these masses is specified from the two variables, θ and ϕ. In fact, letting the origin be located at the point at the top where the pendulum is suspended and assuming the vibration is in a plane,

$$\mathbf{x}_1 = (l_1 \sin\theta, -l_1 \cos\theta)$$

and

$$\mathbf{x}_2 = (l_1 \sin\theta + l_2 \sin\phi, -l_1 \cos\theta - l_2 \cos\phi).$$

Therefore,

$$\begin{aligned} \dot{\mathbf{x}}_1 &= \left(l_1 \dot{\theta} \cos\theta, l_1 \dot{\theta} \sin\theta\right) \\ \dot{\mathbf{x}}_2 &= \left(l_1 \dot{\theta} \cos\theta + l_2 \dot{\phi} \cos\phi, l_1 \dot{\theta} \sin\theta + l_2 \dot{\phi} \sin\phi\right). \end{aligned}$$

It follows the kinetic energy is given by

$$T = \frac{1}{2} m_2 \left(2 l_1 \dot{\theta} (\cos\theta) l_2 \dot{\phi} \cos\phi + l_1^2 (\dot{\theta})^2 + 2 l_1 \dot{\theta} (\sin\theta) l_2 \dot{\phi} \sin\phi + l_2^2 (\dot{\phi})^2\right) + \frac{1}{2} m_1 \left(l_1^2 (\dot{\theta})^2\right).$$

There are forces of constraint acting on these masses and there is the force of gravity acting on them. The force from gravity on m_1 is $-m_1 g$ and the force from gravity on m_2 is $-m_2 g$. Our function, ϕ is just the total potential energy. Thus $\phi(\mathbf{x}_1, \mathbf{x}_2) = m_1 g y_1 + m_2 g y_2$. It follows that $\phi(\mathbf{G}(\mathbf{q})) = m_1 g(-l_1 \cos\theta) + m_2 g(-l_1 \cos\theta - l_2 \cos\phi)$. Therefore, the Lagrangian, L, is

$$\frac{1}{2} m_2 \left(2 l_1 l_2 \dot{\theta} \dot{\phi} (\cos(\phi - \theta)) + l_1^2 (\dot{\theta})^2 + l_2^2 (\dot{\phi})^2\right) + \frac{1}{2} m_1 \left(l_1^2 (\dot{\theta})^2\right)$$

$$- \left[m_1 g(-l_1 \cos\theta) + m_2 g(-l_1 \cos\theta - l_2 \cos\phi)\right].$$

It now becomes an easy task to find the equations of motion in terms of the two angles, θ and ϕ.

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} =$$

$$\theta''(m_1 + m_2) l_1^2 + m_2 l_2 l_1 \cos(\phi - \theta) \phi'' - m_2 l_1 l_2 \sin(\phi - \theta)\left(\phi' - \theta'\right)\phi'$$

$$\begin{aligned}
&+(m_1+m_2)\,gl_1\sin\theta - m_2l_1l_2\theta'\phi'\sin(\phi-\theta)\\
&=\theta''(m_1+m_2)\,l_1^2 + m_2l_2l_1\cos(\phi-\theta)\,\phi'' - m_2l_1l_2\sin(\phi-\theta)\,\phi'^2\\
&+(m_1+m_2)\,gl_1\sin\theta = 0. \qquad (12.34)
\end{aligned}$$

To get the other equation,

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\phi}}\right) - \frac{\partial L}{\partial\phi} =$$

$$\begin{aligned}
&\frac{d}{dt}\left[m_2l_1l_2\dot{\theta}(\cos(\phi-\theta)) + m_2l_2^2\dot{\phi}\right] + m_2gl_2\sin\phi - \left(-m_2l_1l_2\dot{\theta}\dot{\phi}\sin(\phi-\theta)\right)\\
&= m_2l_1l_2\theta''\cos(\phi-\theta) - m_2l_1l_2\theta'\sin(\phi-\theta)\left(\phi'-\theta'\right)\\
&+m_2l_2^2\phi'' + m_2gl_2\sin\phi + m_2l_2l_1\phi'\theta'\sin(\phi-\theta)\\
&= m_2l_1l_2\theta''\cos(\phi-\theta) + m_2l_1l_2\left(\theta'\right)^2\sin(\phi-\theta) + m_2l_2^2\phi'' + m_2gl_2\sin\phi = 0 \qquad (12.35)
\end{aligned}$$

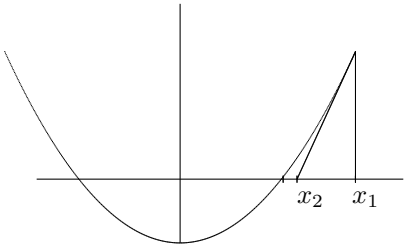
Admittedly, 12.34 and 12.35 are horrific equations, but what would you expect from something as complicated as the double pendulum? They can at least be solved numerically. The conservation of energy gives some idea what is going on. Thus

$$\begin{aligned}
&\frac{1}{2}m_2\left(2l_1l_2\dot{\theta}\dot{\phi}(\cos(\phi-\theta)) + l_1^2(\dot{\theta})^2 + l_2^2(\dot{\phi})^2\right) + \frac{1}{2}m_1\left(l_1^2(\dot{\theta})^2\right)\\
&+\left[m_1g(-l_1\cos\theta) + m_2g(-l_1\cos\theta - l_2\cos\phi)\right] = C.
\end{aligned}$$

12.6 Newton's Method*

12.6.1 The Newton Raphson Method In One Dimension

The Newton Raphson method is a way to get approximations of solutions to various equations. For example, suppose you want to find $\sqrt{2}$. The existence of $\sqrt{2}$ is not difficult to establish by considering the continuous function, $f(x) = x^2 - 2$ which is negative at $x = 0$ and positive at $x = 2$. Therefore, by the intermediate value theorem, there exists $x \in (0,2)$ such that $f(x) = 0$ and this x must equal $\sqrt{2}$. The problem consists of how to find this number, not just to prove it exists. The following picture illustrates the procedure of the Newton Raphson method.



In this picture, a first approximation, denoted in the picture as x_1 is chosen and then the tangent line to the curve $y = f(x)$ at the point $(x_1, f(x_1))$ is obtained. The equation of this tangent line is

$$y - f(x_1) = f'(x_1)(x - x_1).$$

Then extend this tangent line to find where it intersects the x axis. In other words, set $y = 0$ and solve for x. This value of x is denoted by x_2. Thus

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

This second point, x_2 is the second approximation and the same process is done for x_2 that was done for x_1 in order to get the third approximation, x_3. Thus

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}.$$

Continuing this way, yields a sequence of points, $\{x_n\}$ given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{12.36}$$

which hopefully has the property that $\lim_{n\to\infty} x_n = x$ where $f(x) = 0$. You can see from the above picture that this must work out in the case of $f(x) = x^2 - 2$.

Now carry out the computations in the above case for $x_1 = 2$ and $f(x) = x^2 - 2$. From 12.36,

$$x_2 = 2 - \frac{2}{4} = 1.5.$$

Then

$$x_3 = 1.5 - \frac{(1.5)^2 - 2}{2(1.5)} \leq 1.417,$$

$$x_4 = 1.417 - \frac{(1.417)^2 - 2}{2(1.417)} = 1.414\,216\,302\,046\,577,$$

What is the true value of $\sqrt{2}$? To several decimal places this is $\sqrt{2} = 1.414\,213\,562\,373\,095$, showing that the Newton Raphson method has yielded a very good approximation after only a few iterations, even starting with an initial approximation, 2, which was not very good.

This method does not always work. For example, suppose you wanted to find the solution to $f(x) = 0$ where $f(x) = x^{1/3}$. You should check that the sequence of iterates which results does not converge. This is because, starting with x_1 the above procedure yields $x_2 = -2x_1$ and so as the iteration continues, the sequence oscillates between positive and negative values as its absolute value gets larger and larger. The problem is that $f'(0)$ does not exist.

However, if $f(x_0) = 0$ and $f''(x) > 0$ for x near x_0, you can draw a picture to show that the method will yield a sequence which converges to x_0 provided the first approximation, x_1 is taken sufficiently close to x_0. Similarly, if $f''(x) < 0$ for x near x_0, then the method produces a sequence which converges to x_0 provided x_1 is close enough to x_0.

12.6.2 Newton's Method For Nonlinear Systems

The same formula yields a procedure for finding solutions to systems of functions of n variables. This is particularly interesting because you can't make any sense of things from drawing pictures. The technique of graphing and zooming which really works well for functions of one variable is no longer available.

Procedure 12.6.1 *Suppose* $\mathbf{f}$ *is a* C^1 *function of* n *variables and* $\mathbf{f}(\mathbf{z}) = \mathbf{0}$. *Then to find* $\mathbf{z}$, *you use the same iteration which you would use in one dimension,*

$$\mathbf{x}_{k+1} = \mathbf{x}_k - D\mathbf{f}(\mathbf{x}_k)^{-1}\mathbf{f}(\mathbf{x}_k)$$

where $\mathbf{x}_0$ *is an initial approximation chosen close to* $\mathbf{z}$.

Example 12.6.2 *Find a solution to the nonlinear system of equations,*

$$\mathbf{f}(x,y) = \begin{pmatrix} x^3 - 3xy^2 - 3x^2 + 3y^2 + 7x - 5 \\ 3x^2y - y^3 - 6xy + 7y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

You can verify that $(x,y) = (1,2)$, $(1,-2)$, and $(1,0)$ all are solutions to the above system. Suppose then that you didn't know this.

$$D\mathbf{f}(x,y) = \begin{pmatrix} 3x^2 - 3y^2 - 6x + 7 & -6xy + 6y \\ 6xy - 6y & 3x^2 - 3y^2 - 6x + 7 \end{pmatrix}$$

Start with an initial guess $(x_0, y_0) = (1,3)$. Then the next iteration is

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} - \begin{pmatrix} -23 & 0 \\ 0 & -23 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{72}{23} \end{pmatrix}$$

The next iteration is

$$\begin{aligned} &\begin{pmatrix} 1 \\ \frac{72}{23} \end{pmatrix} - \begin{pmatrix} -3.937\,183\,7 \times 10^{-2} & 0 \\ 0 & -3.937\,183\,7 \times 10^{-2} \end{pmatrix} \begin{pmatrix} 0 \\ -18.155\,338 \end{pmatrix} \\ = &\begin{pmatrix} 1.0 \\ 2.415\,625\,8 \end{pmatrix} \end{aligned}$$

I will not bother to use all the decimals in 2.4156258. The next iteration is

$$\begin{aligned} &\begin{pmatrix} 1.0 \\ 2.4 \end{pmatrix} - \begin{pmatrix} -7.530\,120\,5 \times 10^{-2} & 0 \\ 0 & -7.530\,120\,5 \times 10^{-2} \end{pmatrix} \begin{pmatrix} 0 \\ -4.224 \end{pmatrix} \\ = &\begin{pmatrix} 1.0 \\ 2.081\,927\,7 \end{pmatrix}. \end{aligned}$$

Notice how the process is converging to the solution $(x,y) = (1,2)$. If you do one more iteration, you will be really close.

The above was pretty painful because at every step the derivative had to be re evaluated and the inverse taken. It turns out a simpler procedure will work in which you don't have to constantly re evaluate the inverse of the derivative.

Procedure 12.6.3 *Suppose* $\mathbf{f}$ *is a* C^1 *function of* n *variables and* $\mathbf{f}(\mathbf{z}) = \mathbf{0}$. *Then to find* $\mathbf{z}$, *you can use the following iteration procedure*

$$\mathbf{x}_{k+1} = \mathbf{x}_k - D\mathbf{f}(\mathbf{x}_0)^{-1}\mathbf{f}(\mathbf{x}_k)$$

where $\mathbf{x}_0$ *is an initial approximation chosen close to* $\mathbf{z}$.

To illustrate, I will use this new procedure on the same example.

Example 12.6.4 *Find a solution to the nonlinear system of equations,*

$$\mathbf{f}(x,y) = \begin{pmatrix} x^3 - 3xy^2 - 3x^2 + 3y^2 + 7x - 5 \\ 3x^2y - y^3 - 6xy + 7y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

You can verify that $(x,y) = (1,2)$, $(1,-2)$, and $(1,0)$ all are solutions to the above system. Suppose then that you didn't know this. Take $(x_0, y_0) = (1,3)$ as above. Then a little computation will show

$$D\mathbf{f}(1,3)^{-1} = \begin{pmatrix} -\frac{1}{23} & 0 \\ 0 & -\frac{1}{23} \end{pmatrix}$$

The first iteration is then

$$\begin{aligned}\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 3 \end{pmatrix} - \begin{pmatrix} -\frac{1}{23} & 0 \\ 0 & -\frac{1}{23} \end{pmatrix} \begin{pmatrix} 0 \\ -15.0 \end{pmatrix} \\ &= \begin{pmatrix} 1.0 \\ 2.347\,826\,1 \end{pmatrix}\end{aligned}$$

The next iteration is

$$\begin{aligned}\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} &= \begin{pmatrix} 1.0 \\ 2.347\,826\,1 \end{pmatrix} - \begin{pmatrix} -\frac{1}{23} & 0 \\ 0 & -\frac{1}{23} \end{pmatrix} \begin{pmatrix} 0 \\ -3.550\,587\,8 \end{pmatrix} \\ &= \begin{pmatrix} 1.0 \\ 2.193\,452\,7 \end{pmatrix}\end{aligned}$$

The next iteration is

$$\begin{aligned}\begin{pmatrix} x_3 \\ y_3 \end{pmatrix} &= \begin{pmatrix} 1.0 \\ 2.193\,452\,7 \end{pmatrix} - \begin{pmatrix} -\frac{1}{23} & 0 \\ 0 & -\frac{1}{23} \end{pmatrix} \begin{pmatrix} 0 \\ -1.779\,405 \end{pmatrix} \\ &= \begin{pmatrix} 1.0 \\ 2.116\,087\,3 \end{pmatrix}\end{aligned}$$

The next iteration is

$$\begin{aligned}\begin{pmatrix} x_4 \\ y_4 \end{pmatrix} &= \begin{pmatrix} 1.0 \\ 2.116\,087\,3 \end{pmatrix} - \begin{pmatrix} -\frac{1}{23} & 0 \\ 0 & -\frac{1}{23} \end{pmatrix} \begin{pmatrix} 0 \\ -1.011\,120\,4 \end{pmatrix} \\ &= \begin{pmatrix} 1.0 \\ 2.072\,125\,5 \end{pmatrix}.\end{aligned}$$

You see it appears to be converging to a zero of the nonlinear system. It is doing so more slowly than in the case of Newton's method but there is less trouble involved in each step of the iteration.

Of course there is a question about how to choose the initial approximation. There are methods for doing this called homotopy methods which are based on numerical methods for differential equations. The idea for these methods is to consider the problem

$$(1-t)(\mathbf{x}-\mathbf{x}_0) + t\mathbf{f}(\mathbf{x}) = \mathbf{0}.$$

When $t = 0$ this reduces to $\mathbf{x} = \mathbf{x}_0$. Then when $t = 1$, it reduces to $\mathbf{f}(\mathbf{x}) = \mathbf{0}$. The equation specifies $\mathbf{x}$ as a function of t (hopefully). Differentiating with respect to t, you see that $\mathbf{x}$ must solve the following initial value problem,

$$-(\mathbf{x}-\mathbf{x}_0) + (1-t)\mathbf{x}' + \mathbf{f}(\mathbf{x}) + tD\mathbf{f}(\mathbf{x})\mathbf{x}' = \mathbf{0},\ \mathbf{x}(0) = \mathbf{x}_0.$$

where $\mathbf{x}'$ denotes the time derivative of the vector $\mathbf{x}$. Initial value problems of this sort are routinely solvable using standard numerical methods. The idea is you solve it on $[0,1]$ and your zero is $\mathbf{x}(1)$. Because of roundoff error, $\mathbf{x}(1)$ won't be quite right so you use it as an initial guess in Newton's method and find the zero to great accuracy.

12.7 Convergence Questions*

12.7.1 A Fixed Point Theorem

The message of this section is that under reasonable conditions amounting to an assumption that $D\mathbf{f}(\mathbf{z})^{-1}$ exists, Newton's method will converge whenever you take an initial approximation sufficiently close to $\mathbf{z}$. This is just like the situation for the method in one dimension.

The proof of convergence rests on the following lemma which is somewhat more interesting than Newton's method. It is a case of the contraction mapping principle important in differential and integral equations.

Lemma 12.7.1 *Suppose $T : B(\mathbf{x}_0, \delta) \subseteq \mathbb{R}^p \to \mathbb{R}^p$ and it satisfies*

$$|T\mathbf{x} - T\mathbf{y}| \leq \frac{1}{2}|\mathbf{x} - \mathbf{y}| \text{ for all } \mathbf{x}, \mathbf{y} \in B(\mathbf{x}_0, \delta). \tag{12.37}$$

Suppose also that $|T\mathbf{x}_0 - \mathbf{x}_0| < \frac{\delta}{4}$. Then $\{T^n\mathbf{x}_0\}_{n=1}^{\infty}$ converges to a point, $\mathbf{x} \in B(\mathbf{x}_0, \delta)$ such that $T\mathbf{x} = \mathbf{x}$. This point is called a fixed point. Furthermore, there is at most one fixed point on $B(\mathbf{x}_0, \delta)$.

Proof: From the triangle inequality, and the use of 12.37,

$$\begin{aligned} |T^n\mathbf{x}_0 - \mathbf{x}_0| &\leq \sum_{k=1}^{n} \left|T^k\mathbf{x}_0 - T^{k-1}\mathbf{x}_0\right| \\ &\leq \sum_{k=1}^{n} \left(\frac{1}{2}\right)^{k-1} |T\mathbf{x}_0 - \mathbf{x}_0| \\ &\leq 2|T\mathbf{x}_0 - \mathbf{x}_0| < 2\frac{\delta}{4} = \frac{\delta}{2} < \delta. \end{aligned}$$

Thus the sequence remains in the closed ball, $\overline{B(\mathbf{x}_0, \delta/2)} \subseteq B(\mathbf{x}_0, \delta)$. Also, by similar reasoning,

$$|T^n\mathbf{x}_0 - T^m\mathbf{x}_0| \leq \sum_{k=m}^{n} \left|T^{k+1}\mathbf{x}_0 - T^k\mathbf{x}_0\right| \leq \sum_{k=m}^{n} \left(\frac{1}{2}\right)^{k} |T\mathbf{x}_0 - \mathbf{x}_0| \leq \frac{\delta}{4}\frac{1}{2^{m-1}}.$$

It follows, that $\{T^n\mathbf{x}_0\}$ is a Cauchy sequence. Therefore, it converges to a point of

$$\overline{B(\mathbf{x}_0, \delta/2)} \subseteq B(\mathbf{x}_0, \delta).$$

Call this point, $\mathbf{x}$. Then since T is continuous, it follows

$$\mathbf{x} = \lim_{n\to\infty} T^n\mathbf{x}_0 = T \lim_{n\to\infty} T^{n-1}\mathbf{x}_0 = T\mathbf{x}_0.$$

If $T\mathbf{x} = \mathbf{x}$ and $T\mathbf{y} = \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in B(\mathbf{x}_0, \delta)$ then $|\mathbf{x} - \mathbf{y}| = |T\mathbf{x} - T\mathbf{y}| \leq \frac{1}{2}|\mathbf{x} - \mathbf{y}|$ and so $\mathbf{x} = \mathbf{y}$.

12.7.2 The Operator Norm

How do you measure the distance between linear transformations defined on $\mathbb{F}^n$? It turns out there are many ways to do this but I will give the most common one here.

Definition 12.7.2 *$\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ denotes the space of linear transformations mapping $\mathbb{F}^n$ to $\mathbb{F}^m$. For $A \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$, the **operator norm** is defined by*

$$||A|| \equiv \max\{|Ax|_{\mathbb{F}^m} : |x|_{\mathbb{F}^n} \leq 1\} < \infty.$$

Theorem 12.7.3 *Denote by $|\cdot|$ the norm on either $\mathbb{F}^n$ or $\mathbb{F}^m$. Then $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ with this operator norm is a* ***complete normed linear space*** *of dimension nm with*

$$||A\mathbf{x}|| \leq ||A||\,|\mathbf{x}|\,.$$

Here ***Completeness*** *means that every Cauchy sequence converges.*

Proof: It is necessary to show the norm defined on $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ really is a norm. This means it is necessary to verify

$$||A|| \geq 0 \text{ and equals zero if and only if } A = 0.$$

For α a scalar,

$$||\alpha A|| = |\alpha|\,||A||\,,$$

and for $A, B \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$,

$$||A + B|| \leq ||A|| + ||B||$$

The first two properties are obvious but you should verify them. It remains to verify the norm is well defined and also to verify the triangle inequality above. First if $|\mathbf{x}| \leq 1$, and (A_{ij}) is the matrix of the linear transformation with respect to the usual basis vectors, then

$$\begin{aligned} ||A|| &= \max\left\{\left(\sum_i |(A\mathbf{x})_i|^2\right)^{1/2} : |\mathbf{x}| \leq 1\right\} \\ &= \max\left\{\left(\sum_i \left|\sum_j A_{ij}x_j\right|^2\right)^{1/2} : |\mathbf{x}| \leq 1\right\} \end{aligned}$$

which is a finite number by the extreme value theorem.

It is clear that a basis for $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ consists of linear transformations whose matrices are of the form E_{ij} where E_{ij} consists of the $m \times n$ matrix having all zeros except for a 1 in the ij^{th} position. In effect, this considers $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ as $\mathbb{F}^{nm}$. Think of the $m \times n$ matrix as a long vector folded up.

If $\mathbf{x} \neq \mathbf{0}$,

$$|A\mathbf{x}|\,\frac{1}{|\mathbf{x}|} = \left|A\frac{\mathbf{x}}{|\mathbf{x}|}\right| \leq ||A|| \tag{12.38}$$

It only remains to verify completeness. Suppose then that $\{A_k\}$ is a Cauchy sequence in $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$. Then from 12.38 $\{A_k\mathbf{x}\}$ is a Cauchy sequence for each $\mathbf{x} \in \mathbb{F}^n$. This follows because

$$|A_k\mathbf{x} - A_l\mathbf{x}| \leq ||A_k - A_l||\,|\mathbf{x}|$$

which converges to 0 as $k, l \to \infty$. Therefore, by completeness of $\mathbb{F}^m$, there exists $A\mathbf{x}$, the name of the thing to which the sequence, $\{A_k\mathbf{x}\}$ converges such that

$$\lim_{k\to\infty} A_k\mathbf{x} = A\mathbf{x}.$$

Then A is linear because

$$\begin{aligned} A(a\mathbf{x} + b\mathbf{y}) &\equiv \lim_{k\to\infty} A_k(a\mathbf{x} + b\mathbf{y}) \\ &= \lim_{k\to\infty} (aA_k\mathbf{x} + bA_k\mathbf{y}) \\ &= a\lim_{k\to\infty} A_k\mathbf{x} + b\lim_{k\to\infty} A_k\mathbf{y} \\ &= aA\mathbf{x} + bA\mathbf{y}. \end{aligned}$$

By the first part of this argument, $||A|| < \infty$ and so $A \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$. This proves the theorem.

The following is an interesting exercise which is left for you.

Proposition 12.7.4 *Let $A(\mathbf{x}) \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ for each $\mathbf{x} \in U \subseteq \mathbb{F}^p$. Then letting $(A_{ij}(\mathbf{x}))$ denote the matrix of $A(\mathbf{x})$ with respect to the standard basis, it follows A_{ij} is continuous at $\mathbf{x}$ for each i, j if and only if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if $|\mathbf{x} - \mathbf{y}| < \delta$, then $||A(\mathbf{x}) - A(\mathbf{y})|| < \varepsilon$. That is, A is a continuous function having values in $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ at $\mathbf{x}$.*

Proof: Suppose first the second condition holds. Then from the material on linear transformations,

$$\begin{aligned} |A_{ij}(\mathbf{x}) - A_{ij}(\mathbf{y})| &= |\mathbf{e}_i \cdot (A(\mathbf{x}) - A(\mathbf{y}))\,\mathbf{e}_j| \\ &\leq |\mathbf{e}_i|\,|(A(\mathbf{x}) - A(\mathbf{y}))\,\mathbf{e}_j| \\ &\leq ||A(\mathbf{x}) - A(\mathbf{y})||. \end{aligned}$$

Therefore, the second condition implies the first.

Now suppose the first condition holds. That is each A_{ij} is continuous at $\mathbf{x}$. Let $|\mathbf{v}| \leq 1$.

$$\begin{aligned} |(A(\mathbf{x}) - A(\mathbf{y}))(\mathbf{v})| &= \left(\sum_i \left| \sum_j (A_{ij}(\mathbf{x}) - A_{ij}(\mathbf{y}))\, v_j \right|^2 \right)^{1/2} \\ &\leq \left(\sum_i \left(\sum_j |A_{ij}(\mathbf{x}) - A_{ij}(\mathbf{y})|\,|v_j| \right)^2 \right)^{1/2}. \end{aligned} \tag{12.39}$$

By continuity of each A_{ij}, there exists a $\delta > 0$ such that for each i, j

$$|A_{ij}(\mathbf{x}) - A_{ij}(\mathbf{y})| < \frac{\varepsilon}{n\sqrt{m}}$$

whenever $|\mathbf{x} - \mathbf{y}| < \delta$. Then from 12.39, if $|\mathbf{x} - \mathbf{y}| < \delta$,

$$\begin{aligned} |(A(\mathbf{x}) - A(\mathbf{y}))(\mathbf{v})| &< \left(\sum_i \left(\sum_j \frac{\varepsilon}{n\sqrt{m}} |\mathbf{v}| \right)^2 \right)^{1/2} \\ &\leq \left(\sum_i \left(\sum_j \frac{\varepsilon}{n\sqrt{m}} \right)^2 \right)^{1/2} = \varepsilon \end{aligned}$$

This proves the proposition.

The proposition implies that a function is C^1 if and only if the derivative, $D\mathbf{f}$ exists and the function, $\mathbf{x} \to D\mathbf{f}(\mathbf{x})$ is continuous in the usual way. That is, for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $|\mathbf{x} - \mathbf{y}| < \delta$, then $||D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{y})|| < \varepsilon$.

The following is a version of the mean value theorem valid for functions defined on $\mathbb{R}^n$.

Theorem 12.7.5 *Suppose U is an open subset of $\mathbb{R}^p$ and $\mathbf{f} : U \to \mathbb{R}^q$ has the property that $D\mathbf{f}(\mathbf{x})$ exists for all $\mathbf{x}$ in U and that, $\mathbf{x}+t(\mathbf{y} - \mathbf{x}) \in U$ for all $t \in [0, 1]$. (The line*

segment joining the two points lies in U.) Suppose also that for all points on this line segment,

$$||D\mathbf{f}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))|| \leq M.$$

Then

$$||\mathbf{f}(\mathbf{y})-\mathbf{f}(\mathbf{x})|| \leq M||\mathbf{y}-\mathbf{x}||.$$

Proof: Let

$$S \equiv \{t \in [0,1] : \text{ for all } s \in [0,t],$$
$$||\mathbf{f}(\mathbf{x}+s(\mathbf{y}-\mathbf{x}))-\mathbf{f}(\mathbf{x})|| \leq (M+\varepsilon)s||\mathbf{y}-\mathbf{x}||\}.$$

Then $0 \in S$ and by continuity of $\mathbf{f}$, it follows that if $t \equiv \sup S$, then $t \in S$ and if $t < 1$,

$$||\mathbf{f}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))-\mathbf{f}(\mathbf{x})|| = (M+\varepsilon)t||\mathbf{y}-\mathbf{x}||. \tag{12.40}$$

If $t < 1$, then there exists a sequence of positive numbers, $\{h_k\}_{k=1}^{\infty}$ converging to 0 such that

$$||\mathbf{f}(\mathbf{x}+(t+h_k)(\mathbf{y}-\mathbf{x}))-\mathbf{f}(\mathbf{x})|| > (M+\varepsilon)(t+h_k)||\mathbf{y}-\mathbf{x}||$$

which implies that

$$||\mathbf{f}(\mathbf{x}+(t+h_k)(\mathbf{y}-\mathbf{x}))-\mathbf{f}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))||$$
$$+||\mathbf{f}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))-\mathbf{f}(\mathbf{x})|| > (M+\varepsilon)(t+h_k)||\mathbf{y}-\mathbf{x}||.$$

By 12.40, this inequality implies

$$||\mathbf{f}(\mathbf{x}+(t+h_k)(\mathbf{y}-\mathbf{x}))-\mathbf{f}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))|| > (M+\varepsilon)h_k||\mathbf{y}-\mathbf{x}||$$

which yields upon dividing by h_k and taking the limit as $h_k \to 0$,

$$||D\mathbf{f}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x})|| \geq (M+\varepsilon)||\mathbf{y}-\mathbf{x}||.$$

Now by the definition of the norm of a linear operator,

$$\begin{aligned} M||\mathbf{y}-\mathbf{x}|| &\geq ||D\mathbf{f}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))||\,||\mathbf{y}-\mathbf{x}|| \\ &\geq ||D\mathbf{f}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x})|| \geq (M+\varepsilon)||\mathbf{y}-\mathbf{x}||, \end{aligned}$$

a contradiction. Therefore, $t = 1$ and so

$$||\mathbf{f}(\mathbf{x}+(\mathbf{y}-\mathbf{x}))-\mathbf{f}(\mathbf{x})|| \leq (M+\varepsilon)||\mathbf{y}-\mathbf{x}||.$$

Since $\varepsilon > 0$ is arbitrary, this proves the theorem.

12.7.3 A Method For Finding Zeros

Theorem 12.7.6 *Suppose $\mathbf{f} : U \subseteq \mathbb{R}^p \to \mathbb{R}^p$ is a C^1 function and suppose $\mathbf{f}(\mathbf{z}) = \mathbf{0}$. Suppose also that for all $\mathbf{x}$ sufficiently close to $\mathbf{z}$, it follows that $D\mathbf{f}(\mathbf{x})^{-1}$ exists. Let $\delta > 0$ be small enough that for all $\mathbf{x}, \mathbf{x}_0 \in B(\mathbf{z}, 2\delta)$*

$$\left|\left|I - D\mathbf{f}(\mathbf{x}_0)^{-1}D\mathbf{f}(\mathbf{x})\right|\right| < \frac{1}{2}. \tag{12.41}$$

Now pick $\mathbf{x}_0 \in B(\mathbf{z}, \delta)$ also close enough to $\mathbf{z}$ such that

$$\left|\left|D\mathbf{f}(\mathbf{x}_0)^{-1}\right|\right||\mathbf{f}(\mathbf{x}_0)| < \frac{\delta}{4}.$$

Define

$$T\mathbf{x} \equiv \mathbf{x} - D\mathbf{f}(\mathbf{x}_0)^{-1}\mathbf{f}(\mathbf{x}).$$

Then the sequence, $\{T^n\mathbf{x}_0\}_{n=1}^{\infty}$ converges to $\mathbf{z}$.

Proof: First note that $|T\mathbf{x}_0 - \mathbf{x}_0| = \left|D\mathbf{f}(\mathbf{x}_0)^{-1}\mathbf{f}(\mathbf{x}_0)\right| \leq \left|\left|D\mathbf{f}(\mathbf{x}_0)^{-1}\right|\right| |\mathbf{f}(\mathbf{x}_0)| < \frac{\delta}{4}$. Also on $B(\mathbf{x}_0,\delta) \subseteq B(\mathbf{z},2\delta)$ the inequality, 12.41, the chain rule, and Theorem 12.7.5 shows that for $\mathbf{x},\mathbf{y} \in B(\mathbf{x}_0,\delta)$,

$$|T\mathbf{x} - T\mathbf{y}| \leq \frac{1}{2}|\mathbf{x}-\mathbf{y}|.$$

This follows because $DT\mathbf{x} = I - D\mathbf{f}(\mathbf{x}_0)^{-1}\mathbf{f}(\mathbf{x})$. The conclusion now follows from Lemma 12.7.1. This proves the lemma.

12.7.4 Newton's Method

Theorem 12.7.7 *Suppose $\mathbf{f}: U \subseteq \mathbb{R}^p \to \mathbb{R}^p$ is a C^1 function and suppose $\mathbf{f}(\mathbf{z}) = \mathbf{0}$. Suppose that for all $\mathbf{x}$ sufficiently close to $\mathbf{z}$, it follows that $D\mathbf{f}(\mathbf{x})^{-1}$ exists. Suppose also that*[2]

$$\left|\left|D\mathbf{f}(\mathbf{x}_2)^{-1} - D\mathbf{f}(\mathbf{x}_1)^{-1}\right|\right| \leq K|\mathbf{x}_2 - \mathbf{x}_1|. \tag{12.42}$$

Then there exists $\delta > 0$ small enough that for all $\mathbf{x}_1,\mathbf{x}_2 \in B(\mathbf{z},2\delta)$

$$\left|\mathbf{x}_1 - \mathbf{x}_2 - D\mathbf{f}(\mathbf{x}_2)^{-1}(\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2))\right| \leq \frac{1}{4}|\mathbf{x}_1 - \mathbf{x}_2|, \tag{12.43}$$

$$|\mathbf{f}(\mathbf{x}_1)| < \frac{1}{4K}. \tag{12.44}$$

Now pick $\mathbf{x}_0 \in B(\mathbf{z},\delta)$ also close enough to $\mathbf{z}$ such that

$$\left|\left|D\mathbf{f}(\mathbf{x}_0)^{-1}\right|\right| |\mathbf{f}(\mathbf{x}_0)| < \frac{\delta}{4}.$$

Define

$$T\mathbf{x} \equiv \mathbf{x} - D\mathbf{f}(\mathbf{x})^{-1}\mathbf{f}(\mathbf{x}).$$

Then the sequence, $\{T^n\mathbf{x}_0\}_{n=1}^{\infty}$ converges to $\mathbf{z}$.

Proof: The left side of 12.43 equals

$$\begin{aligned}
&\left|\mathbf{x}_1 - \mathbf{x}_2 - D\mathbf{f}(\mathbf{x}_2)^{-1}(D\mathbf{f}(\mathbf{x}_2)(\mathbf{x}_1-\mathbf{x}_2) + \mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2) - D\mathbf{f}(\mathbf{x}_2)(\mathbf{x}_1-\mathbf{x}_2))\right| \\
= &\left|D\mathbf{f}(\mathbf{x}_2)^{-1}(\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2) - D\mathbf{f}(\mathbf{x}_2)(\mathbf{x}_1-\mathbf{x}_2))\right| \\
\leq &\, C|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2) - D\mathbf{f}(\mathbf{x}_2)(\mathbf{x}_1-\mathbf{x}_2)|
\end{aligned}$$

because 12.42 implies $\left|\left|D\mathbf{f}(\mathbf{x})^{-1}\right|\right|$ is bounded for $\mathbf{x} \in B(\mathbf{z},\delta)$. Now use the assumption that $\mathbf{f}$ is C^1 and Proposition 12.7.4 to conclude there exists δ small enough that $||D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{z})|| < \frac{1}{8}$ for all $\mathbf{x} \in B(\mathbf{z},2\delta)$. Then let $\mathbf{x}_1,\mathbf{x}_2 \in B(\mathbf{z},2\delta)$. Define $\mathbf{h}(\mathbf{x}) \equiv \mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_2) - D\mathbf{f}(\mathbf{x}_2)(\mathbf{x}-\mathbf{x}_2)$. Then

$$\begin{aligned}
||D\mathbf{h}(\mathbf{x})|| &= ||D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{x}_2)|| \\
&\leq ||D\mathbf{f}(\mathbf{x}) - D\mathbf{f}(\mathbf{z})|| + ||D\mathbf{f}(\mathbf{z}) - D\mathbf{f}(\mathbf{x}_2)|| \\
&\leq \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.
\end{aligned}$$

[2]The following condition as well as the preceeding can be shown to hold if you simply assume $\mathbf{f}$ is a C^2 function and $D\mathbf{f}(\mathbf{z})^{-1}$ exists. This requires the use of the inverse function theorem, one of the major theorems which should be studied in an advanced calculus class.

It follows from Theorem 12.7.5

$$\begin{aligned} |\mathbf{h}(\mathbf{x}_1) - \mathbf{h}(\mathbf{x}_2)| &= |\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2) - D\mathbf{f}(\mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)| \\ &\leq \frac{1}{4}|\mathbf{x}_1 - \mathbf{x}_2|. \end{aligned}$$

This proves 12.43. 12.44 can be satisfied by taking δ still smaller if necessary and using $\mathbf{f}(\mathbf{z}) = \mathbf{0}$ and the continuity of $\mathbf{f}$.

Now let $\mathbf{x}_0 \in B(\mathbf{z}, \delta)$ be as described. Then

$$|T\mathbf{x}_0 - \mathbf{x}_0| = \left|D\mathbf{f}(\mathbf{x}_0)^{-1}\mathbf{f}(\mathbf{x}_0)\right| \leq \left|\left|D\mathbf{f}(\mathbf{x}_0)^{-1}\right|\right| |\mathbf{f}(\mathbf{x}_0)| < \frac{\delta}{4}.$$

Letting $\mathbf{x}_1, \mathbf{x}_2 \in B(\mathbf{x}_0, \delta) \subseteq B(\mathbf{z}, 2\delta)$,

$$\begin{aligned} |T\mathbf{x}_1 - T\mathbf{x}_2| &= \left|\mathbf{x}_1 - D\mathbf{f}(\mathbf{x}_1)^{-1}\mathbf{f}(\mathbf{x}_1) - \left(\mathbf{x}_2 - D\mathbf{f}(\mathbf{x}_2)^{-1}\mathbf{f}(\mathbf{x}_2)\right)\right| \\ &\leq \left|\mathbf{x}_1 - \mathbf{x}_2 - D\mathbf{f}(\mathbf{x}_2)^{-1}(\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2))\right| + \left|\left(D\mathbf{f}(\mathbf{x}_1)^{-1} - D\mathbf{f}(\mathbf{x}_2)^{-1}\right)\mathbf{f}(\mathbf{x}_1)\right| \\ &\leq \tfrac{1}{4}|\mathbf{x}_1 - \mathbf{x}_2| + K|\mathbf{x}_1 - \mathbf{x}_2||\mathbf{f}(\mathbf{x}_1)| \leq \tfrac{1}{2}|\mathbf{x}_1 - \mathbf{x}_2|. \end{aligned}$$

The desired result now follows from Lemma 12.7.1.

12.8 Exercises

1. Suppose $\mathbf{f} : U \to \mathbb{R}^q$ and let $\mathbf{x} \in U$ and $\mathbf{v}$ be a unit vector. Show $D_{\mathbf{v}}\mathbf{f}(\mathbf{x}) = D\mathbf{f}(\mathbf{x})\mathbf{v}$. Recall that

$$D_{\mathbf{v}}\mathbf{f}(\mathbf{x}) \equiv \lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{v}) - \mathbf{f}(\mathbf{x})}{t}.$$

2. Let $f(x,y) = \begin{cases} xy\sin\left(\frac{1}{x}\right) \text{ if } x \neq 0 \\ 0 \text{ if } x = 0 \end{cases}$. Find where f is differentiable and compute the derivative at all these points.

3. Let

$$f(x,y) = \begin{cases} x \text{ if } |y| > |x| \\ -x \text{ if } |y| \leq |x| \end{cases}.$$

Show f is continuous at $(0,0)$ and that the partial derivatives exist at $(0,0)$ but the function is not differentiable at $(0,0)$.

4. Let

$$\mathbf{f}(x,y,z) = \begin{pmatrix} x^2\sin y + z^3 \\ \sin(x+y) + z^3\cos x \end{pmatrix}.$$

Find $D\mathbf{f}(1,2,3)$.

5. Let

$$\mathbf{f}(x,y,z) = \begin{pmatrix} x\tan y + z^3 \\ \cos(x+y) + z^3\cos x \end{pmatrix}.$$

Find $D\mathbf{f}(1,2,3)$.

6. Let

$$\mathbf{f}(x,y,z) = \begin{pmatrix} x\sin y + z^3 \\ \sin(x+y) + z^3\cos x \\ x^5 + y^2 \end{pmatrix}.$$

Find $D\mathbf{f}(x,y,z)$.

7. Let
$$f(x,y) = \begin{cases} \frac{\left(x^2-y^4\right)^2}{(x^2+y^4)^2} \text{ if } (x,y) \neq (0,0) \\ 1 \text{ if } (x,y) = (0,0) \end{cases}.$$
Show that all directional derivatives of f exist at $(0,0)$, and are all equal to zero but the function is not even continuous at $(0,0)$. Therefore, it is not differentiable. Why?

8. In the example of Problem 7 show the partial derivatives exist but are not continuous.

9. A certain building is shaped like the top half of the ellipsoid, $\frac{x^2}{900} + \frac{y^2}{900} + \frac{z^2}{400} = 1$ determined by letting $z \geq 0$. Here dimensions are measured in meters. The building needs to be painted. The paint, when applied is about .005 meters thick. About how many cubic meters of paint will be needed. **Hint:**This is going to replace the numbers, 900 and 400 with slightly larger numbers when the ellipsoid is fattened slightly by the paint. The volume of the top half of the ellipsoid, $x^2/a^2 + y^2/b^2 + z^2/c^2 \leq 1, z \geq 0$ is $(2/3)\pi abc$.

10. Show carefully that the usual one variable version of the chain rule is a special case of Theorem 12.4.15.

11. Let $z = f(\mathbf{y}) = \left(y_1^2 + \sin y_2 + \tan y_3\right)$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_2 \\ x_2^2 - x_1 + x_2 \\ x_2^2 + x_1 + \sin x_2 \end{pmatrix}$.
Find $D(f \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z}{\partial x_i}$ for $i = 1, 2$.

12. Let $z = f(\mathbf{y}) = \left(y_1^2 + \cot y_2 + \sin y_3\right)$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_4 + x_3 \\ x_2^2 - x_1 + x_2 \\ x_2^2 + x_1 + \sin x_4 \end{pmatrix}$.
Find $D(f \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z}{\partial x_i}$ for $i = 1, 2, 3, 4$.

13. Let $z = f(\mathbf{y}) = \left(y_1^2 + y_2^2 + \sin y_3 + y_4\right)$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_4 + x_3 \\ x_2^2 - x_1 + x_2 \\ x_2^2 + x_1 + \sin x_4 \\ x_4 + x_2 \end{pmatrix}$.
Find $D(f \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z}{\partial x_i}$ for $i = 1, 2, 3, 4$.

14. Let $\mathbf{z} = \mathbf{f}(\mathbf{y}) = \begin{pmatrix} y_1^2 + \sin y_2 + \tan y_3 \\ y_1^2 y_2 + y_3 \end{pmatrix}$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_2 \\ x_2^2 - x_1 + x_2 \\ x_2^2 + x_1 + \sin x_2 \end{pmatrix}$.
Find $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z_k}{\partial x_i}$ for $i = 1, 2$ and $k = 1, 2$.

15. Let $\mathbf{z} = \mathbf{f}(\mathbf{y}) = \begin{pmatrix} y_1^2 + \sin y_2 + \tan y_3 \\ y_1^2 y_2 + y_3 \\ \cos\left(y_1^2\right) + y_2^3 y_3 \end{pmatrix}$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_4 \\ x_2^2 - x_1 + x_3 \\ x_3^2 + x_1 + \sin x_2 \end{pmatrix}$.
Find $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z_k}{\partial x_i}$ for $i = 1, 2, 3, 4$ and $k = 1, 2, 3$.

16. Let $z = \mathbf{f}(\mathbf{y}) = \begin{pmatrix} y_2^2 + \sin y_1 + \sec y_2 + y_4 \\ y_1^2 y_2 + y_3^3 \\ y_2^3 y_4 + y_1 \\ y_1 + y_2 \end{pmatrix}$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + 2x_4 \\ x_2^2 - 2x_1 + x_3 \\ x_3^2 + x_1 + \cos x_1 \\ x_2^2 \end{pmatrix}$.
Find $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z_k}{\partial x_i}$ for $i = 1, 2, 3, 4$ and $k = 1, 2, 3, 4$.

17. Let $\mathbf{f}(\mathbf{y}) = \begin{pmatrix} y_1^2 + \sin y_2 + \tan y_3 \\ y_1^2 y_2 + y_3 \\ \cos\left(y_1^2\right) + y_2^3 y_3 \end{pmatrix}$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_4 \\ x_2^2 - x_1 + x_3 \\ x_3^2 + x_1 + \sin x_2 \end{pmatrix}$. Find $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z_k}{\partial x_i}$ for $i = 1, 2, 3, 4$ and $k = 1, 2, 3$.

18. Suppose $\mathbf{r}_1(t) = (\cos t, \sin t, t)$, $\mathbf{r}_2(t) = (t, 2t, 1)$, and $\mathbf{r}_3(t) = (1, t, 1)$. Find the rate of change with respect to t of the volume of the parallelepiped determined by these three vectors when $t = 1$.

19. A trash compacter is compacting a rectangular block of trash. The width is changing at the rate of -1 inches per second, the length is changing at the rate of -2 inches per second and the height is changing at the rate of -3 inches per second. How fast is the volume changing when the length is 20, the height is 10, and the width is 10.

20. A trash compacter is compacting a rectangular block of trash. The width is changing at the rate of -2 inches per second, the length is changing at the rate of -1 inches per second and the height is changing at the rate of -4 inches per second. How fast is the surface area changing when the length is 20, the height is 10, and the width is 10.

21. The ideal gas law is $PV = kT$ where k is a constant which depends on the number of moles and on the gas being considered. If V is changing at the rate of 2 cubic cm. per second and T is changing at the rate of 3 degrees Kelvin per second, how fast is the pressure changing when $T = 300$ and V equals 400 cubic cm.?

22. Let S denote a level surface of the form $f(x_1, x_2, x_3) = C$. Suppose now that $\mathbf{r}(t)$ is a space curve which lies in this level surface. Thus $f(r_1(t), r_2(t), r_3(t))$. Show using the chain rule that $Df(r_1(t), r_2(t), r_3(t))(r_1'(t), r_2'(t), r_3'(t))^T = 0$. Note that $Df(x_1, x_2, x_3) = (f_{x_1}, f_{x_2}, f_{x_3})$. This is denoted by $\nabla f(x_1, x_2, x_3) = (f_{x_1}, f_{x_2}, f_{x_3})^T$. This 3×1 matrix or column vector is called the gradient vector. Argue that

$$\nabla f(r_1(t), r_2(t), r_3(t)) \cdot (r_1'(t), r_2'(t), r_3'(t))^T = 0.$$

What geometric fact have you just established?

23. Suppose $\mathbf{f}$ is a C^1 function which maps U, an open subset of $\mathbb{R}^n$ one to one and onto V, an open set in $\mathbb{R}^m$ such that the inverse map, $\mathbf{f}^{-1}$ is also C^1. What must be true of m and n? Why? **Hint:** Consider Example 12.4.12 on Page 258. Also you can use the fact that if A is an $m \times n$ matrix which maps $\mathbb{R}^n$ onto $\mathbb{R}^m$, then $m \leq n$.

The Gradient And Optimization

13.0.1 Outcomes

1. Interpret the gradient of a function as a normal to a level curve or a level surface.
2. Find the normal line and tangent plane to a smooth surface at a given point.
3. Find the angles between curves and surfaces.
4. Define what is meant by a local extreme point.
5. Find candidates for local extrema using the gradient.
6. Find the local extreme values and saddle points of a C^2 function.
7. Use the second derivative test to identify the nature of a singluar point.
8. Find the extreme values of a function defined on a closed and bounded region.
9. Solve word problems involving maximum and minimum values.
10. Use the method of Lagrange to determine the extreme values of a function subject to a constraint.
11. Solve word problems using the method of Lagrange multipliers.

Recall the concept of the gradient. This has already been considered in the special case of a C^1 function. However, you do not need so much to define the gradient.

13.1 Fundamental Properties

Let $f : U \to \mathbb{R}$ where U is an open subset of $\mathbb{R}^n$ and suppose f is differentiable on U. Thus if $\mathbf{x} \in U$,

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \sum_{j=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_i} v_i + o(\mathbf{v}). \tag{13.1}$$

Recall Proposition 11.3.6, a more general version of which is stated here for convenience. It is more general because here it is only assumed that f is differentiable, not C^1.

Proposition 13.1.1 *If f is differentiable at $\mathbf{x}$ and for $\mathbf{v}$ a unit vector,*

$$D_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}.$$

Proof:

$$\begin{aligned}\frac{f(\mathbf{x}+t\mathbf{v})-f(\mathbf{x})}{t} &= \frac{1}{t}\left(f(\mathbf{x})+\sum_{j=1}^{n}\frac{\partial f(\mathbf{x})}{\partial x_i}tv_i+o(t\mathbf{v})-f(\mathbf{x})\right)\\ &= \frac{1}{t}\left(\sum_{j=1}^{n}\frac{\partial f(\mathbf{x})}{\partial x_i}tv_i+o(t\mathbf{v})\right)\\ &= \sum_{j=1}^{n}\frac{\partial f(\mathbf{x})}{\partial x_i}v_i+\frac{o(t\mathbf{v})}{t}.\end{aligned}$$

Now $\lim_{t\to 0}\frac{o(t\mathbf{v})}{t}=0$ and so

$$D_{\mathbf{v}}f(\mathbf{x})=\lim_{t\to 0}\frac{f(\mathbf{x}+t\mathbf{v})-f(\mathbf{x})}{t}=\sum_{j=1}^{n}\frac{\partial f(\mathbf{x})}{\partial x_i}v_i=\nabla f(\mathbf{x})\cdot\mathbf{v}$$

as claimed.

Definition 13.1.2 *When f is differentiable, define $\nabla f(\mathbf{x})\equiv\left(\frac{\partial f}{\partial x_1}(\mathbf{x}),\cdots,\frac{\partial f}{\partial x_n}(\mathbf{x})\right)^T$ just as was done in the special case where f is C^1. As before, this vector is called the* ***gradient vector****.*

This defines the gradient for a differentiable scalar valued function. There are ways to define the gradient for vector valued functions but this will not be attempted in this book.

It follows immediately from 13.1 that

$$f(\mathbf{x}+\mathbf{v})=f(\mathbf{x})+\nabla f(\mathbf{x})\cdot\mathbf{v}+o(\mathbf{v}) \tag{13.2}$$

As mentioned above, an important aspect of the gradient is its relation with the directional derivative. A repeat of the above argument gives the following. From 13.2, for $\mathbf{v}$ a unit vector,

$$\begin{aligned}\frac{f(\mathbf{x}+t\mathbf{v})-f(\mathbf{x})}{t} &= \nabla f(\mathbf{x})\cdot\mathbf{v}+\frac{o(t\mathbf{v})}{t}\\ &= \nabla f(\mathbf{x})\cdot\mathbf{v}+\frac{o(t)}{t}.\end{aligned}$$

Therefore, taking $t\to 0$,

$$D_{\mathbf{v}}f(\mathbf{x})=\nabla f(\mathbf{x})\cdot\mathbf{v}. \tag{13.3}$$

Example 13.1.3 *Let $f(x,y,z)=x^2+\sin(xy)+z$. Find $D_{\mathbf{v}}f(1,0,1)$ where*

$$\mathbf{v}=\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right).$$

Note this vector which is given is already a unit vector. Therefore, from the above, it is only necessary to find $\nabla f(1,0,1)$ and take the dot product.

$$\nabla f(x,y,z)=(2x+(\cos xy)\,y,(\cos xy)\,x,1).$$

Therefore, $\nabla f(1,0,1)=(2,1,1)$. Therefore, the directional derivative is

$$(2,1,1)\cdot\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)=\frac{4}{3}\sqrt{3}.$$

Because of 13.3 it is easy to find the largest possible directional derivative and the smallest possible directional derivative. That which follows is a more algebraic treatment of an earlier result with the trigonometry removed.

Proposition 13.1.4 *Let $f : U \to \mathbb{R}$ be a differentiable function and let $\mathbf{x} \in U$. Then*

$$\max\{D_{\mathbf{v}} f(\mathbf{x}) : |\mathbf{v}| = 1\} = |\nabla f(x)| \tag{13.4}$$

and

$$\min\{D_{\mathbf{v}} f(\mathbf{x}) : |\mathbf{v}| = 1\} = -|\nabla f(x)|. \tag{13.5}$$

Furthermore, the maximum in 13.4 occurs when $\mathbf{v} = \nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$ and the minimum in 13.5 occurs when $\mathbf{v} = -\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$.

Proof: From 13.3 and the Cauchy Schwarz inequality,

$$|D_{\mathbf{v}} f(\mathbf{x})| \leq |\nabla f(\mathbf{x})|$$

and so for any choice of $\mathbf{v}$ with $|\mathbf{v}| = 1$,

$$-|\nabla f(\mathbf{x})| \leq D_{\mathbf{v}} f(\mathbf{x}) \leq |\nabla f(\mathbf{x})|.$$

The proposition is proved by noting that if $\mathbf{v} = -\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$, then

$$\begin{aligned} D_{\mathbf{v}} f(\mathbf{x}) &= \nabla f(\mathbf{x}) \cdot (-\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|) \\ &= -|\nabla f(\mathbf{x})|^2 / |\nabla f(\mathbf{x})| = -|\nabla f(\mathbf{x})| \end{aligned}$$

while if $\mathbf{v} = \nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$, then

$$\begin{aligned} D_{\mathbf{v}} f(\mathbf{x}) &= \nabla f(\mathbf{x}) \cdot (\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|) \\ &= |\nabla f(\mathbf{x})|^2 / |\nabla f(\mathbf{x})| = |\nabla f(\mathbf{x})|. \end{aligned}$$

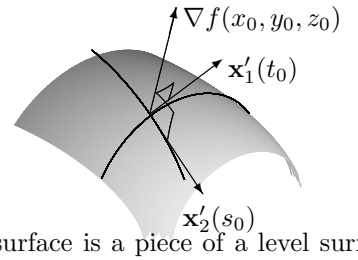
The conclusion of the above proposition is important in many physical models. For example, consider some material which is at various temperatures depending on location. Because it has cool places and hot places, it is expected that the heat will flow from the hot places to the cool places. Consider a small surface having a unit normal, $\mathbf{n}$. Thus $\mathbf{n}$ is a normal to this surface and has unit length. If it is desired to find the rate in calories per second at which heat crosses this little surface in the direction of $\mathbf{n}$ it is defined as $\mathbf{J} \cdot \mathbf{n} A$ where A is the area of the surface and $\mathbf{J}$ is called the heat flux. It is reasonable to suppose the rate at which heat flows across this surface will be largest when $\mathbf{n}$ is in the direction of greatest rate of decrease of the temperature. In other words, heat flows most readily in the direction which involves the maximum rate of decrease in temperature. This expectation will be realized by taking $\mathbf{J} = -K\nabla u$ where K is a positive scalar function which can depend on a variety of things. The above relation between the heat flux and ∇u is usually called the Fourier heat conduction law and the constant, K is known as the coefficient of thermal conductivity. It is a material property, different for iron than for aluminum. In most applications, K is considered to be a constant but this is wrong. Experiments show this scalar should depend on temperature. Nevertheless, things get very difficult if this dependence is allowed. The constant can depend on position in the material or even on time.

An identical relationship is usually postulated for the flow of a diffusing species. In this problem, something like a pollutant diffuses. It may be an insecticide in ground water for example. Like heat, it tries to move from areas of high concentration toward areas of low concentration. In this case $\mathbf{J} = -K\nabla c$ where c is the concentration of the diffusing species. When applied to diffusion, this relationship is known as Fick's law. Mathematically, it is indistinguishable from the problem of heat flow.

Note the importance of the gradient in formulating these models.

13.2 Tangent Planes

The gradient has fundamental geometric significance illustrated by the following picture.



In this picture, the surface is a piece of a level surface of a function of three variables, $f(x,y,z)$. Thus the surface is defined by $f(x,y,z) = c$ or more completely as $\{(x,y,z) : f(x,y,z) = c\}$. For example, if $f(x,y,z) = x^2+y^2+z^2$, this would be a piece of a sphere. There are two smooth curves in this picture which lie in the surface having parameterizations, $\mathbf{x}_1(t) = (x_1(t), y_1(t), z_1(t))$ and $\mathbf{x}_2(s) = (x_2(s), y_2(s), z_2(s))$ which intersect at the point, (x_0, y_0, z_0) on this surface[1]. This intersection occurs when $t = t_0$ and $s = s_0$. Since the points, $\mathbf{x}_1(t)$ for t in an interval lie in the level surface, it follows

$$f(x_1(t), y_1(t), z_1(t)) = c$$

for all t in some interval. Therefore, taking the derivative of both sides and using the chain rule on the left,

$$\frac{\partial f}{\partial x}(x_1(t), y_1(t), z_1(t))\, x_1'(t) +$$

$$\frac{\partial f}{\partial y}(x_1(t), y_1(t), z_1(t))\, y_1'(t) + \frac{\partial f}{\partial z}(x_1(t), y_1(t), z_1(t))\, z_1'(t) = 0.$$

In terms of the gradient, this merely states

$$\nabla f(x_1(t), y_1(t), z_1(t)) \cdot \mathbf{x}_1'(t) = 0.$$

Similarly,

$$\nabla f(x_2(s), y_2(s), z_2(s)) \cdot \mathbf{x}_2'(s) = 0.$$

Letting $s = s_0$ and $t = t_0$, it follows

$$\nabla f(x_0, y_0, z_0) \cdot \mathbf{x}_1'(t_0) = 0,\ \nabla f(x_0, y_0, z_0) \cdot \mathbf{x}_2'(s_0) = 0.$$

It follows $\nabla f(x_0, y_0, z_0)$ is perpendicular to both the direction vectors of the two indicated curves shown. Surely if things are as they should be, these two direction vectors would determine a plane which deserves to be called the tangent plane to the level surface of f at the point (x_0, y_0, z_0) and that $\nabla f(x_0, y_0, z_0)$ is perpendicular to this tangent plane at the point, (x_0, y_0, z_0).

Example 13.2.1 *Find the equation of the tangent plane to the level surface,* $f(x,y,z) = 6$ *of the function,* $f(x,y,z) = x^2 + 2y^2 + 3z^2$ *at the point* $(1,1,1)$.

First note that $(1,1,1)$ is a point on this level surface. To find the desired plane it suffices to find the normal vector to the proposed plane. But $\nabla f(x,y,z) = (2x, 4y, 6z)$

[1]Do there exist any smooth curves which lie in the level surface of f and pass through the point (x_0, y_0, z_0)? It turns out there do if $\nabla f(x_0, y_0, z_0) \neq \mathbf{0}$ and if the function, f, is C^1. However, this is a consequence of the implicit function theorem, one of the greatest theorems in all mathematics and a topic for an advanced calculus class. It is also in an appendix to this book

and so $\nabla f(1,1,1) = (2,4,6)$. Therefore, from this problem, the equation of the plane is

$$(2,4,6) \cdot (x-1, y-1, z-1) = 0$$

or in other words,

$$2x - 12 + 4y + 6z = 0.$$

Example 13.2.2 *The point,* $\left(\sqrt{3}, 1, 4\right)$ *is on both the surfaces,* $z = x^2 + y^2$ *and* $z = 8 - \left(x^2 + y^2\right)$. *Find the cosine of the angle between the two tangent planes at this point.*

Recall this is the same as the angle between two normal vectors. Of course there is some ambiguity here because if $\mathbf{n}$ is a normal vector, then so is $-\mathbf{n}$ and replacing $\mathbf{n}$ with $-\mathbf{n}$ in the formula for the cosine of the angle will change the sign. We agree to look for the acute angle and its cosine rather than the obtuse angle. The normals are $\left(2\sqrt{3}, 2, -1\right)$ and $\left(2\sqrt{3}, 2, 1\right)$. Therefore, the cosine of the angle desired is

$$\frac{\left(2\sqrt{3}\right)^2 + 4 - 1}{17} = \frac{15}{17}.$$

Example 13.2.3 *The point,* $\left(1, \sqrt{3}, 4\right)$ *is on the surface,* $z = x^2 + y^2$. *Find the line perpendicular to the surface at this point.*

All that is needed is the direction vector of this line. The surface is the level surface, $x^2 + y^2 - z = 0$. The normal to this surface is given by the gradient at this point. Thus the desired line is $\left(1, \sqrt{3}, 4\right) + t\left(2, 2\sqrt{3}, -1\right)$.

13.3 Exercises

1. Find the gradient of $f =$
   (a) $x^2 y + z^3$ at $(1,1,2)$
   (b) $z \sin\left(x^2 y\right) + 2^{x+y}$ at $(1,1,0)$
   (c) $u \ln\left(x + y + z^2 + w\right)$ at $(x,y,z,w,u) = (1,1,1,1,2)$

2. Find the directional derivatives of f at the indicated point in the direction, $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right)$.
   (a) $x^2 y + z^3$ at $(1,1,1)$
   (b) $z \sin\left(x^2 y\right) + 2^{x+y}$ at $(1,1,2)$
   (c) $xy + z^2 + 1$ at $(1,2,3)$

3. Find the tangent plane to the indicated level surface at the indicated point.
   (a) $x^2 y + z^3 = 2$ at $(1,1,1)$
   (b) $z \sin\left(x^2 y\right) + 2^{x+y} = 2 \sin 1 + 4$ at $(1,1,2)$
   (c) $\cos(x) + z \sin(x+y) = 1$ at $\left(-\pi, \frac{3\pi}{2}, 2\right)$

4. Explain why the displacement vector of an object from a given point in $\mathbb{R}^3$ is always perpendicular to the velocity vector if the magnitude of the displacement vector is constant.

5. The point $\left(1,1,\sqrt{2}\right)$ is a point on the level surface, $x^2+y^2+z^2=4$. Find the line perpendicular to the surface at this point.

6. The point $\left(1,1,\sqrt{2}\right)$ is a point on the level surface, $x^2+y^2+z^2=4$ and the level surface, $y^2+2z^2=5$. Find the angle between the two tangent planes at this point.

7. The level surfaces $x^2+y^2+z^2=4$ and $z+x^2+y^2=4$ have the point $\left(\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2},1\right)$ in the curve formed by the intersection of these surfaces. Find a direction vector for this curve at this point. **Hint:** Recall the gradients of the two surfaces are perpendicular to the corresponding surfaces at this point. A direction vector for the desired curve should be perpendicular to both of these gradients.

8. In a slightly more general setting, suppose $f_1(x,y,z)=0$ and $f_2(x,y,z)=0$ are two level surfaces which intersect in a curve which has parameterization, $(x(t),y(t),z(t))$. Find a differential equation such that one of its solutions is the above parameterization.

Suppose $f:D(f)\to\mathbb{R}$ where $D(f)\subseteq\mathbb{R}^n$.

13.4 Local Extrema

Definition 13.4.1 *A point $\mathbf{x}\in D(f)\subseteq\mathbb{R}^n$ is called a **local minimum** if $f(\mathbf{x})\leq f(\mathbf{y})$ for all $\mathbf{y}\in D(f)$ sufficiently close to $\mathbf{x}$. A point $\mathbf{x}\in D(f)$ is called a **local maximum** if $f(\mathbf{x})\geq f(\mathbf{y})$ for all $\mathbf{y}\in D(f)$ sufficiently close to $\mathbf{x}$. A **local extremum** is a point of $D(f)$ which is either a local minimum or a local maximum. The plural for extremum is extrema. The plural for minimum is minima and the plural for maximum is maxima.*

Procedure 13.4.2 *To find candidates for local extrema which are interior points of $D(f)$ where f is a differentiable function, you simply identify those points where ∇f equals the zero vector. To justify this, note that the graph of f is the level surface*

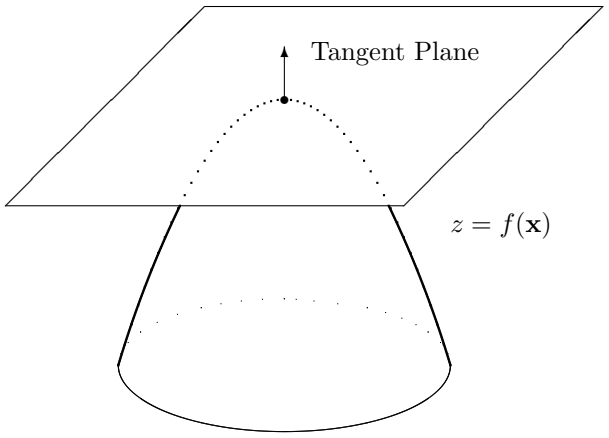
$$F(\mathbf{x},z)\equiv z-f(\mathbf{x})=0$$

and the local extrema at such interior points must have horizontal tangent planes. Therefore, a normal vector at such points must be a multiple of $(0,\cdots,0,1)$. Thus ∇F at such points must be a multiple of this vector. That is, if $\mathbf{x}$ is such a point,

$$k(0,\cdots,0,1)=(-f_{x_1}(\mathbf{x}),\cdots,-f_{x_n}(\mathbf{x}),1).$$

Thus $\nabla f(\mathbf{x})=\mathbf{0}$.

This is illustrated in the following picture.



A more rigorous explanation is as follows. Let $\mathbf{v}$ be any vector in $\mathbb{R}^n$ and suppose $\mathbf{x}$ is a local maximum (minimum) for $\mathbf{f}$. Then consider the real valued function of one variable, $h(t) \equiv f(\mathbf{x}+t\mathbf{v})$ for small $|t|$. Since $\mathbf{f}$ has a local maximum (minimum), it follows that h is a differentiable function of the single variable t for small t which has a local maximum (minimum) when $t=0$. Therefore, $h'(0)=0$. But $h'(t) = Df(\mathbf{x}+t\mathbf{v})\,\mathbf{v}$ by the chain rule. Therefore,

$$h'(0) = Df(\mathbf{x})\,\mathbf{v} = 0$$

and since $\mathbf{v}$ is arbitrary, it follows $Df(\mathbf{x}) = 0$. However,

$$Df(\mathbf{x}) = \begin{pmatrix} f_{x_1}(\mathbf{x}) & \cdots & f_{x_n}(\mathbf{x}) \end{pmatrix}$$

and so $\nabla f(\mathbf{x}) = \mathbf{0}$. This proves the following theorem.

Theorem 13.4.3 *Suppose U is an open set contained in $D(f)$ such that f is C^1 on U and suppose $\mathbf{x} \in U$ is a local minimum or local maximum for f. Then $\nabla f(\mathbf{x}) = \mathbf{0}$.*

A more general result is left for you to do in the exercises.

Definition 13.4.4 *A **singular point** for f is a point $\mathbf{x}$ where $\nabla f(\mathbf{x}) = \mathbf{0}$. This is also called a **critical point**.*

Example 13.4.5 *Find the critical points for the function, $f(x,y) \equiv xy - x - y$ for $x, y > 0$.*

Note that here $D(f)$ is an open set and so every point is an interior point. Where is the gradient equal to zero?

$$f_x = y - 1 = 0,\ f_y = x - 1 = 0$$

and so there is exactly one critical point $(1,1)$.

Example 13.4.6 *Find the volume of the smallest tetrahedron made up of the coordinate planes in the first octant and a plane which is tangent to the sphere $x^2+y^2+z^2=4$.*

The normal to the sphere at a point, (x_0, y_0, z_0) on a point of the sphere is

$$\left(x_0, y_0, \sqrt{4 - x_0^2 - y_0^2}\right)$$

and so the equation of the tangent plane at this point is

$$x_0(x - x_0) + y_0(y - y_0) + \sqrt{4 - x_0^2 - y_0^2}\left(z - \sqrt{4 - x_0^2 - y_0^2}\right) = 0$$

When $x = y = 0$,

$$z = \frac{4}{\sqrt{(4 - x_0^2 - y_0^2)}}$$

When $z = 0 = y$,

$$x = \frac{4}{x_0},$$

and when $z = x = 0$,

$$y = \frac{4}{y_0}.$$

Therefore, the function to minimize is

$$f(x,y) = \frac{1}{6}\frac{64}{xy\sqrt{(4-x^2-y^2)}}$$

This is because in beginning calculus it was shown that the volume of a pyramid is 1/3 the area of the base times the height. Therefore, you simply need to find the gradient of this and set it equal to zero. Thus upon taking the partial derivatives, you need to have

$$\frac{-4+2x^2+y^2}{x^2y\left(-4+x^2+y^2\right)\sqrt{(4-x^2-y^2)}} = 0,$$

and

$$\frac{-4+x^2+2y^2}{xy^2\left(-4+x^2+y^2\right)\sqrt{(4-x^2-y^2)}} = 0.$$

Therefore, $x^2+2y^2 = 4$ and $2x^2+y^2 = 4$. Thus $x = y$ and so $x = y = \frac{2}{\sqrt{3}}$. It follows from the equation for z that $z = \frac{2}{\sqrt{3}}$ also. How do you know this is not the largest tetrahedron?

Example 13.4.7 *An open box is to contain 32 cubic feet. Find the dimensions which will result in the least surface area.*

Let the height of the box be z and the length and width be x and y respectively. Then $xyz = 32$ and so $z = 32/xy$. The total area is $xy + 2xz + 2yz$ and so in terms of the two variables, x and y, the area is

$$A = xy + \frac{64}{y} + \frac{64}{x}$$

To find best dimensions you note these must result in a local minimum.

$$A_x = \frac{yx^2-64}{x^2} = 0,\ A_y = \frac{xy^2-64}{y^2}.$$

Therefore, $yx^2-64 = 0$ and $xy^2-64 = 0$ so $xy^2 = yx^2$. For sure the answer excludes the case where any of the variables equals zero. Therefore, $x = y$ and so $x = 4 = y$. Then $z = 2$ from the requirement that $xyz = 32$. How do you know this gives the least surface area? Why doesn't this give the largest surface area?

13.5 The Second Derivative Test

There is a version of the second derivative test in the case that the function and its first and second partial derivatives are all continuous. The proof of this theorem is dependent on fundamental results in linear algebra which are in an appendix. You can skip the proof if you like. It is given later.

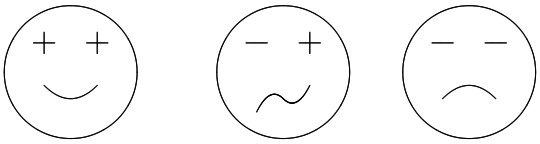
Definition 13.5.1 *The matrix, $H(\mathbf{x})$ whose ij^{th} entry at the point $\mathbf{x}$ is*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$$

is called the ***Hessian matrix****. The eigenvalues of $H(\mathbf{x})$ are the solutions λ to the equation*

$$\det(\lambda I - H(\mathbf{x})) = 0$$

The following theorem says that if all the eigenvalues of the Hessian matrix at a critical point are positive, then the critical point is a local minimum. If all the eigenvalues of the Hessian matrix at a critical point are negative, then the critical point is a local maximum. Finally, if some of the eigenvalues of the Hessian matrix at the critical point are positive and some are negative then the critical point is a saddle point. The following picture illustrates the situation.



Theorem 13.5.2 *Let $f : U \to \mathbb{R}$ for U an open set in $\mathbb{R}^n$ and let f be a C^2 function and suppose that at some $\mathbf{x} \in U$, $\nabla f(\mathbf{x}) = \mathbf{0}$. Also let μ and λ be respectively, the largest and smallest eigenvalues of the matrix, $H(\mathbf{x})$. If $\lambda > 0$ then f has a local minimum at $\mathbf{x}$. If $\mu < 0$ then f has a local maximum at $\mathbf{x}$. If either λ or μ equals zero, the test fails. If $\lambda < 0$ and $\mu > 0$ there exists a direction in which when f is evaluated on the line through the critical point having this direction, the resulting function of one variable has a local minimum and there exists a direction in which when f is evaluated on the line through the critical point having this direction, the resulting function of one variable has a local maximum. This last case is called a* ***saddle point****.*

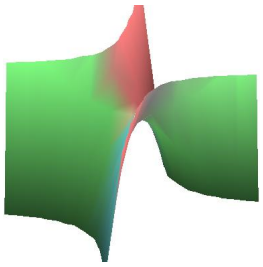
Here is an example.

Example 13.5.3 *Let $f(x,y) = 10xy + y^2$. Find the critical points and determine whether they are local minima, local maxima or saddle points.*

First $\nabla\left(10xy + y^2\right) = (10y, 10x + 2y)$ and so there is one critical point at the point $(0,0)$. What is it? The Hessian matrix is

$$\begin{pmatrix} 0 & 10 \\ 10 & 2 \end{pmatrix}$$

and the eigenvalues are of different signs. Therefore, the critical point $(0,0)$ is a saddle point. Here is a graph drawn by Maple.



Here is another example.

Example 13.5.4 *Let $f(x,y) = 2x^4 - 4x^3 + 14x^2 + 12yx^2 - 12yx - 12x + 2y^2 + 4y + 2$. Find the critical points and determine whether they are local minima, local maxima, or saddle points.*

$f_x(x,y) = 8x^3 - 12x^2 + 28x + 24yx - 12y - 12$ and $f_y(x,y) = 12x^2 - 12x + 4y + 4$. The points at which both f_x and f_y equal zero are $\left(\frac{1}{2}, -\frac{1}{4}\right), (0,-1)$, and $(1,-1)$.

The Hessian matrix is

$$\begin{pmatrix} 24x^2 + 28 + 24y - 24x & 24x - 12 \\ 24x - 12 & 4 \end{pmatrix}$$

and the thing to determine is the sign of its eigenvalues evaluated at the critical points.

First consider the point $\left(\frac{1}{2}, -\frac{1}{4}\right)$. The Hessian matrix is $\begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}$ and its eigenvalues are $16, 4$ showing that this is a local minimum.

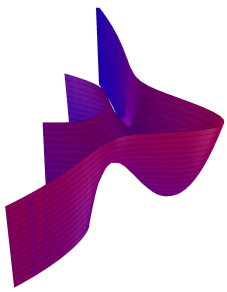
Next consider $(0,-1)$ at this point the Hessian matrix is $\begin{pmatrix} 4 & -12 \\ -12 & 4 \end{pmatrix}$ and the eigenvalues are $16, -8$. Therefore, this point is a saddle point. To determine this, find the eigenvalues.

$$\det\left(\lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 4 & -12 \\ -12 & 4 \end{pmatrix}\right) = \lambda^2 - 8\lambda - 128 = (\lambda + 8)(\lambda - 16)$$

so the eigenvalues are -8 and 16 as claimed.

Finally consider the point $(1,-1)$. At this point the Hessian is $\begin{pmatrix} 4 & 12 \\ 12 & 4 \end{pmatrix}$ and the eigenvalues are $16, -8$ so this point is also a saddle point.

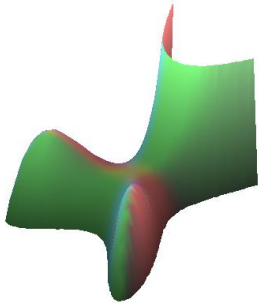
Below is a graph of this function which illustrates the behavior near saddle points.



Or course sometimes the second derivative test is inadequate to determine what is going on. This should be no surprise since this was the case even for a function of one variable. For a function of two variables, a nice example is the Monkey saddle.

Example 13.5.5 *Suppose $f(x,y) = 6xy^2 - 2x^3 - 3y^4$. Show that $(0,0)$ is a critical point for which the second derivative test gives no information.*

Before doing anything it might be interesting to look at the graph of this function of two variables plotted using Maple.



This picture should indicate why this is called a monkey saddle. It is because the monkey can sit in the saddle and have a place for his tail. Now to see $(0,0)$ is a critical point, note that $f_x(0,0) = f_y(0,0) = 0$ because $f_x(x,y) = 6y^2 - 6x^2$, $f_y(x,y) = 12xy - 12y^3$ and so $(0,0)$ is a critical point. So are $(1,1)$ and $(1,-1)$. Now $f_{xx}(0,0) = 0$ and so are $f_{xy}(0,0)$ and $f_{yy}(0,0)$. Therefore, the Hessian matrix is the zero matrix and clearly has only the zero eigenvalue. Therefore, the second derivative test is totally useless at this point.

However, suppose you took $x = t$ and $y = t$ and evaluated this function on this line. This reduces to $h(t) = f(t,t) = 4t^3 - 3t^4$), which is strictly increasing near $t = 0$. This shows the critical point $(0,0)$ of f is neither a local max. nor a local min. Next let $x = 0$ and $y = t$. Then $p(t) \equiv f(0,t) = -3t^4$. Therefore, along the line, $(0,t)$, f has a local maximum at $(0,0)$.

Example 13.5.6 *Find the critical points of the following function of three variables and classify them as local minimums, local maximums or saddle points.*

$$f(x,y,z) = \frac{5}{6}x^2 + 4x + 16 - \frac{7}{3}xy - 4y - \frac{4}{3}xz + 12z + \frac{5}{6}y^2 - \frac{4}{3}zy + \frac{1}{3}z^2$$

First you need to locate the critical points. This involves taking the gradient.

$$\begin{aligned} & \nabla\left(\frac{5}{6}x^2 + 4x + 16 - \frac{7}{3}xy - 4y - \frac{4}{3}xz + 12z + \frac{5}{6}y^2 - \frac{4}{3}zy + \frac{1}{3}z^2\right) \\ = \quad & \left(\frac{5}{3}x + 4 - \frac{7}{3}y - \frac{4}{3}z, -\frac{7}{3}x - 4 + \frac{5}{3}y - \frac{4}{3}z, -\frac{4}{3}x + 12 - \frac{4}{3}y + \frac{2}{3}z\right) \end{aligned}$$

Next you need to set the gradient equal to zero and solve the equations. This yields $y = 5, x = 3, z = -2$. Now to use the second derivative test, you assemble the Hessian matrix which is

$$\begin{pmatrix} \frac{5}{3} & -\frac{7}{3} & -\frac{4}{3} \\ -\frac{7}{3} & \frac{5}{3} & -\frac{4}{3} \\ -\frac{4}{3} & -\frac{4}{3} & \frac{2}{3} \end{pmatrix}.$$

Note that in this simple example, the Hessian matrix is constant and so all that is left is to consider the eigenvalues. Writing the characteristic equation and solving yields the eigenvalues are $2, -2, 4$. Thus the given point is a saddle point.

13.6 Exercises

1. Use the second derivative test on the critical points $(1,1)$, and $(1,-1)$ for Example 13.5.5.

2. If $H = H^T$ and $H\mathbf{x} = \lambda\mathbf{x}$ while $H\mathbf{x} = \mu\mathbf{x}$ for $\lambda \neq \mu$, show $\mathbf{x} \cdot \mathbf{y} = 0$.

3. Show the points $\left(\frac{1}{2}, -\frac{21}{4}\right), (0,-4)$, and $(1,-4)$ are critical points of the following function of two variables and classify them as local minima, local maxima or saddle points.

   $f(x,y) = -x^4 + 2x^3 + 39x^2 + 10yx^2 - 10yx - 40x - y^2 - 8y - 16.$

   Answer:

   The Hessian matrix is

   $$\begin{pmatrix} -12x^2 + 78 + 20y + 12x & 20x - 10 \\ 20x - 10 & -2 \end{pmatrix}$$

   The eigenvalues must be checked at the critical points. First consider the point $\left(\frac{1}{2}, -\frac{21}{4}\right)$. At this point, the Hessian is

   $$\begin{pmatrix} -24 & 0 \\ 0 & -2 \end{pmatrix}$$

   and its eigenvalues are $-24, -2$, both negative. Therefore, the function has a local maximum at this point.

   Next consider $(0,-4)$. At this point the Hessian matrix is

   $$\begin{pmatrix} -2 & -10 \\ -10 & -2 \end{pmatrix}$$

   and the eigenvalues are $8, -12$ so the function has a saddle point.

   Finally consider the point $(1,-4)$. The Hessian equals

   $$\begin{pmatrix} -2 & 10 \\ 10 & -2 \end{pmatrix}$$

   having eigenvalues: $8, -12$ and so there is a saddle point here.

4. Show the points $\left(\frac{1}{2}, -\frac{53}{12}\right), (0,-4)$, and $(1,-4)$ are critical points of the following function of two variables and classify them according to whether they are local minima, local maxima or saddle points.

   $f(x,y) = -3x^4 + 6x^3 + 37x^2 + 10yx^2 - 10yx - 40x - 3y^2 - 24y - 48.$

   Answer:

   The Hessian matrix is

   $$\begin{pmatrix} -36x^2 + 74 + 20y + 36x & 20x - 10 \\ 20x - 10 & -6 \end{pmatrix}.$$

   Check its eigenvalues at the critical points. First consider the point $\left(\frac{1}{2}, -\frac{53}{12}\right)$. At this point the Hessian is

$$\begin{pmatrix} -\frac{16}{3} & 0 \\ 0 & -6 \end{pmatrix}$$

and its eigenvalues are $-\frac{16}{3}, -6$ so there is a local maximum at this point. The same analysis shows there are saddle points at the other two critical points.

5. Show the points $\left(\frac{1}{2}, \frac{37}{20}\right), (0,2)$, and $(1,2)$ are critical points of the following function of two variables and classify them according to whether they are local minima, local maxima or saddle points.

   $f(x,y) = 5x^4 - 10x^3 + 17x^2 - 6yx^2 + 6yx - 12x + 5y^2 - 20y + 20.$

   Answer:

   The Hessian matrix is

   $$\begin{pmatrix} 60x^2 + 34 - 12y - 60x & -12x + 6 \\ -12x + 6 & 10 \end{pmatrix}.$$

   Check its eigenvalues at the critical points. First consider the point $\left(\frac{1}{2}, \frac{37}{20}\right)$. At this point, the Hessian matrix is

   $$\begin{pmatrix} -\frac{16}{5} & 0 \\ 0 & 10 \end{pmatrix}$$

   and its eigenvalues are $-\frac{16}{5}, 10$. Therefore, there is a saddle point.

   Next consider $(0,2)$ at this point the Hessian matrix is

   $$\begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}$$

   and the eigenvalues are $16, 4$. Therefore, there is a local minimum at this point. There is also a local minimum at the critical point, $(1,2)$.

6. Show the points $\left(\frac{1}{2}, -\frac{17}{8}\right), (0,-2)$, and $(1,-2)$ are critical points of the following function of two variables and classify them according to whether they are local minima, local maxima or saddle points.

   $f(x,y) = 4x^4 - 8x^3 - 4yx^2 + 4yx + 8x - 4x^2 + 4y^2 + 16y + 16.$

   Answer:

   The Hessian matrix is $\begin{pmatrix} 48x^2 - 8 - 8y - 48x & -8x + 4 \\ -8x + 4 & 8 \end{pmatrix}$. Check its eigenvalues at the critical points. First consider the point $\left(\frac{1}{2}, -\frac{17}{8}\right)$. This matrix is

   $\begin{pmatrix} -3 & 0 \\ 0 & 8 \end{pmatrix}$ and its eigenvalues are $-3, 8$.

   Next consider $(0,-2)$ at this point the Hessian matrix is

   $\begin{pmatrix} 8 & 4 \\ 4 & 8 \end{pmatrix}$ and the eigenvalues are $12, 4$. Finally consider the point $(1,-2)$.

   $\begin{pmatrix} 8 & -4 \\ -4 & 8 \end{pmatrix}$, eigenvalues: $12, 4$.

   If the eigenvalues are both negative, then local max. If both positive, then local min. Otherwise the test fails.

7. Find the critical points of the following function of three variables and classify them according to whether they are local minima, local maxima or saddle points.
$f(x,y,z) = \frac{1}{3}x^2 + \frac{32}{3}x + \frac{4}{3} - \frac{16}{3}yx - \frac{58}{3}y - \frac{4}{3}zx - \frac{46}{3}z + \frac{1}{3}y^2 - \frac{4}{3}zy - \frac{5}{3}z^2$.
Answer:
The critical point is at $(-2,3,-5)$. The eigenvalues of the Hessian matrix at this point are $-6, -2$, and 6.

8. Find the critical points of the following function of three variables and classify them according to whether they are local minima, local maxima or saddle points.
$f(x,y,z) = -\frac{5}{3}x^2 + \frac{2}{3}x - \frac{2}{3} + \frac{8}{3}yx + \frac{2}{3}y + \frac{14}{3}zx - \frac{28}{3}z - \frac{5}{3}y^2 + \frac{14}{3}zy - \frac{8}{3}z^2$.
Answer:
The eigenvalues are $4, -10$, and -6 and the only critical point is $(1,1,0)$.

9. Find the critical points of the following function of three variables and classify them according to whether they are local minima, local maxima or saddle points.
$f(x,y,z) = -\frac{11}{3}x^2 + \frac{40}{3}x - \frac{56}{3} + \frac{8}{3}yx + \frac{10}{3}y - \frac{4}{3}zx + \frac{22}{3}z - \frac{11}{3}y^2 - \frac{4}{3}zy - \frac{5}{3}z^2$.

10. Find the critical points of the following function of three variables and classify them according to whether they are local minima, local maxima or saddle points.
$f(x,y,z) = -\frac{2}{3}x^2 + \frac{28}{3}x + \frac{37}{3} + \frac{14}{3}yx + \frac{10}{3}y - \frac{4}{3}zx - \frac{26}{3}z - \frac{2}{3}y^2 - \frac{4}{3}zy + \frac{7}{3}z^2$.

11. Show that if f has a critical point and some eigenvalue of the Hessian matrix is positive, then there exists a direction in which when f is evaluated on the line through the critical point having this direction, the resulting function of one variable has a local minimum. State and prove a similar result in the case where some eigenvalue of the Hessian matrix is negative.

12. Suppose $\mu = 0$ but there are negative eigenvalues of the Hessian at a critical point. Show by giving examples that the second derivative tests fails.

13. Show the points $\left(\frac{1}{2}, -\frac{9}{2}\right), (0,-5)$, and $(1,-5)$ are critical points of the following function of two variables and classify them as local minima, local maxima or saddle points.
$f(x,y) = 2x^4 - 4x^3 + 42x^2 + 8yx^2 - 8yx - 40x + 2y^2 + 20y + 50$.

14. Show the points $\left(1, -\frac{11}{2}\right), (0,-5)$, and $(2,-5)$ are critical points of the following function of two variables and classify them as local minima, local maxima or saddle points.
$f(x,y) = 4x^4 - 16x^3 - 4x^2 - 4yx^2 + 8yx + 40x + 4y^2 + 40y + 100$.

15. Show the points $\left(\frac{3}{2}, \frac{27}{20}\right), (0,0)$, and $(3,0)$ are critical points of the following function of two variables and classify them as local minima, local maxima or saddle points.
$f(x,y) = 5x^4 - 30x^3 + 45x^2 + 6yx^2 - 18yx + 5y^2$.

16. Find the critical points of the following function of three variables and classify them as local minima, local maxima or saddle points.
$f(x,y,z) = \frac{10}{3}x^2 - \frac{44}{3}x + \frac{64}{3} - \frac{10}{3}yx + \frac{16}{3}y + \frac{2}{3}zx - \frac{20}{3}z + \frac{10}{3}y^2 + \frac{2}{3}zy + \frac{4}{3}z^2$.

17. Find the critical points of the following function of three variables and classify them as local minima, local maxima or saddle points.
$f(x,y,z) = -\frac{7}{3}x^2 - \frac{146}{3}x + \frac{83}{3} + \frac{16}{3}yx + \frac{4}{3}y - \frac{14}{3}zx + \frac{94}{3}z - \frac{7}{3}y^2 - \frac{14}{3}zy + \frac{8}{3}z^2$.

18. Find the critical points of the following function of three variables and classify them as local minima, local maxima or saddle points.
$f(x,y,z) = \frac{2}{3}x^2 + 4x + 75 - \frac{14}{3}yx - 38y - \frac{8}{3}zx - 2z + \frac{2}{3}y^2 - \frac{8}{3}zy - \frac{1}{3}z^2$.

19. Find the critical points of the following function of three variables and classify them as local minima, local maxima or saddle points.
$f(x,y,z) = 4x^2 - 30x + 510 - 2yx + 60y - 2zx - 70z + 4y^2 - 2zy + 4z^2$.

20. Show the critical points of the following function are points of the form, $(x,y,z) = \left(t, 2t^2 - 10t, -t^2 + 5t\right)$ for $t \in \mathbf{R}$ and classify them as local minima, local maxima or saddle points.
$f(x,y,z) = -\frac{1}{6}x^4 + \frac{5}{3}x^3 - \frac{25}{6}x^2 + \frac{10}{3}yx^2 - \frac{50}{3}yx + \frac{19}{3}zx^2 - \frac{95}{3}zx - \frac{5}{3}y^2 - \frac{10}{3}zy - \frac{1}{6}z^2$.
The verification that the critical points are of the indicated form is left for you.
The Hessian is
$$\begin{pmatrix} -2x^2 + 10x - \frac{25}{3} + \frac{20}{3}y + \frac{38}{3}z & \frac{20}{3}x - \frac{50}{3} & \frac{38}{3}x - \frac{95}{3} \\ \frac{20}{3}x - \frac{50}{3} & -\frac{10}{3} & -\frac{10}{3} \\ \frac{38}{3}x - \frac{95}{3} & -\frac{10}{3} & -\frac{1}{3} \end{pmatrix}$$
at a critical point it is
$$\begin{pmatrix} -\frac{4}{3}t^2 + \frac{20}{3}t - \frac{25}{3} & \frac{20}{3}(t) - \frac{50}{3} & \frac{38}{3}(t) - \frac{95}{3} \\ \frac{20}{3}(t) - \frac{50}{3} & -\frac{10}{3} & -\frac{10}{3} \\ \frac{38}{3}(t) - \frac{95}{3} & -\frac{10}{3} & -\frac{1}{3} \end{pmatrix}.$$
The eigenvalues are
$$0, -\frac{2}{3}t^2 + \frac{10}{3}t - 6 + \frac{2}{3}\sqrt{(t^4 - 10t^3 + 493t^2 - 2340t + 2916)},$$
and
$$-\frac{2}{3}t^2 + \frac{10}{3}t - 6 - \frac{2}{3}\sqrt{(t^4 - 10t^3 + 493t^2 - 2340t + 2916)}.$$
If you graph these functions of t you find the second is always positive and the third is always negative. Therefore, all these critical points are saddle points.

21. Show the critical points of the following function are
$$(0,-3,0), (2,-3,0), \left(1,-3,-\frac{1}{3}\right)$$
and classify them as local minima, local maxima or saddle points.
$f(x,y,z) = -\frac{3}{2}x^4 + 6x^3 - 6x^2 + zx^2 - 2zx - 2y^2 - 12y - 18 - \frac{3}{2}z^2$.
The Hessian is
$$\begin{pmatrix} -12 + 36x + 2z - 18x^2 & 0 & -2 + 2x \\ 0 & -4 & 0 \\ -2 + 2x & 0 & -3 \end{pmatrix}$$
Now consider the critical point, $\left(1,-3,-\frac{1}{3}\right)$. At this point the Hessian matrix equals
$$\begin{pmatrix} \frac{16}{3} & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -3 \end{pmatrix},$$

The eigenvalues are $\frac{16}{3}, -3, -4$ and so this point is a saddle point.

Next consider the critical point, $(2, -3, 0)$. At this point the Hessian matrix is

$$\begin{pmatrix} -12 & 0 & 2 \\ 0 & -4 & 0 \\ 2 & 0 & -3 \end{pmatrix}$$

The eigenvalues are $-4, -\frac{15}{2} + \frac{1}{2}\sqrt{97}, -\frac{15}{2} - \frac{1}{2}\sqrt{97}$, all negative so at this point there is a local max.

Finally consider the critical point, $(0, -3, 0)$. At this point the Hessian is

$$\begin{pmatrix} -12 & 0 & -2 \\ 0 & -4 & 0 \\ -2 & 0 & -3 \end{pmatrix}$$

and the eigenvalues are the same as the above, all negative. Therefore, there is a local maximum at this point.

22. Show the critical points of the following function are points of the form, $(x, y, z) = \left(t, 2t^2 + 6t, -t^2 - 3t\right)$ for $t \in \mathbf{R}$ and classify them as local minima, local maxima or saddle points.

    $f(x, y, z) = -2yx^2 - 6yx - 4zx^2 - 12zx + y^2 + 2yz.$

23. Show the critical points of the following function are $(0, -1, 0), (4, -1, 0)$, and $(2, -1, -12)$ and classify them as local minima, local maxima or saddle points.

    $f(x, y, z) = \frac{1}{2}x^4 - 4x^3 + 8x^2 - 3zx^2 + 12zx + 2y^2 + 4y + 2 + \frac{1}{2}z^2.$

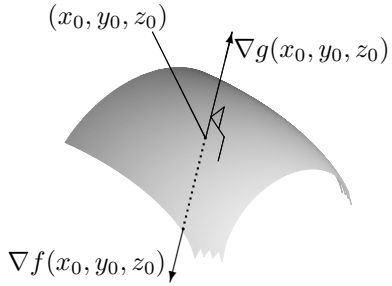
24. Can you establish the following theorem which generalizes Theorem 13.4.3? Suppose U is an open set contained in $D(f)$ such that f is differentiable at $\mathbf{x} \in U$ and $\mathbf{x}$ is either a local minimum or local maximum for f. Then $\nabla f(\mathbf{x}) = \mathbf{0}$. **Hint:** It ought to be this way because it works like this for a function of one variable. Differentiability at the local max. or min. is sufficient. You don't have to know the function is differentiable near the point, only at the point. This is not hard to do if you use the definition of the derivative.

25. Suppose $f(x, y)$, a function of two variables defined on all $\mathbb{R}^n$ has all directional derivatives at $(0, 0)$ and they are all equal to 0 there. Suppose also that for $h(t) \equiv f(tu, tv)$ and (u, v) a unit vector, it follows that $h''(0) > 0$. By the one variable second derivative test, this implies that along every straight line through $(0, 0)$ the function restricted to this line has a local minimum at $(0, 0)$. Can it be concluded that f has a local minimum at $(0, 0)$. In other words, can you conclude a point is a local minimum if it appears to be so along every straight line through the point? **Hint:** Consider $f(x, y) = x^2 + y^2$ for (x, y) not on the curve $y = x^2$ for $x \neq 0$ and on this curve, let $f = -1$.

13.7 Lagrange Multipliers

Lagrange multipliers are used to solve extremum problems for a function defined on a level set of another function. For example, suppose you want to maximize xy given that $x + y = 4$. This is not too hard to do using methods developed earlier. Solve for one of the variables, say y, in the constraint equation, $x + y = 4$ to find $y = 4 - x$. Then the function to maximize is $f(x) = x(4 - x)$ and the answer is clearly $x = 2$.

Thus the two numbers are $x = y = 2$. This was easy because you could easily solve the constraint equation for one of the variables in terms of the other. Now what if you wanted to maximize $f(x,y,z) = xyz$ subject to the constraint that $x^2+y^2+z^2 = 4$? It is still possible to do this using using similar techniques. Solve for one of the variables in the constraint equation, say z, substitute it into f, and then find where the partial derivatives equal zero to find candidates for the extremum. However, it seems you might encounter many cases and it does look a little fussy. However, sometimes you can't solve the constraint equation for one variable in terms of the others. Also, what if you had many constraints. What if you wanted to maximize $f(x,y,z)$ subject to the constraints $x^2+y^2 = 4$ and $z = 2x+3y^2$. Things are clearly getting more involved and messy. It turns out that at an extremum, there is a simple relationship between the gradient of the function to be maximized and the gradient of the constraint function. This relation can be seen geometrically as in the following picture.



In the picture, the surface represents a piece of the level surface of $g(x,y,z) = 0$ and $f(x,y,z)$ is the function of three variables which is being maximized or minimized on the level surface and suppose the extremum of f occurs at the point (x_0,y_0,z_0). As shown above, $\nabla g(x_0,y_0,z_0)$ is perpendicular to the surface or more precisely to the tangent plane. However, if $\mathbf{x}(t) = (x(t), y(t), z(t))$ is a point on a smooth curve which passes through (x_0,y_0,z_0) when $t = t_0$, then the function, $h(t) = f(x(t), y(t), z(t))$ must have either a maximum or a minimum at the point, $t = t_0$. Therefore, $h'(t_0) = 0$. But this means

$$\begin{aligned} 0 &= h'(t_0) = \nabla f(x(t_0), y(t_0), z(t_0)) \cdot \mathbf{x}'(t_0) \\ &= \nabla f(x_0,y_0,z_0) \cdot \mathbf{x}'(t_0) \end{aligned}$$

and since this holds for any such smooth curve, $\nabla f(x_0,y_0,z_0)$ is also perpendicular to the surface. This picture represents a situation in three dimensions and you can see that it is intuitively clear that this implies $\nabla f(x_0,y_0,z_0)$ is some scalar multiple of $\nabla g(x_0,y_0,z_0)$. Thus

$$\nabla f(x_0,y_0,z_0) = \lambda \nabla g(x_0,y_0,z_0)$$

This λ is called a **Lagrange multiplier** after Lagrange who considered such problems in the 1700's.

Of course the above argument is at best only heuristic. It does not deal with the question of existence of smooth curves lying in the constraint surface passing through (x_0,y_0,z_0). Nor does it consider all cases, being essentially confined to three dimensions. In addition to this, it fails to consider the situation in which there are many constraints. However, I think it is likely a geometric notion like that presented above which led Lagrange to formulate the method.

Example 13.7.1 *Maximize xyz subject to $x^2+y^2+z^2 = 27$.*

Here $f(x,y,z) = xyz$ while $g(x,y,z) = x^2+y^2+z^2-27$. Then $\nabla g(x,y,z) = (2x,2y,2z)$ and $\nabla f(x,y,z) = (yz,xz,xy)$. Then at the point which maximizes this

function[2],

$$(yz, xz, xy) = \lambda (2x, 2y, 2z).$$

Therefore, each of $2\lambda x^2, 2\lambda y^2, 2\lambda z^2$ equals xyz. It follows that at any point which maximizes xyz, $|x| = |y| = |z|$. Therefore, the only candidates for the point where the maximum occurs are $(3,3,3), (-3,-3,3)(-3,3,3)$, etc. The maximum occurs at $(3,3,3)$ which can be verified by plugging in to the function which is being maximized.

The method of Lagrange multipliers allows you to consider maximization of functions defined on closed and bounded sets. Recall that any continuous function defined on a closed and bounded set has a maximum and a minimum on the set. Candidates for the extremum on the interior of the set can be located by setting the gradient equal to zero. The consideration of the boundary can then sometimes be handled with the method of Lagrange multipliers.

Example 13.7.2 *Maximize $f(x,y) = xy + y$ subject to the constraint, $x^2 + y^2 \leq 1$.*

Here I know there is a maximum because the set is the closed circle, a closed and bounded set. Therefore, it is just a matter of finding it. Look for singular points on the interior of the circle. $\nabla f(x,y) = (y, x+1) = (0,0)$. There are no points on the interior of the circle where the gradient equals zero. Therefore, the maximum occurs on the boundary of the circle. That is the problem reduces to maximizing $xy + y$ subject to $x^2 + y^2 = 1$. From the above,

$$(y, x+1) - \lambda (2x, 2y) = 0.$$

Hence $y^2 - 2\lambda xy = 0$ and $x(x+1) - 2\lambda xy = 0$ so $y^2 = x(x+1)$. Therefore from the constraint, $x^2 + x(x+1) = 1$ and the solution is $x = -1, x = \frac{1}{2}$. Then the candidates for a solution are $(-1,0), \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right), \left(\frac{1}{2}, \frac{-\sqrt{3}}{2}\right)$. Then

$$f(-1,0) = 0, f\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = \frac{3\sqrt{3}}{4}, f\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) = -\frac{3\sqrt{3}}{4}.$$

It follows the maximum value of this function is $\frac{3\sqrt{3}}{4}$ and it occurs at $\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$. The minimum value is $-\frac{3\sqrt{3}}{4}$ and it occurs at $\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)$.

Example 13.7.3 *Find the maximum and minimum values of the function, $f(x,y) = xy - x^2$ on the set, $\left\{(x,y) : x^2 + 2xy + y^2 \leq 4\right\}$.*

First, the only point where ∇f equals zero is $(x,y) = (0,0)$ and this is in the desired set. In fact it is an interior point of this set. This takes care of the interior points. What about those on the boundary $x^2 + 2xy + y^2 = 4$? The problem is to maximize $xy - x^2$ subject to the constraint, $x^2 + 2xy + y^2 = 4$. The Lagrangian is $xy - x^2 - \lambda\left(x^2 + 2xy + y^2 - 4\right)$ and this yields the following system.

$$\begin{gathered} y - 2x - \lambda(2x + 2y) = 0 \\ x - 2\lambda(x + y) = 0 \\ 2x^2 + 2xy + y^2 = 4 \end{gathered}$$

From the first two equations,

$$\begin{gathered} (2 + 2\lambda)x - (1 - 2\lambda)y = 0 \\ (1 - 2\lambda)x - 2\lambda y = 0 \end{gathered}$$

[2]There exists such a point because the sphere is closed and bounded.

Since not both x and y equal zero, it follows

$$\det \begin{pmatrix} 2+2\lambda & 2\lambda-1 \\ 1-2\lambda & -2\lambda \end{pmatrix} = 0$$

which yields

$$\lambda = 1/8$$

Therefore,

$$y = -\frac{3}{4}x \tag{13.6}$$

From the constraint equation,

$$2x^2 + 2x\left(-\frac{3}{4}x\right) + \left(-\frac{3}{4}x\right)^2 = 4$$

and so

$$x = \frac{8}{17}\sqrt{17} \text{ or } -\frac{8}{17}\sqrt{17}$$

Now from 13.6, the points of interest on the boundary of this set are

$$\left(\frac{8}{17}\sqrt{17}, -\frac{6}{17}\sqrt{17}\right), \text{ and } \left(-\frac{8}{17}\sqrt{17}, \frac{6}{17}\sqrt{17}\right). \tag{13.7}$$

$$\begin{aligned} f\left(\frac{8}{17}\sqrt{17}, -\frac{6}{17}\sqrt{17}\right) &= \left(\frac{8}{17}\sqrt{17}\right)\left(-\frac{6}{17}\sqrt{17}\right) - \left(\frac{8}{17}\sqrt{17}\right)^2 \\ &= -\frac{112}{17} \end{aligned}$$

$$\begin{aligned} f\left(-\frac{8}{17}\sqrt{17}, \frac{6}{17}\sqrt{17}\right) &= \left(-\frac{8}{17}\sqrt{17}\right)\left(\frac{6}{17}\sqrt{17}\right) - \left(-\frac{8}{17}\sqrt{17}\right)^2 \\ &= -\frac{112}{17} \end{aligned}$$

It follows the maximum value of this function on the given set occurs at $(0,0)$ and is equal to zero and the minimum occurs at either of the two points in 13.7 and has the value $-112/17$.

This illustrates how to use the method of Lagrange multipliers to identify the extrema for a function defined on a closed and bounded set. You try and consider the boundary as a level curve or level surface and then use the method of Lagrange multipliers on it and look for singular points on the interior of the set.

There are no magic bullets here. It was still required to solve a system of nonlinear equations to get the answer. However, it does often help to do it this way.

The above generalizes to a general procedure which is described in the following major Theorem. All correct proofs of this theorem will involve some appeal to the implicit function theorem or to fundamental existence theorems from differential equations. A complete proof is very fascinating but it will not come cheap. Good advanced calculus books will usually give a correct proof and there is a proof given in an appendix to this book. First here is a simple definition explaining one of the terms in the statement of this theorem.

Definition 13.7.4 *Let A be an $m \times n$ matrix. A submatrix is any matrix which can be obtained from A by deleting some rows and some columns.*

Theorem 13.7.5 *Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be a C^1 function. Then if $\mathbf{x}_0 \in U$ is either a local maximum or local minimum of f subject to the constraints*

$$g_i(\mathbf{x}) = 0,\ i = 1, \cdots, m \tag{13.8}$$

and if some $m \times m$ submatrix of

$$D\mathbf{g}(\mathbf{x}_0) \equiv \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) & g_{1x_2}(\mathbf{x}_0) & \cdots & g_{1x_n}(\mathbf{x}_0) \\ \vdots & \vdots & & \vdots \\ g_{mx_1}(\mathbf{x}_0) & g_{mx_2}(\mathbf{x}_0) & \cdots & g_{mx_n}(\mathbf{x}_0) \end{pmatrix}$$

has nonzero determinant, then there exist scalars, $\lambda_1, \cdots, \lambda_m$ such that

$$\begin{pmatrix} f_{x_1}(\mathbf{x}_0) \\ \vdots \\ f_{x_n}(\mathbf{x}_0) \end{pmatrix} = \lambda_1 \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix} + \cdots + \lambda_m \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix} \tag{13.9}$$

holds.

To help remember how to use 13.9 it may be helpful to do the following. First write the Lagrangian,

$$L = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

and then proceed to take derivatives with respect to each of the components of $\mathbf{x}$ and also derivatives with respect to each λ_i and set all of these equations equal to 0. The formula 13.9 is what results from taking the derivatives of L with respect to the components of $\mathbf{x}$. When you take the derivatives with respect to the Lagrange multipliers, and set what results equal to 0, you just pick up the constraint equations. This yields $n+m$ equations for the $n+m$ unknowns, $x_1, \cdots, x_n, \lambda_1, \cdots, \lambda_m$. Then you proceed to look for solutions to these equations. Of course these might be impossible to find using methods of algebra, but you just do your best and hope it will work out.

Example 13.7.6 *Minimize xyz subject to the constraints $x^2+y^2+z^2 = 4$ and $x-2y = 0$.*

Form the Lagrangian,

$$L = xyz - \lambda\left(x^2 + y^2 + z^2 - 4\right) - \mu(x - 2y)$$

and proceed to take derivatives with respect to every possible variable, leading to the following system of equations.

$$\begin{aligned} yz - 2\lambda x - \mu &= 0 \\ xz - 2\lambda y + 2\mu &= 0 \\ xy - 2\lambda z &= 0 \\ x^2 + y^2 + z^2 &= 4 \\ x - 2y &= 0 \end{aligned}$$

Now you have to find the solutions to this system of equations. In general, this could be very hard or even impossible. If $\lambda = 0$, then from the third equation, either x or y must equal 0. Therefore, from the first two equations, $\mu = 0$ also. If $\mu = 0$ and $\lambda \neq 0$, then from the first two equations, $xyz = 2\lambda x^2$ and $xyz = 2\lambda y^2$ and so either $x = y$ or $x = -y$, which requires that both x and y equal zero thanks to the last equation. But

then from the fourth equation, $z = \pm 2$ and now this contradicts the third equation. Thus μ and λ are either both equal to zero or neither one is and the expression, xyz equals zero in this case. However, I know this is not the best value for a minimizer because I can take $x = 2\sqrt{\frac{3}{5}}, y = \sqrt{\frac{3}{5}}$, and $z = -1$. This satisfies the constraints and the product of these numbers equals a negative number. Therefore, both μ and λ must be non zero. Now use the last equation eliminate x and write the following system.

$$\begin{array}{c} 5y^2 + z^2 = 4 \\ y^2 - \lambda z = 0 \\ yz - \lambda y + \mu = 0 \\ yz - 4\lambda y - \mu = 0 \end{array}$$

From the last equation, $\mu = (yz - 4\lambda y)$. Substitute this into the third and get

$$\begin{array}{c} 5y^2 + z^2 = 4 \\ y^2 - \lambda z = 0 \\ yz - \lambda y + yz - 4\lambda y = 0 \end{array}$$

$y = 0$ will not yield the minimum value from the above example. Therefore, divide the last equation by y and solve for λ to get $\lambda = (2/5)\, z$. Now put this in the second equation to conclude

$$\begin{array}{c} 5y^2 + z^2 = 4 \\ y^2 - (2/5)\, z^2 = 0 \end{array},$$

a system which is easy to solve. Thus $y^2 = 8/15$ and $z^2 = 4/3$. Therefore, candidates for minima are $\left(2\sqrt{\frac{8}{15}}, \sqrt{\frac{8}{15}}, \pm\sqrt{\frac{4}{3}}\right)$, and $\left(-2\sqrt{\frac{8}{15}}, -\sqrt{\frac{8}{15}}, \pm\sqrt{\frac{4}{3}}\right)$, a choice of 4 points to check. Clearly the one which gives the smallest value is

$$\left(2\sqrt{\frac{8}{15}}, \sqrt{\frac{8}{15}}, -\sqrt{\frac{4}{3}}\right)$$

or $\left(-2\sqrt{\frac{8}{15}}, -\sqrt{\frac{8}{15}}, -\sqrt{\frac{4}{3}}\right)$ and the minimum value of the function subject to the constraints is $-\frac{2}{5}\sqrt{30} - \frac{2}{3}\sqrt{3}$.

You should rework this problem first solving the second easy constraint for x and then producing a simpler problem involving only the variables y and z.

13.8 Exercises

1. Maximize $2x + 3y - 6z$ subject to the constraint, $x^2 + 2y^2 + 3z^2 = 9$.
2. Find the dimensions of the largest rectangle which can be inscribed in a circle of radius r.
3. Maximize $2x + y$ subject to the condition that $\frac{x^2}{4} + \frac{y^2}{9} \leq 1$.
4. Maximize $x + 2y$ subject to the condition that $x^2 + \frac{y^2}{9} \leq 1$.
5. Maximize $x + y$ subject to the condition that $x^2 + \frac{y^2}{9} + z^2 \leq 1$.
6. Maximize $x + y + z$ subject to the condition that $x^2 + \frac{y^2}{9} + z^2 \leq 1$.
7. Find the points on $y^2 x = 9$ which are closest to $(0, 0)$.

8. Find points on $xy = 4$ farthest from $(0,0)$ if any exist. If none exist, tell why. What does this say about the method of Lagrange multipliers?

9. A can is supposed to have a volume of 36π cubic centimeters. Find the dimensions of the can which minimizes the surface area.

10. A can is supposed to have a volume of 36π cubic centimeters. The top and bottom of the can are made of tin costing 4 cents per square centimeter and the sides of the can are made of aluminum costing 5 cents per square centimeter. Find the dimensions of the can which minimizes the cost.

11. Minimize $\sum_{j=1}^{n} x_j$ subject to the constraint $\sum_{j=1}^{n} x_j^2 = a^2$. Your answer should be some function of a which you may assume is a positive number.

12. Find the point, (x,y,z) on the level surface, $4x^2 + y^2 - z^2 = 1$which is closest to $(0,0,0)$.

13. A curve is formed from the intersection of the plane, $2x + 3y + z = 3$ and the cylinder $x^2 + y^2 = 4$. Find the point on this curve which is closest to $(0,0,0)$.

14. A curve is formed from the intersection of the plane, $2x + 3y + z = 3$ and the sphere $x^2 + y^2 + z^2 = 16$. Find the point on this curve which is closest to $(0,0,0)$.

15. Find the point on the plane, $2x + 3y + z = 4$ which is closest to the point $(1,2,3)$.

16. Let $A = (A_{ij})$ be an $n \times n$ matrix which is symmetric. Thus $A_{ij} = A_{ji}$ and recall $(A\mathbf{x})_i = A_{ij}x_j$ where as usual sum over the repeated index. Show $\frac{\partial}{\partial x_i}(A_{ij}x_j x_i) = 2A_{ij}x_j$. Show that when you use the method of Lagrange multipliers to maximize the function, $A_{ij}x_j x_i$ subject to the constraint, $\sum_{j=1}^{n} x_j^2 = 1$, the value of λ which corresponds to the maximum value of this functions is such that $A_{ij}x_j = \lambda x_i$. Thus $A\mathbf{x} = \lambda\mathbf{x}$.

17. Here are two lines. $\mathbf{x} = (1 + 2t, 2 + t, 3 + t)^T$ and $\mathbf{x} = (2 + s, 1 + 2s, 1 + 3s)^T$. Find points $\mathbf{p}_1$ on the first line and $\mathbf{p}_2$ on the second with the property that $|\mathbf{p}_1 - \mathbf{p}_2|$ is at least as small as the distance between any other pair of points, one chosen on one line and the other on the other line.

18. Find the dimensions of the largest triangle which can be inscribed in a circle of radius r.

19. Find the point on the intersection of $z = x^2 + y^2$ and $x + y + z = 1$ which is closest to $(0,0,0)$.

20. Minimize $4x^2 + y^2 + 9z^2$ subject to $x + y - z = 1$ and $x - 2y + z = 0$.

21. Minimize xyz subject to the constraints $x^2 + y^2 + z^2 = r^2$ and $x - y = 0$.

22. Let n be a positive integer. Find n numbers whose sum is $8n$ and the sum of the squares is as small as possible.

23. Find the point on the level surface, $2x^2 + xy + z^2 = 16$ which is closest to $(0,0,0)$.

24. Find the point on $\frac{x^2}{4} + \frac{y^2}{9} + z^2 = 1$ closest to the plane $x + y + z = 10$.

25. Let $x_1, \cdots, x_5$ be 5 positive numbers. Maximize their product subject to the constraint that
$$x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 300.$$

26. Let $f(x_1, \cdots, x_n) = x_1^n x_2^{n-1} \cdots x_n^1$. Then f achieves a maximum on the set,

$$S \equiv \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n i x_i = 1 \text{ and each } x_i \geq 0 \right\}.$$

If $\mathbf{x} \in S$ is the point where this maximum is achieved, find x_1/x_n.

27. Let (x, y) be a point on the ellipse, $x^2/a^2 + y^2/b^2 = 1$ which is in the first quadrant. Extend the tangent line through (x, y) till it intersects the x and y axes and let $A(x, y)$ denote the area of the triangle formed by this line and the two coordinate axes. Find the maximum value of the area of this triangle as a function of a and b.

28. Maximize $\prod_{i=1}^n x_i^2$ $(\equiv x_1^2 \times x_2^2 \times x_3^2 \times \cdots \times x_n^2)$ subject to the constraint, $\sum_{i=1}^n x_i^2 = r^2$. Show the maximum is $\left(r^2/n\right)^n$. Now show from this that

$$\left(\prod_{i=1}^n x_i^2 \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n x_i^2$$

and finally, conclude that if each number $x_i \geq 0$, then

$$\left(\prod_{i=1}^n x_i \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n x_i$$

and there exist values of the x_i for which equality holds. This says the "geometric mean" is always smaller than the arithmetic mean.

29. Maximize $x^2 y^2$ subject to the constraint

$$\frac{x^{2p}}{p} + \frac{y^{2q}}{q} = r^2$$

where p, q are real numbers larger than 1 which have the property that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

show the maximum is achieved when $x^{2p} = y^{2q}$ and equals r^2. Now conclude that if $x, y > 0$, then

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}$$

and there are values of x and y where this inequality is an equation.

13.9 Exercises With Answers

1. Maximize $x + 3y - 6z$ subject to the constraint, $x^2 + 2y^2 + z^2 = 9$.

The Lagrangian is $L = x + 3y - 6z - \lambda\left(x^2 + 2y^2 + z^2 - 9\right)$. Now take the derivative with respect to x. This gives the equation $1 - 2\lambda x = 0$. Next take the derivative with respect to y. This gives the equation $3 - 4\lambda y = 0$. The derivative with respect to z gives $-6 - 2\lambda z = 0$. Clearly $\lambda \neq 0$ since this would contradict the first of these equations. Similarly, none of the variables, x, y, z can equal zero. Solving each of these equations for λ gives $\frac{1}{2x} = \frac{3}{4y} = \frac{-3}{z}$. Thus $y = \frac{3x}{2}$ and $z = -6x$. Now you use the constraint equation plugging in these values for y and

z. $x^2+2\left(\frac{3x}{2}\right)^2+(-6x)^2=9$. This gives the values for x as $x=\frac{3}{83}\sqrt{166}, x=-\frac{3}{83}\sqrt{166}$. From the three equations above, this also determines the values of z and y. $y=\frac{9}{166}\sqrt{166}$ or $-\frac{9}{166}\sqrt{166}$ and $z=-\frac{18}{83}\sqrt{166}$ or $\frac{18}{83}\sqrt{166}$. Thus there are two points to look at. One will give the minimum value and the other will give the maximum value. You know the minimum and maximum exist because of the extreme value theorem. The two points are $\left(\frac{3}{83}\sqrt{166},\frac{9}{166}\sqrt{166},-\frac{18}{83}\sqrt{166}\right)$ and $\left(-\frac{3}{83}\sqrt{166},-\frac{9}{166}\sqrt{166},\frac{18}{83}\sqrt{166}\right)$. Now you just need to find which is the minimum and which is the maximum. Plug these in to the function you are trying to maximize. $\left(\frac{3}{83}\sqrt{166}\right)+3\left(\frac{9}{166}\sqrt{166}\right)-6\left(-\frac{18}{83}\sqrt{166}\right)$ will clearly be the maximum value occuring at $\left(\frac{3}{83}\sqrt{166},\frac{9}{166}\sqrt{166},-\frac{18}{83}\sqrt{166}\right)$. The other point will obviously yield the minimum because this one is positive and the other one is negative. If you use a calculator to compute this you get $\left(\frac{3}{83}\sqrt{166}\right)+3\left(\frac{9}{166}\sqrt{166}\right)-6\left(-\frac{18}{83}\sqrt{166}\right)=19.326$.

2. Find the dimensions of the largest rectangle which can be inscribed in a the ellipse $x^2+4y^2=4$.

   This is one which you could do without Lagrange multipliers. However, it is easier with Lagrange multipliers. Let a corner of the rectangle be at (x,y). Then the area of the rectangle will be $4xy$ and since (x,y) is on the ellipse, you have the constraint $x^2+4y^2=4$. Thus the problem is to maximize $4xy$ subject to $x^2+4y^2=4$. The Lagrangian is then $L=4xy-\lambda\left(x^2+4y^2-4\right)$ and so you get the equations $4y-2\lambda x=0$ and $4x-8\lambda y=0$. You can't have both x and y equal to zero and satisfy the constraint. Therefore, the determinant of the matrix of coefficients must equal zero. Thus $\begin{vmatrix} -2\lambda & 4 \\ 4 & -8\lambda \end{vmatrix}=16\lambda^2-16=0$. This is because the system of equations is of the form

$$\begin{pmatrix} -2\lambda & 4 \\ 4 & -8\lambda \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

   If the matrix has an inverse, then the only solution would be $x=y=0$ which as noted above can't happen. Therefore, $\lambda=\pm1$. First suppose $\lambda=1$. Then the first equation says $2y=x$. Pluggin this in to the constraint equation, $x^2+x^2=4$ and so $x=\pm\sqrt{2}$. Therefore, $y=\pm\frac{\sqrt{2}}{2}$. This yields the dimensions of the largest rectangle to be $2\sqrt{2}\times\sqrt{2}$. You can check all the other cases and see you get the same thing in the other cases as well.

3. Maximize $2x+y$ subject to the condition that $\frac{x^2}{4}+y^2\leq1$.

   The maximum of this function clearly exists because of the extreme value theorem since the condition defines a closed and bounded set in $\mathbb{R}^2$. However, this function does not achieve its maximum on the interior of the given ellipse defined by $\frac{x^2}{4}+y^2\leq1$ because the gradient of the function which is to be maximized is never equal to zero. Therefore, this function must achieve its maximum on the set $\frac{x^2}{4}+y^2=1$. Thus you want to maximuze $2x+y$ subject to $\frac{x^2}{4}+y^2=1$. This is just like Problem 1. You can finish this.

4. Find the points on $y^2x=16$ which are closest to $(0,0)$.

   You want to maximize x^2+y^2 subject to y^2x-16. Of course you really want to maximize $\sqrt{x^2+y^2}$ but the ordered pair which maximized x^2+y^2 is the same as the ordered pair whic maximized $\sqrt{x^2+y^2}$ so it is pointless to drag around the square root. The Lagrangian is $x^2+y^2-\lambda\left(y^2x-16\right)$. Differentiating with respect to x and y gives the equations $2x-\lambda y^2=0$ and $2y-2\lambda yx=0$. Neither

x nor y can equal zero and solve the constraint. Therefore, the second equation implies $\lambda x = 1$. Hence $\lambda = \frac{1}{x} = \frac{2x}{y^2}$. Therefore, $2x^2 = y^2$ and so $2x^3 = 16$ and so $x = 2$. Therefore, $y = \pm 2\sqrt{2}$. The points are $\left(2, 2\sqrt{2}\right)$ and $\left(2, -2\sqrt{2}\right)$. They both give the same answer. Note how ad hoc these procedures are. I can't give you a simple strategy for solving these systems of nonlinear equations by algebra because there is none. Sometimes nothing you do will work.

5. Find points on $xy = 1$ farthest from $(0,0)$ if any exist. If none exist, tell why. What does this say about the method of Lagrange multipliers?

   If you graph $xy = 1$ you see there is no farthest point. However, there is a closest point and the method of Lagrange multipliers will find this closest point. This shows that the answer you get has to be carefully considered to determine whether you have a maximum or a minimum or perhaps neither.

6. A curve is formed from the intersection of the plane, $2x + y + z = 3$ and the cylinder $x^2 + y^2 = 4$. Find the point on this curve which is closest to $(0,0,0)$.

   You want to maximize $x^2 + y^2 + z^2$ subject to the two constraints $2x + y + z = 3$ and $x^2 + y^2 = 4$. This means the Lagrangian will have two multipliers.

$$L = x^2 + y^2 + z^2 - \lambda\left(2x + y + z - 3\right) - \mu\left(x^2 + y^2 - 4\right)$$

   Then this yields the equations $2x - 2\lambda - 2\mu x = 0$, $2y - \lambda - 2\mu y$, and $2z - \lambda = 0$. The last equation says $\lambda = 2z$ and so I will replace λ with $2z$ where ever it occurs. This yields

$$x - 2z - \mu x = 0, 2y - 2z - 2\mu y = 0.$$

   This shows $x\left(1 - \mu\right) = 2y\left(1 - \mu\right)$. First suppose $\mu = 1$. Then from the above equations, $z = 0$ and so the two constraints reduce to $2y + x = 3$ and $x^2 + y^2 = 4$ and $2y + x = 3$. The solutions are $\left(\frac{3}{5} - \frac{2}{5}\sqrt{11}, \frac{6}{5} + \frac{1}{5}\sqrt{11}, 0\right)$, $\left(\frac{3}{5} + \frac{2}{5}\sqrt{11}, \frac{6}{5} - \frac{1}{5}\sqrt{11}, 0\right)$. The other case is that $\mu \neq 1$ in which case $x = 2y$ and the second constraint yields that $y = \pm\frac{2}{\sqrt{5}}$ and $x = \pm\frac{4}{\sqrt{5}}$. Now from the first constraint, $z = -2\sqrt{5} + 3$ in the case where $y = \frac{2}{\sqrt{5}}$ and $z = 2\sqrt{5} + 3$ in the other case. This yields the points $\left(\frac{4}{\sqrt{5}}, \frac{2}{\sqrt{5}}, -2\sqrt{5} + 3\right)$ and $\left(-\frac{4}{\sqrt{5}}, -\frac{2}{\sqrt{5}}, 2\sqrt{5} + 3\right)$. This appears to have exhausted all the possibilities and so it is now just a matter of seeing which of these points gives the best answer. An answer exists because of the extreme value theorem. After all, this constraint set is closed and bounded. The first candidate listed above yields for the answer $\left(\frac{3}{5} - \frac{2}{5}\sqrt{11}\right)^2 + \left(\frac{6}{5} + \frac{1}{5}\sqrt{11}\right)^2 = 4$. The second candidate listed above yields $\left(\frac{3}{5} + \frac{2}{5}\sqrt{11}\right)^2 + \left(\frac{6}{5} - \frac{1}{5}\sqrt{11}\right)^2 = 4$ also. Thus these two give equally good results. Now consider the last two candidates. $\left(\frac{4}{\sqrt{5}}\right)^2 + \left(\frac{2}{\sqrt{5}}\right)^2 + \left(-2\sqrt{5} + 3\right)^2 = 4 + \left(-2\sqrt{5} + 3\right)^2$ which is larger than 4. Finally the last candidate yields $\left(-\frac{4}{\sqrt{5}}\right)^2 + \left(-\frac{2}{\sqrt{5}}\right)^2 + \left(2\sqrt{5} + 3\right)^2 = 4 + \left(2\sqrt{5} + 3\right)^2$ also larger than 4. Therefore, there are two points on the curve of intersection which are closest to the origin, $\left(\frac{3}{5} - \frac{2}{5}\sqrt{11}, \frac{6}{5} + \frac{1}{5}\sqrt{11}, 0\right)$ and $\left(\frac{3}{5} + \frac{2}{5}\sqrt{11}, \frac{6}{5} - \frac{1}{5}\sqrt{11}, 0\right)$. Both are a distance of 4 from the origin.

7. Here are two lines. $\mathbf{x} = \left(1 + 2t, 2 + t, 3 + t\right)^T$ and $\mathbf{x} = \left(2 + s, 1 + 2s, 1 + 3s\right)^T$. Find points $\mathbf{p}_1$ on the first line and $\mathbf{p}_2$ on the second with the property that $|\mathbf{p}_1 - \mathbf{p}_2|$ is at least as small as the distance between any other pair of points, one chosen on one line and the other on the other line.

   Hint: Do you need to use Lagrange multipliers for this?

8. Find the point on $x^2+y^2+z^2=1$ closest to the plane $x+y+z=10$.

   You want to minimize $(x-a)^2+(y-b)^2+(z-c)^2$ subject to the constraints $a+b+c=10$ and $x^2+y^2+z^2=1$. There seem to be a lot of variables in this problem, 6 in all. Start taking derivatives and hope for a miracle. This yields $2(x-a)-2\mu x=0, 2(y-b)-2\mu y=0, 2(z-c)-2\mu z=0$. Also, taking derivatives with respect to a, b, and c you obtain $2(x-a)+\lambda=0, 2(y-b)+\lambda=0, 2(z-c)+\lambda=0$. Comparing the first equations in each list, you see $\lambda=2\mu x$ and then comparing the second two equations in each list, $\lambda=2\mu y$ and similarly, $\lambda=2\mu z$. Therefore, if $\mu\neq 0$, it must follow that $x=y=z$. Now you can see by sketching a rough graph that the answer you want has each of x, y, and z nonnegative. Therefore, using the constraint for these variables, the point desired is $\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right)$ which you could probably see was the answer from the sketch. However, this could be made more difficult rather easily such that the sketch won't help but Lagrange multipliers will.

13.10 Proof Of The Second Derivative Test*

Definition 13.10.1 *The matrix,* $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})\right)$ *is called the Hessian matrix, denoted by* $H(\mathbf{x})$.

Now recall the Taylor formula with the Lagrange form of the remainder. Here is a statement and proof of this important theorem.

Theorem 13.10.2 *Suppose* f *has* $n+1$ *derivatives on an interval,* (a,b) *and let* $c\in(a,b)$. *Then if* $x\in(a,b)$, *there exists* ξ *between* c *and* x *such that*

$$f(x)=f(c)+\sum_{k=1}^{n}\frac{f^{(k)}(c)}{k!}(x-c)^k+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-c)^{n+1}.$$

(In this formula, the symbol $\sum_{k=1}^{0}a_k$ *will denote the number* 0.*)*

Proof: If $n=0$ then the theorem is true because it is just the mean value theorem. Suppose the theorem is true for $n-1, n\geq 1$. It can be assumed $x\neq c$ because if $x=c$ there is nothing to show. Then there exists K such that

$$f(x)-\left(f(c)+\sum_{k=1}^{n}\frac{f^{(k)}(c)}{k!}(x-c)^k+K(x-c)^{n+1}\right)=0 \tag{13.10}$$

In fact,

$$K=\frac{-f(x)+\left(f(c)+\sum_{k=1}^{n}\frac{f^{(k)}(c)}{k!}(x-c)^k\right)}{(x-c)^{n+1}}.$$

Now define $F(t)$ for t in the closed interval determined by x and c by

$$F(t)\equiv f(x)-\left(f(t)+\sum_{k=1}^{n}\frac{f^{(k)}(c)}{k!}(x-t)^k+K(x-t)^{n+1}\right).$$

The c in 13.10 got replaced by t.

Therefore, $F(c)=0$ by the way K was chosen and also $F(x)=0$. By the mean value theorem or Rolle's theorem, there exists t_1 between x and c such that $F'(t_1)=0$.

Therefore,

$$\begin{aligned} 0 &= f'(t_1) - \sum_{k=1}^{n} \frac{f^{(k)}(c)}{k!} k (x-t_1)^{k-1} - K(n+1)(x-t_1)^n \\ &= f'(t_1) - \left(f'(c) + \sum_{k=1}^{n-1} \frac{f^{(k+1)}(c)}{k!} (x-t_1)^k \right) - K(n+1)(x-t_1)^n \\ &= f'(t_1) - \left(f'(c) + \sum_{k=1}^{n-1} \frac{f'^{(k)}(c)}{k!} (x-t_1)^k \right) - K(n+1)(x-t_1)^n \end{aligned}$$

By induction applied to f', there exists ξ between x and t_1 such that the above simplifies to

$$\begin{aligned} 0 &= \frac{f'^{(n)}(\xi)(x-t_1)^n}{n!} - K(n+1)(x-t_1)^n \\ &= \frac{f^{(n+1)}(\xi)(x-t_1)^n}{n!} - K(n+1)(x-t_1)^n \end{aligned}$$

therefore,

$$K = \frac{f^{(n+1)}(\xi)}{(n+1)n!} = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

and the formula is true for n. This proves the theorem.

Now let $f : U \to \mathbb{R}$ where U is an open subset of $\mathbb{R}^n$. Suppose $f \in C^2(U)$. Let $\mathbf{x} \in U$ and let $r > 0$ be such that

$$B(\mathbf{x},r) \subseteq U.$$

Then for $||\mathbf{v}|| < r$ consider

$$f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x}) \equiv h(t)$$

for $t \in [0,1]$. Then from Taylor's theorem for the case where $m = 2$ and the chain rule, using the repeated index summation convention and the chain rule,

$$h'(t) = \frac{\partial f}{\partial x_i}(\mathbf{x}+t\mathbf{v}) v_i,\ h''(t) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}+t\mathbf{v}) v_i v_j.$$

Thus

$$h''(t) = \mathbf{v}^T H(\mathbf{x}+t\mathbf{v})\mathbf{v}.$$

From Theorem 13.10.2 there exists $t \in (0,1)$ such that

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \frac{\partial f}{\partial x_i}(\mathbf{x}) v_i + \frac{1}{2}\mathbf{v}^T H(\mathbf{x}+t\mathbf{v})\mathbf{v}$$

By the continuity of the second partial derivative

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x})\cdot\mathbf{v} + \frac{1}{2}\mathbf{v}^T H(\mathbf{x})\mathbf{v} + \frac{1}{2}\left(\mathbf{v}^T (H(\mathbf{x}+t\mathbf{v}) - H(\mathbf{x}))\mathbf{v}\right) \tag{13.11}$$

where the last term satisfies

$$\lim_{|\mathbf{v}|\to 0} \frac{1}{2} \frac{\left(\mathbf{v}^T (H(\mathbf{x}+t\mathbf{v}) - H(\mathbf{x}))\mathbf{v}\right)}{|\mathbf{v}|^2} = 0 \tag{13.12}$$

because of the continuity of the entries of $H(\mathbf{x})$.

Theorem 13.10.3 *Suppose* $\mathbf{x}$ *is a critical point for* f*. That is, suppose* $\frac{\partial f}{\partial x_i}(\mathbf{x}) = 0$ *for each* i*. Then if* $H(\mathbf{x})$ *has all positive eigenvalues,* $\mathbf{x}$ *is a local minimum. If* $H(\mathbf{x})$ *has all negative eigenvalues, then* $\mathbf{x}$ *is a local maximum. If* $H(\mathbf{x})$ *has a positive eigenvalue, then there exists a direction in which* f *has a local minimum at* $\mathbf{x}$*, while if* $H(\mathbf{x})$ *has a negative eigenvalue, there exists a direction in which* f *has a local maximum at* $\mathbf{x}$*.*

Proof: Since $\nabla f(\mathbf{x}) = \mathbf{0}$, formula 13.11 implies

$$f(\mathbf{x}+\mathbf{v}) = f(\mathbf{x}) + \frac{1}{2}\mathbf{v}^T H(\mathbf{x})\,\mathbf{v} + \frac{1}{2}\left(\mathbf{v}^T\left(H(\mathbf{x}+t\mathbf{v}) - H(\mathbf{x})\right)\mathbf{v}\right) \tag{13.13}$$

and by continuity of the second derivatives, these mixed second derivatives are equal and so $H(\mathbf{x})$ is a symmetric matrix . Thus, by Lemma A.2.27 on Page 425 in the appendix, $H(\mathbf{x})$ has all real eigenvalues. Suppose first that $H(\mathbf{x})$ has all positive eigenvalues and that all are larger than $\delta^2 > 0$. Then by Theorem A.2.29 on Page 425 of the appendix,

$$\mathbf{u}^T H(\mathbf{x})\,\mathbf{u} \geq \delta^2 |\mathbf{u}|^2$$

By continuity of H, if $\mathbf{v}$ is small enough,

$$f(\mathbf{x}+\mathbf{v}) \geq f(\mathbf{x}) + \frac{1}{2}\delta^2|\mathbf{v}|^2 - \frac{1}{4}\delta^2|\mathbf{v}|^2 = f(\mathbf{x}) + \frac{\delta^2}{4}|\mathbf{v}|^2.$$

This shows the first claim of the theorem. The second claim follows from similar reasoning or applying the above to $-f$.

Suppose $H(\mathbf{x})$ has a positive eigenvalue λ^2. Then let $\mathbf{v}$ be an eigenvector for this eigenvalue. Then from 13.13, replacing $\mathbf{v}$ with $s\mathbf{v}$ and letting t depend on s,

$$f(\mathbf{x}+s\mathbf{v}) = f(\mathbf{x}) + \frac{1}{2}s^2\mathbf{v}^T H(\mathbf{x})\,\mathbf{v} +$$

$$\frac{1}{2}s^2\left(\mathbf{v}^T\left(H(\mathbf{x}+ts\mathbf{v}) - H(\mathbf{x})\right)\mathbf{v}\right)$$

which implies

$$\begin{aligned} f(\mathbf{x}+s\mathbf{v}) &= f(\mathbf{x}) + \frac{1}{2}s^2\lambda^2|\mathbf{v}|^2 + \frac{1}{2}s^2\left(\mathbf{v}^T\left(H(\mathbf{x}+ts\mathbf{v}) - H(\mathbf{x})\right)\mathbf{v}\right) \\ &\geq f(\mathbf{x}) + \frac{1}{4}s^2\lambda^2|\mathbf{v}|^2 \end{aligned}$$

whenever s is small enough. Thus in the direction $\mathbf{v}$ the function has a local minimum at $\mathbf{x}$. The assertion about the local maximum in some direction follows similarly. This proves the theorem.

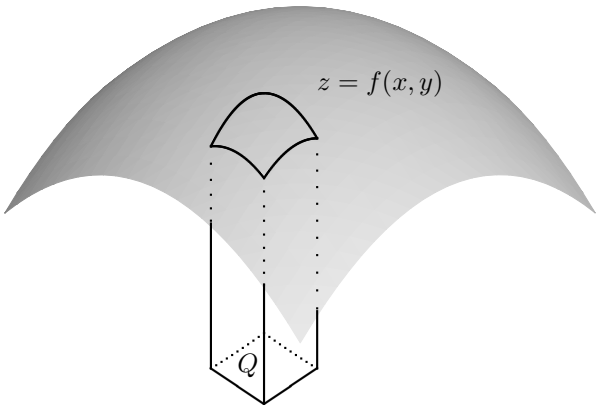
The Riemann Integral On $\mathbb{R}^n$

14.0.1 Outcomes

1. Recall and define the Riemann integral.
2. Recall the relation between iterated integrals and the Riemann integral.
3. Evaluate double integrals over simple regions.
4. Evaluate multiple integrals over simple regions.
5. Use multiple integrals to calculate the volume and mass.

14.1 Methods For Double Integrals

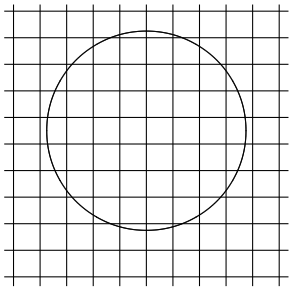
This chapter is on the Riemann integral for a function of n variables. It begins by introducing the basic concepts and applications of the integral. The proofs of the theorems involved are difficult and are left till the end. To begin with consider the problem of finding the volume under a surface of the form $z = f(x, y)$ where $f(x, y) \geq 0$ and $f(x, y) = 0$ for all (x, y) outside of some bounded set. To solve this problem, consider the following picture.



In this picture, the volume of the little prism which lies above the rectangle Q and the graph of the function would lie between $M_Q(f)\, v(Q)$ and $m_Q(f)\, v(Q)$ where

$$M_Q(f) \equiv \sup\{f(\mathbf{x}) : \mathbf{x} \in Q\},\ m_Q(f) \equiv \inf\{f(\mathbf{x}) : \mathbf{x} \in Q\}, \tag{14.1}$$

and $v(Q)$ is defined as the area of Q. Now consider the following picture.

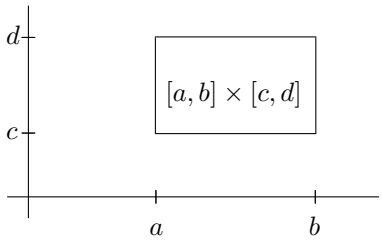


In this picture, it is assumed f equals zero outside the circle and f is a bounded nonnegative function. Then each of those little squares are the base of a prism of the sort in the previous picture and the sum of the volumes of those prisms should be the volume under the surface, $z = f(x,y)$. Therefore, the desired volume must lie between the two numbers,

$$\sum_Q M_Q(f)\, v(Q) \text{ and } \sum_Q m_Q(f)\, v(Q)$$

where the notation, $\sum_Q M_Q(f)\, v(Q)$, means for each Q, take $M_Q(f)$, multiply it by the area of $Q, v(Q)$, and then add all these numbers together. Thus in $\sum_Q M_Q(f)\, v(Q)$, adds numbers which are at least as large as what is desired while in $\sum_Q m_Q(f)\, v(Q)$ numbers are added which are at least as small as what is desired. Note this is a finite sum because by assumption, $f = 0$ except for finitely many Q, namely those which intersect the circle. The sum, $\sum_Q M_Q(f)\, v(Q)$ is called an upper sum, $\sum_Q m_Q(f)\, v(Q)$ is a lower sum, and the desired volume is caught between these upper and lower sums.

None of this depends in any way on the function being nonnegative. It also does not depend in any essential way on the function being defined on $\mathbb{R}^2$, although it is impossible to draw meaningful pictures in higher dimensional cases. To define the Riemann integral, it is necessary to first give a description of something called a **grid**. First you must understand that something like $[a,b] \times [c,d]$ is a rectangle in $\mathbb{R}^2$, having sides parallel to the axes. The situation is illustrated in the following picture.



$(x,y) \in [a,b] \times [c,d]$, means $x \in [a,b]$ and also $y \in [c,d]$ and the points which do this comprise the rectangle just as shown in the picture.

Definition 14.1.1 *For $i = 1, 2$, let $\{\alpha_k^i\}_{k=-\infty}^{\infty}$ be points on $\mathbb{R}$ which satisfy*

$$\lim_{k\to\infty} \alpha_k^i = \infty,\ \lim_{k\to-\infty} \alpha_k^i = -\infty,\ \alpha_k^i < \alpha_{k+1}^i. \tag{14.2}$$

*For such sequences, define a **grid** on $\mathbb{R}^2$ denoted by $\mathcal{G}$ or $\mathcal{F}$ as the collection of rectangles of the form*

$$Q = \left[\alpha_k^1, \alpha_{k+1}^1\right] \times \left[\alpha_l^2, \alpha_{l+1}^2\right]. \tag{14.3}$$

*If $\mathcal{G}$ is a grid, another grid, $\mathcal{F}$ is a **refinement** of $\mathcal{G}$ if every box of $\mathcal{G}$ is the union of boxes of $\mathcal{F}$.*

For $\mathcal{G}$ a grid, the expression,

$$\sum_{Q\in\mathcal{G}} M_Q(f)\, v(Q)$$

is called the upper sum associated with the grid, $\mathcal{G}$ as described above in the discussion of the volume under a surface. Again, this means to take a rectangle from $\mathcal{G}$ multiply $M_Q(f)$ defined in 14.1 by its area, $v(Q)$ and sum all these products for every $Q \in \mathcal{G}$. The symbol,

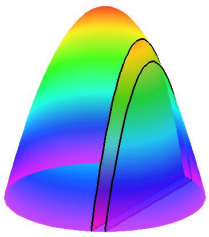
$$\sum_{Q\in\mathcal{G}} m_Q(f)\, v(Q),$$

called a lower sum, is defined similarly. With this preparation it is time to give a definition of the **Riemann integral** of a function of two variables.

Definition 14.1.2 *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a bounded function which equals zero for all (x, y) outside some bounded set. Then $\int f\, dV$ is defined to be the unique number which lies between all upper sums and all lower sums. In the case of $\mathbb{R}^2$, it is common to replace the V with A and write this symbol as $\int f\, dA$ where A stands for area.*

This definition begs a difficult question. For which functions does there exist a unique number between all the upper and lower sums? This interesting and fundamental question is discussed in any advanced calculus book and may be seen in the appendix on the theory of the Riemann integral. It is a hard problem which was only solved in the first part of the twentieth century. When it was solved, it was also realized that the Riemann integral was not the right integral to use.

Consider the question: How can the Riemann integral be computed? Consider the following picture where f is assumed to be 0 outside the base of the solid which is contained in some rectangle $[a, b] \times [c, d]$.



It depicts a slice taken from the solid defined by $\{(x, y) : 0 \le y \le f(x, y)\}$. You see these when you look at a loaf of bread. If you wanted to find the volume of the loaf of bread, and you knew the volume of each slice of bread, you could find the volume of the whole loaf by adding the volumes of individual slices. It is the same here. If you could find the volume of the slice represented in this picture, you could add these up and get the volume of the solid. The slice in the picture corresponds to y and $y + h$ and is assumed to be very thin, having thickness equal to h. Denote the volume of the solid under the graph of $z = f(x, y)$ on $[a, b] \times [c, y]$ by $V(y)$. Then

$$V(y + h) - V(y) \approx h \int_a^b f(x, y)\, dx$$

where the integral is obtained by fixing y and integrating with respect to x and is the area of the cross section corresponding to y. It is hoped that the approximation would be increasingly good as h gets smaller. Thus, dividing by h and taking a limit, it is expected that

$$V'(y) = \int_a^b f(x, y)\, dx,\ V(c) = 0.$$

Therefore, as in the method of cross sections, the volume of the solid under the graph of $z = f(x,y)$ is obtained by doing $\int_c^d$ to both sides,

$$\int_c^d \left(\int_a^b f(x,y)\, dx \right) dy \tag{14.4}$$

but this volume was also the result of $\int f\, dV$. Therefore, it is expected that this is a way to evaluate $\int f\, dV$.

Note what has been gained here. A hard problem, finding $\int f\, dV$, is reduced to a sequence of easier problems. First do

$$\int_a^b f(x,y)\, dx$$

getting a function of y, say $F(y)$ and then do

$$\int_c^d \left(\int_a^b f(x,y)\, dx \right) dy = \int_c^d F(y)\, dy.$$

Of course there is nothing special about fixing y first. The same thing should be obtained from the integral,

$$\int_a^b \left(\int_c^d f(x,y)\, dy \right) dx \tag{14.5}$$

These expressions in 14.4 and 14.5 are called **iterated integrals**. They are tools for evaluating $\int f\, dV$ which would be hard to find otherwise. In practice, the parenthesis is usually omitted in these expressions. Thus

$$\int_a^b \left(\int_c^d f(x,y)\, dy \right) dx = \int_a^b \int_c^d f(x,y)\, dy\, dx$$

and it is understood that you are to do the inside integral first and then when you have done it, obtaining a function of x, you integrate this function of x. Note that this is nothing more than using an integral to compute the area of a cross section and then using this method to find a volume.

However, there is no difference in the general case where f is not necessarily nonnegative as can be seen by applying the method to the nonnegative functions f^+, f^- given by

$$f^+ \equiv \frac{|f| - f}{2},\ f^- \equiv \frac{|f| - f}{2}$$

and then noting that $f = f^+ - f^-$ and the integral is linear. Thus

$$\begin{aligned} \int f dV &= \int f^+ - f^- dV = \int f^+ dV - \int f^- dV \\ &= \int_a^b \int_c^d f^+(x,y)\, dy\, dx - \int_a^b \int_c^d f^-(x,y)\, dy\, dx \\ &= \int_a^b \int_c^d f(x,y)\, dy\, dx \end{aligned}$$

A careful presentation which is not for the faint of heart is in an appendix.

Another aspect of this is the notion of integrating a function which is defined on some set, not on all $\mathbb{R}^2$. For example, suppose f is defined on the set, $S \subseteq \mathbb{R}^2$. What is meant by $\int_S f\, dV$?

Definition 14.1.3 *Let $f : S \to \mathbb{R}$ where S is a subset of $\mathbb{R}^2$. Then denote by f_1 the function defined by*

$$f_1(x,y) \equiv \begin{cases} f(x,y) & \text{if } (x,y) \in S \\ 0 & \text{if } (x,y) \notin S \end{cases}.$$

Then

$$\int_S f\,dV \equiv \int f_1\,dV.$$

Example 14.1.4 *Let $f(x,y) = x^2y + yx$ for $(x,y) \in [0,1] \times [0,2] \equiv R$. Find $\int_R f\,dV$.*

This is done using iterated integrals like those defined above. Thus

$$\int_R f\,dV = \int_0^1 \int_0^2 \left(x^2y + yx\right)\,dy\,dx.$$

The inside integral yields

$$\int_0^2 \left(x^2y + yx\right)\,dy = 2x^2 + 2x$$

and now the process is completed by doing $\int_0^1$ to what was just obtained. Thus

$$\int_0^1 \int_0^2 \left(x^2y + yx\right)\,dy\,dx = \int_0^1 \left(2x^2 + 2x\right)\,dx = \frac{5}{3}.$$

If the integration is done in the opposite order, the same answer should be obtained.

$$\int_0^2 \int_0^1 \left(x^2y + yx\right)\,dx\,dy$$

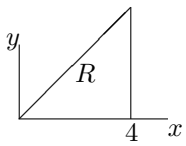
$$\int_0^1 \left(x^2y + yx\right)\,dx = \frac{5}{6}y$$

Now

$$\int_0^2 \int_0^1 \left(x^2y + yx\right)\,dx\,dy = \int_0^2 \left(\frac{5}{6}y\right)\,dy = \frac{5}{3}.$$

If a different answer had been obtained it would have been a sign that a mistake had been made.

Example 14.1.5 *Let $f(x,y) = x^2y + yx$ for $(x,y) \in R$ where R is the triangular region defined to be in the first quadrant, below the line $y = x$ and to the left of the line $x = 4$. Find $\int_R f\,dV$.*



Now from the above discussion,

$$\int_R f\,dV = \int_0^4 \int_0^x \left(x^2y + yx\right)\,dy\,dx$$

The reason for this is that x goes from 0 to 4 and for each fixed x between 0 and 4, y goes from 0 to the slanted line, $y = x$, the function being defined to be 0 for larger

y. Thus y goes from 0 to x. This explains the inside integral. Now $\int_0^x \left(x^2y + yx\right)\, dy = \frac{1}{2}x^4 + \frac{1}{2}x^3$ and so

$$\int_R f\, dV = \int_0^4 \left(\frac{1}{2}x^4 + \frac{1}{2}x^3\right)\, dx = \frac{672}{5}.$$

What of integration in a different order? Lets put the integral with respect to y on the outside and the integral with respect to x on the inside. Then

$$\int_R f\, dV = \int_0^4 \int_y^4 \left(x^2y + yx\right)\, dx\, dy$$

For each y between 0 and 4, the variable x, goes from y to 4.

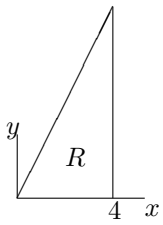
$$\int_y^4 \left(x^2y + yx\right)\, dx = \frac{88}{3}y - \frac{1}{3}y^4 - \frac{1}{2}y^3$$

Now

$$\int_R f\, dV = \int_0^4 \left(\frac{88}{3}y - \frac{1}{3}y^4 - \frac{1}{2}y^3\right)\, dy = \frac{672}{5}.$$

Here is a similar example.

Example 14.1.6 *Let $f(x,y) = x^2y$ for $(x,y) \in R$ where R is the triangular region defined to be in the first quadrant, below the line $y = 2x$ and to the left of the line $x = 4$. Find $\int_R f\, dV$.*



Put the integral with respect to x on the outside first. Then

$$\int_R f\, dV = \int_0^4 \int_0^{2x} \left(x^2y\right)\, dy\, dx$$

because for each $x \in [0,4]$, y goes from 0 to $2x$. Then

$$\int_0^{2x} \left(x^2y\right)\, dy = 2x^4$$

and so

$$\int_R f\, dV = \int_0^4 \left(2x^4\right)\, dx = \frac{2048}{5}$$

Now do the integral in the other order. Here the integral with respect to y will be on the outside. What are the limits of this integral? Look at the triangle and note that x goes from 0 to 4 and so $2x = y$ goes from 0 to 8. Now for fixed y between 0 and 8, where does x go? It goes from the x coordinate on the line $y = 2x$ which corresponds to this y to 4. What is the x coordinate on this line which goes with y? It is $x = y/2$. Therefore, the iterated integral is

$$\int_0^8 \int_{y/2}^4 \left(x^2y\right)\, dx\, dy.$$

Now

$$\int_{y/2}^{4} \left(x^2 y\right)\, dx = \frac{64}{3}y - \frac{1}{24}y^4$$

and so

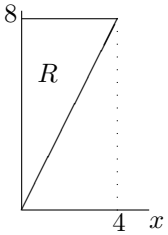
$$\int_R f\, dV = \int_0^8 \left(\frac{64}{3}y - \frac{1}{24}y^4\right)\, dy = \frac{2048}{5}$$

the same answer.

A few observations are in order here. In finding $\int_S f\, dV$ there is no problem in setting things up if S is a rectangle. However, if S is not a rectangle, the procedure **always** is agonizing. A good rule of thumb is that if what you do is easy it will be wrong. There are no shortcuts! There are no quick fixes which require no thought! Pain and suffering is inevitable and you must not expect it to be otherwise. Always draw a picture and then begin **agonizing** over the correct limits. Even when you are careful you will make lots of mistakes until you get used to the process.

Sometimes an integral can be evaluated in one order but not in another.

Example 14.1.7 *For R as shown below, find $\int_R \sin\left(y^2\right)\, dV$.*



Setting this up to have the integral with respect to y on the inside yields

$$\int_0^4 \int_{2x}^8 \sin\left(y^2\right)\, dy\, dx.$$

Unfortunately, there is no antiderivative in terms of elementary functions for $\sin\left(y^2\right)$ so there is an immediate problem in evaluating the inside integral. It doesn't work out so the next step is to do the integration in another order and see if some progress can be made. This yields

$$\int_0^8 \int_0^{y/2} \sin\left(y^2\right)\, dx\, dy = \int_0^8 \frac{y}{2}\sin\left(y^2\right)\, dy$$

and $\int_0^8 \frac{y}{2}\sin\left(y^2\right)\, dy = -\frac{1}{4}\cos 64 + \frac{1}{4}$ which you can verify by making the substitution, $u = y^2$. Thus

$$\int_R \sin\left(y^2\right)\, dy = -\frac{1}{4}\cos 64 + \frac{1}{4}.$$

This illustrates an important idea. The integral $\int_R \sin\left(y^2\right)\, dV$ is defined as a number. It is the unique number between all the upper sums and all the lower sums. Finding it is another matter. In this case it was possible to find it using one order of integration but not the other. The iterated integral in this other order also is defined as a number but it can't be found directly without interchanging the order of integration. Of course sometimes nothing you try will work out.

14.1.1 Density And Mass

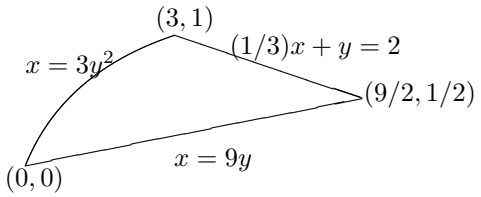
Consider a two dimensional material. Of course there is no such thing but a flat plate might be modeled as one. The density ρ is a function of position and is defined as follows. Consider a small chunk of area, dV located at the point whose Cartesian coordinates are (x, y). Then the mass of this small chunk of material is given by $\rho(x, y)\, dV$. Thus if the material occupies a region in two dimensional space, U, the total mass of this material would be

$$\int_U \rho\, dV$$

In other words you integrate the density to get the mass. Now by letting ρ depend on position, you can include the case where the material is not homogeneous. Here is an example.

Example 14.1.8 *Let $\rho(x, y)$ denote the density of the plane region determined by the curves $\frac{1}{3}x + y = 2, x = 3y^2$, and $x = 9y$. Find the total mass if $\rho(x, y) = y$.*

You need to first draw a picture of the region, R. A rough sketch follows.



This region is in two pieces, one having the graph of $x = 9y$ on the bottom and the graph of $x = 3y^2$ on the top and another piece having the graph of $x = 9y$ on the bottom and the graph of $\frac{1}{3}x + y = 2$ on the top. Therefore, in setting up the integrals, with the integral with respect to x on the outside, the double integral equals the following sum of iterated integrals.

$$\overbrace{\int_0^3 \int_{x/9}^{\sqrt{x/3}} y\, dy\, dx}^{\text{has } x=3y^2 \text{ on top}} + \overbrace{\int_3^{\frac{9}{2}} \int_{x/9}^{2-\frac{1}{3}x} y\, dy\, dx}^{\text{has } \frac{1}{3}x+y=2 \text{ on top}}$$

You notice it is not necessary to have a perfect picture, just one which is good enough to figure out what the limits should be. The dividing line between the two cases is $x = 3$ and this was shown in the picture. Now it is only a matter of evaluating the iterated integrals which in this case is routine and gives 1.

14.2 Exercises

1. Let $\rho(x, y)$ denote the density of the plane region closest to $(0, 0)$ which is between the curves $\frac{1}{4}x + y = 6, x = 4y^2$, and $x = 16y$. Find the total mass if $\rho(x, y) = y$. Your answer should be $\frac{1168}{75}$.

2. Let $\rho(x, y)$ denote the density of the plane region determined by the curves $\frac{1}{5}x + y = 6, x = 5y^2$, and $x = 25y$. Find the total mass if $\rho(x, y) = y + 2x$. Your answer should be $\frac{1735}{3}$.

3. Let $\rho(x,y)$ denote the density of the plane region determined by the curves $y = 3x, y = x, 3x+3y = 9$. Find the total mass if $\rho(x,y) = y+1$. Your answer should be $\frac{81}{32}$.

4. Let $\rho(x,y)$ denote the density of the plane region determined by the curves $y = 3x, y = x, 4x+2y = 8$. Find the total mass if $\rho(x,y) = y+1$.

5. Let $\rho(x,y)$ denote the density of the plane region determined by the curves $y = 3x, y = x, 2x+2y = 4$. Find the total mass if $\rho(x,y) = x+2y$.

6. Let $\rho(x,y)$ denote the density of the plane region determined by the curves $y = 3x, y = x, 5x+2y = 10$. Find the total mass if $\rho(x,y) = y+1$.

7. Find $\int_0^4 \int_{y/2}^2 \frac{1}{x} e^{2\frac{y}{x}}\, dx\, dy$. Your answer should be $e^4 - 1$. You might need to interchange the order of integration.

8. Find $\int_0^8 \int_{y/2}^4 \frac{1}{x} e^{3\frac{y}{x}}\, dx\, dy$.

9. Find $\int_0^8 \int_{y/2}^4 \frac{1}{x} e^{3\frac{y}{x}}\, dx\, dy$.

10. Find $\int_0^4 \int_{y/2}^2 \frac{1}{x} e^{3\frac{y}{x}}\, dx\, dy$.

11. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^4 \int_0^{3y} xy^3\, dx\, dy$. Your answer for the iterated integral should be $\int_0^{12} \int_{\frac{1}{3}x}^4 xy^3\, dy\, dx$.

12. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^3 \int_0^{3y} xy^3\, dx\, dy$.

13. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^2 \int_0^{2y} xy^2\, dx\, dy$.

14. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^3 \int_0^{y} xy^3\, dx\, dy$.

15. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^1 \int_0^{y} xy^2\, dx\, dy$.

16. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^5 \int_0^{3y} xy^2\, dx\, dy$.

17. Find $\int_0^{\frac{1}{3}\pi} \int_x^{\frac{1}{3}\pi} \frac{\sin y}{y}\, dy\, dx$. Your answer should be $\frac{1}{2}$.

18. Find $\int_0^{\frac{1}{2}\pi} \int_x^{\frac{1}{2}\pi} \frac{\sin y}{y}\, dy\, dx$.

19. Find $\int_0^{\pi} \int_x^{\pi} \frac{\sin y}{y}\, dy\, dx$

20. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_{-3}^3 \int_{-x}^x x^2\, dy\, dx$

    Your answer for the iterated integral should be $\int_3^0 \int_{-3}^{-y} x^2\, dx\, dy + \int_0^{-3} \int_{-3}^{y} x^2\, dx\, dy + \int_0^3 \int_y^3 x^2\, dx\, dy + \int_{-3}^0 \int_{-y}^3 x^2\, dx\, dy$. This is a very interesting example which shows that iterated integrals have a life of their own, not just as a method for evaluating double integrals.

21. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_{-2}^2 \int_{-x}^x x^2\, dy\, dx$.

14.3 Methods For Triple Integrals

14.3.1 Definition Of The Integral

The integral of a function of three variables is similar to the integral of a function of two variables.

Definition 14.3.1 *For $i = 1, 2, 3$ let $\left\{\alpha_k^i\right\}_{k=-\infty}^{\infty}$ be points on $\mathbb{R}$ which satisfy*

$$\lim_{k\to\infty} \alpha_k^i = \infty,\ \lim_{k\to-\infty} \alpha_k^i = -\infty,\ \alpha_k^i < \alpha_{k+1}^i. \tag{14.6}$$

*For such sequences, define a **grid** on $\mathbb{R}^3$ denoted by $\mathcal{G}$ or $\mathcal{F}$ as the collection of boxes of the form*

$$Q = \left[\alpha_k^1, \alpha_{k+1}^1\right] \times \left[\alpha_l^2, \alpha_{l+1}^2\right] \times \left[\alpha_p^3, \alpha_{p+1}^3\right]. \tag{14.7}$$

*If $\mathcal{G}$ is a grid, $\mathcal{F}$ is called a **refinement** of $\mathcal{G}$ if every box of $\mathcal{G}$ is the union of boxes of $\mathcal{F}$.*

For $\mathcal{G}$ a grid,

$$\sum_{Q\in\mathcal{G}} M_Q(f)\, v(Q)$$

is the upper sum associated with the grid, $\mathcal{G}$ where

$$M_Q(f) \equiv \sup\{f(\mathbf{x}) : \mathbf{x} \in Q\}$$

and if $Q = [a, b]\times[c, d]\times[e, f]$, then $v(Q)$ is the volume of Q given by $(b-a)(d-c)(f-e)$. Letting

$$m_Q(f) \equiv \inf\{f(\mathbf{x}) : \mathbf{x} \in Q\}$$

the lower sum associated with this partition is

$$\sum_{Q\in\mathcal{G}} m_Q(f)\, v(Q),$$

With this preparation it is time to give a definition of the **Riemann integral** of a function of three variables. This definition is just like the one for a function of two variables.

Definition 14.3.2 *Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a bounded function which equals zero outside some bounded subset of $\mathbb{R}^3$. $\int f\,dV$ is defined as the unique number between all the upper sums and lower sums.*

As in the case of a function of two variables there are all sorts of mathematical questions which are dealt with later.

The way to think of integrals is as follows. Located at a point $\mathbf{x}$, there is an "infinitesimal" chunk of volume, dV. The integral involves taking this little chunk of volume, dV, multiplying it by $f(\mathbf{x})$ and then adding up all such products. Upper sums are too large and lower sums are too small but the unique number between all the lower and upper sums is just right and corresponds to the notion of adding up all the $f(\mathbf{x})\,dV$. Even the notation is suggestive of this concept of sum. It is a long thin S denoting sum. This is the fundamental concept for the integral in any number of dimensions and all the definitions and technicalities are designed to give precision and mathematical respectability to this notion.

Integrals of functions of three variables are also evaluated by using iterated integrals. Imagine a sum of the form $\sum_{ijk} a_{ijk}$ where there are only finitely many choices for $i, j,$

and k and the symbol means you simply add up all the a_{ijk}. By the commutative law of addition, these may be added systematically in the form, $\sum_k \sum_j \sum_i a_{ijk}$. A similar process is used to evaluate triple integrals and since integrals are like sums, you might expect it to be valid. Specifically,

$$\int f\,dV = \int_?^? \int_?^? \int_?^? f(x,y,z)\,dx\,dy\,dz.$$

In words, sum with respect to x and then sum what you get with respect to y and finally, with respect to z. Of course this should hold in any other order such as

$$\int f\,dV = \int_?^? \int_?^? \int_?^? f(x,y,z)\,dz\,dy\,dx.$$

This is proved in an appendix[1].

Having discussed double and triple integrals, the definition of the integral of a function of n variables is accomplished in the same way.

Definition 14.3.3 *For $i = 1, \cdots, n$, let $\left\{\alpha_k^i\right\}_{k=-\infty}^{\infty}$ be points on $\mathbb{R}$ which satisfy*

$$\lim_{k\to\infty} \alpha_k^i = \infty,\ \lim_{k\to-\infty} \alpha_k^i = -\infty,\ \alpha_k^i < \alpha_{k+1}^i. \tag{14.8}$$

For such sequences, define a grid on $\mathbb{R}^n$ denoted by $\mathcal{G}$ or $\mathcal{F}$ as the collection of boxes of the form

$$Q = \prod_{i=1}^n \left[\alpha_{j_i}^i, \alpha_{j_i+1}^i\right]. \tag{14.9}$$

If $\mathcal{G}$ is a grid, $\mathcal{F}$ is called a refinement of $\mathcal{G}$ if every box of $\mathcal{G}$ is the union of boxes of $\mathcal{F}$.

Definition 14.3.4 *Let f be a bounded function which equals zero off a bounded set, D, and let $\mathcal{G}$ be a grid. For $Q \in \mathcal{G}$, define*

$$M_Q(f) \equiv \sup\{f(\mathbf{x}) : \mathbf{x} \in Q\},\ m_Q(f) \equiv \inf\{f(\mathbf{x}) : \mathbf{x} \in Q\}. \tag{14.10}$$

Also define for Q a box, the volume of Q, denoted by $v(Q)$ by

$$v(Q) \equiv \prod_{i=1}^n (b_i - a_i),\ Q \equiv \prod_{i=1}^n [a_i, b_i].$$

Now define upper sums, $\mathcal{U}_{\mathcal{G}}(f)$ and lower sums, $\mathcal{L}_{\mathcal{G}}(f)$ with respect to the indicated grid, by the formulas

$$\mathcal{U}_{\mathcal{G}}(f) \equiv \sum_{Q\in\mathcal{G}} M_Q(f)\,v(Q),\ \mathcal{L}_{\mathcal{G}}(f) \equiv \sum_{Q\in\mathcal{G}} m_Q(f)\,v(Q).$$

Then a function of n variables is Riemann integrable if there is a unique number between all the upper and lower sums. This number is the value of the integral.

In this book most integrals will involve no more than three variables. However, this does not mean an integral of a function of more than three variables is unimportant. Therefore, I will begin to refer to the general case when theorems are stated.

Definition 14.3.5 *For $E \subseteq \mathbb{R}^n$,*

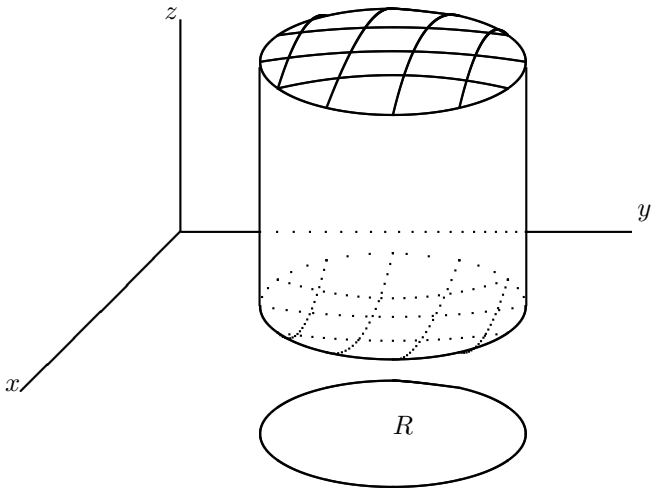
$$\mathcal{X}_E(\mathbf{x}) \equiv \begin{cases} 1 \text{ if } \mathbf{x} \in E \\ 0 \text{ if } \mathbf{x} \notin E \end{cases}.$$

Define $\int_E f\,dV \equiv \int \mathcal{X}_E f\,dV$ when $f\mathcal{X}_E \in \mathcal{R}(\mathbb{R}^n)$.

[1] All of these fundamental questions about integrals can be considered more easily in the context of the Lebesgue integral. However, this integral is more abstract than the Riemann integral.

14.3.2 Iterated Integrals

As before, the integral is often computed by using an iterated integral. In general it is impossible to set up an iterated integral for finding $\int_E f dV$ for arbitrary regions, E but when the region is sufficiently simple, one can make progress. Suppose the region, E over which the integral is to be taken is of the form $E = \{(x,y,z) : a(x,y) \leq z \leq b(x,y)\}$ for $(x,y) \in R$, a two dimensional region. This is illustrated in the following picture in which the bottom surface is the graph of $z = a(x,y)$ and the top is the graph of $z = b(x,y)$.



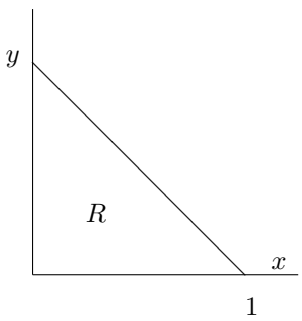
Then

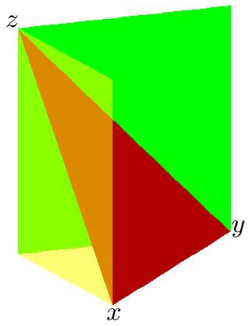
$$\int_E f dV = \int_R \int_{a(x,y)}^{b(x,y)} f(x,y,z)\, dz dA$$

It might be helpful to think of $dV = dzdA$. Now $\int_{a(x,y)}^{b(x,y)} f(x,y,z)\, dz$ is a function of x and y and so you have reduced the triple integral to a double integral over R of this function of x and y. Similar reasoning would apply if the region in $\mathbb{R}^3$ were of the form $\{(x,y,z) : a(y,z) \leq x \leq b(y,z)\}$ or $\{(x,y,z) : a(x,z) \leq y \leq b(x,z)\}$.

Example 14.3.6 *Find the volume of the region, E in the first octant between $z = 1-(x+y)$ and $z = 0$.*

In this case, R is the region shown.





Thus the region, E is between the plane $z = 1-(x+y)$ on the top, $z = 0$ on the bottom, and over R shown above. Thus

$$\begin{aligned}\int_E 1dV &= \int_R \int_0^{1-(x+y)} dzdA \\ &= \int_0^1 \int_0^{1-x} \int_0^{1-(x+y)} dzdydx = \frac{1}{6}\end{aligned}$$

Of course iterated integrals have a life of their own although this will not be explored here. You can just write them down and go to work on them. Here are some examples.

Example 14.3.7 *Find* $\int_2^3 \int_3^x \int_{3y}^x (x-y)\ dz\,dy\,dx$.

The inside integral yields $\int_{3y}^x (x-y)\ dz = x^2 - 4xy + 3y^2$. Next this must be integrated with respect to y to give $\int_3^x \left(x^2 - 4xy + 3y^2\right)\ dy = -3x^2 + 18x - 27$. Finally the third integral gives

$$\int_2^3 \int_3^x \int_{3y}^x (x-y)\ dz\,dy\,dx = \int_2^3 \left(-3x^2 + 18x - 27\right)\ dx = -1.$$

Example 14.3.8 *Find* $\int_0^\pi \int_0^{3y} \int_0^{y+z} \cos(x+y)\ dx\,dz\,dy$.

The inside integral is $\int_0^{y+z} \cos(x+y)\ dx = 2\cos z \sin y \cos y + 2\sin z \cos^2 y - \sin z - \sin y$. Now this has to be integrated.

$$\begin{aligned}\int_0^{3y} \int_0^{y+z} \cos(x+y)\ dx\,dz &= \int_0^{3y} \left(2\cos z \sin y \cos y + 2\sin z \cos^2 y - \sin z - \sin y\right)\ dz \\ &= -1 - 16\cos^5 y + 20\cos^3 y - 5\cos y - 3(\sin y)\, y + 2\cos^2 y.\end{aligned}$$

Finally, this last expression must be integrated from 0 to π. Thus

$$\int_0^\pi \int_0^{3y} \int_0^{y+z} \cos(x+y)\ dx\,dz\,dy$$

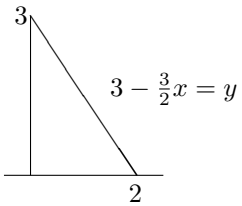
$$\begin{aligned}&= \int_0^\pi \left(-1 - 16\cos^5 y + 20\cos^3 y - 5\cos y - 3(\sin y)\, y + 2\cos^2 y\right)\ dy \\ &= -3\pi\end{aligned}$$

Example 14.3.9 *Here is an iterated integral:* $\int_0^2 \int_0^{3-\frac{3}{2}x} \int_0^{x^2}\ dz\,dy\,dx$. *Write as an iterated integral in the order* $dz\,dx\,dy$.

The inside integral is just a function of x and y. (In fact, only a function of x.) The order of the last two integrals must be interchanged. Thus the iterated integral which needs to be done in a different order is

$$\int_0^2 \int_0^{3-\frac{3}{2}x} f(x,y)\ dy\,dx.$$

As usual, it is important to draw a picture and then go from there.



Thus this double integral equals

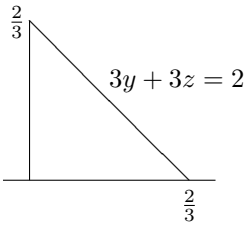
$$\int_0^3 \int_0^{\frac{2}{3}(3-y)} f(x,y)\, dx\, dy.$$

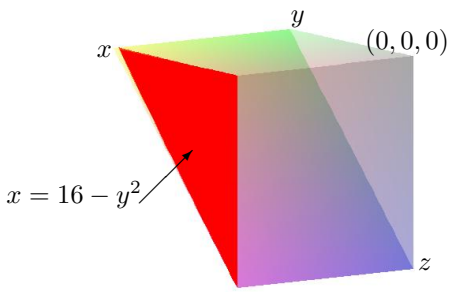
Now substituting in for $f(x,y)$,

$$\int_0^3 \int_0^{\frac{2}{3}(3-y)} \int_0^{x^2} dz\, dx\, dy.$$

Example 14.3.10 *Find the volume of the bounded region determined by* $3y + 3z = 2, x = 16 - y^2, y = 0, x = 0$.

In the yz plane, the first of the following pictures corresponds to $x = 0$.





Therefore, the outside integrals taken with respect to z and y are of the form $\int_0^{\frac{2}{3}} \int_0^{\frac{2}{3}-y} dz\, dy$ and now for any choice of (y,z) in the above triangular region, x goes from 0 to $16 - y^2$. Therefore, the iterated integral is

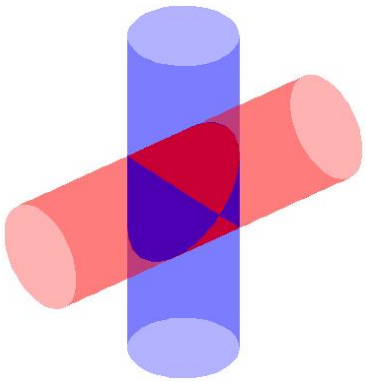
$$\int_0^{\frac{2}{3}} \int_0^{\frac{2}{3}-y} \int_0^{16-y^2} dx\, dz\, dy = \frac{860}{243}$$

Example 14.3.11 *Find the volume of the region determined by the intersection of the two cylinders,* $x^2 + y^2 \leq 9$ *and* $y^2 + z^2 \leq 9$.

The first listed cylinder intersects the xy plane in the disk, $x^2 + y^2 \leq 9$. What is the volume of the three dimensional region which is between this disk and the two surfaces, $z = \sqrt{9 - y^2}$ and $z = -\sqrt{9 - y^2}$? An iterated integral for the volume is

$$\int_{-3}^3 \int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}} \int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}} dz\, dx\, dy = 144.$$

Note I drew no picture of the three dimensional region. If you are interested, here it is.



One of the cylinders is parallel to the z axis, $x^2+y^2 \leq 9$ and the other is parallel to the x axis, $y^2+z^2 \leq 9$. I did not need to be able to draw such a nice picture in order to work this problem. This is the key to doing these. Draw pictures in two dimensions and reason from the two dimensional pictures rather than attempt to wax artistic and consider all three dimensions at once. These problems are hard enough without making them even harder by attempting to be an artist.

14.3.3 Mass And Density

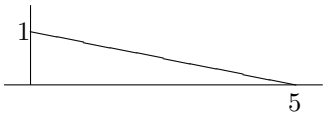
As an example of the use of triple integrals, consider a solid occupying a set of points, $U \subseteq \mathbb{R}^3$ having density ρ. Thus ρ is a function of position and the total mass of the solid equals

$$\int_U \rho \, dV.$$

This is just like the two dimensional case. The mass of an infinitesimal chunk of the solid located at $\mathbf{x}$ would be $\rho(\mathbf{x})\, dV$ and so the total mass is just the sum of all these, $\int_U \rho(\mathbf{x})\, dV$.

Example 14.3.12 *Find the volume of R where R is the bounded region formed by the plane $\frac{1}{5}x+y+\frac{1}{5}z=1$ and the planes $x=0, y=0, z=0$.*

When $z=0$, the plane becomes $\frac{1}{5}x+y=1$. Thus the intersection of this plane with the xy plane is this line shown in the following picture.



Therefore, the bounded region is between the triangle formed in the above picture by the x axis, the y axis and the above line and the surface given by $\frac{1}{5}x+y+\frac{1}{5}z=1$ or $z=5\left(1-\left(\frac{1}{5}x+y\right)\right)=5-x-5y$. Therefore, an iterated integral which yields the volume is

$$\int_0^5 \int_0^{1-\frac{1}{5}x} \int_0^{5-x-5y} dz\, dy\, dx = \frac{25}{6}.$$

Example 14.3.13 *Find the mass of the bounded region, R formed by the plane $\frac{1}{3}x+\frac{1}{3}y+\frac{1}{5}z=1$ and the planes $x=0, y=0, z=0$ if the density is $\rho(x,y,z)=z$.*

This is done just like the previous example except in this case there is a function to integrate. Thus the answer is

$$\int_0^3 \int_0^{3-x} \int_0^{5-\frac{5}{3}x-\frac{5}{3}y} z \; dz\, dy\, dx = \frac{75}{8}.$$

Example 14.3.14 *Find the total mass of the bounded solid determined by $z=9-x^2-y^2$ and $x,y,z \geq 0$ if the mass is given by $\rho(x,y,z)=z$*

When $z=0$ the surface, $z=9-x^2-y^2$ intersects the xy plane in a circle of radius 3 centered at $(0,0)$. Since $x,y \geq 0$, it is only a quarter of a circle of interest, the part where both these variables are nonnegative. For each (x,y) inside this quarter circle, z goes from 0 to $9-x^2-y^2$. Therefore, the iterated integral is of the form,

$$\int_0^3 \int_0^{\sqrt{(9-x^2)}} \int_0^{9-x^2-y^2} z \; dz\, dy\, dx = \frac{243}{8}\pi$$

Example 14.3.15 *Find the volume of the bounded region determined by* $x \geq 0, y \geq 0, z \geq 0$, *and* $\frac{1}{7}x + y + \frac{1}{4}z = 1$, *and* $x + \frac{1}{7}y + \frac{1}{4}z = 1$.

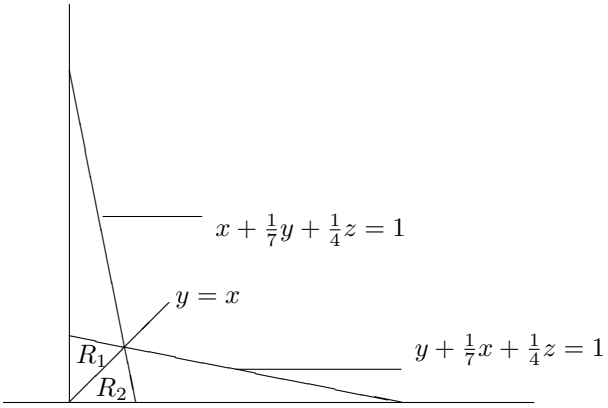
When $z = 0$, the plane $\frac{1}{7}x + y + \frac{1}{4}z = 1$ intersects the xy plane in the line whose equation is

$$\frac{1}{7}x + y = 1$$

while the plane, $x + \frac{1}{7}y + \frac{1}{4}z = 1$ intersects the xy plane in the line whose equation is

$$x + \frac{1}{7}y = 1.$$

Furthermore, the two planes intersect when $x = y$ as can be seen from the equations, $x + \frac{1}{7}y = 1 - \frac{z}{4}$ and $\frac{1}{7}x + y = 1 - \frac{z}{4}$ which imply $x = y$. Thus the two dimensional picture to look at is depicted in the following picture.



You see in this picture, the base of the region in the xy plane is the union of the two triangles, R_1 and R_2. For $(x, y) \in R_1$, z goes from 0 to what it needs to be to be on the plane, $\frac{1}{7}x + y + \frac{1}{4}z = 1$. Thus z goes from 0 to $4\left(1 - \frac{1}{7}x - y\right)$. Similarly, on R_2, z goes from 0 to $4\left(1 - \frac{1}{7}y - x\right)$. Therefore, the integral needed is

$$\int_{R_1} \int_0^{4\left(1-\frac{1}{7}x-y\right)} dz\, dV + \int_{R_2} \int_0^{4\left(1-\frac{1}{7}y-x\right)} dz\, dV$$

and now it only remains to consider $\int_{R_1} dV$ and $\int_{R_2} dV$. The point of intersection of these lines shown in the above picture is $\left(\frac{7}{8}, \frac{7}{8}\right)$ and so an iterated integral is

$$\int_0^{7/8} \int_x^{1-\frac{x}{7}} \int_0^{4\left(1-\frac{1}{7}x-y\right)} dz\, dy\, dx + \int_0^{7/8} \int_y^{1-\frac{y}{7}} \int_0^{4\left(1-\frac{1}{7}y-x\right)} dz\, dx\, dy = \frac{7}{6}$$

14.4 Exercises

1. Evaluate the integral $\int_2^4 \int_2^{2x} \int_{2y}^{x} dz\, dy\, dx$
2. Find $\int_0^3 \int_0^{2-5x} \int_0^{2-x-2y} 2x\ dz\, dy\, dx$
3. Find $\int_0^2 \int_0^{1-3x} \int_0^{3-3x-2y} x\ dz\, dy\, dx$
4. Evaluate the integral $\int_2^5 \int_4^{3x} \int_{4y}^{x} (x - y)\ dz\, dy\, dx$

5. Evaluate the integral $\int_0^{\pi}\int_0^{3y}\int_0^{y+z}\cos(x+y)\,dx\,dz\,dy$

6. Evaluate the integral $\int_0^{\pi}\int_0^{4y}\int_0^{y+z}\sin(x+y)\,dx\,dz\,dy$

7. Fill in the missing limits. $\int_0^1\int_0^z\int_0^z f(x,y,z)\,dx\,dy\,dz=\int_?^?\int_?^?\int_?^? f(x,y,z)\,dx\,dz\,dy$,
$\int_0^1\int_0^z\int_0^{2z} f(x,y,z)\,dx\,dy\,dz=\int_?^?\int_?^?\int_?^? f(x,y,z)\,dy\,dz\,dx$,
$\int_0^1\int_0^z\int_0^z f(x,y,z)\,dx\,dy\,dz=\int_?^?\int_?^?\int_?^? f(x,y,z)\,dz\,dy\,dx$,
$\int_0^1\int_{z/2}^{\sqrt{z}}\int_0^{y+z} f(x,y,z)\,dx\,dy\,dz=\int_?^?\int_?^?\int_?^? f(x,y,z)\,dx\,dz\,dy$,
$\int_4^6\int_2^6\int_0^4 f(x,y,z)\,dx\,dy\,dz=\int_?^?\int_?^?\int_?^? f(x,y,z)\,dz\,dy\,dx$.

8. Find the volume of R where R is the bounded region formed by the plane $\frac{1}{5}x+\frac{1}{3}y+\frac{1}{4}z=1$ and the planes $x=0,y=0,z=0$.

9. Find the volume of R where R is the bounded region formed by the plane $\frac{1}{4}x+\frac{1}{2}y+\frac{1}{4}z=1$ and the planes $x=0,y=0,z=0$.

10. Find the mass of the bounded region, R formed by the plane $\frac{1}{4}x+\frac{1}{3}y+\frac{1}{2}z=1$ and the planes $x=0,y=0,z=0$ if the density is $\rho(x,y,z)=y+z$

11. Find the mass of the bounded region, R formed by the plane $\frac{1}{4}x+\frac{1}{2}y+\frac{1}{5}z=1$ and the planes $x=0,y=0,z=0$ if the density is $\rho(x,y,z)=y$

12. Here is an iterated integral: $\int_0^2\int_0^{1-\frac{1}{2}x}\int_0^{x^2}dz\,dy\,dx$. Write as an iterated integral in the following orders: $dz\,dx\,dy$, $dx\,dz\,dy$, $dx\,dy\,dz$, $dy\,dx\,dz$, $dy\,dz\,dx$.

13. Find the volume of the bounded region determined by $2y+z=3, x=9-y^2, y=0, x=0, z=0$.

14. Find the volume of the bounded region determined by $3y+2z=5, x=9-y^2, y=0, x=0$.

    Your answer should be $\frac{11\,525}{648}$

15. Find the volume of the bounded region determined by $5y+2z=3, x=9-y^2, y=0, x=0$.

16. Find the volume of the region bounded by $x^2+y^2=25, z=x, z=0$, and $x\geq 0$.

    Your answer should be $\frac{250}{3}$.

17. Find the volume of the region bounded by $x^2+y^2=9, z=3x, z=0$, and $x\geq 0$.

18. Find the volume of the region determined by the intersection of the two cylinders, $x^2+y^2\leq 16$ and $y^2+z^2\leq 16$.

19. Find the total mass of the bounded solid determined by $z=4-x^2-y^2$ and $x,y,z\geq 0$ if the mass is given by $\rho(x,y,z)=y$

20. Find the total mass of the bounded solid determined by $z=9-x^2-y^2$ and $x,y,z\geq 0$ if the mass is given by $\rho(x,y,z)=z^2$

21. Find the volume of the region bounded by $x^2+y^2=4, z=0, z=5-y$

22. Find the volume of the bounded region determined by $x\geq 0, y\geq 0, z\geq 0$, and $\frac{1}{7}x+\frac{1}{3}y+\frac{1}{3}z=1$, and $\frac{1}{3}x+\frac{1}{7}y+\frac{1}{3}z=1$.

23. Find the volume of the bounded region determined by $x \geq 0, y \geq 0, z \geq 0$, and $\frac{1}{5}x + \frac{1}{3}y + z = 1$, and $\frac{1}{3}x + \frac{1}{5}y + z = 1$.

24. Find the mass of the solid determined by $16x^2 + 4y^2 \leq 9, z \geq 0$, and $z = x + 2$ if the density is $\rho(x, y, z) = z$.

25. Find $\int_0^2 \int_0^{6-2z} \int_{\frac{1}{2}x}^{3-z} (3 - z) \cos\left(y^2\right)\, dy\, dx\, dz$.

26. Find $\int_0^1 \int_0^{18-3z} \int_{\frac{1}{3}x}^{6-z} (6 - z) \exp\left(y^2\right)\, dy\, dx\, dz$.

27. Find $\int_0^2 \int_0^{24-4z} \int_{\frac{1}{4}y}^{6-z} (6 - z) \exp\left(x^2\right)\, dx\, dy\, dz$.

28. Find $\int_0^1 \int_0^{12-4z} \int_{\frac{1}{4}y}^{3-z} \frac{\sin x}{x}\, dx\, dy\, dz$.

29. Find $\int_0^{20} \int_0^1 \int_{\frac{1}{5}y}^{5-z} \frac{\sin x}{x}\, dx\, dz\, dy + \int_{20}^{25} \int_0^{5-\frac{1}{5}y} \int_{\frac{1}{5}y}^{5-z} \frac{\sin x}{x}\, dx\, dz\, dy$. **Hint:** You might try doing it in the order, $dy\, dx\, dz$

14.5 Exercises With Answers

1. Evaluate the integral $\int_4^7 \int_5^{3x} \int_{5y}^{x} dz\, dy\, dx$

   Answer:

   $-\frac{3417}{2}$

2. Find $\int_0^4 \int_0^{2-5x} \int_0^{4-2x-y} (2x)\, dz\, dy\, dx$

   Answer:

   $-\frac{2464}{3}$

3. Find $\int_0^2 \int_0^{2-5x} \int_0^{1-4x-3y} (2x)\, dz\, dy\, dx$

   Answer:

   $-\frac{196}{3}$

4. Evaluate the integral $\int_5^8 \int_4^{3x} \int_{4y}^{x} (x - y)\, dz\, dy\, dx$

   Answer:

   $\frac{114\,607}{8}$

5. Evaluate the integral $\int_0^{\pi} \int_0^{4y} \int_0^{y+z} \cos(x + y)\, dx\, dz\, dy$

   Answer:

   -4π

6. Evaluate the integral $\int_0^{\pi} \int_0^{2y} \int_0^{y+z} \sin(x + y)\, dx\, dz\, dy$

   Answer:

   $-\frac{19}{4}$

7. Fill in the missing limits. $\int_0^1 \int_0^z \int_0^z f(x, y, z)\, dx\, dy\, dz = \int_?^? \int_?^? \int_?^? f(x, y, z)\, dx\, dz\, dy$,
   $\int_0^1 \int_0^z \int_0^{2z} f(x, y, z)\, dx\, dy\, dz = \int_?^? \int_?^? \int_?^? f(x, y, z)\, dy\, dz\, dx$,
   $\int_0^1 \int_0^z \int_0^z f(x, y, z)\, dx\, dy\, dz = \int_?^? \int_?^? \int_?^? f(x, y, z)\, dz\, dy\, dx$,
   $\int_0^1 \int_{z/2}^{\sqrt{z}} \int_0^{y+z} f(x, y, z)\, dx\, dy\, dz = \int_?^? \int_?^? \int_?^? f(x, y, z)\, dx\, dz\, dy$,

$\int_5^7 \int_2^5 \int_0^3 f(x,y,z)\, dx\, dy\, dz = \int_?^? \int_?^? \int_?^? f(x,y,z)\, dz\, dy\, dx.$

Answer:

$\int_0^1 \int_0^z \int_0^z f(x,y,z)\, dx\, dy\, dz = \int_0^1 \int_y^1 \int_0^z f(x,y,z)\, dx\, dz\, dy,$

$\int_0^1 \int_0^z \int_0^{2z} f(x,y,z)\, dx\, dy\, dz = \int_0^2 \int_{x/2}^1 \int_0^z f(x,y,z)\, dy\, dz\, dx,$

$\int_0^1 \int_0^z \int_0^z f(x,y,z)\, dx\, dy\, dz = \int_0^1 \left[\int_0^x \int_x^1 f(x,y,z)\, dz\, dy + \int_x^1 \int_y^1 f(x,y,z)\, dz\, dy \right] dx,$

$\int_0^1 \int_{z/2}^{\sqrt{z}} \int_0^{y+z} f(x,y,z)\, dx\, dy\, dz =$

$\int_0^{1/2} \int_{y^2}^{2y} \int_0^{y+z} f(x,y,z)\, dx\, dz\, dy + \int_{1/2}^1 \int_{y^2}^1 \int_0^{y+z} f(x,y,z)\, dx\, dz\, dy$

$\int_5^7 \int_2^5 \int_0^3 f(x,y,z)\, dx\, dy\, dz = \int_0^3 \int_2^5 \int_5^7 f(x,y,z)\, dz\, dy\, dx$

8. Find the volume of R where R is the bounded region formed by the plane $\frac{1}{5}x + y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$.

   Answer:

   $\int_0^5 \int_0^{1-\frac{1}{5}x} \int_0^{4-\frac{4}{5}x-4y} dz\, dy\, dx = \frac{10}{3}$

9. Find the volume of R where R is the bounded region formed by the plane $\frac{1}{5}x + \frac{1}{2}y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$.

   Answer:

   $\int_0^5 \int_0^{2-\frac{2}{5}x} \int_0^{4-\frac{4}{5}x-2y} dz\, dy\, dx = \frac{20}{3}$

10. Find the mass of the bounded region, R formed by the plane $\frac{1}{4}x + \frac{1}{2}y + \frac{1}{3}z = 1$ and the planes $x = 0, y = 0, z = 0$ if the density is $\rho(x,y,z) = y$

    Answer:

    $\int_0^4 \int_0^{2-\frac{1}{2}x} \int_0^{3-\frac{3}{4}x-\frac{3}{2}y} (y)\, dz\, dy\, dx = 2$

11. Find the mass of the bounded region, R formed by the plane $\frac{1}{2}x + \frac{1}{2}y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$ if the density is $\rho(x,y,z) = z^2$

    Answer:

    $\int_0^2 \int_0^{2-x} \int_0^{4-2x-2y} \left(z^2\right) dz\, dy\, dx = \frac{64}{15}$

12. Here is an iterated integral: $\int_0^3 \int_0^{3-x} \int_0^{x^2} dz\, dy\, dx$. Write as an iterated integral in the following orders: $dz\, dx\, dy$, $dx\, dz\, dy$, $dx\, dy\, dz$, $dy\, dx\, dz$, $dy\, dz\, dx$.

    Answer:

    $$\int_0^3 \int_0^{x^2} \int_0^{3-x} dy\, dz\, dx, \int_0^9 \int_{\sqrt{z}}^3 \int_0^{3-x} dy\, dx\, dz, \int_0^9 \int_0^{3-\sqrt{z}} \int_{\sqrt{z}}^{3-y} dx\, dy\, dz,$$
    $$\int_0^3 \int_0^{3-y} \int_0^{x^2} dz\, dx\, dy, \int_0^3 \int_0^{(3-y)^2} \int_{\sqrt{z}}^{3-y} dx\, dz\, dy$$

13. Find the volume of the bounded region determined by $5y + 2z = 4, x = 4 - y^2, y = 0, x = 0$.

    Answer:

    $\int_0^{\frac{4}{5}} \int_0^{2-\frac{5}{2}y} \int_0^{4-y^2} dx\, dz\, dy = \frac{1168}{375}$

14. Find the volume of the bounded region determined by $4y+3z=3, x=4-y^2, y=0, x=0$.

    Answer:
    $\int_0^{\frac{3}{4}} \int_0^{1-\frac{4}{3}y} \int_0^{4-y^2} \, dx\, dz\, dy = \frac{375}{256}$

15. Find the volume of the bounded region determined by $3y+z=3, x=4-y^2, y=0, x=0$.

    Answer:
    $\int_0^1 \int_0^{3-3y} \int_0^{4-y^2} \, dx\, dz\, dy = \frac{23}{4}$

16. Find the volume of the region bounded by $x^2+y^2=16, z=3x, z=0$, and $x \geq 0$.

    Answer:
    $\int_0^4 \int_{-\sqrt{(16-x^2)}}^{\sqrt{(16-x^2)}} \int_0^{3x} \, dz\, dy\, dx = 128$

17. Find the volume of the region bounded by $x^2+y^2=25, z=2x, z=0$, and $x \geq 0$.

    Answer:
    $\int_0^5 \int_{-\sqrt{(25-x^2)}}^{\sqrt{(25-x^2)}} \int_0^{2x} \, dz\, dy\, dx = \frac{500}{3}$

18. Find the volume of the region determined by the intersection of the two cylinders, $x^2+y^2 \leq 9$ and $y^2+z^2 \leq 9$.

    Answer:
    $8 \int_0^3 \int_0^{\sqrt{(9-y^2)}} \int_0^{\sqrt{(9-y^2)}} \, dz\, dx\, dy = 144$

19. Find the total mass of the bounded solid determined by $z = a^2 - x^2 - y^2$ and $x, y, z \geq 0$ if the mass is given by $\rho(x,y,z) = z$

    Answer:
    $\int_0^4 \int_0^{\sqrt{(16-x^2)}} \int_0^{16-x^2-y^2} (z) \, dz\, dy\, dx = \frac{512}{3}\pi$

20. Find the total mass of the bounded solid determined by $z = a^2 - x^2 - y^2$ and $x, y, z \geq 0$ if the mass is given by $\rho(x,y,z) = x+1$

    Answer:
    $\int_0^5 \int_0^{\sqrt{(25-x^2)}} \int_0^{25-x^2-y^2} (x+1) \, dz\, dy\, dx = \frac{625}{8}\pi + \frac{1250}{3}$

21. Find the volume of the region bounded by $x^2+y^2=9, z=0, z=5-y$

    Answer:
    $\int_{-3}^3 \int_{-\sqrt{(9-x^2)}}^{\sqrt{(9-x^2)}} \int_0^{5-y} \, dz\, dy\, dx = 45\pi$

22. Find the volume of the bounded region determined by $x \geq 0, y \geq 0, z \geq 0$, and $\frac{1}{2}x + y + \frac{1}{2}z = 1$, and $x + \frac{1}{2}y + \frac{1}{2}z = 1$.

    Answer:
    $\int_0^{\frac{2}{3}} \int_x^{1-\frac{1}{2}x} \int_0^{2-x-2y} \, dz\, dy\, dx + \int_0^{\frac{2}{3}} \int_y^{1-\frac{1}{2}y} \int_0^{2-2x-y} \, dz\, dx\, dy = \frac{4}{9}$

23. Find the volume of the bounded region determined by $x \geq 0, y \geq 0, z \geq 0$, and $\frac{1}{7}x + y + \frac{1}{3}z = 1$, and $x + \frac{1}{7}y + \frac{1}{3}z = 1$.

    Answer:
    $\int_0^{\frac{7}{8}} \int_x^{1-\frac{1}{7}x} \int_0^{3-\frac{3}{7}x-3y} \, dz\, dy\, dx + \int_0^{\frac{7}{8}} \int_y^{1-\frac{1}{7}y} \int_0^{3-3x-\frac{3}{7}y} \, dz\, dx\, dy = \frac{7}{8}$

24. Find the mass of the solid determined by $25x^2 + 4y^2 \leq 9, z \geq 0$, and $z = x + 2$ if the density is $\rho(x, y, z) = x$.

    Answer:
    $\int_{-\frac{3}{5}}^{\frac{3}{5}} \int_{-\frac{1}{2}\sqrt{(9-25x^2)}}^{\frac{1}{2}\sqrt{(9-25x^2)}} \int_0^{x+2} (x)\, dz\, dy\, dx = \frac{81}{1000}\pi$

25. Find $\int_0^1 \int_0^{35-5z} \int_{\frac{1}{5}x}^{7-z} (7-z)\cos\left(y^2\right)\, dy\, dx\, dz$.

    Answer:

    You need to interchange the order of integration. $\int_0^1 \int_0^{7-z} \int_0^{5y} (7-z)\cos\left(y^2\right)\, dx\, dy\, dz = \frac{5}{4}\cos 36 - \frac{5}{4}\cos 49$

26. Find $\int_0^2 \int_0^{12-3z} \int_{\frac{1}{3}x}^{4-z} (4-z)\exp\left(y^2\right)\, dy\, dx\, dz$.

    Answer:

    You need to interchange the order of integration. $\int_0^2 \int_0^{4-z} \int_0^{3y} (4-z)\exp\left(y^2\right)\, dx\, dy\, dz = -\frac{3}{4}e^4 - 9 + \frac{3}{4}e^{16}$

27. Find $\int_0^2 \int_0^{25-5z} \int_{\frac{1}{5}y}^{5-z} (5-z)\exp\left(x^2\right)\, dx\, dy\, dz$.

    Answer:

    You need to interchange the order of integration.

$$\int_0^2 \int_0^{5-z} \int_0^{5x} (5-z)\exp\left(x^2\right)\, dy\, dx\, dz = -\frac{5}{4}e^9 - 20 + \frac{5}{4}e^{25}$$

28. Find $\int_0^1 \int_0^{10-2z} \int_{\frac{1}{2}y}^{5-z} \frac{\sin x}{x}\, dx\, dy\, dz$.

    Answer:

    You need to interchange the order of integration.

$$\int_0^1 \int_0^{5-z} \int_0^{2x} \frac{\sin x}{x}\, dy\, dx\, dz =$$

$$-2\sin 1\cos 5 + 2\cos 1\sin 5 + 2 - 2\sin 5$$

29. Find $\int_0^{20} \int_0^2 \int_{\frac{1}{5}y}^{6-z} \frac{\sin x}{x}\, dx\, dz\, dy + \int_{20}^{30} \int_0^{6-\frac{1}{5}y} \int_{\frac{1}{5}y}^{6-z} \frac{\sin x}{x}\, dx\, dz\, dy$.

    Answer:

    You need to interchange the order of integration.

$$\int_0^2 \int_0^{30-5z} \int_{\frac{1}{5}y}^{6-z} \frac{\sin x}{x}\, dx\, dy\, dz = \int_0^2 \int_0^{6-z} \int_0^{5x} \frac{\sin x}{x}\, dy\, dx\, dz$$

$$= -5\sin 2\cos 6 + 5\cos 2\sin 6 + 10 - 5\sin 6$$

The Integral In Other Coordinates

15.0.1 Outcomes

1. Represent a region in polar coordinates and use to evaluate integrals.
2. Represent a region in spherical or cylindrical coordinates and use to evaluate integrals.
3. Convert integrals in rectangular coordinates to integrals in polar coordinates and use to evaluate the integral.
4. Evaluate integrals in any coordinate system using the Jacobian.
5. Evaluate areas and volumes using another coordinate system.
6. Understand the transformation equations between spherical, polar and cylindrical coordinates and be able to change algebraic expressions from one system to another.
7. Use multiple integrals in an appropriate coordinate system to calculate the volume, mass, moments, center of gravity and moment of inertia.

15.1 Different Coordinates

As mentioned above, the fundamental concept of an integral is a sum of things of the form $f(\mathbf{x})\,dV$ where dV is an "infinitesimal" chunk of volume located at the point, $\mathbf{x}$. Up to now, this infinitesimal chunk of volume has had the form of a box with sides $dx_1, \cdots, dx_n$ so $dV = dx_1\,dx_2 \cdots dx_n$ but its form is not important. It could just as well be an infinitesimal parallelepiped or parallelogram for example. In what follows, this is what it will be.

First recall the definition of the box product given in Definition 5.5.17 on Page 113. The absolute value of the box product of three vectors gave the volume of the parallelepiped determined by the three vectors.

Definition 15.1.1 *Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ be vectors in $\mathbb{R}^3$. The parallelepiped determined by these vectors will be denoted by $P(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ and it is defined as*

$$P(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \equiv \left\{ \sum_{j=1}^{3} s_j \mathbf{u}_j : s_j \in [0,1] \right\}.$$

Lemma 15.1.2 *The volume of the parallelepiped, $P(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ is given by $\left|\det\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}\right|$ where $\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}$ is the matrix having columns $\mathbf{u}_1, \mathbf{u}_2,$ and $\mathbf{u}_3$.*

Proof: Recall from the discussion of the box product or triple product,

$$\text{volume of } P(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \equiv |[\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]| = \left|\det\begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{pmatrix}\right|$$

where $\begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{pmatrix}$ is the matrix having rows equal to the vectors, $\mathbf{u}_1, \mathbf{u}_2$ and $\mathbf{u}_3$ arranged horizontally. Since the determinant of a matrix equals the determinant of its transpose,

$$\text{volume of } P(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = |[\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]| = \left|\det\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}\right|.$$

This proves the lemma.

Definition 15.1.3 *In the case of two vectors, $P(\mathbf{u}_1, \mathbf{u}_2)$ will denote the parallelogram determined by $\mathbf{u}_1$ and $\mathbf{u}_2$. Thus*

$$P(\mathbf{u}_1, \mathbf{u}_2) \equiv \left\{ \sum_{j=1}^{2} s_j \mathbf{u}_j : s_j \in [0,1] \right\}.$$

Lemma 15.1.4 *The area of the parallelogram, $P(\mathbf{u}_1, \mathbf{u}_2)$ is given by $\left|\det\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix}\right|$ where $\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix}$ is the matrix having columns $\mathbf{u}_1$ and $\mathbf{u}_2$.*

Proof: Letting $\mathbf{u}_1 = (a,b)^T$ and $\mathbf{u}_2 = (c,d)^T$, consider the vectors in $\mathbb{R}^3$ defined by $\widehat{\mathbf{u}}_1 \equiv (a,b,0)^T$ and $\widehat{\mathbf{u}}_2 \equiv (c,d,0)^T$. Then the area of the parallelogram determined by the vectors, $\widehat{\mathbf{u}}_1$ and $\widehat{\mathbf{u}}_2$ is the norm of the cross product of $\widehat{\mathbf{u}}_1$ and $\widehat{\mathbf{u}}_2$. This follows directly from the geometric definition of the cross product given in Definition 5.5.2 on Page 106. But this is the same as the area of the parallelogram determined by the vectors $\mathbf{u}_1, \mathbf{u}_2$. Taking the cross product of $\widehat{\mathbf{u}}_1$ and $\widehat{\mathbf{u}}_2$ yields $\mathbf{k}(ad - bc)$. Therefore, the norm of this cross product is

$$|ad - bc|$$

which is the same as

$$\left|\det\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix}\right|$$

where $\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix}$ denotes the matrix having the two vectors $\mathbf{u}_1, \mathbf{u}_2$ as columns. This proves the lemma.

It always works this way. The n dimensional volume of the n dimensional parallelepiped determined by the vectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is always

$$\left|\det\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix}\right|$$

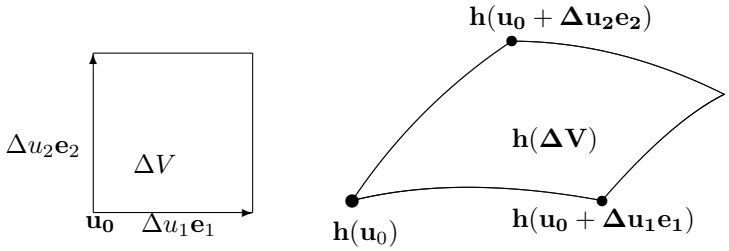
This general fact will not be used in what follows.

15.1.1 Two Dimensional Coordinates

Suppose U is a set in $\mathbb{R}^2$ and $\mathbf{h}$ is a C^1 function[1] mapping U one to one onto $\mathbf{h}(U)$, a set in $\mathbb{R}^2$. Consider a small square inside U. The following picture is of such a square

[1] By this is meant $\mathbf{h}$ is the restriction to U of a function defined on an open set containing U which is C^1. If you like, you can assume U is open but this is not necessary. Neither is C^1.

having a corner at the point, $\mathbf{u}_0$ and sides as indicated. The image of this square is also represented.



For small Δu_i you would expect the sides going from $\mathbf{h}(\mathbf{u}_0)$ to $\mathbf{h}(\mathbf{u}_0 + \Delta u_1 \mathbf{e}_1)$ and from $\mathbf{h}(\mathbf{u}_0)$ to $\mathbf{h}(\mathbf{u}_0 + \Delta u_2 \mathbf{e}_2)$ to be almost the same as the vectors, $\mathbf{h}(\mathbf{u}_0 + \Delta u_1 \mathbf{e}_1) - \mathbf{h}(\mathbf{u}_0)$ and $\mathbf{h}(\mathbf{u}_0 + \Delta u_2 \mathbf{e}_2) - \mathbf{h}(\mathbf{u}_0)$ which are approximately equal to $\frac{\partial \mathbf{h}}{\partial u_1}(\mathbf{u}_0)\Delta u_1$ and $\frac{\partial \mathbf{h}}{\partial u_2}(\mathbf{u}_0)\Delta u_2$ respectively. Therefore, the area of $\mathbf{h}(\Delta V)$ for small Δu_i is essentially equal to the area of the parallelogram determined by the two vectors, $\frac{\partial \mathbf{h}}{\partial u_1}(\mathbf{u}_0)\Delta u_1$ and $\frac{\partial \mathbf{h}}{\partial u_2}(\mathbf{u}_0)\Delta u_2$. By Lemma 15.1.4 this equals

$$\left|\det\left(\begin{array}{cc} \frac{\partial \mathbf{h}}{\partial u_1}(\mathbf{u}_0)\Delta u_1 & \frac{\partial \mathbf{h}}{\partial u_2}(\mathbf{u}_0)\Delta u_2 \end{array}\right)\right| = \left|\det\left(\begin{array}{cc} \frac{\partial \mathbf{h}}{\partial u_1}(\mathbf{u}_0) & \frac{\partial \mathbf{h}}{\partial u_2}(\mathbf{u}_0) \end{array}\right)\right| \Delta u_1 \Delta u_2$$

Thus an infinitesimal chunk of area in $\mathbf{h}(U)$ located at $\mathbf{u}_0$ is of the form

$$\left|\det\left(\begin{array}{cc} \frac{\partial \mathbf{h}}{\partial u_1}(\mathbf{u}_0) & \frac{\partial \mathbf{h}}{\partial u_2}(\mathbf{u}_0) \end{array}\right)\right| dV$$

where dV is a corresponding chunk of area located at the point $\mathbf{u}_0$. This shows the following change of variables formula is reasonable.

$$\int_{\mathbf{h}(U)} f(\mathbf{x})\, dV(\mathbf{x}) = \int_U f(\mathbf{h}(\mathbf{u})) \left|\det\left(\begin{array}{cc} \frac{\partial \mathbf{h}}{\partial u_1}(\mathbf{u}) & \frac{\partial \mathbf{h}}{\partial u_2}(\mathbf{u}) \end{array}\right)\right| dV(\mathbf{u})$$

Definition 15.1.5 *Let $\mathbf{h} : U \to \mathbf{h}(U)$ be a one to one and C^1 mapping. The (volume) area element in terms of $\mathbf{u}$ is defined as $\left|\det\left(\begin{array}{cc} \frac{\partial \mathbf{h}}{\partial u_1}(\mathbf{u}) & \frac{\partial \mathbf{h}}{\partial u_2}(\mathbf{u}) \end{array}\right)\right| dV(\mathbf{u})$. The factor, $\left|\det\left(\begin{array}{cc} \frac{\partial \mathbf{h}}{\partial u_1}(\mathbf{u}) & \frac{\partial \mathbf{h}}{\partial u_2}(\mathbf{u}) \end{array}\right)\right|$ is called the Jacobian. It equals*

$$\left|\det\begin{pmatrix} \frac{\partial h_1}{\partial u_1}(u_1,u_2) & \frac{\partial h_1}{\partial u_2}(u_1,u_2) \\ \frac{\partial h_2}{\partial u_1}(u_1,u_2) & \frac{\partial h_2}{\partial u_2}(u_1,u_2) \end{pmatrix}\right|.$$

It is traditional to call two dimensional volumes area. However, it is probably better to simply always refer to it as volume. Thus there is 2 dimensional volume, 3 dimensional volume, etc. Sometimes you can get confused by too many different words to describe things which are really not essentially different.

Example 15.1.6 *Find the area element for polar coordinates.*

Here the $\mathbf{u}$ coordinates are θ and r. The polar coordinate transformations are

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}$$

Therefore, the volume (area) element is

$$\left|\det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}\right| d\theta dr = r d\theta dr.$$

15.1.2 Three Dimensions

The situation is no different for coordinate systems in any number of dimensions although I will concentrate here on three dimensions. $\mathbf{x} = \mathbf{f}(\mathbf{u})$ where $\mathbf{u} \in U$, a subset of $\mathbb{R}^3$ and $\mathbf{x}$ is a point in $V = \mathbf{f}(U)$, a subset of 3 dimensional space. Thus, letting the Cartesian coordinates of $\mathbf{x}$ be given by $\mathbf{x} = (x_1, x_2, x_3)^T$, each x_i being a function of $\mathbf{u}$, an infinitesimal box located at $\mathbf{u}_0$ corresponds to an infinitesimal parallelepiped located at $\mathbf{f}(\mathbf{u}_0)$ which is determined by the 3 vectors $\left\{\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} du_i\right\}_{i=1}^{3}$. From Lemma 15.1.2, the volume of this infinitesimal parallelepiped located at $\mathbf{f}(\mathbf{u}_0)$ is given by

$$\left|\left[\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1} du_1, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2} du_2, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_3} du_3\right]\right| = \left|\left[\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1}, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2}, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_3}\right]\right| du_1 du_2 du_3 \tag{15.1}$$

$$= \left|\det\left(\begin{array}{ccc} \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1} & \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2} & \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_3} \end{array}\right)\right| du_1 du_2 du_3$$

There is also no change in going to higher dimensions than 3.

Definition 15.1.7 *Let $\mathbf{x} = \mathbf{f}(\mathbf{u})$ be as described above. Then for $n = 2, 3$, the symbol, $\frac{\partial(x_1, \cdots x_n)}{\partial(u_1, \cdots, u_n)}$, called the Jacobian determinant, is defined by*

$$\det\left(\begin{array}{ccc} \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1} & \cdots & \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_n} \end{array}\right) \equiv \frac{\partial(x_1, \cdots, x_n)}{\partial(u_1, \cdots, u_n)}.$$

Also, the symbol, $\left|\frac{\partial(x_1, \cdots, x_n)}{\partial(u_1, \cdots, u_n)}\right| du_1 \cdots du_n$ is called the volume element.

This has given motivation for the following fundamental procedure often called the **change of variables formula** which holds under fairly general conditions.

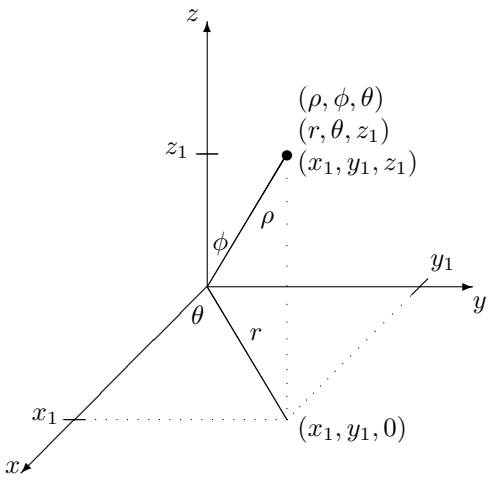
Procedure 15.1.8 *Suppose U is an open subset of $\mathbb{R}^n$ for $n = 2, 3$ and suppose $\mathbf{f} : U \to \mathbf{f}(U)$ is a C^1 function which is one to one, $\mathbf{x} = \mathbf{f}(\mathbf{u})$.[2] Then if $h : \mathbf{f}(U) \to \mathbb{R}$,*

$$\int_U h(\mathbf{f}(\mathbf{u})) \left|\frac{\partial(x_1, \cdots, x_n)}{\partial(u_1, \cdots, u_n)}\right| dV = \int_{\mathbf{f}(U)} h(\mathbf{x})\, dV.$$

[2]This will cause non overlapping infinitesimal boxes in U to be mapped to non overlapping infinitesimal parallelepipeds in V.

Also, in the context of the Riemann integral we should say more about the set U in any case the function, h. These conditions are mainly technical however, and since a mathematically respectable treatment will not be attempted for this theorem, I think it best to give a memorable version of it which is essentially correct in all examples of interest.

Now consider spherical coordinates. Recall the geometrical meaning of these coordinates illustrated in the following picture.



Thus there is a relationship between these coordinates and rectangular coordinates given by

$$x = \rho \sin\phi \cos\theta, y = \rho \sin\phi \sin\theta,\ z = \rho \cos\phi \tag{15.2}$$

where $\phi \in [0,\pi], \theta \in [0,2\pi)$, and $\rho > 0$. Thus (ρ,ϕ,θ) is a point in $\mathbb{R}^3$, more specifically in the set

$$U = (0,\infty) \times [0,\pi] \times [0,2\pi)$$

and corresponding to such a $(\rho,\phi,\theta) \in U$ there exists a unique point, $(x,y,z) \in V$ where V consists of all points of $\mathbb{R}^3$ other than the origin, $(0,0,0)$. This (x,y,z) determines a unique point in three dimensional space as mentioned earlier. From the above argument, the volume element is

$$\left|\det\left(\begin{array}{ccc} \frac{\partial \mathbf{x}(\rho_0,\phi_0,\theta_0)}{\partial\rho} & \frac{\partial \mathbf{x}(\rho_0,\phi_0,\theta_0)}{\partial\phi} & \frac{\partial \mathbf{x}(\rho_0,\phi_0,\theta_0)}{\partial\theta} \end{array}\right)\right| d\rho d\theta d\phi.$$

The mapping between spherical and rectangular coordinates is written as

$$\mathbf{x} = \left(\begin{array}{c} x \\ y \\ z \end{array}\right) = \left(\begin{array}{c} \rho\sin\phi\cos\theta \\ \rho\sin\phi\sin\theta \\ \rho\cos\phi \end{array}\right) = \mathbf{f}(\rho,\phi,\theta) \tag{15.3}$$

Therefore, $\det\left(\frac{\partial \mathbf{x}(\rho_0,\phi_0,\theta_0)}{\partial\rho}, \frac{\partial \mathbf{x}(\rho_0,\phi_0,\theta_0)}{\partial\phi}, \frac{\partial \mathbf{x}(\rho_0,\phi_0,\theta_0)}{\partial\theta}\right) =$

$$\det\left(\begin{array}{ccc} \sin\phi\cos\theta & \rho\cos\phi\cos\theta & -\rho\sin\phi\sin\theta \\ \sin\phi\sin\theta & \rho\cos\phi\sin\theta & \rho\sin\phi\cos\theta \\ \cos\phi & -\rho\sin\phi & 0 \end{array}\right) = \rho^2\sin\phi$$

which is positive because $\phi \in [0,\pi]$.

Example 15.1.9 *Find the volume of a ball, B_R of radius R.*

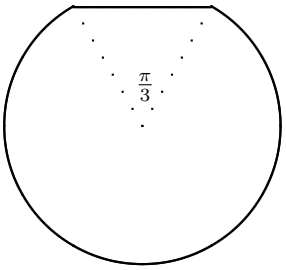
In this case, $U = (0,R] \times [0,\pi] \times [0,2\pi)$ and use spherical coordinates. Then 15.3 yields a set in $\mathbb{R}^3$ which clearly differs from the ball of radius R only by a set having

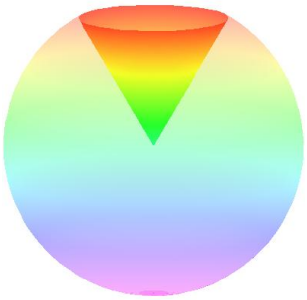
volume equal to zero. It leaves out the point at the origin is all. Therefore, the volume of the ball is

$$\begin{aligned}\int_{B_R} 1\,dV &= \int_U \rho^2 \sin\phi\, dV \\ &= \int_0^R \int_0^\pi \int_0^{2\pi} \rho^2 \sin\phi \;\, d\theta\, d\phi\, d\rho = \frac{4}{3}R^3\pi.\end{aligned}$$

The reason this was effortless, is that the ball, B_R is realized as a box in terms of the spherical coordinates. Remember what was pointed out earlier about setting up iterated integrals over boxes.

Example 15.1.10 *A cone is cut out of a ball of radius R as shown in the following picture, the diagram on the left being a side view. The angle of the cone is $\pi/3$. Find the volume of what is left.*

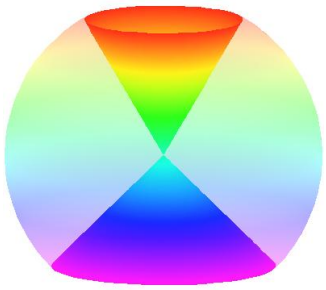




Use spherical coordinates. This volume is then

$$\int_{\pi/6}^{\pi} \int_0^{2\pi} \int_0^R \rho^2 \sin(\phi)\, d\rho d\theta d\phi = \frac{2}{3}\pi R^3 + \frac{1}{3}\sqrt{3}\pi R^3$$

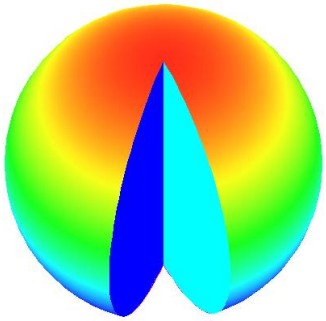
Now change the example a little by cutting out a cone at the bottom which has an angle of $\pi/2$ as shown. What is the volume of what is left?



This time you would have the volume equals

$$\int_{\pi/6}^{3\pi/4} \int_0^{2\pi} \int_0^R \rho^2 \sin(\phi)\, d\rho d\theta d\phi = \frac{1}{3}\sqrt{2}\pi R^3 + \frac{1}{3}\sqrt{3}\pi R^3$$

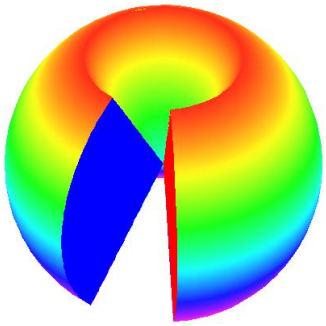
Example 15.1.11 *Next suppose the ball of radius R is a sort of an orange and you remove a slice as shown in the picture. What is the volume of what is left? Assume the slice is formed by the two half planes $\theta = 0$ and $\theta = \pi/4$.*



Using spherical coordinates, this gives for the volume

$$\int_0^{\pi}\int_{\pi/4}^{2\pi}\int_0^R \rho^2 \sin(\phi)\, d\rho d\theta d\phi = \frac{7}{6}\pi R^3$$

Example 15.1.12 *Now remove the same two cones as in the above examples along with the same slice and find the volume of what is left.*

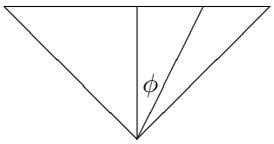


This time you need

$$\int_{\pi/6}^{3\pi/4}\int_{\pi/4}^{2\pi}\int_0^R \rho^2 \sin(\phi)\, d\rho d\theta d\phi = \frac{7}{24}\sqrt{2}\pi R^3 + \frac{7}{24}\sqrt{3}\pi R^3$$

Example 15.1.13 *Set up the integrals to find the volume of the cone* $0 \le z \le 4, z = \sqrt{x^2+y^2}$.

This is entirely the wrong coordinate system to use for this problem but it is a good exercise. Here is a side view.



You need to figure out what ρ is as a function of ϕ which goes from 0 to $\pi/4$. You should get

$$\int_0^{2\pi}\int_0^{\pi/4}\int_0^{4\sec(\phi)} \rho^2 \sin(\phi)\, d\rho d\phi d\theta = \frac{64}{3}\pi$$

Example 15.1.14 *Find the volume element for cylindrical coordinates.*

In cylindrical coordinates,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \\ z \end{pmatrix}$$

Therefore, the Jacobian determinant is

$$\det\begin{pmatrix}\cos\theta & -r\sin\theta & 0\\ \sin\theta & r\cos\theta & 0\\ 0 & 0 & 1\end{pmatrix} = r.$$

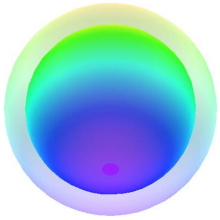
It follows the volume element in cylindrical coordinates is $r\,d\theta\,dr\,dz$.

Example 15.1.15 *In the cone of Example 15.1.13 set up the integrals for finding the volume in cylindrical coordinates.*

This is a better coordinate system for this example than spherical coordinates. This time you should get

$$\int_0^{2\pi}\int_0^4\int_r^4 r\,dz\,dr\,d\theta = \frac{64}{3}\pi$$

Example 15.1.16 *This example uses spherical coordinates to verify an important conclusion about gravitational force. Let the hollow sphere, H be defined by $a^2 < x^2+y^2+z^2 < b^2$*



and suppose this hollow sphere has constant density taken to equal 1. Now place a unit mass at the point $(0,0,z_0)$ where $|z_0| \in [a,b]$. Show the force of gravity acting on this unit mass is $\left(\alpha G\int_H \frac{(z-z_0)}{\left[x^2+y^2+(z-z_0)^2\right]^{3/2}}\,dV\right)\mathbf{k}$ and then show that if $|z_0| > b$ then the force of gravity acting on this point mass is the same as if the entire mass of the hollow sphere were placed at the origin, while if $|z_0| < a$, the total force acting on the point mass from gravity equals zero. Here G is the gravitation constant and α is the density. In particular, this shows that the force a planet exerts on an object is as though the entire mass of the planet were situated at its center[3].

Without loss of generality, assume $z_0 > 0$. Let dV be a little chunk of material located at the point (x,y,z) of H the hollow sphere. Then according to Newton's law of gravity, the force this small chunk of material exerts on the given point mass equals

$$\frac{x\mathbf{i}+y\mathbf{j}+(z-z_0)\,\mathbf{k}}{|x\mathbf{i}+y\mathbf{j}+(z-z_0)\,\mathbf{k}|}\frac{1}{\left(x^2+y^2+(z-z_0)^2\right)}G\alpha\,dV =$$

$$(x\mathbf{i}+y\mathbf{j}+(z-z_0)\,\mathbf{k})\frac{1}{\left(x^2+y^2+(z-z_0)^2\right)^{3/2}}G\alpha\,dV$$

Therefore, the total force is

$$\int_H (x\mathbf{i}+y\mathbf{j}+(z-z_0)\,\mathbf{k})\frac{1}{\left(x^2+y^2+(z-z_0)^2\right)^{3/2}}G\alpha\,dV.$$

[3]This was shown by Newton in 1685 and allowed him to assert his law of gravitation applied to the planets as though they were point masses. It was a major accomplishment.

By the symmetry of the sphere, the **i** and **j** components will cancel out when the integral is taken. This is because there is the same amount of stuff for negative x and y as there is for positive x and y. Hence what remains is

$$\alpha G\mathbf{k}\int_H \frac{(z-z_0)}{\left[x^2+y^2+(z-z_0)^2\right]^{3/2}}\,dV$$

as claimed. Now for the interesting part, the integral is evaluated. In spherical coordinates this integral is.

$$\int_0^{2\pi}\int_a^b\int_0^{\pi} \frac{(\rho\cos\phi - z_0)\,\rho^2\sin\phi}{\left(\rho^2+z_0^2-2\rho z_0\cos\phi\right)^{3/2}}\,d\phi\,d\rho\,d\theta. \tag{15.4}$$

Rewrite the inside integral and use integration by parts to obtain this inside integral equals

$$\frac{1}{2z_0}\int_0^{\pi}\left(\rho^2\cos\phi-\rho z_0\right)\frac{(2z_0\rho\sin\phi)}{\left(\rho^2+z_0^2-2\rho z_0\cos\phi\right)^{3/2}}\,d\phi =$$

$$\frac{1}{2z_0}\left(-2\frac{-\rho^2-\rho z_0}{\sqrt{(\rho^2+z_0^2+2\rho z_0)}}+2\frac{\rho^2-\rho z_0}{\sqrt{(\rho^2+z_0^2-2\rho z_0)}}\right.$$

$$\left.-\int_0^{\pi}2\rho^2\frac{\sin\phi}{\sqrt{(\rho^2+z_0^2-2\rho z_0\cos\phi)}}\,d\phi\right). \tag{15.5}$$

There are some cases to consider here.

First suppose $z_0 < a$ so the point is on the inside of the hollow sphere and it is always the case that $\rho > z_0$. Then in this case, the two first terms reduce to

$$\frac{2\rho\,(\rho+z_0)}{\sqrt{(\rho+z_0)^2}}+\frac{2\rho\,(\rho-z_0)}{\sqrt{(\rho-z_0)^2}}=\frac{2\rho\,(\rho+z_0)}{(\rho+z_0)}+\frac{2\rho\,(\rho-z_0)}{\rho-z_0}=4\rho$$

and so the expression in 15.5 equals

$$\frac{1}{2z_0}\left(4\rho-\int_0^{\pi}2\rho^2\frac{\sin\phi}{\sqrt{(\rho^2+z_0^2-2\rho z_0\cos\phi)}}\,d\phi\right)$$

$$=\frac{1}{2z_0}\left(4\rho-\frac{1}{z_0}\int_0^{\pi}\rho\frac{2\rho z_0\sin\phi}{\sqrt{(\rho^2+z_0^2-2\rho z_0\cos\phi)}}\,d\phi\right)$$

$$=\frac{1}{2z_0}\left(4\rho-\frac{2\rho}{z_0}\left(\rho^2+z_0^2-2\rho z_0\cos\phi\right)^{1/2}|_0^{\pi}\right)$$

$$=\frac{1}{2z_0}\left(4\rho-\frac{2\rho}{z_0}\left[(\rho+z_0)-(\rho-z_0)\right]\right)=0.$$

Therefore, in this case the inner integral of 15.4 equals zero and so the original integral will also be zero.

The other case is when $z_0 > b$ and so it is always the case that $z_0 > \rho$. In this case the first two terms of 15.5 are

$$\frac{2\rho\,(\rho+z_0)}{\sqrt{(\rho+z_0)^2}}+\frac{2\rho\,(\rho-z_0)}{\sqrt{(\rho-z_0)^2}}=\frac{2\rho\,(\rho+z_0)}{(\rho+z_0)}+\frac{2\rho\,(\rho-z_0)}{z_0-\rho}=0.$$

Therefore in this case, 15.5 equals

$$\begin{aligned} & \frac{1}{2z_0}\left(-\int_0^\pi 2\rho^2 \frac{\sin\phi}{\sqrt{(\rho^2+z_0^2-2\rho z_0\cos\phi)}}\,d\phi\right) \\ = & \frac{-\rho}{2z_0^2}\left(\int_0^\pi \frac{2\rho z_0\sin\phi}{\sqrt{(\rho^2+z_0^2-2\rho z_0\cos\phi)}}\,d\phi\right) \end{aligned}$$

which equals

$$\begin{aligned} & \frac{-\rho}{z_0^2}\left(\left(\rho^2+z_0^2-2\rho z_0\cos\phi\right)^{1/2}|_0^\pi\right) \\ = & \frac{-\rho}{z_0^2}\left[(\rho+z_0)-(z_0-\rho)\right] = -\frac{2\rho^2}{z_0^2}. \end{aligned}$$

Thus the inner integral of 15.4 reduces to the above simple expression. Therefore, 15.4 equals

$$\int_0^{2\pi}\int_a^b\left(-\frac{2}{z_0^2}\rho^2\right)\,d\rho\,d\theta = -\frac{4}{3}\pi\frac{b^3-a^3}{z_0^2}$$

and so

$$\begin{aligned} & \alpha G\mathbf{k}\int_H \frac{(z-z_0)}{\left[x^2+y^2+(z-z_0)^2\right]^{3/2}}\,dV \\ = & \alpha G\mathbf{k}\left(-\frac{4}{3}\pi\frac{b^3-a^3}{z_0^2}\right) = -\mathbf{k}G\frac{\text{total mass}}{z_0^2}. \end{aligned}$$

15.2 Exercises

1. Find the area of the bounded region, R, determined by $5x+y=2, 5x+y=8, y=2x$, and $y=6x$.

2. Find the area of the bounded region, R, determined by $y+2x=6, y+2x=10, y=3x$, and $y=4x$.

3. A solid, R is determined by $3x+y=2, 3x+y=4, y=2x$, and $y=6x$ and the density is $\rho=x$. Find the total mass of R.

4. A solid, R is determined by $4x+2y=5, 4x+2y=6, y=5x$, and $y=7x$ and the density is $\rho=y$. Find the total mass of R.

5. A solid, R is determined by $3x+y=3, 3x+y=10, y=3x$, and $y=5x$ and the density is $\rho=y^{-1}$. Find the total mass of R.

6. Find the volume of the region, E, bounded by the ellipsoid, $\frac{1}{4}x^2+y^2+z^2=1$.

7. Here are three vectors. $(4,1,2)^T, (5,0,2)^T$, and $(3,1,3)^T$. These vectors determine a parallelepiped, R, which is occupied by a solid having density $\rho=x$. Find the mass of this solid.

8. Here are three vectors. $(5,1,6)^T, (6,0,6)^T$, and $(4,1,7)^T$. These vectors determine a parallelepiped, R, which is occupied by a solid having density $\rho=y$. Find the mass of this solid.

9. Here are three vectors. $(5,2,9)^T, (6,1,9)^T$, and $(4,2,10)^T$. These vectors determine a parallelepiped, R, which is occupied by a solid having density $\rho = y + x$. Find the mass of this solid.

10. Let $D = \{(x,y) : x^2 + y^2 \leq 25\}$. Find $\int_D e^{25x^2+25y^2}\, dx\, dy$.

11. Let $D = \{(x,y) : x^2 + y^2 \leq 16\}$. Find $\int_D \cos\left(9x^2 + 9y^2\right)\, dx\, dy$.

12. The ice cream in a sugar cone is described in spherical coordinates by $\rho \in [0,10], \phi \in \left[0, \frac{1}{3}\pi\right], \theta \in [0, 2\pi]$. If the units are in centimeters, find the total volume in cubic centimeters of this ice cream.

13. Find the volume between $z = 5 - x^2 - y^2$ and $z = 2\sqrt{x^2 + y^2}$.

14. A ball of radius 3 is placed in a drill press and a hole of radius 2 is drilled out with the center of the hole a diameter of the ball. What is the volume of the material which remains?

15. A ball of radius 9 has density equal to $\sqrt{x^2 + y^2 + z^2}$ in rectangular coordinates. The top of this ball is sliced off by a plane of the form $z = 2$. What is the mass of what remains?

16. Find $\int_S \frac{y}{x}\, dV$ where S is described in polar coordinates as $1 \leq r \leq 2$ and $0 \leq \theta \leq \pi/4$.

17. Find $\int_S \left(\left(\frac{y}{x}\right)^2 + 1\right)\, dV$ where S is given in polar coordinates as $1 \leq r \leq 2$ and $0 \leq \theta \leq \frac{1}{6}\pi$.

18. Use polar coordinates to evaluate the following integral. Here S is given in terms of the polar coordinates. $\int_S \sin\left(2x^2 + 2y^2\right)\, dV$ where $r \leq 2$ and $0 \leq \theta \leq \frac{3}{2}\pi$.

19. Find $\int_S e^{2x^2+2y^2}\, dV$ where S is given in terms of the polar coordinates, $r \leq 2$ and $0 \leq \theta \leq \pi$.

20. Compute the volume of a sphere of radius R using cylindrical coordinates.

21. In Example 15.1.16 on Page 338 check out all the details by working the integrals to be sure the steps are right.

22. What if the hollow sphere in Example 15.1.16 were in two dimensions and everything, including Newton's law still held? Would similar conclusions hold? Explain.

23. Fill in all details for the following argument that $\int_0^\infty e^{-x^2} dx = \frac{1}{2}\sqrt{\pi}$. Let $I = \int_0^\infty e^{-x^2} dx$. Then

$$I^2 = \int_0^\infty \int_0^\infty e^{-\left(x^2+y^2\right)} dx\, dy = \int_0^{\pi/2} \int_0^\infty r e^{-r^2} dr\, d\theta = \frac{1}{4}\pi$$

from which the result follows.

24. Show $\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$. Here σ is a positive number called the standard deviation and μ is a number called the mean.

25. Show using Problem 23 $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$,. Recall $\Gamma(\alpha) \equiv \int_0^\infty e^{-t} t^{\alpha-1} dt$.

26. Let $p, q > 0$ and define $B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}$. Show

$$\Gamma(p)\Gamma(q) = B(p,q)\Gamma(p+q)$$

. **Hint:** It is fairly routine if you start with the left side and proceed to change variables.

15.3 Exercises With Answers

1. Find the area of the bounded region, R, determined by $3x+3y=1, 3x+3y=8, y=3x$, and $y=4x$.

   Answer:

   Let $u=\frac{y}{x}, v=3x+3y$. Then solving these equations for x and y yields

   $$\left\{x=\frac{1}{3}\frac{v}{1+u}, y=\frac{1}{3}u\frac{v}{1+u}\right\}.$$

   Now

   $$\frac{\partial(x,y)}{\partial(u,v)}=\det\begin{pmatrix} -\frac{1}{3}\frac{v}{(1+u)^2} & \frac{1}{3+3u} \\ \frac{1}{3}\frac{v}{(1+u)^2} & \frac{1}{3}\frac{u}{1+u}\end{pmatrix}=-\frac{1}{9}\frac{v}{(1+u)^2}.$$

   Also, $u\in[3,4]$ while $v\in[1,8]$. Therefore,

   $$\int_R dV=\int_3^4\int_1^8\left|-\frac{1}{9}\frac{v}{(1+u)^2}\right|dv\,du=$$

   $$\int_3^4\int_1^8\frac{1}{9}\frac{v}{(1+u)^2}\,dv\,du=\frac{7}{40}$$

2. Find the area of the bounded region, R, determined by $5x+y=1, 5x+y=9, y=2x$, and $y=5x$.

   Answer:

   Let $u=\frac{y}{x}, v=5x+y$. Then solving these equations for x and y yields

   $$\left\{x=\frac{v}{5+u}, y=u\frac{v}{5+u}\right\}.$$

   Now

   $$\frac{\partial(x,y)}{\partial(u,v)}=\det\begin{pmatrix} -\frac{v}{(5+u)^2} & \frac{1}{5+u} \\ 5\frac{v}{(5+u)^2} & \frac{u}{5+u}\end{pmatrix}=-\frac{v}{(5+u)^2}.$$

   Also, $u\in[2,5]$ while $v\in[1,9]$. Therefore,

   $$\int_R dV=\int_2^5\int_1^9\left|-\frac{v}{(5+u)^2}\right|dv\,du=\int_2^5\int_1^9\frac{v}{(5+u)^2}\,dv\,du=\frac{12}{7}$$

3. A solid, R is determined by $5x+3y=4, 5x+3y=9, y=2x$, and $y=5x$ and the density is $\rho=x$. Find the total mass of R.

   Answer:

   Let $u=\frac{y}{x}, v=5x+3y$. Then solving these equations for x and y yields

   $$\left\{x=\frac{v}{5+3u}, y=u\frac{v}{5+3u}\right\}.$$

   Now

$$\frac{\partial(x,y)}{\partial(u,v)} = \det\begin{pmatrix} -3\frac{v}{(5+3u)^2} & \frac{1}{5+3u} \\ 5\frac{v}{(5+3u)^2} & \frac{u}{5+3u} \end{pmatrix} = -\frac{v}{(5+3u)^2}.$$

Also, $u \in [2,5]$ while $v \in [4,9]$. Therefore,

$$\int_R \rho\, dV = \int_2^5 \int_4^9 \frac{v}{5+3u}\left|-\frac{v}{(5+3u)^2}\right| dv\, du =$$

$$\int_2^5 \int_4^9 \left(\frac{v}{5+3u}\right)\left(\frac{v}{(5+3u)^2}\right) dv\, du = \frac{4123}{19\,360}.$$

4. A solid, R is determined by $2x+2y=1, 2x+2y=10, y=4x$, and $y=5x$ and the density is $\rho = x+1$. Find the total mass of R.

   Answer:

   Let $u=\frac{y}{x}, v=2x+2y$. Then solving these equations for x and y yields

   $$\left\{x = \frac{1}{2}\frac{v}{1+u}, y = \frac{1}{2}u\frac{v}{1+u}\right\}.$$

   Now

   $$\frac{\partial(x,y)}{\partial(u,v)} = \det\begin{pmatrix} -\frac{1}{2}\frac{v}{(1+u)^2} & \frac{1}{2+2u} \\ \frac{1}{2}\frac{v}{(1+u)^2} & \frac{1}{2}\frac{u}{1+u} \end{pmatrix} = -\frac{1}{4}\frac{v}{(1+u)^2}.$$

   Also, $u \in [4,5]$ while $v \in [1,10]$. Therefore,

   $$\begin{aligned} \int_R \rho\, dV &= \int_4^5 \int_1^{10} (x+1)\left|-\frac{1}{4}\frac{v}{(1+u)^2}\right| dv\, du \\ &= \int_4^5 \int_1^{10} (x+1)\left(\frac{1}{4}\frac{v}{(1+u)^2}\right) dv\, du \end{aligned}$$

5. A solid, R is determined by $4x+2y=1, 4x+2y=9, y=x$, and $y=6x$ and the density is $\rho = y^{-1}$. Find the total mass of R.

   Answer:

   Let $u=\frac{y}{x}, v=4x+2y$. Then solving these equations for x and y yields

   $$\left\{x = \frac{1}{2}\frac{v}{2+u}, y = \frac{1}{2}u\frac{v}{2+u}\right\}.$$

   Now

   $$\frac{\partial(x,y)}{\partial(u,v)} = \det\begin{pmatrix} -\frac{1}{2}\frac{v}{(2+u)^2} & \frac{1}{4+2u} \\ \frac{v}{(2+u)^2} & \frac{1}{2}\frac{u}{2+u} \end{pmatrix} = -\frac{1}{4}\frac{v}{(2+u)^2}.$$

   Also, $u \in [1,6]$ while $v \in [1,9]$. Therefore,

   $$\int_R \rho\, dV = \int_1^6 \int_1^9 \left(\frac{1}{2}u\frac{v}{2+u}\right)^{-1}\left|-\frac{1}{4}\frac{v}{(2+u)^2}\right| dv\, du = -4\ln 2 + 4\ln 3$$

6. Find the volume of the region, E, bounded by the ellipsoid, $\frac{1}{4}x^2+\frac{1}{9}y^2+\frac{1}{49}z^2=1$.

   Answer:

   Let $u=\frac{1}{2}x, v=\frac{1}{3}y, w=\frac{1}{7}z$. Then (u,v,w) is a point in the unit ball, B. Therefore,

$$\int_B \frac{\partial(x,y,z)}{\partial(u,v,w)}\,dV=\int_E dV.$$

   But $\frac{\partial(x,y,z)}{\partial(u,v,w)}=42$ and so the answer is

$$(volume\ of\ B)\times 42=\frac{4}{3}\pi 42=56\pi.$$

7. Here are three vectors. $(4,1,4)^T$, $(5,0,4)^T$, and$(3,1,5)^T$. These vectors determine a parallelepiped, R, which is occupied by a solid having density $\rho=x$. Find the mass of this solid.

   Answer:

   Let $\begin{pmatrix}4&5&3\\1&0&1\\4&4&5\end{pmatrix}\begin{pmatrix}u\\v\\w\end{pmatrix}=\begin{pmatrix}x\\y\\z\end{pmatrix}$. Then this maps the unit cube,

$$Q\equiv[0,1]\times[0,1]\times[0,1]$$

   onto R and

$$\frac{\partial(x,y,z)}{\partial(u,v,w)}=\left|\det\begin{pmatrix}4&5&3\\1&0&1\\4&4&5\end{pmatrix}\right|=|-9|=9$$

   so the mass is

$$\int_R x\,dV=\int_Q(4u+5v+3w)(9)\,dV$$

$$=\int_0^1\int_0^1\int_0^1(4u+5v+3w)(9)\,du\,dv\,dw=54$$

8. Here are three vectors. $(3,2,6)^T$, $(4,1,6)^T$, and $(2,2,7)^T$. These vectors determine a parallelepiped, R, which is occupied by a solid having density $\rho=y$. Find the mass of this solid.

   Answer:

   Let $\begin{pmatrix}3&4&2\\2&1&2\\6&6&7\end{pmatrix}\begin{pmatrix}u\\v\\w\end{pmatrix}=\begin{pmatrix}x\\y\\z\end{pmatrix}$. Then this maps the unit cube,

$$Q\equiv[0,1]\times[0,1]\times[0,1]$$

   onto R and

$$\frac{\partial(x,y,z)}{\partial(u,v,w)}=\left|\det\begin{pmatrix}3&4&2\\2&1&2\\6&6&7\end{pmatrix}\right|=|-11|=11$$

   and so the mass is

$$\int_R x\,dV = \int_Q (2u+v+2w)\,(11)\;dV$$

$$= \int_0^1\int_0^1\int_0^1 (2u+v+2w)\,(11)\;du\,dv\,dw = \frac{55}{2}.$$

9. Here are three vectors. $(2,2,4)^T, (3,1,4)^T$, and $(1,2,5)^T$. These vectors determine a parallelepiped, R, which is occupied by a solid having density $\rho = y+x$. Find the mass of this solid.

   Answer:

   Let $\begin{pmatrix} 2 & 3 & 1 \\ 2 & 1 & 2 \\ 4 & 4 & 5 \end{pmatrix}\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Then this maps the unit cube,

   $$Q \equiv [0,1]\times[0,1]\times[0,1]$$

   onto R and

   $$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \left|\det\begin{pmatrix} 2 & 3 & 1 \\ 2 & 1 & 2 \\ 4 & 4 & 5 \end{pmatrix}\right| = |-8| = 8$$

   and so the mass is $2u+3v+w$

   $$\int_R x\,dV = \int_Q (4u+4v+3w)\,(8)\;dV$$

   $$= \int_0^1\int_0^1\int_0^1 (4u+4v+3w)\,(8)\;du\,dv\,dw = 44.$$

10. Let $D = \left\{(x,y) : x^2+y^2 \le 25\right\}$. Find $\int_D e^{36x^2+36y^2}\,dx\,dy$.

    Answer:

    This is easy in polar coordinates. $x = r\cos\theta, y = r\sin\theta$. Thus $\frac{\partial(x,y)}{\partial(r,\theta)} = r$ and in terms of these new coordinates, the disk, D, is the rectangle,

    $$R = \{(r,\theta) \in [0,5]\times[0,2\pi]\}.$$

    Therefore,

    $$\int_D e^{36x^2+36y^2}\,dV = \int_R e^{36r^2} r\,dV =$$

    $$\int_0^5\int_0^{2\pi} e^{36r^2} r\,d\theta\,dr = \frac{1}{36}\pi\left(e^{900}-1\right).$$

    Note you wouldn't get very far without changing the variables in this.

11. Let $D = \left\{(x,y) : x^2+y^2 \le 9\right\}$. Find $\int_D \cos\left(36x^2+36y^2\right)\,dx\,dy$.

    Answer:

    This is easy in polar coordinates. $x = r\cos\theta, y = r\sin\theta$. Thus $\frac{\partial(x,y)}{\partial(r,\theta)} = r$ and in terms of these new coordinates, the disk, D, is the rectangle,

    $$R = \{(r,\theta) \in [0,3]\times[0,2\pi]\}.$$

Therefore,

$$\int_D \cos\left(36x^2+36y^2\right)\,dV = \int_R \cos\left(36r^2\right) r\,dV =$$

$$\int_0^3 \int_0^{2\pi} \cos\left(36r^2\right) r\,d\theta\,dr = \frac{1}{36}\left(\sin 324\right)\pi.$$

12. The ice cream in a sugar cone is described in spherical coordinates by $\rho \in [0,8]\,, \phi \in \left[0, \frac{1}{4}\pi\right], \theta \in [0, 2\pi]$. If the units are in centimeters, find the total volume in cubic centimeters of this ice cream.

    Answer:

    Remember that in spherical coordinates, the volume element is $\rho^2 \sin\phi\, dV$ and so the total volume of this is $\int_0^8 \int_0^{\frac{1}{4}\pi} \int_0^{2\pi} \rho^2 \sin\phi\, d\theta\, d\phi\, d\rho = -\frac{512}{3}\sqrt{2}\pi + \frac{1024}{3}\pi$.

13. Find the volume between $z = 5 - x^2 - y^2$ and $z = \sqrt{(x^2+y^2)}$.

    Answer:

    Use cylindrical coordinates. In terms of these coordinates the shape is

    $$h - r^2 \geq z \geq r, r \in \left[0, \frac{1}{2}\sqrt{21} - \frac{1}{2}\right], \theta \in [0, 2\pi]\,.$$

    Also, $\frac{\partial(x,y,z)}{\partial(r,\theta,z)} = r$. Therefore, the volume is

    $$\int_0^{2\pi} \int_0^{\frac{1}{2}\sqrt{21}-\frac{1}{2}} \int_0^{5-r^2} r\,dz\,dr\,d\theta = \frac{39}{4}\pi + \frac{1}{4}\pi\sqrt{21}$$

14. A ball of radius 12 is placed in a drill press and a hole of radius 4 is drilled out with the center of the hole a diameter of the ball. What is the volume of the material which remains?

    Answer:

    You know the formula for the volume of a sphere and so if you find out how much stuff is taken away, then it will be easy to find what is left. To find the volume of what is removed, it is easiest to use cylindrical coordinates. This volume is

    $$\int_0^4 \int_0^{2\pi} \int_{-\sqrt{(144-r^2)}}^{\sqrt{(144-r^2)}} r\,dz\,d\theta\,dr = -\frac{4096}{3}\sqrt{2}\pi + 2304\pi.$$

    Therefore, the volume of what remains is $\frac{4}{3}\pi\left(12\right)^3$ minus the above. Thus the volume of what remains is

    $$\frac{4096}{3}\sqrt{2}\pi.$$

15. A ball of radius 11 has density equal to $\sqrt{x^2+y^2+z^2}$ in rectangular coordinates. The top of this ball is sliced off by a plane of the form $z = 1$. What is the mass of what remains?

    Answer:

$$\int_0^{2\pi} \int_0^{\arcsin\left(\frac{2}{11}\sqrt{30}\right)} \int_0^{\sec\phi} \rho^3 \sin\phi\, d\rho\, d\phi\, d\theta + \int_0^{2\pi} \int_{\arcsin\left(\frac{2}{11}\sqrt{30}\right)}^{\pi} \int_0^{11} \rho^3 \sin\phi\, d\rho\, d\phi\, d\theta$$

$$= \frac{24\,623}{3}\pi$$

16. Find $\int_S \frac{y}{x}\,dV$ where S is described in polar coordinates as $1 \le r \le 2$ and $0 \le \theta \le \pi/4$.

    Answer:

    Use $x = r\cos\theta$ and $y = r\sin\theta$. Then the integral in polar coordinates is

$$\int_0^{\pi/4}\int_1^2 (r\tan\theta)\; dr\,d\theta = \frac{3}{4}\ln 2.$$

17. Find $\int_S \left(\left(\frac{y}{x}\right)^2 + 1\right)\,dV$ where S is given in polar coordinates as $1 \le r \le 2$ and $0 \le \theta \le \frac{1}{4}\pi$.

    Answer:

    Use $x = r\cos\theta$ and $y = r\sin\theta$. Then the integral in polar coordinates is

$$\int_0^{\frac{1}{4}\pi}\int_1^2 \left(1 + \tan^2\theta\right) r\,dr\,d\theta.$$

18. Use polar coordinates to evaluate the following integral. Here S is given in terms of the polar coordinates. $\int_S \sin\left(4x^2 + 4y^2\right)\,dV$ where $r \le 2$ and $0 \le \theta \le \frac{1}{6}\pi$.

    Answer:

$$\int_0^{\frac{1}{6}\pi}\int_0^2 \sin\left(4r^2\right) r\,dr\,d\theta.$$

19. Find $\int_S e^{2x^2+2y^2}\,dV$ where S is given in terms of the polar coordinates, $r \le 2$ and $0 \le \theta \le \frac{1}{3}\pi$.

    Answer:

    The integral is

$$\int_0^{\frac{1}{3}\pi}\int_0^2 re^{2r^2}\,dr\,d\theta = \frac{1}{12}\pi\left(e^8 - 1\right).$$

20. Compute the volume of a sphere of radius R using cylindrical coordinates.

    Answer:

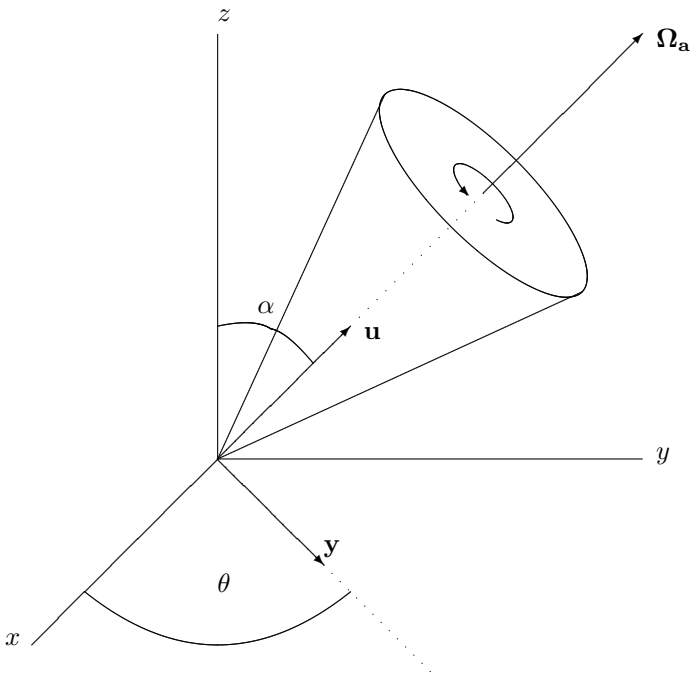
    Using cylindrical coordinates, the integral is $\int_0^{2\pi}\int_0^R \int_{-\sqrt{R^2-r^2}}^{\sqrt{R^2-r^2}} r\,dz\,dr\,d\theta = \frac{4}{3}\pi R^3$.

15.4 The Moment Of Inertia

In order to appreciate the importance of this concept, it is necessary to discuss its physical significance.

15.4.1 The Spinning Top

To begin with consider a spinning top as illustrated in the following picture.



For the purpose of this discussion, consider the top as a large number of point masses, m_i, located at the positions, $\mathbf{r}_i(t)$ for $i = 1, 2, \cdots, N$ and these masses are symmetrically arranged relative to the axis of the top. As the top spins, the axis of symmetry is observed to move around the z axis. This is called precession and you will see it occur whenever you spin a top. What is the speed of this precession? In other words, what is θ'? The following discussion follows one given in Sears and Zemansky [26].

Imagine a coordinate system which is fixed relative to the moving top. Thus in this coordinate system the points of the top are fixed. Let the standard unit vectors of the coordinate system moving with the top be denoted by $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$. From Theorem 8.4.2 on Page 163, there exists an angular velocity vector $\mathbf{\Omega}(t)$ such that if $\mathbf{u}(t)$ is the position vector of a point fixed in the top, $(\mathbf{u}(t) = u_1\mathbf{i}(t) + u_2\mathbf{j}(t) + u_3\mathbf{k}(t))$,

$$\mathbf{u}'(t) = \mathbf{\Omega}(t) \times \mathbf{u}(t).$$

The vector $\mathbf{\Omega}_a$ shown in the picture is the vector for which

$$\mathbf{r}_i'(t) \equiv \mathbf{\Omega}_a \times \mathbf{r}_i(t)$$

is the velocity of the i^{th} point mass due to rotation about the axis of the top. Thus $\mathbf{\Omega}(t) = \mathbf{\Omega}_a(t) + \mathbf{\Omega}_p(t)$ and it is assumed $\mathbf{\Omega}_p(t)$ is very small relative to $\mathbf{\Omega}_a$. In other words, it is assumed the axis of the top moves very slowly relative to the speed of the points in the top which are spinning very fast around the axis of the top. The angular momentum, $\mathbf{L}$ is defined by

$$\mathbf{L} \equiv \sum_{i=1}^{N} \mathbf{r}_i \times m_i\mathbf{v}_i \tag{15.6}$$

where $\mathbf{v}_i$ equals the velocity of the i^{th} point mass. Thus $\mathbf{v}_i = \boldsymbol{\Omega}(t) \times \mathbf{r}_i$ and from the above assumption, $\mathbf{v}_i$ may be taken equal to $\boldsymbol{\Omega}_a \times \mathbf{r}_i$. Therefore, $\mathbf{L}$ is essentially given by

$$\begin{aligned} \mathbf{L} &\equiv \sum_{i=1}^{N} m_i \mathbf{r}_i \times (\boldsymbol{\Omega}_a \times \mathbf{r}_i) \\ &= \sum_{i=1}^{N} m_i \left(|\mathbf{r}_i|^2 \boldsymbol{\Omega}_a - (\mathbf{r}_i \cdot \boldsymbol{\Omega}_a) \mathbf{r}_i \right). \end{aligned}$$

By symmetry of the top, this last expression equals a multiple of $\boldsymbol{\Omega}_a$. Thus $\mathbf{L}$ is parallel to $\boldsymbol{\Omega}_a$. Also,

$$\begin{aligned} \mathbf{L} \cdot \Omega_a &= \sum_{i=1}^{N} m_i \boldsymbol{\Omega}_a \cdot \mathbf{r}_i \times (\boldsymbol{\Omega}_a \times \mathbf{r}_i) \\ &= \sum_{i=1}^{N} m_i (\boldsymbol{\Omega}_a \times \mathbf{r}_i) \cdot (\boldsymbol{\Omega}_a \times \mathbf{r}_i) \\ &= \sum_{i=1}^{N} m_i |\boldsymbol{\Omega}_a \times \mathbf{r}_i|^2 = \sum_{i=1}^{N} m_i |\boldsymbol{\Omega}_a|^2 |\mathbf{r}_i|^2 \sin^2(\beta_i) \end{aligned}$$

where β_i denotes the angle between the position vector of the i^{th} point mass and the axis of the top. Since this expression is positive, this also shows $\mathbf{L}$ has the same direction as $\boldsymbol{\Omega}_a$. Let $\omega \equiv |\boldsymbol{\Omega}_a|$. Then the above expression is of the form

$$\mathbf{L} \cdot \boldsymbol{\Omega}_a = I\omega^2,$$

where

$$I \equiv \sum_{i=1}^{N} m_i |\mathbf{r}_i|^2 \sin^2(\beta_i).$$

Thus, to get I you take the mass of the i^{th} point mass, multiply it by the square of its distance to the axis of the top and add all these up. This is defined as the moment of inertia of the top about the axis of the top. Letting $\mathbf{u}$ denote a unit vector in the direction of the axis of the top, this implies

$$\mathbf{L} = I\omega\mathbf{u}. \tag{15.7}$$

Note the simple description of the angular momentum in terms of the moment of inertia. Referring to the above picture, define the vector, $\mathbf{y}$ to be the projection of the vector, $\mathbf{u}$ on the xy plane. Thus

$$\mathbf{y} = \mathbf{u} - (\mathbf{u} \cdot \mathbf{k})\mathbf{k}$$

and

$$(\mathbf{u} \cdot \mathbf{i}) = (\mathbf{y} \cdot \mathbf{i}) = \sin\alpha \cos\theta. \tag{15.8}$$

Now also from 15.6,

$$\begin{aligned} \frac{d\mathbf{L}}{dt} &= \sum_{i=1}^{N} m_i \overbrace{\mathbf{r}_i' \times \mathbf{v}_i}^{=0} + \mathbf{r}_i \times m_i \mathbf{v}_i' \\ &= \sum_{i=1}^{N} \mathbf{r}_i \times m_i \mathbf{v}_i' = -\sum_{i=1}^{N} \mathbf{r}_i \times m_i g\mathbf{k} \end{aligned}$$

where g is the acceleration of gravity. From 15.7, 15.8, and the above,

$$\begin{aligned}\frac{d\mathbf{L}}{dt}\cdot\mathbf{i} &= I\omega\left(\frac{d\mathbf{u}}{dt}\cdot\mathbf{i}\right) = I\omega\left(\frac{d\mathbf{y}}{dt}\cdot\mathbf{i}\right)\\ &= \left(-I\omega\sin\alpha\sin\theta\right)\theta' = -\sum_{i=1}^{N}\mathbf{r}_i\times m_i g\mathbf{k}\cdot\mathbf{i}\\ &= -\sum_{i=1}^{N} m_i g\mathbf{r}_i\cdot\mathbf{k}\times\mathbf{i} = -\sum_{i=1}^{N} m_i g\mathbf{r}_i\cdot\mathbf{j}. \qquad (15.9)\end{aligned}$$

To simplify this further, recall the following definition of the center of mass.

Definition 15.4.1 *Define the total mass, M by*

$$M = \sum_{i=1}^{N} m_i$$

and the center of mass, $\mathbf{r}_0$ by

$$\mathbf{r}_0 \equiv \frac{\sum_{i=1}^{N}\mathbf{r}_i m_i}{M}. \qquad (15.10)$$

In terms of the center of mass, the last expression equals

$$\begin{aligned}-Mg\mathbf{r}_0\cdot\mathbf{j} &= -Mg\left(\mathbf{r}_0 - (\mathbf{r}_0\cdot\mathbf{k})\,\mathbf{k} + (\mathbf{r}_0\cdot\mathbf{k})\,\mathbf{k}\right)\cdot\mathbf{j}\\ &= -Mg\left(\mathbf{r}_0 - (\mathbf{r}_0\cdot\mathbf{k})\,\mathbf{k}\right)\cdot\mathbf{j}\\ &= -Mg\left|\mathbf{r}_0 - (\mathbf{r}_0\cdot\mathbf{k})\,\mathbf{k}\right|\cos\theta\\ &= -Mg\left|\mathbf{r}_0\right|\sin\alpha\cos\left(\frac{\pi}{2}-\theta\right).\end{aligned}$$

Note that by symmetry, $\mathbf{r}_0(t)$ is on the axis of the top, is in the same direction as $\mathbf{L}, \mathbf{u}$, and $\boldsymbol{\Omega}_a$, and also $|\mathbf{r}_0|$ is independent of t. Therefore, from the second line of 15.9,

$$\left(-I\omega\sin\alpha\sin\theta\right)\theta' = -Mg\left|\mathbf{r}_0\right|\sin\alpha\sin\theta.$$

which shows

$$\theta' = \frac{Mg\left|\mathbf{r}_0\right|}{I\omega}. \qquad (15.11)$$

From 15.11, the angular velocity of precession does not depend on α in the picture. It also is slower when ω is large and I is large.

The above discussion is a considerable simplification of the problem of a spinning top obtained from an assumption that $\boldsymbol{\Omega}_a$ is approximately equal to $\boldsymbol{\Omega}$. It also leaves out all considerations of friction and the observation that the axis of symmetry wobbles. This is wobbling is called **nutation**. The full mathematical treatment of this problem involves the Euler angles and some fairly complicated differential equations obtained using techniques discussed in advanced physics classes. Lagrange studied these types of problems back in the 1700's.

15.4.2 Kinetic Energy

The next problem is that of understanding the total kinetic energy of a collection of moving point masses. Consider a possibly large number of point masses, m_i located at

the positions $\mathbf{r}_i$ for $i = 1, 2, \cdots, N$. Thus the velocity of the i^{th} point mass is $\mathbf{r}_i' = \mathbf{v}_i$. The kinetic energy of the mass m_i is defined by

$$\frac{1}{2} m_i \left|\mathbf{r}_i'\right|^2.$$

(This is a very good time to review the presentation on kinetic energy given on Page 170.) The total kinetic energy of the collection of masses is then

$$E = \sum_{i=1}^{N} \frac{1}{2} m_i \left|\mathbf{r}_i'\right|^2. \tag{15.12}$$

As these masses move about, so does the center of mass, $\mathbf{r}_0$. Thus $\mathbf{r}_0$ is a function of t just as the other $\mathbf{r}_i$. From 15.12 the total kinetic energy is

$$\begin{aligned} E &= \sum_{i=1}^{N} \frac{1}{2} m_i \left|\mathbf{r}_i' - \mathbf{r}_0' + \mathbf{r}_0'\right|^2 \\ &= \sum_{i=1}^{N} \frac{1}{2} m_i \left[\left|\mathbf{r}_i' - \mathbf{r}_0'\right|^2 + \left|\mathbf{r}_0'\right|^2 + 2\left(\mathbf{r}_i' - \mathbf{r}_0' \cdot \mathbf{r}_0'\right)\right]. \end{aligned} \tag{15.13}$$

Now

$$\begin{aligned} \sum_{i=1}^{N} m_i \left(\mathbf{r}_i' - \mathbf{r}_0' \cdot \mathbf{r}_0'\right) &= \left(\sum_{i=1}^{N} m_i \left(\mathbf{r}_i - \mathbf{r}_0\right)\right)' \cdot \mathbf{r}_0' \\ &= 0 \end{aligned}$$

because from 15.10

$$\begin{aligned} \sum_{i=1}^{N} m_i \left(\mathbf{r}_i - \mathbf{r}_0\right) &= \sum_{i=1}^{N} m_i \mathbf{r}_i - \sum_{i=1}^{N} m_i \mathbf{r}_0 \\ &= \sum_{i=1}^{N} m_i \mathbf{r}_i - \sum_{i=1}^{N} m_i \left(\frac{\sum_{i=1}^{N} \mathbf{r}_i m_i}{\sum_{i=1}^{N} m_i}\right) = \mathbf{0}. \end{aligned}$$

Let $M \equiv \sum_{i=1}^{N} m_i$ be the total mass. Then 15.13 reduces to

$$\begin{aligned} E &= \sum_{i=1}^{N} \frac{1}{2} m_i \left[\left|\mathbf{r}_i' - \mathbf{r}_0'\right|^2 + \left|\mathbf{r}_0'\right|^2\right] \\ &= \frac{1}{2} M \left|\mathbf{r}_0'\right|^2 + \sum_{i=1}^{N} \frac{1}{2} m_i \left|\mathbf{r}_i' - \mathbf{r}_0'\right|^2. \end{aligned} \tag{15.14}$$

The first term is just the kinetic energy of a point mass equal to the sum of all the masses involved, located at the center of mass of the system of masses while the second term represents kinetic energy which comes from the relative velocities of the masses taken with respect to the center of mass. It is this term which is considered more carefully in the case where the system of masses maintain distance between each other.

To illustrate the contrast between the case where the masses maintain a constant distance and one in which they don't, take a hard boiled egg and spin it and then take a raw egg and give it a spin. You will certainly feel a big difference in the way the two eggs respond. Incidentally, this is a good way to tell whether the egg has been hard boiled or is raw and can be used to prevent messiness which could occur if you think it is hard boiled and it really isn't.

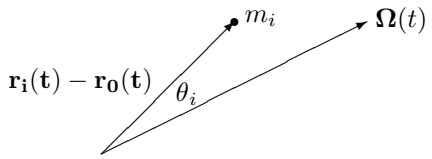
Now let $\mathbf{e}_1(t), \mathbf{e}_2(t)$, and $\mathbf{e}_3(t)$ be an orthonormal set of vectors which is fixed in the body undergoing rigid body motion. This means that $\mathbf{r}_i(t) - \mathbf{r}_0(t)$ has components which are constant in t with respect to the vectors, $\mathbf{e}_i(t)$. By Theorem 8.4.2 on Page 163 there exists a vector, $\mathbf{\Omega}(t)$ which does not depend on i such that

$$\mathbf{r}_i'(t) - \mathbf{r}_0'(t) = \mathbf{\Omega}(t) \times (\mathbf{r}_i(t) - \mathbf{r}_0(t)).$$

Now using this in 15.14,

$$\begin{aligned} E &= \frac{1}{2}M|\mathbf{r}_0'|^2 + \sum_{i=1}^{N}\frac{1}{2}m_i|\mathbf{\Omega}(t) \times (\mathbf{r}_i(t) - \mathbf{r}_0(t))|^2 \\ &= \frac{1}{2}M|\mathbf{r}_0'|^2 + \frac{1}{2}\left(\sum_{i=1}^{N} m_i|\mathbf{r}_i(t) - \mathbf{r}_0(t)|^2 \sin^2\theta_i\right)|\mathbf{\Omega}(t)|^2 \\ &= \frac{1}{2}M|\mathbf{r}_0'|^2 + \frac{1}{2}\left(\sum_{i=1}^{N} m_i|\mathbf{r}_i(0) - \mathbf{r}_0(0)|^2 \sin^2\theta_i\right)|\mathbf{\Omega}(t)|^2 \end{aligned}$$

where θ_i is the angle between $\mathbf{\Omega}(t)$ and the vector, $\mathbf{r}_i(t) - \mathbf{r}_0(t)$. Therefore, $|\mathbf{r}_i(t) - \mathbf{r}_0(t)|\sin\theta_i$ is the distance between the point mass, m_i located at $\mathbf{r}_i$ and a line through the center of mass, $\mathbf{r}_0$ with direction, $\mathbf{\Omega}$ as indicated in the following picture.



Thus the expression, $\sum_{i=1}^{N} m_i|\mathbf{r}_i(0) - \mathbf{r}_0(0)|^2 \sin^2\theta_i$ plays the role of a mass in the definition of kinetic energy except instead of the speed, substitute the angular speed, $|\mathbf{\Omega}(t)|$. It is this expression which is called the moment of inertia about the line whose direction is $\mathbf{\Omega}(t)$.

In both of these examples, the center of mass and the moment of inertia occurred in a natural way.

15.4.3 Finding The Moment Of Inertia And Center Of Mass

The methods used to evaluate multiple integrals make possible the determination of centers of mass and moments of inertia. In the case of a solid material rather than finitely many point masses, you replace the sums with integrals. The sums are essentially approximations of the integrals which result. This leads to the following definition.

Definition 15.4.2 *Let a solid occupy a region R such that its density is $\rho(\mathbf{x})$ for $\mathbf{x}$ a point in R and let L be a line. For $\mathbf{x} \in R$, let $l(\mathbf{x})$ be the distance from the point, $\mathbf{x}$ to the line L. The moment of inertia of the solid is defined as*

$$\int_R l(\mathbf{x})^2 \rho(\mathbf{x})\, dV.$$

Letting $(\overline{\mathbf{x}},\overline{\mathbf{y}},\overline{\mathbf{z}})$ denote the Cartesian coordinates of the center of mass,

$$\begin{aligned} \overline{\mathbf{x}} &= \frac{\int_R x\rho(\mathbf{x})\, dV}{\int_R \rho(\mathbf{x})\, dV},\ \overline{\mathbf{y}} = \frac{\int_R y\rho(\mathbf{x})\, dV}{\int_R \rho(\mathbf{x})\, dV}, \\ \overline{\mathbf{z}} &= \frac{\int_R z\rho(\mathbf{x})\, dV}{\int_R \rho(\mathbf{x})\, dV} \end{aligned}$$

where x, y, z are the Cartesian coordinates of the point at $\mathbf{x}$.

Example 15.4.3 *Let a solid occupy the three dimensional region R and suppose the density is ρ. What is the moment of inertia of this solid about the z axis? What is the center of mass?*

Here the little masses would be of the form $\rho(\mathbf{x})\,dV$ where $\mathbf{x}$ is a point of R. Therefore, the contribution of this mass to the moment of inertia would be

$$\left(x^2+y^2\right)\rho(\mathbf{x})\,dV$$

where the Cartesian coordinates of the point $\mathbf{x}$ are (x,y,z). Then summing these up as an integral, yields the following for the moment of inertia.

$$\int_R \left(x^2+y^2\right)\rho(\mathbf{x})\;dV. \tag{15.15}$$

To find the center of mass, sum up $\mathbf{r}\rho\,dV$ for the points in R and divide by the total mass. In Cartesian coordinates, where $\mathbf{r}=(x,y,z)$, this means to sum up vectors of the form $(x\rho\,dV, y\rho\,dV, z\rho\,dV)$ and divide by the total mass. Thus the Cartesian coordinates of the center of mass are

$$\left(\frac{\int_R x\rho\,dV}{\int_R \rho\,dV}, \frac{\int_R y\rho\,dV}{\int_R \rho\,dV}, \frac{\int_R z\rho\,dV}{\int_R \rho\,dV}\right) \equiv \frac{\int_R \mathbf{r}\rho\,dV}{\int_R \rho\,dV}.$$

Here is a specific example.

Example 15.4.4 *Find the moment of inertia about the z axis and center of mass of the solid which occupies the region, R defined by $9-\left(x^2+y^2\right)\geq z\geq 0$ if the density is $\rho(x,y,z)=\sqrt{x^2+y^2}$.*

This moment of inertia is $\int_R \left(x^2+y^2\right)\sqrt{x^2+y^2}\,dV$ and the easiest way to find this integral is to use cylindrical coordinates. Thus the answer is

$$\int_0^{2\pi}\int_0^3\int_0^{9-r^2} r^3 r\,dz\,dr\,d\theta = \frac{8748}{35}\pi.$$

To find the center of mass, note the x and y coordinates of the center of mass,

$$\frac{\int_R x\rho\,dV}{\int_R \rho\,dV}, \frac{\int_R y\rho\,dV}{\int_R \rho\,dV}$$

both equal zero because the above shape is symmetric about the z axis and ρ is also symmetric in its values. Thus $x\rho\,dV$ will cancel with $-x\rho\,dV$ and a similar conclusion will hold for the y coordinate. It only remains to find the z coordinate of the center of mass, $\overline{z}$. In polar coordinates, $\rho=r$ and so,

$$\overline{z}=\frac{\int_R z\rho\,dV}{\int_R \rho\,dV}=\frac{\int_0^{2\pi}\int_0^3\int_0^{9-r^2} zr^2\,dz\,dr\,d\theta}{\int_0^{2\pi}\int_0^3\int_0^{9-r^2} r^2\,dz\,dr\,d\theta}=\frac{18}{7}.$$

Thus the center of mass will be $\left(0,0,\frac{18}{7}\right)$.

15.5 Exercises

1. Let R denote the finite region bounded by $z = 4 - x^2 - y^2$ and the xy plane. Find z_c, the z coordinate of the center of mass if the density, σ is a constant.

2. Let R denote the finite region bounded by $z = 4 - x^2 - y^2$ and the xy plane. Find z_c, the z coordinate of the center of mass if the density, σ is equals $\sigma(x, y, z) = z$.

3. Find the mass and center of mass of the region between the surfaces $z = -y^2 + 8$ and $z = 2x^2 + y^2$ if the density equals $\sigma = 1$.

4. Find the mass and center of mass of the region between the surfaces $z = -y^2 + 8$ and $z = 2x^2 + y^2$ if the density equals $\sigma(x, y, z) = x^2$.

5. The two cylinders, $x^2 + y^2 = 4$ and $y^2 + z^2 = 4$ intersect in a region, R. Find the mass and center of mass if the density, σ, is given by $\sigma(x, y, z) = z^2$.

6. The two cylinders, $x^2 + y^2 = 4$ and $y^2 + z^2 = 4$ intersect in a region, R. Find the mass and center of mass if the density, σ, is given by $\sigma(x, y, z) = 4 + z$.

7. Find the mass and center of mass of the set, (x, y, z) such that $\frac{x^2}{4} + \frac{y^2}{9} + z^2 \leq 1$ if the density is $\sigma(x, y, z) = 4 + y + z$.

8. Let R denote the finite region bounded by $z = 9 - x^2 - y^2$ and the xy plane. Find the moment of inertia of this shape about the z axis given the density equals 1.

9. Let R denote the finite region bounded by $z = 9 - x^2 - y^2$ and the xy plane. Find the moment of inertia of this shape about the x axis given the density equals 1.

10. Let B be a solid ball of constant density and radius R. Find the moment of inertia about a line through a diameter of the ball. You should get $\frac{2}{5}R^2M$ where M equals the mass.

11. Let B be a solid ball of density, $\sigma = \rho$ where ρ is the distance to the center of the ball which has radius R. Find the moment of inertia about a line through a diameter of the ball. Write your answer in terms of the total mass and the radius as was done in the constant density case.

12. Let C be a solid cylinder of constant density and radius R. Find the moment of inertia about the axis of the cylinder

    You should get $\frac{1}{2}R^2M$ where M is the mass.

13. Let C be a solid cylinder of constant density and radius R and mass M and let B be a solid ball of radius R and mass M. The cylinder and the sphere are placed on the top of an inclined plane and allowed to roll to the bottom. Which one will arrive first and why?

14. Suppose a solid of mass M occupying the region, B has moment of inertia, I_l about a line, l which passes through the center of mass of M and let l_1 be another line parallel to l and at a distance of a from l. Then the parallel axis theorem states $I_{l_1} = I_l + a^2M$. Prove the parallel axis theorem. **Hint:** Choose axes such that the z axis is l and l_1 passes through the point $(a, 0)$ in the xy plane.

15. Using the parallel axis theorem find the moment of inertia of a solid ball of radius R and mass M about an axis located at a distance of a from the center of the ball. Your answer should be $Ma^2 + \frac{2}{5}MR^2$.

16. Consider all axes in computing the moment of inertia of a solid. Will the smallest possible moment of inertia always result from using an axis which goes through the center of mass?

17. Find the moment of inertia of a solid thin rod of length l, mass M, and constant density about an axis through the center of the rod perpendicular to the axis of the rod. You should get $\frac{1}{12}l^2M$.

18. Using the parallel axis theorem, find the moment of inertia of a solid thin rod of length l, mass M, and constant density about an axis through an end of the rod perpendicular to the axis of the rod. You should get $\frac{1}{3}l^2M$.

19. Let the angle between the z axis and the sides of a right circular cone be α. Also assume the height of this cone is h. Find the z coordinate of the center of mass of this cone in terms of α and h assuming the density is constant.

20. Let the angle between the z axis and the sides of a right circular cone be α. Also assume the height of this cone is h. Assuming the density is $\sigma = 1$, find the moment of inertia about the z axis in terms of α and h.

21. Let R denote the part of the solid ball, $x^2 + y^2 + z^2 \leq R^2$ which lies in the first octant. That is $x, y, z \geq 0$. Find the coordinates of the center of mass if the density is constant. Your answer for one of the coordinates for the center of mass should be $(3/8)\, R$.

22. Show that in general for $\mathbf{L}$ angular momentum,

$$\frac{d\mathbf{L}}{dt} = \mathbf{\Gamma}$$

where $\mathbf{\Gamma}$ is the total torque,

$$\mathbf{\Gamma} \equiv \sum \mathbf{r}_i \times \mathbf{F}_i$$

where $\mathbf{F}_i$ is the force on the i^{th} point mass.

The Integral On Two Dimensional Surfaces In $\mathbb{R}^3$

16.0.1 Outcomes

1. Find the area of a surface.

2. Define and compute integrals over surfaces given parametrically.

16.1 The Two Dimensional Area In $\mathbb{R}^3$

Consider the boundary of some three dimensional region such that a function, f is defined on this boundary. Imagine taking the value of this function at a point, multiplying this value by the area of an infinitesimal chunk of area located at this point and then adding these up. This is just the notion of the integral presented earlier only now there is a difference because this infinitesimal chunk of area should be considered as two dimensional even though it is in three dimensions. However, it is not really all that different from what was done earlier. It all depends on the following fundamental definition which is just a review of the fact presented earlier that the area of a parallelogram determined by two vectors in $\mathbb{R}^3$ is the norm of the cross product of the two vectors.

Definition 16.1.1 *Let $\mathbf{u}_1, \mathbf{u}_2$ be vectors in $\mathbb{R}^3$. The 2 dimensional parallelogram determined by these vectors will be denoted by $P(\mathbf{u}_1, \mathbf{u}_2)$ and it is defined as*

$$P(\mathbf{u}_1, \mathbf{u}_2) \equiv \left\{ \sum_{j=1}^{2} s_j \mathbf{u}_j : s_j \in [0,1] \right\}.$$

Then the area of this parallelogram is

$$\text{area } P(\mathbf{u}_1, \mathbf{u}_2) \equiv |\mathbf{u}_1 \times \mathbf{u}_2|.$$

Suppose then that $\mathbf{x} = \mathbf{f}(\mathbf{u})$ where $\mathbf{u} \in U$, a subset of $\mathbb{R}^2$ and $\mathbf{x}$ is a point in V, a subset of 3 dimensional space. Thus, letting the Cartesian coordinates of $\mathbf{x}$ be given by $\mathbf{x} = (x_1, x_2, x_3)^T$, each x_i being a function of $\mathbf{u}$, an infinitesimal rectangle located at $\mathbf{u}_0$ corresponds to an infinitesimal parallelogram located at $\mathbf{f}(\mathbf{u}_0)$ which is determined by the 2 vectors $\left\{ \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} du_i \right\}_{i=1}^{2}$, each of which is tangent to the surface defined by $\mathbf{x} = \mathbf{f}(\mathbf{u})$. (No sum on the repeated index.) From Definition 16.1.1, the volume of this

infinitesimal parallelepiped located at $\mathbf{f}(\mathbf{u}_0)$ is given by

$$\left|\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1}\,du_1 \times \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2}\,du_2\right| = \left|\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1} \times \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2}\right| du_1 du_2 \tag{16.1}$$

$$= |\mathbf{f}_{u_1} \times \mathbf{f}_{u_2}|\, du_1 du_2 \tag{16.2}$$

It might help to think of a lizard. The infinitesimal parallelepiped is like a very small scale on a lizard. This is the essence of the idea. To define the area of the lizard sum up areas of individual scales[1]. If the scales are small enough, their sum would serve as a good approximation to the area of the lizard.

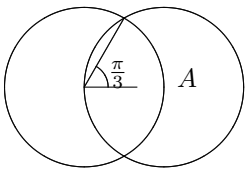
This motivates the following fundamental procedure which I hope is extremely familiar from the earlier material.

Procedure 16.1.2 *Suppose U is a subset of $\mathbb{R}^2$ and suppose $\mathbf{f}: U \to \mathbf{f}(U) \subseteq \mathbb{R}^3$ is a one to one and C^1 function. Then if $h: \mathbf{f}(U) \to \mathbb{R}$, define the 2 dimensional surface integral, $\int_{\mathbf{f}(U)} h(\mathbf{x})\, dA$ according to the following formula.*

$$\int_{\mathbf{f}(U)} h(\mathbf{x})\, dA \equiv \int_U h(\mathbf{f}(\mathbf{u}))\, |\mathbf{f}_{u_1}(\mathbf{u}) \times \mathbf{f}_{u_2}(\mathbf{u})|\, du_1 du_2.$$

Definition 16.1.3 *It is customary to write $|\mathbf{f}_{u_1}(\mathbf{u}) \times \mathbf{f}_{u_2}(\mathbf{u})| = \frac{\partial(x_1,x_2,x_3)}{\partial(u_1,u_2)}$ because this new notation generalizes to far more general situations for which the cross product is not defined. For example, one can consider three dimensional surfaces in $\mathbb{R}^8$.*

Example 16.1.4 *Find the area of the region labeled A in the following picture. The two circles are of radius 1, one has center $(0,0)$ and the other has center $(1,0)$.*



The circles bounding these disks are $x^2+y^2=1$ and $(x-1)^2+y^2 = x^2+y^2-2x+1 = 1$. Therefore, in polar coordinates these are of the form $r=1$ and $r = 2\cos\theta$.

[1]This beautiful lizard is a *Sceloporus magister*. It was photographed by C. Riley Nelson who is in the Zoology department at Brigham Young University © 2004 in Kane Co. Utah. The lizard is a little less than one foot in length.

The set A corresponds to the set U, in the (θ, r) plane determined by $\theta \in \left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$ and for each value of θ in this interval, r goes from 1 up to $2\cos\theta$. Therefore, the area of this region is of the form,

$$\int_U 1\, dV = \int_{-\pi/3}^{\pi/3} \int_1^{2\cos\theta} \frac{\partial(x_1, x_2, x_3)}{\partial(\theta, r)}\, dr\, d\theta.$$

It is necessary to find $\frac{\partial(x_1,x_2)}{\partial(\theta,r)}$. The mapping $\mathbf{f} : U \to \mathbb{R}^2$ takes the form $\mathbf{f}(\theta, r) = (r\cos\theta, r\sin\theta)^T$. Here $x_3 = 0$ and so

$$\frac{\partial(x_1, x_2, x_3)}{\partial(\theta, r)} = \left\| \begin{array}{ccc} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial x_1}{\partial \theta} & \frac{\partial x_2}{\partial \theta} & \frac{\partial x_3}{\partial \theta} \\ \frac{\partial x_1}{\partial r} & \frac{\partial x_2}{\partial r} & \frac{\partial x_3}{\partial r} \end{array} \right\| = \left\| \begin{array}{ccc} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -r\sin\theta & r\cos\theta & 0 \\ \cos\theta & \sin\theta & 0 \end{array} \right\| = r$$

Therefore, the area element is $r\, dr\, d\theta$. It follows the desired area is

$$\int_{-\pi/3}^{\pi/3} \int_1^{2\cos\theta} r\, dr\, d\theta = \frac{1}{2}\sqrt{3} + \frac{1}{3}\pi.$$

Example 16.1.5 *Consider the surface given by $z = x^2$ for $(x, y) \in [0, 1] \times [0, 1] = U$. Find the surface area of this surface.*

The first step in using the above is to write this surface in the form $\mathbf{x} = \mathbf{f}(\mathbf{u})$. This is easy to do if you let $\mathbf{u} = (x, y)$. Then $\mathbf{f}(x, y) = \left(x, y, x^2\right)$. If you like, let $x = u_1$ and $y = u_2$. What is $\frac{\partial(x_1,x_2,x_3)}{\partial(x,y)} = |\mathbf{f}_x \times \mathbf{f}_y|$?

$$\mathbf{f}_x = \begin{pmatrix} 1 \\ 0 \\ 2x \end{pmatrix}, \mathbf{f}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

and so

$$|\mathbf{f}_x \times \mathbf{f}_y| = \left| \begin{pmatrix} 1 \\ 0 \\ 2x \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right| = \sqrt{1 + 4x^2}$$

and so the area element is $\sqrt{1 + 4x^2}\, dx\, dy$ and the surface area is obtained by integrating the function, $h(\mathbf{x}) \equiv 1$. Therefore, this area is

$$\int_{\mathbf{f}(U)} dA = \int_0^1 \int_0^1 \sqrt{1 + 4x^2}\, dx\, dy = \frac{1}{2}\sqrt{5} - \frac{1}{4}\ln\left(-2 + \sqrt{5}\right)$$

which can be obtained by using the trig. substitution, $2x = \tan\theta$ on the inside integral.

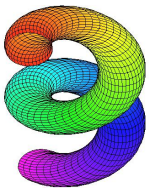
Note this all depends on being able to write the surface in the form, $\mathbf{x} = \mathbf{f}(\mathbf{u})$ for $\mathbf{u} \in U \subseteq \mathbb{R}^p$. Surfaces obtained in this form are called parametrically defined surfaces. These are best but sometimes you have some other description of a surface and in these cases things can get pretty intractable. For example, you might have a level surface of the form $3x^2 + 4y^4 + z^6 = 10$. In this case, you could solve for z using methods of algebra. Thus $z = \sqrt[6]{10 - 3x^2 - 4y^4}$ and a parametric description of part of this level surface is $\left(x, y, \sqrt[6]{10 - 3x^2 - 4y^4}\right)$ for $(x, y) \in U$ where $U = \left\{(x, y) : 3x^2 + 4y^4 \leq 10\right\}$. But what if the level surface was something like

$$\sin\left(x^2 + \ln\left(7 + y^2\sin x\right)\right) + \sin(zx)\, e^z = 11\sin(xyz)?$$

I really don't see how to use methods of algebra to solve for some variable in terms of the others. It isn't even clear to me whether there are any points $(x, y, z) \in \mathbb{R}^3$ satisfying

this particular relation. However, if a point satisfying this relation can be identified, the implicit function theorem from advanced calculus can usually be used to assert one of the variables is a function of the others, proving the existence of a parameterization at least locally. The problem is, this theorem doesn't give the answer in terms of known functions so this isn't much help. Finding a parametric description of a surface is a hard problem and there are no easy answers. This is a good example which illustrates the gulf between theory and practice.

Example 16.1.6 *Let* $U = [0,12] \times [0,2\pi]$ *and let* $\mathbf{f} : U \to \mathbb{R}^3$ *be given by* $\mathbf{f}(t,s) \equiv (2\cos t + \cos s, 2\sin t + \sin s, t)^T$. *Find a double integral for the surface area. A graph of this surface is drawn below.*



It looks like something you would use to make sausages[2]. Anyway,

$$\mathbf{f}_t = \begin{pmatrix} -2\sin t \\ 2\cos t \\ 1 \end{pmatrix}, \mathbf{f}_s = \begin{pmatrix} -\sin s \\ \cos s \\ 0 \end{pmatrix}$$

and

$$\mathbf{f}_t \times \mathbf{f}_s = \begin{pmatrix} -\cos s \\ -\sin s \\ -2\sin t\cos s + 2\cos t \sin s \end{pmatrix}$$

and so

$$\frac{\partial(x_1,x_2,x_3)}{\partial(t,s)} = |\mathbf{f}_t \times \mathbf{f}_s| = \sqrt{5 - 4\sin^2 t \sin^2 s - 8\sin t \sin s \cos t \cos s - 4\cos^2 t \cos^2 s}.$$

Therefore, the desired integral giving the area is

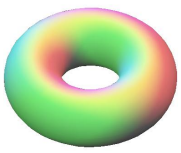
$$\int_0^{2\pi}\int_0^{12} \sqrt{5 - 4\sin^2 t \sin^2 s - 8\sin t \sin s \cos t \cos s - 4\cos^2 t \cos^2 s}\, dt\, ds.$$

If you really needed to find the number this equals, how would you go about finding it? This is an interesting question and there is no single right answer. You should think about this. Here is an example for which you will be able to find the integrals.

Example 16.1.7 *Let* $U = [0,2\pi] \times [0,2\pi]$ *and for* $(t,s) \in U$, *let*

$$\mathbf{f}(t,s) = (2\cos t + \cos t \cos s, -2\sin t - \sin t \cos s, \sin s)^T.$$

Find the area of $\mathbf{f}(U)$. *This is the surface of a donut shown below. The fancy name for this shape is a torus.*



[2]At Volwerth's in Hancock Michigan, they make excellent sausages and hot dogs. The best are made from "natural casings" which are the linings of intestines.

To find its area,

$$\mathbf{f}_t = \begin{pmatrix} -2\sin t - \sin t \cos s \\ -2\cos t - \cos t \cos s \\ 0 \end{pmatrix}, \mathbf{f}_s = \begin{pmatrix} -\cos t \sin s \\ \sin t \sin s \\ \cos s \end{pmatrix}$$

and so $|\mathbf{f}_t \times \mathbf{f}_s| = (\cos s + 2)$ so the area element is $(\cos s + 2)\, ds\, dt$ and the area is

$$\int_0^{2\pi} \int_0^{2\pi} (\cos s + 2)\, ds\, dt = 8\pi^2$$

Example 16.1.8 *Let $U = [0, 2\pi] \times [0, 2\pi]$ and for $(t, s) \in U$, let*

$$\mathbf{f}(t, s) = (2\cos t + \cos t \cos s, -2\sin t - \sin t \cos s, \sin s)^T.$$

Find

$$\int_{\mathbf{f}(U)} h\, dV$$

where $h(x, y, z) = x^2$.

Everything is the same as the preceding example except this time it is an integral of a function. The area element is $(\cos s + 2)\, ds\, dt$ and so the integral called for is

$$\int_{\mathbf{f}(U)} h\, dA = \int_0^{2\pi} \int_0^{2\pi} \left(\overbrace{2\cos t + \cos t \cos s}^{x \text{ on the surface}} \right)^2 (\cos s + 2)\, ds\, dt = 22\pi^2$$

16.1.1 Surfaces Of The Form $z = f(x, y)$

The special case where a surface is in the form $z = f(x, y), (x, y) \in U$, yields a simple formula which is used most often in this situation. You write the surface parametrically in the form $\mathbf{f}(x, y) = (x, y, f(x, y))^T$ such that $(x, y) \in U$. Then

$$\mathbf{f}_x = \begin{pmatrix} 1 \\ 0 \\ f_x \end{pmatrix}, \mathbf{f}_y = \begin{pmatrix} 0 \\ 1 \\ f_y \end{pmatrix}$$

and

$$|\mathbf{f}_x \times \mathbf{f}_y| = \sqrt{1 + f_y^2 + f_x^2}$$

so the area element is

$$\sqrt{1 + f_y^2 + f_x^2}\, dx\, dy.$$

When the surface of interest comes in this simple form, people generally use this area element directly rather than worrying about a parameterization and taking cross products.

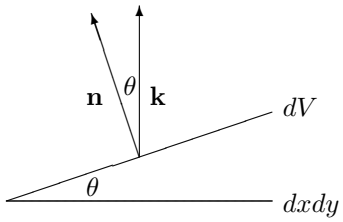
In the case where the surface is of the form $x = f(y, z)$ for $(y, z) \in U$, the area element is obtained similarly and is

$$\sqrt{1 + f_y^2 + f_z^2}\, dy\, dz.$$

I think you can guess what the area element is if $y = f(x, z)$.

There is also a simple geometric description of these area elements. Consider the surface $z = f(x, y)$. This is a level surface of the function of three variables $z - f(x, y)$. In fact the surface is simply $z - f(x, y) = 0$. Now consider the gradient of this function

of three variables. The gradient is perpendicular to the surface and the third component is positive in this case. This gradient is $(-f_x, -f_y, 1)$ and so the unit upward normal is just $\frac{1}{\sqrt{1+f_x^2+f_y^2}}(-f_x, -f_y, 1)$. Now consider the following picture.



In this picture, you are looking at a chunk of area on the surface seen on edge and so it seems reasonable to expect to have $dx\,dy = dV\cos\theta$. But it is easy to find $\cos\theta$ from the picture and the properties of the dot product.

$$\cos\theta = \frac{\mathbf{n}\cdot\mathbf{k}}{|\mathbf{n}|\,|\mathbf{k}|} = \frac{1}{\sqrt{1+f_x^2+f_y^2}}.$$

Therefore, $dA = \sqrt{1+f_x^2+f_y^2}\,dx\,dy$ as claimed. In this context, the surface involved is referred to as S because the vector valued function, $\mathbf{f}$ giving the parameterization will not have been identified.

Example 16.1.9 *Let $z = \sqrt{x^2+y^2}$ where $(x,y) \in U$ for $U = \{(x,y) : x^2+y^2 \le 4\}$ Find*

$$\int_S h\,dS$$

where $h(x,y,z) = x+z$ and S is the surface described as $\left(x, y, \sqrt{x^2+y^2}\right)$ for $(x,y) \in U$.

Here you can see directly the angle in the above picture is $\frac{\pi}{4}$ and so $dV = \sqrt{2}\,dx\,dy$. If you don't see this or if it is unclear, simply compute $\sqrt{1+f_x^2+f_y^2}$ and you will find it is $\sqrt{2}$. Therefore, using polar coordinates,

$$\begin{aligned}\int_S h\,dS &= \int_U \left(x+\sqrt{x^2+y^2}\right)\sqrt{2}\,dA \\ &= \sqrt{2}\int_0^{2\pi}\int_0^2 (r\cos\theta + r)\,r\,dr\,d\theta \\ &= \frac{16}{3}\sqrt{2}\pi.\end{aligned}$$

One other issue is worth mentioning. Suppose $\mathbf{f}_i : U_i \to \mathbb{R}^3$ where U_i are sets in $\mathbb{R}^2$ and suppose $\mathbf{f}_1(U_1)$ intersects $\mathbf{f}_2(U_2)$ along C where $C = \mathbf{h}(V)$ for $V \subseteq \mathbb{R}^1$. Then define integrals and areas over $\mathbf{f}_1(U_1) \cup \mathbf{f}_2(U_2)$ as follows.

$$\int_{\mathbf{f}_1(U_1)\cup\mathbf{f}_2(U_2)} g\,dA \equiv \int_{\mathbf{f}_1(U_1)} g\,dA + \int_{\mathbf{f}_2(U_2)} g\,dA.$$

Admittedly, the set C gets added in twice but this doesn't matter because its 2 dimensional volume equals zero and therefore, the integrals over this set will also be zero.

I have been purposely vague about precise mathematical conditions necessary for the above procedures. This is because the precise mathematical conditions which are usually cited are very technical and at the same time far too restrictive. The most

general conditions under which these sorts of procedures are valid include things like Lipschitz functions defined on very general sets. These are functions satisfying a Lipschitz condition of the form $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|$. For example, $y = |x|$ is Lipschitz continuous. However, this function does not have a derivative at every point. So it is with Lipschitz functions. However, it turns out these functions have derivatives at enough points to push everything through but this requires considerations involving the Lebesgue integral. Lipschitz functions are also not the most general kind of function for which the above is valid.

16.2 Exercises

1. Find a parameterization for the intersection of the planes $4x + 2y + 4z = 3$ and $6x - 2y = -1$.
2. Find a parameterization for the intersection of the plane $3x + y + z = 1$ and the circular cylinder $x^2 + y^2 = 1$.
3. Find a parameterization for the intersection of the plane $3x + 2y + 4z = 4$ and the elliptic cylinder $x^2 + 4z^2 = 16$.
4. Find a parameterization for the straight line joining $(1, 3, 1)$ and $(-2, 5, 3)$.
5. Find a parameterization for the intersection of the surfaces $4y + 3z = 3x^2 + 2$ and $3y + 2z = -x + 3$.
6. Find the area of S if S is the part of the circular cylinder $x^2 + y^2 = 4$ which lies between $z = 0$ and $z = 2 + y$.
7. Find the area of S if S is the part of the cone $x^2 + y^2 = 16z^2$ between $z = 0$ and $z = h$.
8. Parametrizing the cylinder $x^2 + y^2 = a^2$ by $x = a\cos v, y = a\sin v, z = u$, show that the area element is $dA = a\,du\,dv$
9. Find the area enclosed by the limacon $r = 2 + \cos\theta$.
10. Find the surface area of the paraboloid $z = h\left(1 - x^2 - y^2\right)$ between $z = 0$ and $z = h$.
11. Evaluate $\int_S (1 + x)\,dA$ where S is the part of the plane $4x + y + 3z = 12$ which is in the first octant.
12. Evaluate $\int_S (1 + x)\,dA$ where S is the part of the cylinder $x^2 + y^2 = 9$ between $z = 0$ and $z = h$.
13. Evaluate $\int_S (1 + x)\,dA$ where S is the hemisphere $x^2 + y^2 + z^2 = 4$ between $x = 0$ and $x = 2$.
14. For $(\theta, \alpha) \in [0, 2\pi] \times [0, 2\pi]$, let $\mathbf{f}(\theta, \alpha) \equiv (\cos\theta(4 + \cos\alpha), -\sin\theta(4 + \cos\alpha), \sin\alpha)^T$. Find the area of $\mathbf{f}([0, 2\pi] \times [0, 2\pi])$.
15. For $(\theta, \alpha) \in [0, 2\pi] \times [0, 2\pi]$, let $\mathbf{f}(\theta, \alpha) \equiv$
$$(\cos\theta(3 + 2\cos\alpha), -\sin\theta(3 + 2\cos\alpha), 2\sin\alpha)^T.$$
Also let $h(\mathbf{x}) = \cos\alpha$ where α is such that
$$\mathbf{x} = (\cos\theta(3 + 2\cos\alpha), -\sin\theta(3 + 2\cos\alpha), 2\sin\alpha)^T.$$
Find $\int_{\mathbf{f}([0,2\pi]\times[0,2\pi])} h\,dA$.

16. For $(\theta,\alpha)\in[0,2\pi]\times[0,2\pi]$, let $\mathbf{f}(\theta,\alpha)\equiv$

$$(\cos\theta(4+3\cos\alpha),-\sin\theta(4+3\cos\alpha),3\sin\alpha)^T.$$

Also let $h(\mathbf{x})=\cos^2\theta$ where θ is such that

$$\mathbf{x}=(\cos\theta(4+3\cos\alpha),-\sin\theta(4+3\cos\alpha),3\sin\alpha)^T.$$

Find $\int_{\mathbf{f}([0,2\pi]\times[0,2\pi])}h\,dA$.

17. For $(\theta,\alpha)\in[0,28]\times[0,2\pi]$, let $\mathbf{f}(\theta,\alpha)\equiv$

$$(\cos\theta(4+2\cos\alpha),-\sin\theta(4+2\cos\alpha),2\sin\alpha+\theta)^T.$$

Find a double integral which gives the area of $\mathbf{f}([0,28]\times[0,2\pi])$.

18. For $(\theta,\alpha)\in[0,2\pi]\times[0,2\pi]$, and β a fixed real number, define $\mathbf{f}(\theta,\alpha)\equiv$

$$\left(\begin{array}{c}\cos\theta(3+2\cos\alpha),-\cos\beta\sin\theta(3+2\cos\alpha)+\\ 2\sin\beta\sin\alpha,\sin\beta\sin\theta(3+2\cos\alpha)+2\cos\beta\sin\alpha\end{array}\right)^T.$$

Find a double integral which gives the area of $\mathbf{f}([0,2\pi]\times[0,2\pi])$.

19. In spherical coordinates, $\phi=c,\rho\in[0,R]$ determines a cone. Find the area of this cone without doing any work involving Jacobians and such.

16.3 Exercises With Answers

1. Find a parameterization for the intersection of the planes $x+y+2z=-3$ and $2x-y+z=-4$.

   Answer:

   $(x,y,z)=\left(-t-\frac{7}{3},-t-\frac{2}{3},t\right)$

2. Find a parameterization for the intersection of the plane $4x+2y+4z=0$ and the circular cylinder $x^2+y^2=16$.

   Answer:

   The cylinder is of the form $x=4\cos t, y=4\sin t$ and $z=z$. Therefore, from the equation of the plane, $16\cos t+8\sin t+4z=0$. Therefore, $z=-16\cos t-8\sin t$ and this shows the parameterization is of the form $(x,y,z)=(4\cos t,4\sin t,-16\cos t-8\sin t)$ where $t\in[0,2\pi]$.

3. Find a parameterization for the intersection of the plane $3x+2y+z=4$ and the elliptic cylinder $x^2+4z^2=1$.

   Answer:

   The cylinder is of the form $x=\cos t, 2z=\sin t$ and $y=y$. Therefore, from the equation of the plane, $3\cos t+2y+\frac{1}{2}\sin t=4$. Therefore, $y=2-\frac{3}{2}\cos t-\frac{1}{4}\sin t$ and this shows the parameterization is of the form $(x,y,z)=\left(\cos t,2-\frac{3}{2}\cos t-\frac{1}{4}\sin t,\frac{1}{2}\sin t\right)$ where $t\in[0,2\pi]$.

4. Find a parameterization for the straight line joining $(4,3,2)$ and $(1,7,6)$.

   Answer:

   $(x,y,z)=(4,3,2)+t(-3,4,4)=(4-3t,3+4t,2+4t)$ where $t\in[0,1]$.

5. Find a parameterization for the intersection of the surfaces $y + 3z = 4x^2 + 4$ and $4y + 4z = 2x + 4$.

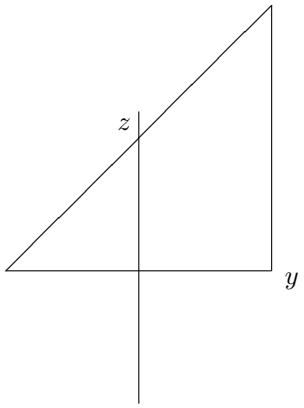
   Answer:

   This is an application of Cramer's rule. $y = -2x^2 - \frac{1}{2} + \frac{3}{4}x, z = -\frac{1}{4}x + \frac{3}{2} + 2x^2$. Therefore, the parameterization is $(x, y, z) = \left(t, -2t^2 - \frac{1}{2} + \frac{3}{4}t, -\frac{1}{4}t + \frac{3}{2} + 2t^2\right)$.

6. Find the area of S if S is the part of the circular cylinder $x^2 + y^2 = 16$ which lies between $z = 0$ and $z = 4 + y$.

   Answer:

   Use the parameterization, $x = 4\cos v, y = 4\sin v$ and $z = u$ with the parameter domain described as follows. The parameter, v goes from $-\frac{\pi}{2}$ to $\frac{3\pi}{2}$ and for each v in this interval, u should go from 0 to $4 + 4\sin v$. To see this observe that the cylinder has its axis parallel to the z axis and if you look at a side view of the surface you would see something like this:



   The positive x axis is coming out of the paper toward you in the above picture and the angle v is the usual angle measured from the positive x axis. Therefore, the area is just $A = \int_{-\pi/2}^{3\pi/2} \int_0^{4+4\sin v} 4\, du\, dv = 32\pi$.

7. Find the area of S if S is the part of the cone $x^2 + y^2 = 9z^2$ between $z = 0$ and $z = h$.

   Answer:

   When $z = h$, $x^2 + y^2 = 9h^2$ which is the boundary of a circle of radius ah. A parameterization of this surface is $x = u, y = v, z = \frac{1}{3}\sqrt{(u^2 + v^2)}$ where $(u, v) \in D$, a disk centered at the origin having radius ha. Therefore, the volume is just $\int_D \sqrt{1 + z_u^2 + z_v^2}\, dA = \int_{-ha}^{ha} \int_{-\sqrt{(9h^2-u^2)}}^{\sqrt{(9h^2-u^2)}} \frac{1}{3}\sqrt{10}\, dv\, du = 3\pi h^2\sqrt{10}$

8. Parametrizing the cylinder $x^2 + y^2 = 4$ by $x = 2\cos v, y = 2\sin v, z = u$, show that the area element is $dA = 2\, du\, dv$

   Answer:

   It is necessary to compute

$$|\mathbf{f}_u \times \mathbf{f}_v| = \left| \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \times \begin{pmatrix} -2\sin v \\ 2\cos v \\ 0 \end{pmatrix} \right| = 2.$$

   and so the area element is as described.

9. Find the area enclosed by the limacon $r = 2 + \cos\theta$.

   Answer:

   You can graph this region and you see it is sort of an oval shape and that $\theta \in [0, 2\pi]$ while r goes from 0 up to $2+\cos\theta$. Now $x = r\cos\theta$ and $y = r\sin\theta$ are the x and y coordinates corresponding to r and θ in the above parameter domain. Therefore, the area of the limacon equals $\int_P \left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| dr\, d\theta = \int_0^{2\pi}\int_0^{2+\cos\theta} r\, dr\, d\theta$ because the Jacobian equals r in this case. Therefore, the area equals $\int_0^{2\pi}\int_0^{2+\cos\theta} r\, dr\, d\theta = \frac{9}{2}\pi$.

10. Find the surface area of the paraboloid $z = h\left(1 - x^2 - y^2\right)$ between $z = 0$ and $z = h$.

    Answer:

    Let R denote the unit circle. Then the area of the surface above this circle would be $\int_R \sqrt{1 + 4x^2h^2 + 4y^2h^2}\, dA$. Changing to polar coordinates, this becomes

    $\int_0^{2\pi}\int_0^1 \left(\sqrt{1 + 4h^2r^2}\right) r\, dr\, d\theta = \frac{\pi}{6h^2}\left(\left(1 + 4h^2\right)^{3/2} - 1\right)$.

11. Evaluate $\int_S (1 + x)\, dA$ where S is the part of the plane $2x + 3y + 3z = 18$ which is in the first octant.

    Answer:

    $\int_0^6 \int_0^{6-\frac{2}{3}x} (1 + x)\, \frac{1}{3}\sqrt{22}\, dy\, dx = 28\sqrt{22}$

12. Evaluate $\int_S (1 + x)\, dA$ where S is the part of the cylinder $x^2 + y^2 = 16$ between $z = 0$ and $z = h$.

    Answer:

    Parametrize the cylinder as $x = 4\cos\theta$ and $y = 4\sin\theta$ while $z = t$ and the parameter domain is just $[0, 2\pi] \times [0, h]$. Then the integral to evaluate would be

    $$\int_0^{2\pi}\int_0^h (1 + 4\cos\theta)\, 4\, dt\, d\theta = 8h\pi.$$

    Note how $4\cos\theta$ was substituted for x and the area element is $4\, dt\, d\theta$.

13. Evaluate $\int_S (1 + x)\, dA$ where S is the hemisphere $x^2 + y^2 + z^2 = 16$ between $x = 0$ and $x = 4$.

    Answer:

    Parametrize the sphere as $x = 4\sin\phi\cos\theta, y = 4\sin\phi\sin\theta$, and $z = 4\cos\phi$ and consider the values of the parameters. Since it is referred to as a hemisphere and involves $x > 0$, $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ and $\phi \in [0, \pi]$. Then the area element is $\sqrt{a^4\sin\phi}\, d\theta\, d\phi$ and so the integral to evaluate is

    $$\int_0^{\pi}\int_{-\pi/2}^{\pi/2} (1 + 4\sin\phi\cos\theta)\, 16\sin\phi\, d\theta\, d\phi = 96\pi$$

14. For $(\theta, \alpha) \in [0, 2\pi] \times [0, 2\pi]$, let

    $$\mathbf{f}(\theta, \alpha) \equiv (\cos\theta\,(2 + \cos\alpha), -\sin\theta\,(2 + \cos\alpha), \sin\alpha)^T.$$

    Find the area of $\mathbf{f}([0, 2\pi] \times [0, 2\pi])$.

Answer:

$$\begin{aligned} |\mathbf{f}_\theta \times \mathbf{f}_\alpha| &= \left| \begin{pmatrix} -\sin(\theta)(2+\cos\alpha) \\ -\cos(\theta)(2+\cos\alpha) \\ 0 \end{pmatrix} \times \begin{pmatrix} -\cos\theta\sin\alpha \\ \sin\theta\sin\alpha \\ \cos\alpha \end{pmatrix} \right| \\ &= \left(4+4\cos\alpha+\cos^2\alpha\right)^{1/2} \end{aligned}$$

and so the area element is

$$\left(4+4\cos\alpha+\cos^2\alpha\right)^{1/2}\, d\theta\, d\alpha.$$

Therefore, the area is

$$\int_0^{2\pi}\int_0^{2\pi} \left(4+4\cos\alpha+\cos^2\alpha\right)^{1/2}\, d\theta\, d\alpha = \int_0^{2\pi}\int_0^{2\pi} (2+\cos\alpha)\, d\theta\, d\alpha = 8\pi^2.$$

15. For $(\theta,\alpha)\in[0,2\pi]\times[0,2\pi]$, let

$$\mathbf{f}(\theta,\alpha) \equiv (\cos\theta(4+2\cos\alpha), -\sin\theta(4+2\cos\alpha), 2\sin\alpha)^T.$$

Also let $h(\mathbf{x}) = \cos\alpha$ where α is such that

$$\mathbf{x} = (\cos\theta(4+2\cos\alpha), -\sin\theta(4+2\cos\alpha), 2\sin\alpha)^T.$$

Find $\int_{\mathbf{f}([0,2\pi]\times[0,2\pi])} h\, dA$.

Answer:

$$\begin{aligned} |\mathbf{f}_\theta \times \mathbf{f}_\alpha| &= \left| \begin{pmatrix} -\sin(\theta)(4+2\cos\alpha) \\ -\cos(\theta)(4+2\cos\alpha) \\ 0 \end{pmatrix} \times \begin{pmatrix} -2\cos\theta\sin\alpha \\ 2\sin\theta\sin\alpha \\ 2\cos\alpha \end{pmatrix} \right| \\ &= \left(64+64\cos\alpha+16\cos^2\alpha\right)^{1/2} \end{aligned}$$

and so the area element is

$$\left(64+64\cos\alpha+16\cos^2\alpha\right)^{1/2}\, d\theta\, d\alpha.$$

Therefore, the desired integral is

$$\begin{aligned} &\int_0^{2\pi}\int_0^{2\pi} (\cos\alpha)\left(64+64\cos\alpha+16\cos^2\alpha\right)^{1/2}\, d\theta\, d\alpha \\ &= \int_0^{2\pi}\int_0^{2\pi} (\cos\alpha)(8+4\cos\alpha)\, d\theta\, d\alpha = 8\pi^2 \end{aligned}$$

16. For $(\theta,\alpha)\in[0,2\pi]\times[0,2\pi]$, let

$$\mathbf{f}(\theta,\alpha) \equiv (\cos\theta(3+\cos\alpha), -\sin\theta(3+\cos\alpha), \sin\alpha)^T.$$

Also let $h(\mathbf{x}) = \cos^2\theta$ where θ is such that

$$\mathbf{x} = (\cos\theta(3+\cos\alpha), -\sin\theta(3+\cos\alpha), \sin\alpha)^T.$$

Find $\int_{\mathbf{f}([0,2\pi]\times[0,2\pi])} h\, dV$.

Answer:

The area element is

$$\left(9+6\cos\alpha+\cos^2\alpha\right)^{1/2}\,d\theta\,d\alpha.$$

Therefore, the desired integral is

$$\int_0^{2\pi}\int_0^{2\pi}\left(\cos^2\theta\right)\left(9+6\cos\alpha+\cos^2\alpha\right)^{1/2}\,d\theta\,d\alpha$$

$$=\int_0^{2\pi}\int_0^{2\pi}\left(\cos^2\theta\right)\left(3+\cos\alpha\right)\,d\theta\,d\alpha=6\pi^2$$

17. For $(\theta,\alpha)\in[0,25]\times[0,2\pi]$, let

$$\mathbf{f}(\theta,\alpha)\equiv\left(\cos\theta\left(4+2\cos\alpha\right),-\sin\theta\left(4+2\cos\alpha\right),2\sin\alpha+\theta\right)^T.$$

Find a double integral which gives the area of $\mathbf{f}\left([0,25]\times[0,2\pi]\right)$.

Answer:

In this case, the area element is

$$\left(68+64\cos\alpha+12\cos^2\alpha\right)^{1/2}\,d\theta\,d\alpha$$

and so the surface area is

$$\int_0^{2\pi}\int_0^{2\pi}\left(68+64\cos\alpha+12\cos^2\alpha\right)^{1/2}\,d\theta\,d\alpha.$$

18. For $(\theta,\alpha)\in[0,2\pi]\times[0,2\pi]$, and β a fixed real number, define $\mathbf{f}(\theta,\alpha)\equiv$

$$\left(\cos\theta\left(2+\cos\alpha\right),-\cos\beta\sin\theta\left(2+\cos\alpha\right)+\sin\beta\sin\alpha,\right.$$
$$\left.\sin\beta\sin\theta\left(2+\cos\alpha\right)+\cos\beta\sin\alpha\right)^T.$$

Find a double integral which gives the area of $\mathbf{f}\left([0,2\pi]\times[0,2\pi]\right)$.

Answer:

After many computations, the area element is $\left(4+4\cos\alpha+\cos^2\alpha\right)^{1/2}\,d\theta\,d\alpha$. Therefore, the area is $\int_0^{2\pi}\int_0^{2\pi}\left(2+\cos\alpha\right)\,d\theta\,d\alpha=8\pi^2$.

Calculus Of Vector Fields

17.0.1 Outcomes

1. Define and evaluate the divergence of a vector field in terms of Cartesian coordinates.
2. Define and evaluate the Curl of a vector field in Cartesian coordinates.
3. Discover vector identities involving the gradient, divergence, and curl.
4. Recall and verify the divergence theorem.
5. Apply the divergence theorem.

17.1 Divergence And Curl Of A Vector Field

Here the important concepts of divergence and curl are defined.

Definition 17.1.1 *Let $\mathbf{f} : U \to \mathbb{R}^p$ for $U \subseteq \mathbb{R}^p$ denote a vector field. A scalar valued function is called a **scalar field**. The function, $\mathbf{f}$ is called a C^k **vector field** if the function, $\mathbf{f}$ is a C^k function. For a C^1 vector field, as just described $\nabla \cdot \mathbf{f}(\mathbf{x}) \equiv \operatorname{div} \mathbf{f}(\mathbf{x})$ known as the **divergence**, is defined as*

$$\nabla \cdot \mathbf{f}(\mathbf{x}) \equiv \operatorname{div} \mathbf{f}(\mathbf{x}) \equiv \sum_{i=1}^{p} \frac{\partial f_i}{\partial x_i}(\mathbf{x}).$$

Using the repeated summation convention, this is often written as

$$f_{i,i}(\mathbf{x}) \equiv \partial_i f_i(\mathbf{x})$$

*where the comma indicates a partial derivative is being taken with respect to the i^{th} variable and ∂_i denotes differentiation with respect to the i^{th} variable. In words, the divergence is the sum of the i^{th} derivative of the i^{th} component function of $\mathbf{f}$ for all values of i. If $p = 3$, the **curl** of the vector field yields another vector field and it is defined as follows.*

$$(\operatorname{curl}(\mathbf{f})(\mathbf{x}))_i \equiv (\nabla \times \mathbf{f}(\mathbf{x}))_i \equiv \varepsilon_{ijk} \partial_j f_k(\mathbf{x})$$

where here ∂_j means the partial derivative with respect to x_j and the subscript of i in $(\operatorname{curl}(\mathbf{f})(\mathbf{x}))_i$ means the i^{th} Cartesian component of the vector, $\operatorname{curl}(\mathbf{f})(\mathbf{x})$. Thus the curl is evaluated by expanding the following determinant along the top row.

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f_1(x,y,z) & f_2(x,y,z) & f_3(x,y,z) \end{vmatrix}.$$

Note the similarity with the cross product. Sometimes the curl is called rot. (Short for rotation not decay.) Also

$$\nabla^2 f \equiv \nabla \cdot (\nabla f).$$

This last symbol is important enough that it is given a name, the ***Laplacian****.It is also denoted by* Δ*. Thus* $\nabla^2 f = \Delta f$*. In addition for* $\mathbf{f}$ *a vector field, the symbol* $\mathbf{f} \cdot \nabla$ *is defined as a "differential operator" in the following way.*

$$\mathbf{f} \cdot \nabla (\mathbf{g}) \equiv f_1 (\mathbf{x}) \frac{\partial \mathbf{g} (\mathbf{x})}{\partial x_1} + f_2 (\mathbf{x}) \frac{\partial \mathbf{g} (\mathbf{x})}{\partial x_2} + \cdots + f_p (\mathbf{x}) \frac{\partial \mathbf{g} (\mathbf{x})}{\partial x_p}.$$

Thus $\mathbf{f} \cdot \nabla$ *takes vector fields and makes them into new vector fields.*

This definition is in terms of a given coordinate system but later coordinate free definitions of the curl and div are presented. For now, everything is defined in terms of a given Cartesian coordinate system. The divergence and curl have profound physical significance and this will be discussed later. For now it is important to understand their definition in terms of coordinates. Be sure you understand that for $\mathbf{f}$ a vector field, $\operatorname{div} \mathbf{f}$ is a scalar field meaning it is a scalar valued function of three variables. For a scalar field, f, ∇f is a vector field described earlier on Page 282. For $\mathbf{f}$ a vector field having values in $\mathbb{R}^3$, $\operatorname{curl} \mathbf{f}$ is another vector field.

Example 17.1.2 *Let* $\mathbf{f} (\mathbf{x}) = xy\mathbf{i} + (z - y)\,\mathbf{j} + (\sin (x) + z)\,\mathbf{k}$*. Find* $\operatorname{div} \mathbf{f}$ *and* $\operatorname{curl} \mathbf{f}$*.*

First the divergence of $\mathbf{f}$ is

$$\frac{\partial (xy)}{\partial x} + \frac{\partial (z - y)}{\partial y} + \frac{\partial (\sin (x) + z)}{\partial z} = y + (-1) + 1 = y.$$

Now $\operatorname{curl} \mathbf{f}$ is obtained by evaluating

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ xy & z - y & \sin (x) + z \end{vmatrix} =$$

$$\mathbf{i} \left(\frac{\partial}{\partial y} (\sin (x) + z) - \frac{\partial}{\partial z} (z - y) \right) - \mathbf{j} \left(\frac{\partial}{\partial x} (\sin (x) + z) - \frac{\partial}{\partial z} (xy) \right) +$$

$$\mathbf{k} \left(\frac{\partial}{\partial x} (z - y) - \frac{\partial}{\partial y} (xy) \right) = -\mathbf{i} - \cos (x)\,\mathbf{j} - x\mathbf{k}.$$

17.1.1 Vector Identities

There are many interesting identities which relate the gradient, divergence and curl.

Theorem 17.1.3 *Assuming* $\mathbf{f}, \mathbf{g}$ *are a* C^2 *vector fields whenever necessary, the following identities are valid.*

1. $\nabla \cdot (\nabla \times \mathbf{f}) = 0$
2. $\nabla \times \nabla \phi = \mathbf{0}$
3. $\nabla \times (\nabla \times \mathbf{f}) = \nabla (\nabla \cdot \mathbf{f}) - \nabla^2 \mathbf{f}$ *where* $\nabla^2 \mathbf{f}$ *is a vector field whose* i^{th} *component is* $\nabla^2 f_i$*.*
4. $\nabla \cdot (\mathbf{f} \times \mathbf{g}) = \mathbf{g} \cdot (\nabla \times \mathbf{f}) - \mathbf{f} \cdot (\nabla \times \mathbf{g})$
5. $\nabla \times (\mathbf{f} \times \mathbf{g}) = (\nabla \cdot \mathbf{g})\,\mathbf{f} - (\nabla \cdot \mathbf{f})\,\mathbf{g} + (\mathbf{g} \cdot \nabla)\,\mathbf{f} - (\mathbf{f} \cdot \nabla)\,\mathbf{g}$

Proof: These are all easy to establish if you use the repeated index summation convention and the reduction identities discussed on Page 116.

$$\begin{aligned}
\nabla \cdot (\nabla \times \mathbf{f}) &= \partial_i (\nabla \times \mathbf{f})_i \\
&= \partial_i (\varepsilon_{ijk} \partial_j f_k) \\
&= \varepsilon_{ijk} \partial_i (\partial_j f_k) \\
&= \varepsilon_{jik} \partial_j (\partial_i f_k) \\
&= -\varepsilon_{ijk} \partial_j (\partial_i f_k) \\
&= -\varepsilon_{ijk} \partial_i (\partial_j f_k) \\
&= -\nabla \cdot (\nabla \times \mathbf{f}) .
\end{aligned}$$

This establishes the first formula. The second formula is done similarly. Now consider the third.

$$\begin{aligned}
(\nabla \times (\nabla \times \mathbf{f}))_i &= \varepsilon_{ijk} \partial_j (\nabla \times \mathbf{f})_k \\
&= \varepsilon_{ijk} \partial_j (\varepsilon_{krs} \partial_r f_s) \\
&= \overbrace{\varepsilon_{kij}}^{=\varepsilon_{ijk}} \varepsilon_{krs} \partial_j (\partial_r f_s) \\
&= (\delta_{ir}\delta_{js} - \delta_{is}\delta_{jr}) \partial_j (\partial_r f_s) \\
&= \partial_j (\partial_i f_j) - \partial_j (\partial_j f_i) \\
&= \partial_i (\partial_j f_j) - \partial_j (\partial_j f_i) \\
&= \left(\nabla (\nabla \cdot \mathbf{f}) - \nabla^2 \mathbf{f}\right)_i
\end{aligned}$$

This establishes the third identity.

Consider the fourth identity.

$$\begin{aligned}
\nabla \cdot (\mathbf{f} \times \mathbf{g}) &= \partial_i (\mathbf{f} \times \mathbf{g})_i \\
&= \partial_i \varepsilon_{ijk} f_j g_k \\
&= \varepsilon_{ijk} (\partial_i f_j) g_k + \varepsilon_{ijk} f_j (\partial_i g_k) \\
&= (\varepsilon_{kij} \partial_i f_j) g_k - (\varepsilon_{jik} \partial_i g_k) f_k \\
&= \nabla \times \mathbf{f} \cdot \mathbf{g} - \nabla \times \mathbf{g} \cdot \mathbf{f}.
\end{aligned}$$

This proves the fourth identity.

Consider the fifth.

$$\begin{aligned}
(\nabla \times (\mathbf{f} \times \mathbf{g}))_i &= \varepsilon_{ijk} \partial_j (\mathbf{f} \times \mathbf{g})_k \\
&= \varepsilon_{ijk} \partial_j \varepsilon_{krs} f_r g_s \\
&= \varepsilon_{kij} \varepsilon_{krs} \partial_j (f_r g_s) \\
&= (\delta_{ir}\delta_{js} - \delta_{is}\delta_{jr}) \partial_j (f_r g_s) \\
&= \partial_j (f_i g_j) - \partial_j (f_j g_i) \\
&= (\partial_j g_j) f_i + g_j \partial_j f_i - (\partial_j f_j) g_i - f_j (\partial_j g_i) \\
&= ((\nabla \cdot \mathbf{g}) \mathbf{f} + (\mathbf{g} \cdot \nabla) (\mathbf{f}) - (\nabla \cdot \mathbf{f}) \mathbf{g} - (\mathbf{f} \cdot \nabla) (\mathbf{g}))_i
\end{aligned}$$

and this establishes the fifth identity.

I think the important thing about the above is not that these identities can be proved and are valid as much as the method by which they were proved. The reduction identities on Page 116 were used to discover the identities. There is a difference between proving something someone tells you about and both discovering what should be proved and proving it. This notation and the reduction identity make the discovery of vector identities fairly routine and this is why these things are of great significance.

17.1.2 Vector Potentials

One of the above identities says $\nabla \cdot (\nabla \times \mathbf{f}) = 0$. Suppose now $\nabla \cdot \mathbf{g} = 0$. Does it follow that there exists $\mathbf{f}$ such that $\mathbf{g} = \nabla \times \mathbf{f}$? It turns out that this is usually the case and when such an $\mathbf{f}$ exists, it is called a **vector potential.** Here is one way to do it, assuming everything is defined so the following formulas make sense.

$$\mathbf{f}(x,y,z) = \left(\int_0^z g_2(x,y,t)\, dt, -\int_0^z g_1(x,y,t)\, dt + \int_0^x g_3(t,y,0)\, dt, 0 \right)^T. \quad (17.1)$$

In verifying this you need to use the following manipulation which will generally hold under reasonable conditions but which has not been carefully shown yet.

$$\frac{\partial}{\partial x} \int_a^b h(x,t)\, dt = \int_a^b \frac{\partial h}{\partial x}(x,t)\, dt. \quad (17.2)$$

The above formula seems plausible because the integral is a sort of a sum and the derivative of a sum is the sum of the derivatives. However, this sort of sloppy reasoning will get you into all sorts of trouble. The formula involves the interchange of two limit operations, the integral and the limit of a difference quotient. Such an interchange can only be accomplished through a theorem. The following gives the necessary result. This lemma is stated without proof.

Lemma 17.1.4 *Suppose h and $\frac{\partial h}{\partial x}$ are continuous on the rectangle $R = [c,d] \times [a,b]$. Then 17.2 holds.*

The second formula of Theorem 17.1.3 states $\nabla \times \nabla \phi = \mathbf{0}$. This suggests the following question: Suppose $\nabla \times \mathbf{f} = \mathbf{0}$, does it follow there exists ϕ, a scalar field such that $\nabla \phi = \mathbf{f}$? The answer to this is often yes and a theorem will be given and proved after the presentation of Stoke's theorem. This scalar field, ϕ, is called a **scalar potential** for $\mathbf{f}$.

17.1.3 The Weak Maximum Principle

There is also a fundamental result having great significance which involves ∇^2 called the maximum principle. This principle says that if $\nabla^2 u \geq 0$ on a bounded open set, U, then u achieves its maximum value on the boundary of U.

Theorem 17.1.5 *Let U be a bounded open set in $\mathbb{R}^n$ and suppose $u \in C^2(U) \cap C(\overline{U})$ such that $\nabla^2 u \geq 0$ in U. Then letting $\partial U = \overline{U} \setminus U$, it follows that* $\max\{u(\mathbf{x}) : \mathbf{x} \in \overline{U}\} = \max\{u(\mathbf{x}) : x \in \partial U\}$.

Proof: If this is not so, there exists $\mathbf{x}_0 \in U$ such that $u(\mathbf{x}_0) > \max\{u(\mathbf{x}) : x \in \partial U\} \equiv M$. Since U is bounded, there exists $\varepsilon > 0$ such that

$$u(\mathbf{x}_0) > \max\left\{ u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2 : \mathbf{x} \in \partial U \right\}.$$

Therefore, $u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2$ also has its maximum in U because for ε small enough,

$$u(\mathbf{x}_0) + \varepsilon |\mathbf{x}_0|^2 > u(\mathbf{x}_0) > \max\left\{ u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2 : \mathbf{x} \in \partial U \right\}$$

for all $\mathbf{x} \in \partial U$.

Now let $\mathbf{x}_1$ be the point in U at which $u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2$ achieves its maximum. As an exercise you should show that $\nabla^2(f+g) = \nabla^2 f + \nabla^2 g$ and therefore, $\nabla^2\left(u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2\right) = \nabla^2 u(\mathbf{x}) + 2n\varepsilon$. (Why?) Therefore,

$$0 \geq \nabla^2 u(\mathbf{x}_1) + 2n\varepsilon \geq 2n\varepsilon,$$

a contradiction. This proves the theorem.

17.2 Exercises

1. Find $\operatorname{div}\mathbf{f}$ and $\operatorname{curl}\mathbf{f}$ where $\mathbf{f}$ is

   (a) $\left(xyz, x^2+\ln(xy), \sin x^2+z\right)^T$

   (b) $(\sin x, \sin y, \sin z)^T$

   (c) $(f(x), g(y), h(z))^T$

   (d) $(x-2, y-3, z-6)^T$

   (e) $\left(y^2, 2xy, \cos z\right)^T$

   (f) $(f(y,z), g(x,z), h(y,z))^T$

2. Prove formula 2 of Theorem 17.1.3.

3. Show that if u and v are C^2 functions, then $\operatorname{curl}(u\nabla v) = \nabla u \times \nabla v$.

4. Simplify the expression $\mathbf{f}\times(\nabla\times\mathbf{g}) + \mathbf{g}\times(\nabla\times\mathbf{f}) + (\mathbf{f}\cdot\nabla)\mathbf{g} + (\mathbf{g}\cdot\nabla)\mathbf{f}$.

5. Simplify $\nabla\times(\mathbf{v}\times\mathbf{r})$ where $\mathbf{r} = (x,y,z)^T = x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$ and $\mathbf{v}$ is a constant vector.

6. Discover a formula which simplifies $\nabla\cdot(v\nabla u)$.

7. Verify that $\nabla\cdot(u\nabla v) - \nabla\cdot(v\nabla u) = u\nabla^2 v - v\nabla^2 u$.

8. Verify that $\nabla^2(uv) = v\nabla^2 u + 2(\nabla u\cdot\nabla v) + u\nabla^2 v$.

9. Functions, u, which satisfy $\nabla^2 u = 0$ are called harmonic functions. Show the following functions are harmonic where ever they are defined.

   (a) $2xy$

   (b) $x^2 - y^2$

   (c) $\sin x \cosh y$

   (d) $\ln\left(x^2+y^2\right)$

   (e) $1/\sqrt{x^2+y^2+z^2}$

10. Verify the formula given in 17.1 is a vector potential for $\mathbf{g}$ assuming that $\operatorname{div}\mathbf{g} = 0$.

11. Show that if $\nabla^2 u_k = 0$ for each $k = 1, 2, \cdots, m$, and c_k is a constant, then $\nabla^2\left(\sum_{k=1}^m c_k u_k\right) = 0$ also.

12. In Theorem 17.1.5 why is $\nabla^2\left(\varepsilon|\mathbf{x}|^2\right) = 2n\varepsilon$?

13. Using Theorem 17.1.5 prove the following: Let $f \in C(\partial U)$ (f is continuous on ∂U.) where U is a bounded open set. Then there exists at most one solution, $u \in C^2(U)\cap C\left(\overline{U}\right)$ and $\nabla^2 u = 0$ in U with $u = f$ on ∂U. **Hint:** Suppose there are two solutions, u_i, $i = 1, 2$ and let $w = u_1 - u_2$. Then use the maximum principle.

14. Suppose $\mathbf{B}$ is a vector field and $\nabla\times\mathbf{A} = \mathbf{B}$. Thus $\mathbf{A}$ is a vector potential for $\mathbf{B}$. Show that $\mathbf{A}+\nabla\phi$ is also a vector potential for $\mathbf{B}$. Here ϕ is just a C^2 scalar field. Thus the vector potential is not unique.

17.3 The Divergence Theorem

The divergence theorem relates an integral over a set to one on the boundary of the set. It is also called Gauss's theorem.

Definition 17.3.1 *A subset, V of $\mathbb{R}^3$ is called cylindrical in the x direction if it is of the form*

$$V = \{(x,y,z) : \phi(y,z) \leq x \leq \psi(y,z) \text{ for } (y,z) \in D\}$$

where D is a subset of the yz plane. V is cylindrical in the z direction if

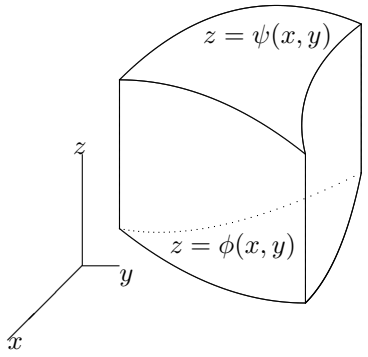
$$V = \{(x,y,z) : \phi(x,y) \leq z \leq \psi(x,y) \text{ for } (x,y) \in D\}$$

where D is a subset of the xy plane, and V is cylindrical in the y direction if

$$V = \{(x,y,z) : \phi(x,z) \leq y \leq \psi(x,z) \text{ for } (x,z) \in D\}$$

where D is a subset of the xz plane. If V is cylindrical in the z direction, denote by ∂V the boundary of V defined to be the points of the form $(x,y,\phi(x,y)), (x,y,\psi(x,y))$ for $(x,y) \in D$, along with points of the form (x,y,z) where $(x,y) \in \partial D$ and $\phi(x,y) \leq z \leq \psi(x,y)$. Points on ∂D are defined to be those for which every open ball contains points which are in D as well as points which are not in D. A similar definition holds for ∂V in the case that V is cylindrical in one of the other directions.

The following picture illustrates the above definition in the case of V cylindrical in the z direction.



Of course, many three dimensional sets are cylindrical in each of the coordinate directions. For example, a ball or a rectangle or a tetrahedron are all cylindrical in each direction. The following lemma allows the exchange of the volume integral of a partial derivative for an area integral in which the derivative is replaced with multiplication by an appropriate component of the unit exterior normal.

Lemma 17.3.2 *Suppose V is cylindrical in the z direction and that ϕ and ψ are the functions in the above definition. Assume ϕ and ψ are C^1 functions and suppose F is a C^1 function defined on V. Also, let $\mathbf{n} = (n_x, n_y, n_z)$ be the unit exterior normal to ∂V. Then*

$$\int_V \frac{\partial F}{\partial z}(x,y,z)\, dV = \int_{\partial V} F n_z \, dA.$$

Proof: From the fundamental theorem of calculus,

$$\begin{aligned} \int_V \frac{\partial F}{\partial z}(x,y,z)\, dV &= \int_D \int_{\phi(x,y)}^{\psi(x,y)} \frac{\partial F}{\partial z}(x,y,z)\, dz\, dx\, dy \qquad (17.3) \\ &= \int_D \left[F(x,y,\psi(x,y)) - F(x,y,\phi(x,y))\right] dx\, dy \end{aligned}$$

Now the unit exterior normal on the top of V, the surface $(x, y, \psi(x, y))$ is

$$\frac{1}{\sqrt{\psi_x^2 + \psi_y^2 + 1}} \left(-\psi_x, -\psi_y, 1\right).$$

This follows from the observation that the top surface is the level surface, $z - \psi(x, y) = 0$ and so the gradient of this function of three variables is perpendicular to the level surface. It points in the correct direction because the z component is positive. Therefore, on the top surface,

$$n_z = \frac{1}{\sqrt{\psi_x^2 + \psi_y^2 + 1}}$$

Similarly, the unit normal to the surface on the bottom is

$$\frac{1}{\sqrt{\phi_x^2 + \phi_y^2 + 1}} \left(\phi_x, \phi_y, -1\right)$$

and so on the bottom surface,

$$n_z = \frac{-1}{\sqrt{\phi_x^2 + \phi_y^2 + 1}}$$

Note that here the z component is negative because since it is the outer normal it must point down. On the lateral surface, the one where $(x, y) \in \partial D$ and $z \in [\phi(x, y), \psi(x, y)]$, $n_z = 0$.

The area element on the top surface is $dA = \sqrt{\psi_x^2 + \psi_y^2 + 1}\, dx\, dy$ while the area element on the bottom surface is $\sqrt{\phi_x^2 + \phi_y^2 + 1}\, dx\, dy$. Therefore, the last expression in 17.3 is of the form,

$$\int_D F(x, y, \psi(x, y)) \overbrace{\frac{1}{\sqrt{\psi_x^2 + \psi_y^2 + 1}}}^{n_z} \overbrace{\sqrt{\psi_x^2 + \psi_y^2 + 1}\, dx\, dy}^{dA} +$$

$$\int_D F(x, y, \phi(x, y)) \overbrace{\left(\frac{-1}{\sqrt{\phi_x^2 + \phi_y^2 + 1}}\right)}^{n_z} \overbrace{\sqrt{\phi_x^2 + \phi_y^2 + 1}\, dx\, dy}^{dA}$$

$$+ \int_{\text{Lateral surface}} F n_z\, dA,$$

the last term equaling zero because on the lateral surface, $n_z = 0$. Therefore, this reduces to $\int_{\partial V} F n_z\, dA$ as claimed.

The following corollary is entirely similar to the above.

Corollary 17.3.3 *If V is cylindrical in the y direction, then*

$$\int_V \frac{\partial F}{\partial y}\, dV = \int_{\partial V} F n_y\, dA$$

and if V is cylindrical in the x direction, then

$$\int_V \frac{\partial F}{\partial x}\, dV = \int_{\partial V} F n_x\, dA$$

With this corollary, here is a proof of the divergence theorem.

Theorem 17.3.4 *Let V be cylindrical in each of the coordinate directions and let $\mathbf{F}$ be a C^1 vector field defined on V. Then*

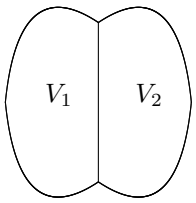
$$\int_V \nabla \cdot \mathbf{F}\, dV = \int_{\partial V} \mathbf{F} \cdot \mathbf{n}\, dA.$$

Proof: From the above lemma and corollary,

$$\begin{aligned} \int_V \nabla \cdot \mathbf{F}\, dV &= \int_V \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial y}\, dV \\ &= \int_{\partial V} (F_1 n_x + F_2 n_y + F_3 n_z)\, dA \\ &= \int_{\partial V} \mathbf{F} \cdot \mathbf{n}\, dA. \end{aligned}$$

This proves the theorem.

The divergence theorem holds for much more general regions than this. Suppose for example you have a complicated region which is the union of finitely many disjoint regions of the sort just described which are cylindrical in each of the coordinate directions. Then the volume integral over the union of these would equal the sum of the integrals over the disjoint regions. If the boundaries of two of these regions intersect, then the area integrals will cancel out on the intersection because the unit exterior normals will point in opposite directions. Therefore, the sum of the integrals over the boundaries of these disjoint regions will reduce to an integral over the boundary of the union of these. Hence the divergence theorem will continue to hold. For example, consider the following picture. If the divergence theorem holds for each V_i in the following picture, then it holds for the union of these two.



General formulations of the divergence theorem involve Hausdorff measures and the Lebesgue integral, a better integral than the old fashioned Riemann integral which has been obsolete now for almost 100 years. When all is said and done, one finds that the conclusion of the divergence theorem is usually true and the theorem can be used with confidence.

Example 17.3.5 *Let $V = [0,1] \times [0,1] \times [0,1]$. That is, V is the cube in the first octant having the lower left corner at $(0,0,0)$ and the sides of length 1. Let $\mathbf{F}(x,y,z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. Find the flux integral in which $\mathbf{n}$ is the unit exterior normal.*

$$\int_{\partial V} \mathbf{F} \cdot \mathbf{n} dS$$

You can certainly inflict much suffering on yourself by breaking the surface up into 6 pieces corresponding to the 6 sides of the cube, finding a parameterization for each face and adding up the appropriate flux integrals. For example, $\mathbf{n} = \mathbf{k}$ on the top face and $\mathbf{n} = -\mathbf{k}$ on the bottom face. On the top face, a parameterization is $(x, y, 1) : (x, y) \in$

$[0,1]\times[0,1]$. The area element is just $dxdy$. It isn't really all that hard to do it this way but it is much easier to use the divergence theorem. The above integral equals

$$\int_V \operatorname{div}(\mathbf{F})\,dV = \int_V 3dV = 3.$$

Example 17.3.6 *This time, let V be the unit ball, $\left\{(x,y,z) : x^2+y^2+z^2 \leq 1\right\}$ and let $\mathbf{F}(x,y,z) = x^2\mathbf{i} + y\mathbf{j} + (z-1)\mathbf{k}$. Find*

$$\int_{\partial V} \mathbf{F}\cdot\mathbf{n}dS.$$

As in the above you could do this by brute force. A parameterization of the ∂V is obtained as

$$x = \sin\phi\cos\theta,\ y = \sin\phi\sin\theta,\ z = \cos\phi$$

where $(\phi,\theta)\in(0,\pi)\times(0,2\pi]$. Now this does not include all the ball but it includes all but the point at the top and at the bottom. As far as the flux integral is concerned these points contribute nothing to the integral so you can neglect them. Then you can grind away and get the flux integral which is desired. However, it is so much easier to use the divergence theorem! Using spherical coordinates,

$$\begin{aligned}\int_{\partial V}\mathbf{F}\cdot\mathbf{n}dS &= \int_V \operatorname{div}(\mathbf{F})\,dV = \int_V (2x+1+1)\,dV \\ &= \int_0^{\pi}\int_0^{2\pi}\int_0^1 (2+2\rho\sin(\phi)\cos\theta)\,\rho^2\sin(\phi)\,d\rho d\theta d\phi = \frac{8}{3}\pi\end{aligned}$$

Example 17.3.7 *Suppose V is an open set in $\mathbb{R}^3$ for which the divergence theorem holds. Let $\mathbf{F}(x,y,z) = x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$. Then show*

$$\int_{\partial V}\mathbf{F}\cdot\mathbf{n}dS = 3\times \textit{volume(V)}.$$

This follows from the divergenc theorem.

$$\int_{\partial V}\mathbf{F}\cdot\mathbf{n}dS = \int_V \operatorname{div}(\mathbf{F})\,dV = 3\int_V dV = 3\times\operatorname{volume}(V).$$

The message of the divergence theorem is the relation between the volume integral and an area integral. This is the exciting thing about this marvelous theorem. It is not its utility as a method for evaluations of boring problems. This will be shown in the examples of its use which follow.

17.3.1 Coordinate Free Concept Of Divergence

The divergence theorem also makes possible a coordinate free definition of the divergence.

Theorem 17.3.8 *Let $B(\mathbf{x},\delta)$ be the ball centered at $\mathbf{x}$ having radius δ and let $\mathbf{F}$ be a C^1 vector field. Then letting $v(B(\mathbf{x},\delta))$ denote the volume of $B(\mathbf{x},\delta)$ given by*

$$\int_{B(\mathbf{x},\delta)} dV,$$

it follows

$$\operatorname{div}\mathbf{F}(\mathbf{x}) = \lim_{\delta\to 0+}\frac{1}{v(B(\mathbf{x},\delta))}\int_{\partial B(\mathbf{x},\delta)}\mathbf{F}\cdot\mathbf{n}\,dA. \tag{17.4}$$

Proof: The divergence theorem holds for balls because they are cylindrical in every direction. Therefore,

$$\frac{1}{v\left(B\left(\mathbf{x},\delta\right)\right)}\int_{\partial B(\mathbf{x},\delta)}\mathbf{F}\cdot\mathbf{n}\,dA=\frac{1}{v\left(B\left(\mathbf{x},\delta\right)\right)}\int_{B(\mathbf{x},\delta)}\operatorname{div}\mathbf{F}\left(\mathbf{y}\right)\,dV.$$

Therefore, since $\operatorname{div}\mathbf{F}\left(\mathbf{x}\right)$ is a constant,

$$\begin{aligned}&\left|\operatorname{div}\mathbf{F}\left(\mathbf{x}\right)-\frac{1}{v\left(B\left(\mathbf{x},\delta\right)\right)}\int_{\partial B(\mathbf{x},\delta)}\mathbf{F}\cdot\mathbf{n}\,dA\right|\\ =&\left|\operatorname{div}\mathbf{F}\left(\mathbf{x}\right)-\frac{1}{v\left(B\left(\mathbf{x},\delta\right)\right)}\int_{B(\mathbf{x},\delta)}\operatorname{div}\mathbf{F}\left(\mathbf{y}\right)\,dV\right|\\ =&\left|\frac{1}{v\left(B\left(\mathbf{x},\delta\right)\right)}\int_{B(\mathbf{x},\delta)}\left(\operatorname{div}\mathbf{F}\left(\mathbf{x}\right)-\operatorname{div}\mathbf{F}\left(\mathbf{y}\right)\right)\,dV\right|\\ \leq&\frac{1}{v\left(B\left(\mathbf{x},\delta\right)\right)}\int_{B(\mathbf{x},\delta)}\left|\operatorname{div}\mathbf{F}\left(\mathbf{x}\right)-\operatorname{div}\mathbf{F}\left(\mathbf{y}\right)\right|\,dV\\ \leq&\frac{1}{v\left(B\left(\mathbf{x},\delta\right)\right)}\int_{B(\mathbf{x},\delta)}\frac{\varepsilon}{2}\,dV<\varepsilon\end{aligned}$$

whenever ε is small enough due to the continuity of $\operatorname{div}\mathbf{F}$. Since ε is arbitrary, this shows 17.4.

How is this definition independent of coordinates? It only involves geometrical notions of volume and dot product. This is why. Imagine rotating the coordinate axes, keeping all distances the same and expressing everything in terms of the new coordinates. The divergence would still have the same value because of this theorem.

17.4 Some Applications Of The Divergence Theorem

17.4.1 Hydrostatic Pressure

Imagine a fluid which does not move which is acted on by an acceleration, $\mathbf{g}$. Of course the acceleration is usually the acceleration of gravity. Also let the density of the fluid be ρ, a function of position. What can be said about the pressure, p, in the fluid? Let $B\left(\mathbf{x},\varepsilon\right)$ be a small ball centered at the point, $\mathbf{x}$. Then the force the fluid exerts on this ball would equal

$$-\int_{\partial B(\mathbf{x},\varepsilon)}p\mathbf{n}\,dA.$$

Here $\mathbf{n}$ is the unit exterior normal at a small piece of $\partial B\left(\mathbf{x},\varepsilon\right)$ having area dA. By the divergence theorem, (see Problem 1 on Page 390) this integral equals

$$-\int_{B(\mathbf{x},\varepsilon)}\nabla p\,dV.$$

Also the force acting on this small ball of fluid is

$$\int_{B(\mathbf{x},\varepsilon)}\rho\mathbf{g}\,dV.$$

Since it is given that the fluid does not move, the sum of these forces must equal zero. Thus

$$\int_{B(\mathbf{x},\varepsilon)} \rho\mathbf{g}\,dV = \int_{B(\mathbf{x},\varepsilon)} \nabla p\,dV.$$

Since this must hold for any ball in the fluid of any radius, it must be that

$$\nabla p = \rho\mathbf{g}. \tag{17.5}$$

It turns out that the pressure in a lake at depth z is equal to $62.5z$. This is easy to see from 17.5. In this case, $\mathbf{g} = g\mathbf{k}$ where $g = 32$ feet/sec^2. The weight of a cubic foot of water is 62.5 pounds. Therefore, the mass in slugs of this water is $62.5/32$. Since it is a cubic foot, this is also the density of the water in slugs per cubic foot. Also, it is normally assumed that water is incompressible[1]. Therefore, this is the mass of water at any depth. Therefore,

$$\frac{\partial p}{\partial x}\mathbf{i}+\frac{\partial p}{\partial y}\mathbf{j}+\frac{\partial p}{\partial z}\mathbf{k} = \frac{62.5}{32}\times 32\mathbf{k}.$$

and so p does not depend on x and y and is only a function of z. It follows $p(0) = 0$, and $p'(z) = 62.5$. Therefore, $p(x,y,z) = 62.5z$. This establishes the claim. This is interesting but 17.5 is more interesting because it does not require ρ to be constant.

17.4.2 Archimedes Law Of Buoyancy

Archimedes principle states that when a solid body is immersed in a fluid the net force acting on the body by the fluid is directly up and equals the total weight of the fluid displaced.

Denote the set of points in three dimensions occupied by the body as V. Then for dA an increment of area on the surface of this body, the force acting on this increment of area would equal $-p\,dA\mathbf{n}$ where $\mathbf{n}$ is the exterior unit normal. Therefore, since the fluid does not move,

$$\int_{\partial V} -p\mathbf{n}\,dA = \int_V -\nabla p\,dV = \int_V \rho g\,dV\mathbf{k}$$

Which equals the total weight of the displaced fluid and you note the force is directed upward as claimed. Here ρ is the density and 17.5 is being used. There is an interesting point in the above explanation. Why does the second equation hold? Imagine that V were filled with fluid. Then the equation follows from 17.5 because in this equation $\mathbf{g} = -g\mathbf{k}$.

17.4.3 Equations Of Heat And Diffusion

Let $\mathbf{x}$ be a point in three dimensional space and let (x_1, x_2, x_3) be Cartesian coordinates of this point. Let there be a three dimensional body having density, $\rho = \rho(\mathbf{x}, t)$.

The heat flux, $\mathbf{J}$, in the body is defined as a vector which has the following property.

$$\text{Rate at which heat crosses } S = \int_S \mathbf{J}\cdot\mathbf{n}\,dA$$

where $\mathbf{n}$ is the unit normal in the desired direction. Thus if V is a three dimensional body,

$$\text{Rate at which heat leaves } V = \int_{\partial V} \mathbf{J}\cdot\mathbf{n}\,dA$$

[1]There is no such thing as an incompressible fluid but this doesn't stop people from making this assumption.

where $\mathbf{n}$ is the unit exterior normal.

Fourier's law of heat conduction states that the heat flux, $\mathbf{J}$ satisfies $\mathbf{J} = -k\nabla(u)$ where u is the temperature and $k = k(u, \mathbf{x}, t)$ is called the coefficient of thermal conductivity. This changes depending on the material. It also can be shown by experiment to change with temperature. This equation for the heat flux states that the heat flows from hot places toward colder places in the direction of greatest rate of decrease in temperature. Let $c(\mathbf{x}, t)$ denote the specific heat of the material in the body. This means the amount of heat within V is given by the formula $\int_V \rho(\mathbf{x}, t)\, c(\mathbf{x}, t)\, u(\mathbf{x}, t)\ dV$. Suppose also there are sources for the heat within the material given by $f(\mathbf{x},u,t)$. If f is positive, the heat is increasing while if f is negative the heat is decreasing. For example such sources could result from a chemical reaction taking place. Then the divergence theorem can be used to verify the following equation for u. Such an equation is called a reaction diffusion equation.

$$\frac{\partial}{\partial t}(\rho(\mathbf{x}, t)\, c(\mathbf{x}, t)\, u(\mathbf{x}, t)) = \nabla \cdot (k(u, \mathbf{x}, t)\, \nabla u(\mathbf{x}, t)) + f(\mathbf{x}, u, t). \tag{17.6}$$

Take an arbitrary V for which the divergence theorem holds. Then the time rate of change of the heat in V is

$$\frac{d}{dt}\int_V \rho(\mathbf{x}, t)\, c(\mathbf{x}, t)\, u(\mathbf{x}, t)\ dV = \int_V \frac{\partial(\rho(\mathbf{x}, t)\, c(\mathbf{x}, t)\, u(\mathbf{x}, t))}{\partial t}\, dV$$

where, as in the preceding example, this is a physical derivation so the consideration of hard mathematics is not necessary. Therefore, from the Fourier law of heat conduction, $\frac{d}{dt}\int_V \rho(\mathbf{x}, t)\, c(\mathbf{x}, t)\, u(\mathbf{x}, t)\ dV =$

$$\int_V \frac{\partial(\rho(\mathbf{x}, t)\, c(\mathbf{x}, t)\, u(\mathbf{x}, t))}{\partial t}\, dV = \overbrace{\int_{\partial V} -\mathbf{J}\cdot\mathbf{n}\, dA}^{\text{rate at which heat enters}} + \int_V f(\mathbf{x}, u, t)\ dV$$

$$= \int_{\partial V} k\nabla(u)\cdot\mathbf{n}\, dA + \int_V f(\mathbf{x}, u, t)\ dV = \int_V (\nabla\cdot(k\nabla(u)) + f)\ dV.$$

Since this holds for every sample volume, V it must be the case that the above reaction diffusion equation, 17.6 holds. Note that more interesting equations can be obtained by letting more of the quantities in the equation depend on temperature. However, the above is a fairly hard equation and people usually assume the coefficient of thermal conductivity depends only on $\mathbf{x}$ and that the reaction term, f depends only on $\mathbf{x}$ and t and that ρ and c are constant. Then it reduces to the much easier equation,

$$\frac{\partial}{\partial t}u(\mathbf{x}, t) = \frac{1}{\rho c}\nabla\cdot(k(\mathbf{x})\,\nabla u(\mathbf{x}, t)) + f(\mathbf{x},t). \tag{17.7}$$

This is often referred to as the heat equation. Sometimes there are modifications of this in which k is not just a scalar but a matrix to account for different heat flow properties in different directions. However, they are not much harder than the above. The major mathematical difficulties result from allowing k to depend on temperature.

It is known that the heat equation is not correct even if the thermal conductivity did not depend on u because it implies infinite speed of propagation of heat. However, this does not prevent people from using it.

17.4.4 Balance Of Mass

Let $\mathbf{y}$ be a point in three dimensional space and let (y_1, y_2, y_3) be Cartesian coordinates of this point. Let V be a region in three dimensional space and suppose a fluid having

density, $\rho(\mathbf{y},t)$ and velocity, $\mathbf{v}(\mathbf{y},t)$ is flowing through this region. Then the mass of fluid leaving V per unit time is given by the area integral, $\int_{\partial V}\rho(\mathbf{y},t)\,\mathbf{v}(\mathbf{y},t)\cdot\mathbf{n}\,dA$ while the total mass of the fluid enclosed in V at a given time is $\int_V \rho(\mathbf{y},t)\,dV$. Also suppose mass originates at the rate $f(\mathbf{y},t)$ per cubic unit per unit time within this fluid. Then the conclusion which can be drawn through the use of the divergence theorem is the following fundamental equation known as the mass balance equation.

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho\mathbf{v})=f(\mathbf{y},t) \tag{17.8}$$

To see this is so, take an arbitrary V for which the divergence theorem holds. Then the time rate of change of the mass in V is

$$\frac{\partial}{\partial t}\int_V \rho(\mathbf{y},t)\,dV=\int_V \frac{\partial\rho(\mathbf{y},t)}{\partial t}\,dV$$

where the derivative was taken under the integral sign with respect to t. (This is a physical derivation and therefore, it is not necessary to fuss with the hard mathematics related to the change of limit operations. You should expect this to be true under fairly general conditions because the integral is a sort of sum and the derivative of a sum is the sum of the derivatives.) Therefore, the rate of change of mass, $\frac{\partial}{\partial t}\int_V \rho(\mathbf{y},t)\,dV$, equals

$$\begin{aligned}\int_V \frac{\partial\rho(\mathbf{y},t)}{\partial t}\,dV &= \overbrace{-\int_{\partial V}\rho(\mathbf{y},t)\,\mathbf{v}(\mathbf{y},t)\cdot\mathbf{n}\,dA}^{\text{rate at which mass enters}}+\int_V f(\mathbf{y},t)\,dV\\ &= -\int_V\left(\nabla\cdot(\rho(\mathbf{y},t)\,\mathbf{v}(\mathbf{y},t))+f(\mathbf{y},t)\right)dV.\end{aligned}$$

Since this holds for every sample volume, V it must be the case that the equation of continuity holds. Again, there are interesting mathematical questions here which can be explored but since it is a physical derivation, it is not necessary to dwell too much on them. If all the functions involved are continuous, it is certainly true but it is true under far more general conditions than that.

Also note this equation applies to many situations and f might depend on more than just $\mathbf{y}$ and t. In particular, f might depend also on temperature and the density, ρ. This would be the case for example if you were considering the mass of some chemical and f represented a chemical reaction. Mass balance is a general sort of equation valid in many contexts.

17.4.5 Balance Of Momentum

This example is a little more substantial than the above. It concerns the balance of momentum for a continuum. To see a full description of all the physics involved, you should consult a book on continuum mechanics. The situation is of a material in three dimensions and it deforms and moves about in three dimensions. This means this material is not a rigid body. Let B_0 denote an open set identifying a chunk of this material at time $t=0$ and let B_t be an open set which identifies the same chunk of material at time $t>0$.

Let $\mathbf{y}(t,\mathbf{x})=(y_1(t,\mathbf{x}),y_2(t,\mathbf{x}),y_3(t,\mathbf{x}))$ denote the position with respect to Cartesian coordinates at time t of the point whose position at time $t=0$ is $\mathbf{x}=(x_1,x_2,x_3)$. The coordinates, $\mathbf{x}$ are sometimes called the reference coordinates and sometimes the material coordinates and sometimes the Lagrangian coordinates. The coordinates, $\mathbf{y}$ are

called the Eulerian coordinates or sometimes the spacial coordinates and the function, $(t, \mathbf{x}) \to \mathbf{y}(t, \mathbf{x})$ is called the motion. Thus

$$\mathbf{y}(0, \mathbf{x}) = \mathbf{x}. \tag{17.9}$$

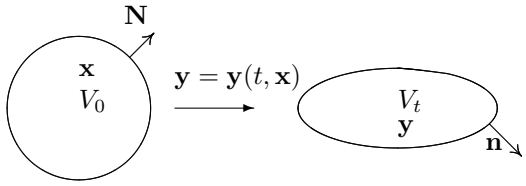
The derivative,

$$D_2\mathbf{y}(t, \mathbf{x}) \equiv D_{\mathbf{x}}\mathbf{y}(t, \mathbf{x})$$

is called the deformation gradient. Recall the notation means you fix t and consider the function, $\mathbf{x} \to \mathbf{y}(t, \mathbf{x})$, taking its derivative. Since it is a linear transformation, it is represented by the usual matrix, whose ij^{th} entry is given by

$$F_{ij}(\mathbf{x}) = \frac{\partial y_i(t, \mathbf{x})}{\partial x_j}.$$

Let $\rho(t, \mathbf{y})$ denote the density of the material at time t at the point, $\mathbf{y}$ and let $\rho_0(\mathbf{x})$ denote the density of the material at the point, $\mathbf{x}$. Thus $\rho_0(\mathbf{x}) = \rho(0, \mathbf{x}) = \rho(0, \mathbf{y}(0, \mathbf{x}))$. The first task is to consider the relationship between $\rho(t, \mathbf{y})$ and $\rho_0(\mathbf{x})$. The following picture is useful to illustrate the ideas.



Lemma 17.4.1 *$\rho_0(\mathbf{x}) = \rho(t, \mathbf{y}(t, \mathbf{x})) \det(F)$ and in any reasonable physical motion,* $\det(F) > 0$.

Proof: Let V_0 represent a small chunk of material at $t = 0$ and let V_t represent the same chunk of material at time t. I will be a little sloppy and refer to V_0 as the small chunk of material at time $t = 0$ and V_t as the chunk of material at time t rather than an open set representing the chunk of material. Then by the change of variables formula for multiple integrals,

$$\int_{V_t} dV = \int_{V_0} |\det(F)| \, dV.$$

If $\det(F) = 0$ for some t the above formula shows that the chunk of material went from positive volume to zero volume and this is not physically possible. Therefore, it is impossible that $\det(F)$ can equal zero. However, at $t = 0$, $F = I$, the identity because of 17.9. Therefore, $\det(F) = 1$ at $t = 0$ and if it is assumed $t \to \det(F)$ is continuous it follows by the intermediate value theorem that $\det(F) > 0$ for all t. Of course it is not known for sure this function is continuous but the above shows why it is at least reasonable to expect $\det(F) > 0$.

Now using the change of variables formula,

$$\begin{aligned} \text{mass of } V_t &= \int_{V_t} \rho(t, \mathbf{y}) \, dV = \int_{V_0} \rho(t, \mathbf{y}(t, \mathbf{x})) \det(F) \, dV \\ &= \text{mass of } V_0 = \int_{V_0} \rho_0(\mathbf{x}) \, dV. \end{aligned}$$

Since V_0 is arbitrary, it follows

$$\rho_0(\mathbf{x}) = \rho(t, \mathbf{y}(t, \mathbf{x})) \det(F)$$

as claimed. Note this shows that $\det(F)$ is a magnification factor for the density.

Now consider a small chunk of material, V_t at time t which corresponds to V_0 at time $t=0$. The total linear momentum of this material at time t is

$$\int_{V_t} \rho(t,\mathbf{y})\,\mathbf{v}(t,\mathbf{y})\;dV$$

where $\mathbf{v}$ is the velocity. By Newton's second law, the time rate of change of this linear momentum should equal the total force acting on the chunk of material. In the following derivation, $dV(\mathbf{y})$ will indicate the integration is taking place with respect to the variable, $\mathbf{y}$. By Lemma 17.4.1 and the change of variables formula for multiple integrals

$$\begin{aligned}
\frac{d}{dt}\left(\int_{V_t}\rho(t,\mathbf{y})\,\mathbf{v}(t,\mathbf{y})\;dV(\mathbf{y})\right) &= \frac{d}{dt}\left(\int_{V_0}\rho(t,\mathbf{y}(t,\mathbf{x}))\,\mathbf{v}(t,\mathbf{y}(t,\mathbf{x}))\det(F)\;dV(\mathbf{x})\right)\\
&= \frac{d}{dt}\left(\int_{V_0}\rho_0(\mathbf{x})\,\mathbf{v}(t,\mathbf{y}(t,\mathbf{x}))\;dV(\mathbf{x})\right)\\
&= \int_{V_0}\rho_0(\mathbf{x})\left[\frac{\partial\mathbf{v}}{\partial t}+\frac{\partial\mathbf{v}}{\partial y_i}\frac{\partial y_i}{\partial t}\right]dV(\mathbf{x})\\
&= \int_{V_t}\overbrace{\rho(t,\mathbf{y})\det(F)}^{\rho_0(\mathbf{x})}\left[\frac{\partial\mathbf{v}}{\partial t}+\frac{\partial\mathbf{v}}{\partial y_i}\frac{\partial y_i}{\partial t}\right]\frac{1}{\det(F)}\,dV(\mathbf{y})\\
&= \int_{V_t}\rho(t,\mathbf{y})\left[\frac{\partial\mathbf{v}}{\partial t}+\frac{\partial\mathbf{v}}{\partial y_i}\frac{\partial y_i}{\partial t}\right]dV(\mathbf{y}).
\end{aligned}$$

Having taken the derivative of the total momentum, it is time to consider the total force acting on the chunk of material.

The force comes from two sources, a body force, $\mathbf{b}$ and a force which acts on the boundary of the chunk of material called a traction force. Typically, the body force is something like gravity in which case, $\mathbf{b}=-g\rho\mathbf{k}$, assuming the Cartesian coordinate system has been chosen in the usual manner. The traction force is of the form

$$\int_{\partial V_t}\mathbf{s}(t,\mathbf{y},\mathbf{n})\;dA$$

where $\mathbf{n}$ is the unit exterior normal. Thus the traction force depends on position, time, and the orientation of the boundary of V_t. Cauchy showed the existence of a linear transformation, $T(t,\mathbf{y})$ such that $T(t,\mathbf{y})\,\mathbf{n}=\mathbf{s}(t,\mathbf{y},\mathbf{n})$. It follows there is a matrix, $T_{ij}(t,\mathbf{y})$ such that the i^{th} component of $\mathbf{s}$ is given by $\mathbf{s}_i(t,\mathbf{y},\mathbf{n})=T_{ij}(t,\mathbf{y})\,n_j$. Cauchy also showed this matrix is symmetric, $T_{ij}=T_{ji}$. It is called the Cauchy stress. Using Newton's second law to equate the time derivative of the total linear momentum with the applied forces and using the usual repeated index summation convention,

$$\int_{V_t}\rho(t,\mathbf{y})\left[\frac{\partial\mathbf{v}}{\partial t}+\frac{\partial\mathbf{v}}{\partial y_i}\frac{\partial y_i}{\partial t}\right]dV(\mathbf{y})=\int_{V_t}\mathbf{b}(t,\mathbf{y})\;dV(\mathbf{y})+\int_{\partial B_t}T_{ij}(t,\mathbf{y})\,n_j\,dA.$$

Here is where the divergence theorem is used. In the last integral, the multiplication by n_j is exchanged for the j^{th} partial derivative and an integral over V_t. Thus

$$\int_{V_t}\rho(t,\mathbf{y})\left[\frac{\partial\mathbf{v}}{\partial t}+\frac{\partial\mathbf{v}}{\partial y_i}\frac{\partial y_i}{\partial t}\right]dV(\mathbf{y})=\int_{V_t}\mathbf{b}(t,\mathbf{y})\;dV(\mathbf{y})+\int_{V_t}\frac{\partial\left(T_{ij}(t,\mathbf{y})\right)}{\partial y_j}\,dV(\mathbf{y}).$$

Since V_t was arbitrary, it follows

$$\begin{aligned}
\rho(t,\mathbf{y})\left[\frac{\partial\mathbf{v}}{\partial t}+\frac{\partial\mathbf{v}}{\partial y_i}\frac{\partial y_i}{\partial t}\right] &= \mathbf{b}(t,\mathbf{y})+\frac{\partial\left(T_{ij}(t,\mathbf{y})\right)}{\partial y_j}\\
&\equiv \mathbf{b}(t,\mathbf{y})+\operatorname{div}(T)
\end{aligned}$$

where here $\operatorname{div} T$ is a vector whose i^{th} component is given by

$$(\operatorname{div} T)_i = \frac{\partial T_{ij}}{\partial y_j}.$$

The term, $\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i}\frac{\partial y_i}{\partial t}$, is the total derivative with respect to t of the velocity $\mathbf{v}$. Thus you might see this written as

$$\rho \dot{\mathbf{v}} = \mathbf{b} + \operatorname{div}(T).$$

The above formulation of the balance of momentum involves the spatial coordinates, $\mathbf{y}$ but people also like to formulate momentum balance in terms of the material coordinates, $\mathbf{x}$. Of course this changes everything.

The momentum in terms of the material coordinates is

$$\int_{V_0} \rho_0(\mathbf{x})\, \mathbf{v}(t, \mathbf{x})\, dV$$

and so, since $\mathbf{x}$ does not depend on t,

$$\frac{d}{dt}\left(\int_{V_0} \rho_0(\mathbf{x})\, \mathbf{v}(t, \mathbf{x})\, dV\right) = \int_{V_0} \rho_0(\mathbf{x})\, \mathbf{v}_t(t, \mathbf{x})\, dV.$$

As indicated earlier, this is a physical derivation and so the mathematical questions related to interchange of limit operations are ignored. This must equal the total applied force. Thus

$$\int_{V_0} \rho_0(\mathbf{x})\, \mathbf{v}_t(t, \mathbf{x})\, dV = \int_{V_0} \mathbf{b}_0(t, \mathbf{x})\, dV + \int_{\partial V_t} T_{ij} n_j dA, \tag{17.10}$$

the first term on the right being the contribution of the body force given per unit volume in the material coordinates and the last term being the traction force discussed earlier. The task is to write this last integral as one over ∂V_0. For $\mathbf{y} \in \partial V_t$ there is a unit outer normal, $\mathbf{n}$. Here $\mathbf{y} = \mathbf{y}(t, \mathbf{x})$ for $\mathbf{x} \in \partial V_0$. Then define $\mathbf{N}$ to be the unit outer normal to V_0 at the point, $\mathbf{x}$. Near the point $\mathbf{y} \in \partial V_t$ the surface, ∂V_t is given parametrically in the form $\mathbf{y} = \mathbf{y}(s, t)$ for $(s, t) \in D \subseteq \mathbb{R}^2$ and it can be assumed the unit normal to ∂V_t near this point is

$$\mathbf{n} = \frac{\mathbf{y}_s(s,t) \times \mathbf{y}_t(s,t)}{|\mathbf{y}_s(s,t) \times \mathbf{y}_t(s,t)|}$$

with the area element given by $|\mathbf{y}_s(s,t) \times \mathbf{y}_t(s,t)|\, ds\, dt$. This is true for $\mathbf{y} \in P_t \subseteq \partial V_t$, a small piece of ∂V_t. Therefore, the last integral in 17.10 is the sum of integrals over small pieces of the form

$$\int_{P_t} T_{ij} n_j dA \tag{17.11}$$

where P_t is parametrized by $\mathbf{y}(s, t)$, $(s, t) \in D$. Thus the integral in 17.11 is of the form

$$\int_D T_{ij}(\mathbf{y}(s,t))\left(\mathbf{y}_s(s,t) \times \mathbf{y}_t(s,t)\right)_j\, ds\, dt.$$

By the chain rule this equals

$$\int_D T_{ij}(\mathbf{y}(s,t))\left(\frac{\partial \mathbf{y}}{\partial x_\alpha}\frac{\partial x_\alpha}{\partial s} \times \frac{\partial \mathbf{y}}{\partial x_\beta}\frac{\partial x_\beta}{\partial t}\right)_j ds\, dt.$$

Remember $\mathbf{y} = \mathbf{y}(t, \mathbf{x})$ and it is always assumed the mapping $\mathbf{x} \to \mathbf{y}(t, \mathbf{x})$ is one to one and so, since on the surface ∂V_t near $\mathbf{y}$, the points are functions of (s, t), it follows $\mathbf{x}$ is

also a function of (s,t). Now by the properties of the cross product, this last integral equals

$$\int_D T_{ij}\left(\mathbf{x}\left(s,t\right)\right)\frac{\partial x_\alpha}{\partial s}\frac{\partial x_\beta}{\partial t}\left(\frac{\partial \mathbf{y}}{\partial x_\alpha}\times\frac{\partial \mathbf{y}}{\partial x_\beta}\right)_j ds\,dt \tag{17.12}$$

where here $\mathbf{x}(s,t)$ is the point of ∂V_0 which corresponds with $\mathbf{y}(s,t)\in\partial V_t$. Thus

$$T_{ij}\left(\mathbf{x}\left(s,t\right)\right)=T_{ij}\left(\mathbf{y}\left(s,t\right)\right).$$

(Perhaps this is a slight abuse of notation because T_{ij} is defined on ∂V_t, not on ∂V_0, but it avoids introducing extra symbols.) Next 17.12 equals

$$\int_D T_{ij}\left(\mathbf{x}\left(s,t\right)\right)\frac{\partial x_\alpha}{\partial s}\frac{\partial x_\beta}{\partial t}\varepsilon_{jab}\frac{\partial y_a}{\partial x_\alpha}\frac{\partial y_b}{\partial x_\beta}\,ds\,dt$$

$$=\int_D T_{ij}\left(\mathbf{x}\left(s,t\right)\right)\frac{\partial x_\alpha}{\partial s}\frac{\partial x_\beta}{\partial t}\varepsilon_{cab}\delta_{jc}\frac{\partial y_a}{\partial x_\alpha}\frac{\partial y_b}{\partial x_\beta}\,ds\,dt$$

$$=\int_D T_{ij}\left(\mathbf{x}\left(s,t\right)\right)\frac{\partial x_\alpha}{\partial s}\frac{\partial x_\beta}{\partial t}\varepsilon_{cab}\overbrace{\frac{\partial y_c}{\partial x_p}\frac{\partial x_p}{\partial y_j}}^{=\delta_{jc}}\frac{\partial y_a}{\partial x_\alpha}\frac{\partial y_b}{\partial x_\beta}\,ds\,dt$$

$$=\int_D T_{ij}\left(\mathbf{x}\left(s,t\right)\right)\frac{\partial x_\alpha}{\partial s}\frac{\partial x_\beta}{\partial t}\frac{\partial x_p}{\partial y_j}\overbrace{\varepsilon_{cab}\frac{\partial y_c}{\partial x_p}\frac{\partial y_a}{\partial x_\alpha}\frac{\partial y_b}{\partial x_\beta}}^{=\varepsilon_{p\alpha\beta}\det(F)}\,ds\,dt$$

$$=\int_D\left(\det F\right)T_{ij}\left(\mathbf{x}\left(s,t\right)\right)\varepsilon_{p\alpha\beta}\frac{\partial x_\alpha}{\partial s}\frac{\partial x_\beta}{\partial t}\frac{\partial x_p}{\partial y_j}\,ds\,dt.$$

Now $\frac{\partial x_p}{\partial y_j}=F^{-1}_{pj}$ and also

$$\varepsilon_{p\alpha\beta}\frac{\partial x_\alpha}{\partial s}\frac{\partial x_\beta}{\partial t}=\left(\mathbf{x}_s\times\mathbf{x}_t\right)_p$$

so the result just obtained is of the form

$$\int_D\left(\det F\right)F^{-1}_{pj}T_{ij}\left(\mathbf{x}\left(s,t\right)\right)\left(\mathbf{x}_s\times\mathbf{x}_t\right)_p\,ds\,dt=$$

$$\int_D\left(\det F\right)T_{ij}\left(\mathbf{x}\left(s,t\right)\right)\left(F^{-T}\right)_{jp}\left(\mathbf{x}_s\times\mathbf{x}_t\right)_p\,ds\,dt.$$

This has transformed the integral over P_t to one over P_0, the part of ∂V_0 which corresponds with P_t. Thus the last integral is of the form

$$\int_{P_0}\det\left(F\right)\left(TF^{-T}\right)_{ip}N_p dA$$

Summing these up over the pieces of ∂V_t and ∂V_0 yields the last integral in 17.10 equals

$$\int_{\partial V_0}\det\left(F\right)\left(TF^{-T}\right)_{ip}N_p dA$$

and so the balance of momentum in terms of the material coordinates becomes

$$\int_{V_0}\rho_0\left(\mathbf{x}\right)\mathbf{v}_t\left(t,\mathbf{x}\right)\,dV=\int_{V_0}\mathbf{b}_0\left(t,\mathbf{x}\right)\,dV+\int_{\partial V_0}\det\left(F\right)\left(TF^{-T}\right)_{ip}N_p dA$$

The matrix, $\det(F)\left(TF^{-T}\right)_{ip}$ is called the Piola Kirchhoff stress, S. An application of the divergence theorem yields

$$\int_{V_0}\rho_0(\mathbf{x})\,\mathbf{v}_t(t,\mathbf{x})\,dV=\int_{V_0}\mathbf{b}_0(t,\mathbf{x})\,dV+\int_{V_0}\frac{\partial\left(\det(F)\left(TF^{-T}\right)_{ip}\right)}{\partial x_p}\,dV.$$

Since V_0 is arbitrary, a balance law for momentum in terms of the material coordinates is obtained

$$\begin{aligned}\rho_0(\mathbf{x})\,\mathbf{v}_t(t,\mathbf{x}) &= \mathbf{b}_0(t,\mathbf{x})+\frac{\partial\left(\det(F)\left(TF^{-T}\right)_{ip}\right)}{\partial x_p}\\ &= \mathbf{b}_0(t,\mathbf{x})+\operatorname{div}\left(\det(F)\left(TF^{-T}\right)\right)\\ &= \mathbf{b}_0(t,\mathbf{x})+\operatorname{div}S. \qquad (17.13)\end{aligned}$$

As just shown, the relation between the Cauchy stress and the Piola Kirchhoff stress is

$$S=\det(F)\left(TF^{-T}\right), \qquad (17.14)$$

perhaps not the first thing you would think of.

The main purpose of this presentation is to show how the divergence theorem is used in a significant way to obtain balance laws and to indicate a very interesting direction for further study. To continue, one needs to specify T or S as an appropriate function of things related to the motion, $\mathbf{y}$. Often the thing related to the motion is something called the strain and such relationships are known as constitutive laws.

17.4.6 Frame Indifference

The proper formulation of constitutive laws involves more physical considerations such as frame indifference in which it is required the response of the system cannot depend on the manner in which the Cartesian coordinate system for the spacial coordinates was chosen.

For $Q(t)$ an orthogonal transformation and

$$\mathbf{y}'=\mathbf{q}(t)+Q(t)\,\mathbf{y},\ \mathbf{n}'=Q\mathbf{n},$$

the new spacial coordinates are denoted by $\mathbf{y}'$. Recall an orthogonal transformation is just one which satisfies

$$Q(t)^T Q(t)=Q(t)\,Q(t)^T=I.$$

The stress has to do with the traction force area density produced by internal changes in the body and has nothing to do with the way the body is observed. Therefore, it is required that

$$T'\mathbf{n}'=QT\mathbf{n}$$

Thus

$$T'Q\mathbf{n}=QT\mathbf{n}$$

Since this is true for any $\mathbf{n}$ normal to the boundary of any piece of the material considered, it must be the case that

$$T'Q=QT$$

and so

$$T'=QTQ^T.$$

This is called frame indifference.

By 17.14, the Piola Kirchhoff stress, S is related to T by

$$S = \det(F) TF^{-T},\ F \equiv D_{\mathbf{x}}\mathbf{y}.$$

This stress involves the use of the material coordinates and a normal $\mathbf{N}$ to a piece of the body in reference configuration. Thus $S\mathbf{N}$ gives the force on a part of ∂V_t per unit area on ∂V_0. Then for a different choice of spacial coordinates, $\mathbf{y}' = \mathbf{q}(t) + Q(t)\mathbf{y}$,

$$S' = \det(F') T' (F')^{-T}$$

but

$$F' = D_{\mathbf{x}}\mathbf{y}' = Q(t) D_{\mathbf{x}}\mathbf{y} = QF$$

and so frame indifference in terms of S is

$$\begin{aligned} S' &= \det(F) QTQ^T (QF)^{-T} \\ &= \det(F) QTQ^T QF^{-T} \\ &= QS \end{aligned}$$

This principle of frame indifference is sometimes ignored and there are certainly interesting mathematical models which have resulted from doing this, but such things cannot be considered physically acceptable.

There are also many other physical properties which can be included and which require a certain form for the constitutive equations. These considerations are outside the scope of this book and require a considerable amount of linear algebra.

There are also balance laws for energy which you may study later but these are more problematic than the balance laws for mass and momentum. However, the divergence theorem is used in these also.

17.4.7 Bernoulli's Principle

Consider a possibly moving fluid with constant density, ρ and let P denote the pressure in this fluid. If B is a part of this fluid the force exerted on B by the rest of the fluid is $\int_{\partial B} -P\mathbf{n}dA$ where $\mathbf{n}$ is the outer normal from B. Assume this is the only force which matters so for example there is no viscosity in the fluid. Thus the Cauchy stress in rectangular coordinates should be

$$T = \begin{pmatrix} -P & 0 & 0 \\ 0 & -P & 0 \\ 0 & 0 & -P \end{pmatrix}.$$

Then

$$\operatorname{div} T = -\nabla P.$$

Also suppose the only body force is from gravity, a force of the form

$$-\rho g\mathbf{k}$$

and so from the balance of momentum

$$\rho\dot{\mathbf{v}} = -\rho g\mathbf{k} - \nabla P(\mathbf{x}). \tag{17.15}$$

Now in all this the coordinates are the spacial coordinates and it is assumed they are rectangular. Thus

$$\mathbf{x} = (x, y, z)^T$$

and $\mathbf{v}$ is the velocity while $\dot{\mathbf{v}}$ is the total derivative of $\mathbf{v} = (v_1, v_2, v_3)^T$ given by $\mathbf{v}_t + v_i \mathbf{v}_{,i}$. Take the dot product of both sides of 17.15 with $\mathbf{v}$. This yields

$$(\rho/2) \frac{d}{dt} |\mathbf{v}|^2 = -\rho g \frac{dz}{dt} - \frac{d}{dt} P(\mathbf{x}).$$

Therefore,

$$\frac{d}{dt} \left(\frac{\rho |\mathbf{v}|^2}{2} + \rho g z + P(\mathbf{x}) \right) = 0$$

and so there is a constant, C' such that

$$\frac{\rho |\mathbf{v}|^2}{2} + \rho g z + P(\mathbf{x}) = C'$$

For convenience define γ to be the weight density of this fluid. Thus $\gamma = \rho g$. Divide by γ. Then

$$\frac{|\mathbf{v}|^2}{2g} + z + \frac{P(\mathbf{x})}{\gamma} = C.$$

this is Bernoulli's[2] principle. Note how if you keep the height the same, then if you raise $|\mathbf{v}|$, it follows the pressure drops.

This is often used to explain the lift of an airplane wing. The top surface is curved which forces the air to go faster over the top of the wing causing a drop in pressure which creates lift. It is also used to explain the concept of a venturi tube in which the air loses pressure due to being pinched which causes it to flow faster. In many of these applications, the assumptions used in which ρ is constant and there is no other contribution to the traction force on ∂B than pressure so in particular, there is no viscosity, are not correct. However, it is hoped that the effects of these deviations from the ideal situation above are small enough that the conclusions are still roughly true. You can see how using balance of momentum can be used to consider more difficult situations. For example, you might have a body force which is more involved than gravity.

17.4.8 The Wave Equation

As an example of how the balance law of momentum is used to obtain an important equation of mathematical physics, suppose $S = kF$ where k is a constant and F is the deformation gradient and let $\mathbf{u} \equiv \mathbf{y} - \mathbf{x}$. Thus $\mathbf{u}$ is the displacement. Then from 17.13 you can verify the following holds.

$$\rho_0(\mathbf{x}) \mathbf{u}_{tt}(t, \mathbf{x}) = \mathbf{b}_0(t, \mathbf{x}) + k \Delta \mathbf{u}(t, \mathbf{x}) \tag{17.16}$$

In the case where ρ_0 is a constant and $\mathbf{b}_0 = 0$, this yields

$$\mathbf{u}_{tt} - c \Delta \mathbf{u} = \mathbf{0}.$$

The wave equation is $u_{tt} - c\Delta u = 0$ and so the above gives three wave equations, one for each component.

[2]There were many Bernoullis. This is Daniel Bernoulli. He seems to have been nicer than some of the others. Daniel was actually a doctor who was interested in mathematics.He lived from 1700-1782.

17.4.9 A Negative Observation

Many of the above applications of the divergence theorem are based on the assumption that matter is continuously distributed in a way that the above arguments are correct. In other words, a continuum. However, there is no such thing as a continuum. It has been known for some time now that matter is composed of atoms. It is not continuously distributed through some region of space as it is in the above. Apologists for this contradiction with reality sometimes say to consider enough of the material in question that it is reasonable to think of it as a continuum. This mystical reasoning is then violated as soon as they go from the integral form of the balance laws to the differential equations expressing the traditional formulation of these laws. See Problem 9 below, for example. However, these laws continue to be used and seem to lead to useful physical models which have value in predicting the behavior of physical systems. This is what justifies their use, not any fundamental truth.

17.4.10 Electrostatics

Coloumb's law says that the electric field intensity at $\mathbf{x}$ of a charge q located at point, $\mathbf{x}_0$ is given by

$$\mathbf{E} = k\frac{q\left(\mathbf{x}-\mathbf{x}_0\right)}{\left|\mathbf{x}-\mathbf{x}_0\right|^3}$$

where the electric field intensity is defined to be the force experienced by a unit positive charge placed at the point, $\mathbf{x}$. Note that this is a vector and that its direction depends on the sign of q. It points away from $\mathbf{x}_0$ if q is positive and points toward $\mathbf{x}_0$ if q is negative. The constant, k is a physical constant like the gravitation constant. It has been computed through careful experiments similar to those used with the calculation of the gravitation constant.

The interesting thing about Coloumb's law is that $\mathbf{E}$ is the gradient of a function. In fact,

$$\mathbf{E} = \nabla\left(qk\frac{1}{\left|\mathbf{x}-\mathbf{x}_0\right|}\right).$$

The other thing which is significant about this is that in three dimensions and for $\mathbf{x} \neq \mathbf{x}_0$,

$$\nabla\cdot\nabla\left(qk\frac{1}{\left|\mathbf{x}-\mathbf{x}_0\right|}\right) = \nabla\cdot\mathbf{E} = 0. \tag{17.17}$$

This is left as an exercise for you to verify.

These observations will be used to derive a very important formula for the integral,

$$\int_{\partial U}\mathbf{E}\cdot\mathbf{n}dS$$

where $\mathbf{E}$ is the electric field intensity due to a charge, q located at the point, $\mathbf{x}_0 \in U$, a bounded open set for which the divergence theorem holds.

Let U_ε denote the open set obtained by removing the open ball centered at $\mathbf{x}_0$ which has radius ε where ε is small enough that the following picture is a correct representation of the situation.

Then on the boundary of B_ε the unit outer normal to U_ε is $-\frac{\mathbf{x}-\mathbf{x}_0}{|\mathbf{x}-\mathbf{x}_0|}$. Therefore,

$$\begin{aligned}\int_{\partial B_\varepsilon} \mathbf{E}\cdot\mathbf{n}dS &= -\int_{\partial B_\varepsilon} k\frac{q(\mathbf{x}-\mathbf{x}_0)}{|\mathbf{x}-\mathbf{x}_0|^3}\cdot\frac{\mathbf{x}-\mathbf{x}_0}{|\mathbf{x}-\mathbf{x}_0|}dS \\ &= -kq\int_{\partial B_\varepsilon}\frac{1}{|\mathbf{x}-\mathbf{x}_0|^2}dS = \frac{-kq}{\varepsilon^2}\int_{\partial B_\varepsilon} dS \\ &= \frac{-kq}{\varepsilon^2}4\pi\varepsilon^2 = -4\pi kq.\end{aligned}$$

Therefore, from the divergence theorem and observation 17.17,

$$-4\pi kq + \int_{\partial U}\mathbf{E}\cdot\mathbf{n}dS = \int_{\partial U_\varepsilon}\mathbf{E}\cdot\mathbf{n}dS = \int_{U_\varepsilon}\nabla\cdot\mathbf{E}dV = 0.$$

It follows that

$$4\pi kq = \int_{\partial U}\mathbf{E}\cdot\mathbf{n}dS.$$

If there are several charges located inside U, say $q_1, q_2, \cdots, q_n$, then letting $\mathbf{E}_i$ denote the electric field intensity of the i^{th} charge and $\mathbf{E}$ denoting the total resulting electric field intensity due to all these charges,

$$\begin{aligned}\int_{\partial U}\mathbf{E}\cdot\mathbf{n}dS &= \sum_{i=1}^n\int_{\partial U}\mathbf{E}_i\cdot\mathbf{n}dS \\ &= \sum_{i=1}^n 4\pi kq_i = 4\pi k\sum_{i=1}^n q_i.\end{aligned}$$

This is known as Gauss's law and it is the fundamental result in electrostatics.

17.5 Exercises

1. To prove the divergence theorem, it was shown first that the spacial partial derivative in the volume integral could be exchanged for multiplication by an appropriate component of the exterior normal. This problem starts with the divergence theorem and goes the other direction. Assuming the divergence theorem, holds for a region, V, show that $\int_{\partial V}\mathbf{n}u\,dA = \int_V \nabla u\,dV$. Note this implies $\int_V \frac{\partial u}{\partial x}\,dV = \int_{\partial V} n_1 u\,dA$.

2. Let V be such that the divergence theorem holds. Show that $\int_V \nabla\cdot(u\nabla v)\,dV = \int_{\partial V} u\frac{\partial v}{\partial n}\,dA$ where $\mathbf{n}$ is the exterior normal and $\frac{\partial v}{\partial n}$ denotes the directional derivative of v in the direction $\mathbf{n}$.

3. Let V be such that the divergence theorem holds. Show that $\int_V \left(v\nabla^2 u - u\nabla^2 v\right)\,dV = \int_{\partial V}\left(v\frac{\partial u}{\partial n} - u\frac{\partial v}{\partial n}\right)\,dA$ where $\mathbf{n}$ is the exterior normal and $\frac{\partial u}{\partial n}$ is defined in Problem 2.

4. Let V be a ball and suppose $\nabla^2 u = f$ in V while $u = g$ on ∂V. Show there is at most one solution to this boundary value problem which is C^2 in V and continuous on V with its boundary. **Hint:** You might consider $w = u - v$ where u and v are solutions to the problem. Then use the result of Problem 2 and the identity

$$w\nabla^2 w = \nabla\cdot(w\nabla w) - \nabla w\cdot\nabla w$$

to conclude $\nabla w = 0$. Then show this implies w must be a constant by considering $h(t) = w(t\mathbf{x} + (1-t)\mathbf{y})$ and showing h is a constant. Alternatively, you might consider the maximum principle.

5. Show that $\int_{\partial V} \nabla \times \mathbf{v} \cdot \mathbf{n}\, dA = 0$ where V is a region for which the divergence theorem holds and $\mathbf{v}$ is a C^2 vector field.

6. Let $\mathbf{F}(x,y,z) = (x,y,z)$ be a vector field in $\mathbb{R}^3$ and let V be a three dimensional shape and let $\mathbf{n} = (n_1, n_2, n_3)$. Show $\int_{\partial V} (xn_1 + yn_2 + zn_3)\, dA = 3\times$ volume of V.

7. Does the divergence theorem hold for higher dimensions? If so, explain why it does. How about two dimensions?

8. Let $\mathbf{F} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and let V denote the tetrahedron formed by the planes, $x = 0, y = 0, z = 0$, and $\frac{1}{3}x + \frac{1}{3}y + \frac{1}{5}z = 1$. Verify the divergence theorem for this example.

9. Suppose $f : U \to \mathbb{R}$ is continuous where U is some open set and for all $B \subseteq U$ where B is a ball, $\int_B f(\mathbf{x})\, dV = 0$. Show this implies $f(\mathbf{x}) = 0$ for all $\mathbf{x} \in U$.

10. Let U denote the box centered at $(0,0,0)$ with sides parallel to the coordinate planes which has width 4, length 2 and height 3. Find the flux integral $\int_{\partial U} \mathbf{F} \cdot \mathbf{n}\, dS$ where $\mathbf{F} = (x+3, 2y, 3z)$. **Hint:** If you like, you might want to use the divergence theorem.

11. Verify 17.16 from 17.13 and the assumption that $S = kF$.

12. Fick's law for diffusion states the flux of a diffusing species, $\mathbf{J}$ is proportional to the gradient of the concentration, c. Write this law getting the sign right for the constant of proportionality and derive an equation similar to the heat equation for the concentration, c. Typically, c is the concentration of some sort of pollutant or a chemical.

13. Show that if $u_k, k = 1, 2, \cdots, n$ each satisfies 17.7 then for any choice of constants, $c_1, \cdots, c_n$, so does
$$\sum_{k=1}^{n} c_k u_k.$$

14. Suppose $k(\mathbf{x}) = k$, a constant and $f = 0$. Then in one dimension, the heat equation is of the form $u_t = \alpha u_{xx}$. Show $u(x,t) = e^{-\alpha n^2 t} \sin(nx)$ satisfies the heat equation[3].

15. In a linear, viscous, incompressible fluid, the Cauchy stress is of the form
$$T_{ij}(t,\mathbf{y}) = \lambda \left(\frac{v_{i,j}(t,\mathbf{y}) + v_{j,i}(t,\mathbf{y})}{2} \right) - p\delta_{ij}$$
where p is the pressure, δ_{ij} equals 0 if $i \neq j$ and 1 if $i = j$, and the comma followed by an index indicates the partial derivative with respect to that variable and $\mathbf{v}$ is the velocity. Thus
$$v_{i,j} = \frac{\partial v_i}{\partial y_j}$$

[3]Fourier, an officer in Napoleon's army studied solutions to the heat equation back in 1813. He was interested in heat flow in cannons. He sought to find solutions by adding up infinitely many solutions of this form. Actually, it was a little more complicated because cannons are not one dimensional but it was the beginning of the study of Fourier series, a topic which fascinated mathematicians for the next 150 years and motivated the development of analysis.

Also, p denotes the pressure. Show, using the balance of mass equation that incompressible implies $\operatorname{div} \mathbf{v} = 0$. Next show the balance of momentum equation requires

$$\rho\dot{\mathbf{v}} - \frac{\lambda}{2}\Delta\mathbf{v} = \rho\left[\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i}v_i\right] - \frac{\lambda}{2}\Delta\mathbf{v} = \mathbf{b} - \nabla p.$$

This is the famous Navier Stokes equation for incompressible viscous linear fluids. There are still open questions related to this equation, one of which is worth $1,000,000 at this time.

Stokes And Green's Theorems

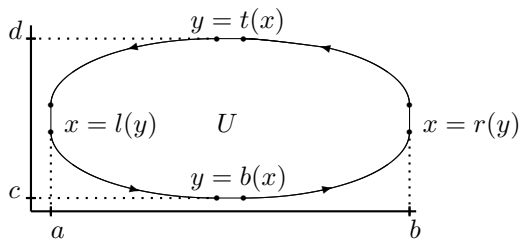
18.0.1 Outcomes

1. Recall and verify Green's theorem.
2. Apply Green's theorem to evaluate line integrals.
3. Apply Green's theorem to find the area of a region.
4. Explain what is meant by the curl of a vector field.
5. Evaluate the curl of a vector field.
6. Derive and apply formulas involving divergence, gradient and curl.
7. Recall and use Stoke's theorem.
8. Apply Stoke's theorem to calculate the circulation or work of a vector field around a simple closed curve.
9. Recall and apply the fundamental theorem for line integrals.
10. Determine whether a vector field is a gradient using the curl test.
11. Recover a function from its gradient when possible.

18.1 Green's Theorem

Green's theorem is an important theorem which relates line integrals to integrals over a surface in the plane. It can be used to establish the seemingly more general Stoke's theorem but is interesting for it's own sake. Historically, theorems like it were important in the development of complex analysis. I will first establish Green's theorem for regions of a particular sort and then show that the theorem holds for many other regions also. Suppose a region is of the form indicated in the following picture in which

$$\begin{aligned} U &= \{(x,y) : x \in (a,b) \text{ and } y \in (b(x), t(x))\} \\ &= \{(x,y) : y \in (c,d) \text{ and } x \in (l(y), r(y))\}. \end{aligned}$$



I will refer to such a region as being convex in both the x and y directions.

Lemma 18.1.1 *Let $\mathbf{F}(x,y) \equiv (P(x,y), Q(x,y))$ be a C^1 vector field defined near U where U is a region of the sort indicated in the above picture which is convex in both the x and y directions. Suppose also that the functions, r, l, t, and b in the above picture are all C^1 functions and denote by ∂U the boundary of U oriented such that the direction of motion is counter clockwise. (As you walk around U on ∂U, the points of U are on your left.) Then*

$$\int_{\partial U} Pdx + Qdy \equiv$$

$$\int_{\partial U} \mathbf{F}\cdot d\mathbf{R} = \int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA. \tag{18.1}$$

Proof: First consider the right side of 18.1.

$$\int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA$$

$$\begin{aligned} &= \int_c^d \int_{l(y)}^{r(y)} \frac{\partial Q}{\partial x} dx dy - \int_a^b \int_{b(x)}^{t(x)} \frac{\partial P}{\partial y} dy dx \\ &= \int_c^d \left(Q(r(y), y) - Q(l(y), y) \right) dy + \int_a^b \left(P(x, b(x)) \right) - P(x, t(x)) \, dx. \end{aligned} \tag{18.2}$$

Now consider the left side of 18.1. Denote by V the vertical parts of ∂U and by H the horizontal parts.

$$\int_{\partial U} \mathbf{F}\cdot d\mathbf{R} =$$

$$\begin{aligned} &= \int_{\partial U} ((0,Q) + (P,0)) \cdot d\mathbf{R} \\ &= \int_c^d (0, Q(r(s), s)) \cdot (r'(s), 1) \, ds + \int_H (0, Q(r(s), s)) \cdot (\pm 1, 0) \, ds \\ &\quad - \int_c^d (0, Q(l(s), s)) \cdot (l'(s), 1) \, ds + \int_a^b (P(s, b(s)), 0) \cdot (1, b'(s)) \, ds \\ &\quad + \int_V (P(s, b(s)), 0) \cdot (0, \pm 1) \, ds - \int_a^b (P(s, t(s)), 0) \cdot (1, t'(s)) \, ds \\ &= \int_c^d Q(r(s), s) \, ds - \int_c^d Q(l(s), s) \, ds + \int_a^b P(s, b(s)) \, ds - \int_a^b P(s, t(s)) \, ds \end{aligned}$$

which coincides with 18.2. This proves the lemma.

Corollary 18.1.2 *Let everything be the same as in Lemma 18.1.1 but only assume the functions r, l, t, and b are continuous and piecewise C^1 functions. Then the conclusion this lemma is still valid.*

Proof: The details are left for you. All you have to do is to break up the various line integrals into the sum of integrals over sub intervals on which the function of interest is C^1.

From this corollary, it follows 18.1 is valid for any triangle for example.

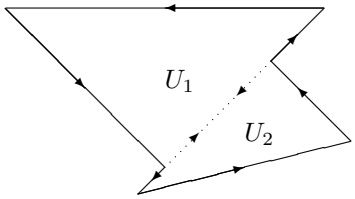
Now suppose 18.1 holds for $U_1, U_2, \cdots, U_m$ and the open sets, U_k have the property that no two have nonempty intersection and their boundaries intersect only in a finite

number of piecewise smooth curves. Then 18.1 must hold for $U \equiv \cup_{i=1}^m U_i$, the union of these sets. This is because

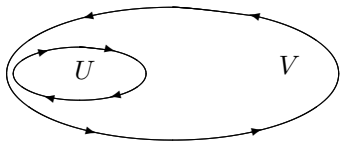
$$\int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA =$$

$$\begin{aligned} &= \sum_{k=1}^m \int_{U_k} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \\ &= \sum_{k=1}^m \int_{\partial U_k} \mathbf{F} \cdot d\mathbf{R} = \int_{\partial U} \mathbf{F} \cdot d\mathbf{R} \end{aligned}$$

because if $\Gamma = \partial U_k \cap \partial U_j$, then its orientation as a part of ∂U_k is opposite to its orientation as a part of ∂U_j and consequently the line integrals over Γ will cancel, points of Γ also not being in ∂U. As an illustration, consider the following picture for two such U_k.



Similarly, if $U \subseteq V$ and if also $\partial U \subseteq V$ and both U and V are open sets for which 18.1 holds, then the open set, $V \setminus (U \cup \partial U)$ consisting of what is left in V after deleting U along with its boundary also satisfies 18.1. Roughly speaking, you can drill holes in a region for which 18.1 holds and get another region for which this continues to hold provided 18.1 holds for the holes. To see why this is so, consider the following picture which typifies the situation just described.



Then

$$\int_{\partial V} \mathbf{F} \cdot d\mathbf{R} = \int_V \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA$$

$$\begin{aligned} &= \int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA + \int_{V \setminus U} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \\ &= \int_{\partial U} \mathbf{F} \cdot d\mathbf{R} + \int_{V \setminus U} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \end{aligned}$$

and so

$$\int_{V \setminus U} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA = \int_{\partial V} \mathbf{F} \cdot d\mathbf{R} - \int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$$

which equals

$$\int_{\partial (V \setminus U)} \mathbf{F} \cdot d\mathbf{R}$$

where ∂V is oriented as shown in the picture. (If you walk around the region, $V \setminus U$ with the area on the left, you get the indicated orientation for this curve.)

You can see that 18.1 is valid quite generally. This verifies the following theorem.

Theorem 18.1.3 *(Green's Theorem) Let U be an open set in the plane and let ∂U be piecewise smooth and let $\mathbf{F}(x,y) = (P(x,y), Q(x,y))$ be a C^1 vector field defined near U. Then it is often[1] the case that*

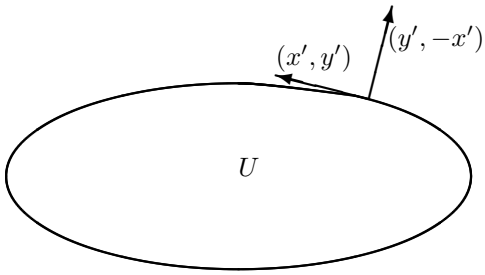
$$\int_{\partial U} \mathbf{F} \cdot d\mathbf{R} = \int_U \left(\frac{\partial Q}{\partial x}(x,y) - \frac{\partial P}{\partial y}(x,y) \right) dA.$$

Here is an alternate proof of Green's theorem from the divergence theorem.

Theorem 18.1.4 *(Green's Theorem) Let U be an open set in the plane and let ∂U be piecewise smooth and let $\mathbf{F}(x,y) = (P(x,y), Q(x,y))$ be a C^1 vector field defined near U. Then it is often the case that*

$$\int_{\partial U} \mathbf{F} \cdot d\mathbf{R} = \int_U \left(\frac{\partial Q}{\partial x}(x,y) - \frac{\partial P}{\partial y}(x,y) \right) dA.$$

Proof: Suppose the divergence theorem holds for U. Consider the following picture.



Since it is assumed that motion around U is counter clockwise, the tangent vector, (x', y') is as shown. Now the unit exterior normal is either

$$\frac{1}{\sqrt{(x')^2 + (y')^2}} (-y', x')$$

or

$$\frac{1}{\sqrt{(x')^2 + (y')^2}} (y', -x')$$

Again, the counter clockwise motion shows the correct unit exterior normal is the second of the above. To see this note that since the area should be on the left as you walk around the edge, you need to have the unit normal point in the direction of $(x', y', 0) \times \mathbf{k}$ which equals $(y', -x', 0)$. Now let $\mathbf{F}(x,y) = (Q(x,y), -P(x,y))$. Also note the area element on ∂U is $\sqrt{(x')^2 + (y')^2}dt$. Suppose the boundary of U consists of m smooth curves, the i^{th} of which is parameterized by (x_i, y_i) with the parameter, $t \in [a_i, b_i]$. Then by the divergence theorem,

$$\int_U (Q_x - P_y)\, dA = \int_U \operatorname{div}(\mathbf{F})\, dA = \int_{\partial U} \mathbf{F} \cdot \mathbf{n} dS$$

[1]For a general version see the advanced calculus book by Apostol. The general versions involve the concept of a rectifiable (finite length) Jordan curve.

$$\begin{aligned}
&= \sum_{i=1}^{m}\int_{a_i}^{b_i}\left(Q\left(x_i\left(t\right),y_i\left(t\right)\right),-P\left(x_i\left(t\right),y_i\left(t\right)\right)\right)\\
&\quad\cdot\frac{1}{\sqrt{\left(x_i'\right)^2+\left(y_i'\right)^2}}\left(y_i',-x_i'\right)\overbrace{\sqrt{\left(x_i'\right)^2+\left(y_i'\right)^2}dt}^{dS}\\
&= \sum_{i=1}^{m}\int_{a_i}^{b_i}\left(Q\left(x_i\left(t\right),y_i\left(t\right)\right),-P\left(x_i\left(t\right),y_i\left(t\right)\right)\right)\cdot\left(y_i',-x_i'\right)dt\\
&= \sum_{i=1}^{m}\int_{a_i}^{b_i}Q\left(x_i\left(t\right),y_i\left(t\right)\right)y_i'\left(t\right)+P\left(x_i\left(t\right),y_i\left(t\right)\right)x_i'\left(t\right)dt\equiv\int_{\partial U}Pdx+Qdy
\end{aligned}$$

This proves Green's theorem from the divergence theorem.

Proposition 18.1.5 *Let U be an open set in $\mathbb{R}^2$ for which Green's theorem holds. Then*

$$\text{Area of } U=\int_{\partial U}\mathbf{F}\cdot d\mathbf{R}$$

where $\mathbf{F}\left(x,y\right)=\frac{1}{2}\left(-y,x\right),\left(0,x\right),$ or $\left(-y,0\right).$

Proof: This follows immediately from Green's theorem.

Example 18.1.6 *Use Proposition 18.1.5 to find the area of the ellipse*

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}\leq 1.$$

You can parameterize the boundary of this ellipse as

$$x=a\cos t,\ y=b\sin t,\ t\in\left[0,2\pi\right].$$

Then from Proposition 18.1.5,

$$\begin{aligned}
\text{Area equals} &= \frac{1}{2}\int_0^{2\pi}\left(-b\sin t,a\cos t\right)\cdot\left(-a\sin t,b\cos t\right)dt\\
&= \frac{1}{2}\int_0^{2\pi}\left(ab\right)dt=\pi ab.
\end{aligned}$$

Example 18.1.7 *Find $\int_{\partial U}\mathbf{F}\cdot d\mathbf{R}$ where U is the set,*

$$\left\{\left(x,y\right):x^2+3y^2\leq 9\right\}$$

and $\mathbf{F}\left(x,y\right)=\left(y,-x\right).$

One way to do this is to parameterize the boundary of U and then compute the line integral directly. It is easier to use Green's theorem. The desired line integral equals

$$\int_U\left(\left(-1\right)-1\right)dA=-2\int_U dA.$$

Now U is an ellipse having area equal to $3\sqrt{3}$ and so the answer is $-6\sqrt{3}$.

Example 18.1.8 *Find $\int_{\partial U}\mathbf{F}\cdot d\mathbf{R}$ where U is the set, $\{(x,y):2\leq x\leq 4,0\leq y\leq 3\}$ and $\mathbf{F}\left(x,y\right)=\left(x\sin y,y^3\cos x\right).$*

From Green's theorem this line integral equals

$$\begin{aligned}&\int_2^4 \int_0^3 \left(-y^3 \sin x - x\cos y\right) dydx \\ = \quad & \frac{81}{4}\cos 4 - 6\sin 3 - \frac{81}{4}\cos 2.\end{aligned}$$

This is much easier than computing the line integral because you don't have to break the boundary in pieces and consider each separately.

Example 18.1.9 *Find* $\int_{\partial U} \mathbf{F}\cdot d\mathbf{R}$ *where* U *is the set,*

$$\{(x,y) : 2 \le x \le 4, x \le y \le 3\}$$

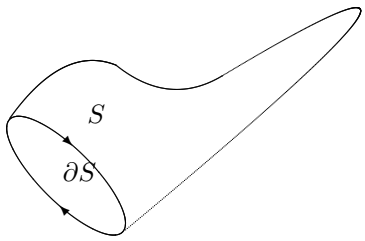
and $\mathbf{F}(x,y) = (x\sin y, y\sin x)$.

From Green's theorem this line integral equals

$$\begin{aligned}&\int_2^4 \int_x^3 (y\cos x - x\cos y)\, dydx \\ = \quad & -\frac{3}{2}\sin 4 - 6\sin 3 - 8\cos 4 - \frac{9}{2}\sin 2 + 4\cos 2.\end{aligned}$$

18.2 Stoke's Theorem From Green's Theorem

Stoke's theorem is a generalization of Green's theorem which relates the integral over a surface to the integral around the boundary of the surface. These terms are a little different from what occurs in $\mathbb{R}^2$. To describe this, consider a sock. The surface is the sock and its boundary will be the edge of the opening of the sock in which you place your foot. Another way to think of this is to imagine a region in $\mathbb{R}^2$ of the sort discussed above for Green's theorem. Suppose it is on a sheet of rubber and the sheet of rubber is stretched in three dimensions. The boundary of the resulting surface is the result of the stretching applied to the boundary of the original region in $\mathbb{R}^2$. Here is a picture describing the situation.



Recall the following definition of the curl of a vector field.

Definition 18.2.1 *Let*

$$\mathbf{F}(x,y,z) = (F_1(x,y,z), F_2(x,y,z), F_3(x,y,z))$$

be a C^1 *vector field defined on an open set,* V *in* $\mathbb{R}^3$. *Then*

$$\nabla\times\mathbf{F} \equiv \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_1 & F_2 & F_3 \end{vmatrix}$$

$$\equiv \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right)\mathbf{i} + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}\right)\mathbf{j} + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\mathbf{k}.$$

This is also called $\operatorname{curl}(\mathbf{F})$ *and written as indicated,* $\nabla\times\mathbf{F}$.

The following lemma gives the fundamental identity which will be used in the proof of Stoke's theorem.

Lemma 18.2.2 *Let $\mathbf{R} : U \to V \subseteq \mathbb{R}^3$ where U is an open subset of $\mathbb{R}^2$ and V is an open subset of $\mathbb{R}^3$. Suppose $\mathbf{R}$ is C^2 and let $\mathbf{F}$ be a C^1 vector field defined in V.*

$$(\mathbf{R}_u \times \mathbf{R}_v) \cdot (\nabla \times \mathbf{F})(\mathbf{R}(u,v)) = ((\mathbf{F} \circ \mathbf{R})_u \cdot \mathbf{R}_v - (\mathbf{F} \circ \mathbf{R})_v \cdot \mathbf{R}_u)(u,v). \tag{18.3}$$

Proof: Start with the left side and let $x_i = R_i(u,v)$ for short.

$$\begin{aligned}
(\mathbf{R}_u \times \mathbf{R}_v) \cdot (\nabla \times \mathbf{F})(\mathbf{R}(u,v)) &= \varepsilon_{ijk} x_{ju} x_{kv} \varepsilon_{irs} \frac{\partial F_s}{\partial x_r} \\
&= (\delta_{jr}\delta_{ks} - \delta_{js}\delta_{kr}) x_{ju} x_{kv} \frac{\partial F_s}{\partial x_r} \\
&= x_{ju} x_{kv} \frac{\partial F_k}{\partial x_j} - x_{ju} x_{kv} \frac{\partial F_j}{\partial x_k} \\
&= \mathbf{R}_v \cdot \frac{\partial (\mathbf{F} \circ \mathbf{R})}{\partial u} - \mathbf{R}_u \cdot \frac{\partial (\mathbf{F} \circ \mathbf{R})}{\partial v}
\end{aligned}$$

which proves 18.3.

The proof of Stoke's theorem given next follows [7]. First, it is convenient to give a definition.

Definition 18.2.3 *A vector valued function, $\mathbf{R} : U \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is said to be in $C^k\left(\overline{U}, \mathbb{R}^n\right)$ if it is the restriction to $\overline{U}$ of a vector valued function which is defined on $\mathbb{R}^m$ and is C^k. That is, this function has continuous partial derivatives up to order k.*

Theorem 18.2.4 *(Stoke's Theorem) Let U be any region in $\mathbb{R}^2$ for which the conclusion of Green's theorem holds and let $\mathbf{R} \in C^2\left(\overline{U}, \mathbb{R}^3\right)$ be a one to one function satisfying $|(\mathbf{R}_u \times \mathbf{R}_v)(u,v)| \neq 0$ for all $(u,v) \in U$ and let S denote the surface,*

$$\begin{aligned}
S &\equiv \{\mathbf{R}(u,v) : (u,v) \in U\}, \\
\partial S &\equiv \{\mathbf{R}(u,v) : (u,v) \in \partial U\}
\end{aligned}$$

where the orientation on ∂S is consistent with the counter clockwise orientation on ∂U (U is on the left as you walk around ∂U). Then for $\mathbf{F}$ a C^1 vector field defined near S,

$$\int_{\partial S} \mathbf{F} \cdot d\mathbf{R} = \int_S \operatorname{curl}(\mathbf{F}) \cdot \mathbf{n} dS$$

where $\mathbf{n}$ is the normal to S defined by

$$\mathbf{n} \equiv \frac{\mathbf{R}_u \times \mathbf{R}_v}{|\mathbf{R}_u \times \mathbf{R}_v|}.$$

Proof: Letting C be an oriented part of ∂U having parameterization,

$$\mathbf{r}(t) \equiv (u(t), v(t))$$

for $t \in [\alpha, \beta]$ and letting $\mathbf{R}(C)$ denote the oriented part of ∂S corresponding to C,

$$\int_{\mathbf{R}(C)} \mathbf{F} \cdot d\mathbf{R} =$$

$$\begin{aligned}
&= \int_\alpha^\beta \mathbf{F}\left(\mathbf{R}\left(u\left(t\right),v\left(t\right)\right)\right)\cdot\left(\mathbf{R}_u u'\left(t\right)+\mathbf{R}_v v'\left(t\right)\right)dt\\
&= \int_\alpha^\beta \mathbf{F}\left(\mathbf{R}\left(u\left(t\right),v\left(t\right)\right)\right)\mathbf{R}_u\left(u\left(t\right),v\left(t\right)\right)u'\left(t\right)dt\\
&\quad+\int_\alpha^\beta \mathbf{F}\left(\mathbf{R}\left(u\left(t\right),v\left(t\right)\right)\right)\mathbf{R}_v\left(u\left(t\right),v\left(t\right)\right)v'\left(t\right)dt\\
&= \int_C \left(\left(\mathbf{F}\circ\mathbf{R}\right)\cdot\mathbf{R}_u,\left(\mathbf{F}\circ\mathbf{R}\right)\cdot\mathbf{R}_v\right)\cdot d\mathbf{r}.
\end{aligned}$$

Since this holds for each such piece of ∂U, it follows

$$\int_{\partial S}\mathbf{F}\cdot d\mathbf{R}=\int_{\partial U}\left(\left(\mathbf{F}\circ\mathbf{R}\right)\cdot\mathbf{R}_u,\left(\mathbf{F}\circ\mathbf{R}\right)\cdot\mathbf{R}_v\right)\cdot d\mathbf{r}.$$

By the assumption that the conclusion of Green's theorem holds for U, this equals

$$\begin{aligned}
&\int_U\left[\left(\left(\mathbf{F}\circ\mathbf{R}\right)\cdot\mathbf{R}_v\right)_u-\left(\left(\mathbf{F}\circ\mathbf{R}\right)\cdot\mathbf{R}_u\right)_v\right]dA\\
&= \int_U\left[\left(\mathbf{F}\circ\mathbf{R}\right)_u\cdot\mathbf{R}_v+\left(\mathbf{F}\circ\mathbf{R}\right)\cdot\mathbf{R}_{vu}-\left(\mathbf{F}\circ\mathbf{R}\right)\cdot\mathbf{R}_{uv}-\left(\mathbf{F}\circ\mathbf{R}\right)_v\cdot\mathbf{R}_u\right]dA\\
&= \int_U\left[\left(\mathbf{F}\circ\mathbf{R}\right)_u\cdot\mathbf{R}_v-\left(\mathbf{F}\circ\mathbf{R}\right)_v\cdot\mathbf{R}_u\right]dA
\end{aligned}$$

the last step holding by equality of mixed partial derivatives, a result of the assumption that $\mathbf{R}$ is C^2. Now by Lemma 18.2.2, this equals

$$\begin{aligned}
&\int_U\left(\mathbf{R}_u\times\mathbf{R}_v\right)\cdot\left(\nabla\times\mathbf{F}\right)dA\\
&= \int_U\nabla\times\mathbf{F}\cdot\left(\mathbf{R}_u\times\mathbf{R}_v\right)dA\\
&= \int_S\nabla\times\mathbf{F}\cdot\mathbf{n}dS
\end{aligned}$$

because $dS=\left|\left(\mathbf{R}_u\times\mathbf{R}_v\right)\right|dA$ and $\mathbf{n}=\frac{\left(\mathbf{R}_u\times\mathbf{R}_v\right)}{\left|\left(\mathbf{R}_u\times\mathbf{R}_v\right)\right|}$. Thus

$$\begin{aligned}
\left(\mathbf{R}_u\times\mathbf{R}_v\right)dA &= \frac{\left(\mathbf{R}_u\times\mathbf{R}_v\right)}{\left|\left(\mathbf{R}_u\times\mathbf{R}_v\right)\right|}\left|\left(\mathbf{R}_u\times\mathbf{R}_v\right)\right|dA\\
&= \mathbf{n}dS.
\end{aligned}$$

This proves Stoke's theorem.

Note that there is no mention made in the final result that $\mathbf{R}$ is C^2. Therefore, it is not surprising that versions of this theorem are valid in which this assumption is not present. It is possible to obtain extremely general versions of Stoke's theorem if you use the Lebesgue integral.

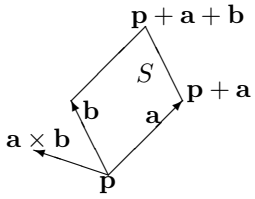
18.2.1 The Normal And The Orientation

Stoke's theorem as just presented needs no apology. However, it is helpful in applications to have some additional geometric insight.

To begin with, suppose the surface S of interest is a parallelogram in $\mathbb{R}^3$ determined by the two vectors $\mathbf{a},\mathbf{b}$. Thus $S=\mathbf{R}\left(Q\right)$ where $Q=[0,1]\times[0,1]$ is the unit square and for $(u,v)\in Q$,

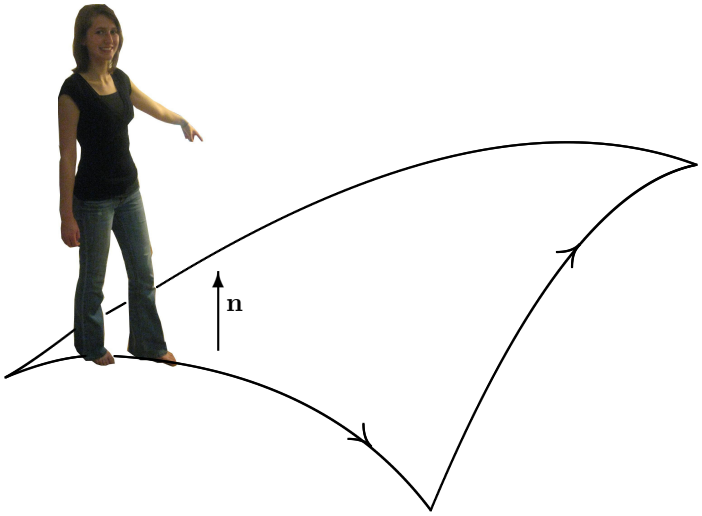
$$\mathbf{R}\left(u,v\right)\equiv u\mathbf{a}+v\mathbf{b}+\mathbf{p},$$

the point $\mathbf{p}$ being a corner of the parallelogram S. Then orient ∂S consistent with the counter clockwise orientation on ∂Q. Thus, following this orientation on S you go from $\mathbf{p}$ to $\mathbf{p}+\mathbf{a}$ to $\mathbf{p}+\mathbf{a}+\mathbf{b}$ to $\mathbf{p}+\mathbf{b}$ to $\mathbf{p}$. Then Stoke's theorem implies that with this orientation on ∂S, $\int_{\partial S}\mathbf{F}\cdot\mathbf{dR}=\int_S \nabla\times\mathbf{F}\cdot\mathbf{n}ds$ where $\mathbf{n}=\mathbf{R}_u\times\mathbf{R}_v/\left|\mathbf{R}_u\times\mathbf{R}_v\right|=\mathbf{a}\times\mathbf{b}/\left|\mathbf{a}\times\mathbf{b}\right|$. Now recall $\mathbf{a},\mathbf{b},\mathbf{a}\times\mathbf{b}$ forms a right hand system.



Thus, if you were walking around ∂S in the direction of the orientation with your left hand over the surface S, the normal vector $\mathbf{a}\times\mathbf{b}$ would be pointing in the direction of your head.

More generally, if S is a surface which is not necessarily a parallelogram but is instead as described in Theorem 18.2.4, you could consider a **small** rectangle Q contained in U and orient the boundary of $\mathbf{R}(Q)$ as described in that theorem. Then if the rectangle is small enough, as you walk around $\partial\mathbf{R}(Q)$ in the direction of the described orientation, your head would point roughly in the direction of $\mathbf{R}_u\times\mathbf{R}_v$. This is because for small enough Q, the normal to the tangent parallelogram would point in roughly the same direction as $\mathbf{R}_u\times\mathbf{R}_v$ at each point of $\mathbf{R}(Q)$ and your head would also point roughly in the same direction if you were on $\mathbf{R}(Q)$ or the tangent parallelogram. You can imagine essentially filling U with non overlapping rectangles, Q_i. Then orienting $\partial\mathbf{R}(Q_i)$ consistent with the counter clockwise orientation on Q_i and adding the resulting line integrals, the line integrals over the common sides cancel and the result is essentially the line integral over ∂S.Thus there is a simple relation between the field of normal vectors on S and the orientation of ∂S. It is simply this. If you walk along ∂S in the direction mandated by the orientation, with your left hand over the surface, the nearby normal vectors in Stoke's theorem will point roughly in the direction of your head.

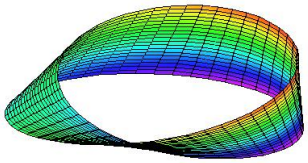


This also illustrates that you can **define** an orientation for ∂S by specifying a field of normal vectors for the surface which varies continuously over the surface, and require that the motion over the boundary of the surface is such that your head points roughly in the direction of nearby normal vectors and your left hand is over the surface. The existence of such a continuous field of normal vectors is what constitutes an orientable surface.

18.2.2 The Mobeus Band

It turns out there are more general formulations of Stoke's theorem than what is presented above. However, it is always necessary for the surface, S to be **orientable**. This means it is possible to obtain a vector field for a unit normal to the surface which is a continuous function of position on S.

An example of a surface which is not orientable is the famous Mobeus band, obtained by taking a long rectangular piece of paper and glueing the ends together after putting a twist in it. Here is a picture of one.



There is something quite interesting about this Mobeus band and this is that it can be written parametrically with a simple parameter domain. The picture above is a maple graph of the parametrically defined surface

$$\mathbf{R}(\theta, v) \equiv \begin{cases} x = 4\cos\theta + v\cos\frac{\theta}{2} \\ y = 4\sin\theta + v\cos\frac{\theta}{2} \\ z = v\sin\frac{\theta}{2} \end{cases}, \theta \in [0, 2\pi], v \in [-1, 1].$$

An obvious question is why the normal vector, $\mathbf{R}_{,\theta} \times \mathbf{R}_{,v} / |\mathbf{R}_{,\theta} \times \mathbf{R}_{,v}|$ is not a continuous function of position on S. You can see easily that it is a continuous function of both θ and v. However, the map, $\mathbf{R}$ is not one to one. In fact, $\mathbf{R}(0,0) = \mathbf{R}(2\pi, 0)$. Therefore, near this point on S, there are two different values for the above normal vector. In fact, a tedious computation will show this normal vector is

$$\frac{\left(4\sin\frac{1}{2}\theta\cos\theta - \frac{1}{2}v, 4\sin\frac{1}{2}\theta\sin\theta + \frac{1}{2}v, -8\cos^2\frac{1}{2}\theta\sin\frac{1}{2}\theta - 8\cos^3\frac{1}{2}\theta + 4\cos\frac{1}{2}\theta\right)}{D}$$

where

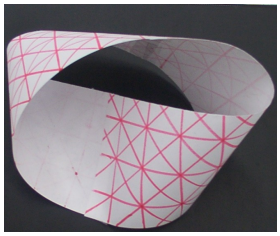
$$\begin{aligned} D &= \left(16\sin^2\left(\frac{\theta}{2}\right) + \frac{v^2}{2} + 4\sin\left(\frac{\theta}{2}\right)v(\sin\theta - \cos\theta)\right. \\ &\quad \left. + 4^3\cos^2\left(\frac{\theta}{2}\right)\left(\cos\left(\frac{1}{2}\theta\right)\sin\left(\frac{1}{2}\theta\right) + \cos^2\left(\frac{1}{2}\theta\right) - \frac{1}{2}\right)^2\right) \end{aligned}$$

and you can verify that the denominator will not vanish. Letting $v = 0$ and $\theta = 0$ and 2π yields the two vectors, $(0,0,-1), (0,0,1)$ so there is a discontinuity. This is why I was careful to say in the statement of Stoke's theorem given above that $\mathbf{R}$ is one to one.

The Mobeus band has some usefulness. In old machine shops the equipment was run by a belt which was given a twist to spread the surface wear on the belt over twice the area.

The above explanation shows that $\mathbf{R}_{,\theta} \times \mathbf{R}_{,v} / |\mathbf{R}_{,\theta} \times \mathbf{R}_{,v}|$ fails to deliver an orientation for the Mobeus band. However, this does not answer the question whether there

is some orientation for it other than this one. In fact there is none. You can see this by looking at the first of the two pictures below or by making one and tracing it with a pencil. There is only one side to the Mobeus band. An oriented surface must have two sides, one side identified by the given unit normal which varies continuously over the surface and the other side identified by the negative of this normal. The second picture below was taken by Ouyang when he was at meetings in Paris and saw it at a museum.



18.2.3 Conservative Vector Fields

Definition 18.2.5 *A vector field,* $\mathbf{F}$ *defined in a three dimensional region is said to be* ***conservative***[2] *if for every piecewise smooth closed curve,* C*, it follows* $\int_C \mathbf{F}\cdot d\mathbf{R} = 0$.

Definition 18.2.6 *Let* $(\mathbf{x}, \mathbf{p}_1, \cdots, \mathbf{p}_n, \mathbf{y})$ *be an ordered list of points in* $\mathbb{R}^p$*. Let*

$$\mathbf{p}(\mathbf{x}, \mathbf{p}_1, \cdots, \mathbf{p}_n, \mathbf{y})$$

denote the piecewise smooth curve consisting of a straight line segment from $\mathbf{x}$ *to* $\mathbf{p}_1$ *and then the straight line segment from* $\mathbf{p}_1$ *to* $\mathbf{p}_2 \cdots$ *and finally the straight line segment from* $\mathbf{p}_n$ *to* $\mathbf{y}$*. This is called a* ***polygonal curve****. An open set in* $\mathbb{R}^p$*,* U*, is said to be a* ***region*** *if it has the property that for any two points,* $\mathbf{x}, \mathbf{y} \in U$*, there exists a polygonal curve joining the two points.*

Conservative vector fields are important because of the following theorem, sometimes called the fundamental theorem for line integrals.

Theorem 18.2.7 *Let* U *be a region in* $\mathbb{R}^p$ *and let* $\mathbf{F} : U \to \mathbb{R}^p$ *be a continuous vector field. Then* $\mathbf{F}$ *is conservative if and only if there exists a scalar valued function of* p *variables,* ϕ *such that* $\mathbf{F} = \nabla\phi$*. Furthermore, if* C *is an oriented curve which goes from* $\mathbf{x}$ *to* $\mathbf{y}$ *in* U*, then*

$$\int_C \mathbf{F}\cdot d\mathbf{R} = \phi(\mathbf{y}) - \phi(\mathbf{x}). \tag{18.4}$$

Thus the line integral is path independent in this case. This function, ϕ *is called a* ***scalar potential*** *for* $\mathbf{F}$*.*

Proof: To save space and fussing over things which are unimportant, denote by $\mathbf{p}(\mathbf{x}_0, \mathbf{x})$ a polygonal curve from $\mathbf{x}_0$ to $\mathbf{x}$. Thus the orientation is such that it goes from $\mathbf{x}_0$ to $\mathbf{x}$. The curve $\mathbf{p}(\mathbf{x}, \mathbf{x}_0)$ denotes the same set of points but in the opposite order. Suppose first $\mathbf{F}$ is conservative. Fix $\mathbf{x}_0 \in U$ and let

$$\phi(\mathbf{x}) \equiv \int_{\mathbf{p}(\mathbf{x}_0,\mathbf{x})} \mathbf{F}\cdot d\mathbf{R}.$$

This is well defined because if $\mathbf{q}(\mathbf{x}_0, \mathbf{x})$ is another polygonal curve joining $\mathbf{x}_0$ to $\mathbf{x}$, Then the curve obtained by following $\mathbf{p}(\mathbf{x}_0, \mathbf{x})$ from $\mathbf{x}_0$ to $\mathbf{x}$ and then from $\mathbf{x}$ to $\mathbf{x}_0$ along

[2] There is no such thing as a liberal vector field.

$\mathbf{q}(\mathbf{x},\mathbf{x}_0)$ is a closed piecewise smooth curve and so by assumption, the line integral along this closed curve equals 0. However, this integral is just

$$\int_{\mathbf{p}(\mathbf{x}_0,\mathbf{x})} \mathbf{F}\cdot d\mathbf{R} + \int_{\mathbf{q}(\mathbf{x},\mathbf{x}_0)} \mathbf{F}\cdot d\mathbf{R} = \int_{\mathbf{p}(\mathbf{x}_0,\mathbf{x})} \mathbf{F}\cdot d\mathbf{R} - \int_{\mathbf{q}(\mathbf{x}_0,\mathbf{x})} \mathbf{F}\cdot d\mathbf{R}$$

which shows

$$\int_{\mathbf{p}(\mathbf{x}_0,\mathbf{x})} \mathbf{F}\cdot d\mathbf{R} = \int_{\mathbf{q}(\mathbf{x}_0,\mathbf{x})} \mathbf{F}\cdot d\mathbf{R}$$

and that ϕ is well defined. For small t,

$$\begin{aligned}\frac{\phi(\mathbf{x}+t\mathbf{e}_i)-\phi(\mathbf{x})}{t} &= \frac{\int_{\mathbf{p}(\mathbf{x}_0,\mathbf{x}+t\mathbf{e}_i)}\mathbf{F}\cdot d\mathbf{R} - \int_{\mathbf{p}(\mathbf{x}_0,\mathbf{x})}\mathbf{F}\cdot d\mathbf{R}}{t}\\ &= \frac{\int_{\mathbf{p}(\mathbf{x}_0,\mathbf{x})}\mathbf{F}\cdot d\mathbf{R} + \int_{\mathbf{p}(\mathbf{x},\mathbf{x}+t\mathbf{e}_i)}\mathbf{F}\cdot d\mathbf{R} - \int_{\mathbf{p}(\mathbf{x}_0,\mathbf{x})}\mathbf{F}\cdot d\mathbf{R}}{t}.\end{aligned}$$

Since U is open, for small t, the ball of radius $|t|$ centered at $\mathbf{x}$ is contained in U. Therefore, the line segment from $\mathbf{x}$ to $\mathbf{x}+t\mathbf{e}_i$ is also contained in U and so one can take $\mathbf{p}(\mathbf{x},\mathbf{x}+t\mathbf{e}_i)(s) = \mathbf{x}+s(t\mathbf{e}_i)$ for $s\in[0,1]$. Therefore, the above difference quotient reduces to

$$\begin{aligned}\frac{1}{t}\int_0^1 \mathbf{F}(\mathbf{x}+s(t\mathbf{e}_i))\cdot t\mathbf{e}_i\, ds &= \int_0^1 F_i(\mathbf{x}+s(t\mathbf{e}_i))\, ds\\ &= F_i(\mathbf{x}+s_t(t\mathbf{e}_i))\end{aligned}$$

by the mean value theorem for integrals. Here s_t is some number between 0 and 1. By continuity of $\mathbf{F}$, this converges to $F_i(\mathbf{x})$ as $t\to 0$. Therefore, $\nabla\phi = \mathbf{F}$ as claimed.

Conversely, if $\nabla\phi = \mathbf{F}$, then if $\mathbf{R}:[a,b]\to\mathbb{R}^p$ is any C^1 curve joining $\mathbf{x}$ to $\mathbf{y}$,

$$\begin{aligned}\int_a^b \mathbf{F}(\mathbf{R}(t))\cdot\mathbf{R}'(t)\, dt &= \int_a^b \nabla\phi(\mathbf{R}(t))\cdot\mathbf{R}'(t)\, dt\\ &= \int_a^b \frac{d}{dt}(\phi(\mathbf{R}(t)))\, dt\\ &= \phi(\mathbf{R}(b))-\phi(\mathbf{R}(a))\\ &= \phi(\mathbf{y})-\phi(\mathbf{x})\end{aligned}$$

and this verifies 18.4 in the case where the curve joining the two points is smooth. The general case follows immediately from this by using this result on each of the pieces of the piecewise smooth curve. For example if the curve goes from $\mathbf{x}$ to $\mathbf{p}$ and then from $\mathbf{p}$ to $\mathbf{y}$, the above would imply the integral over the curve from $\mathbf{x}$ to $\mathbf{p}$ is $\phi(\mathbf{p})-\phi(\mathbf{x})$ while from $\mathbf{p}$ to $\mathbf{y}$ the integral would yield $\phi(\mathbf{y})-\phi(\mathbf{p})$. Adding these gives $\phi(\mathbf{y})-\phi(\mathbf{x})$. The formula 18.4 implies the line integral over any closed curve equals zero because the starting and ending points of such a curve are the same. This proves the theorem.

Example 18.2.8 *Let* $\mathbf{F}(x,y,z) = (\cos x - yz\sin(xz), \cos(xz), -yx\sin(xz))$. *Let* C *be a piecewise smooth curve which goes from* $(\pi,1,1)$ *to* $\left(\frac{\pi}{2},3,2\right)$. *Find* $\int_C \mathbf{F}\cdot d\mathbf{R}$.

The specifics of the curve are not given so the problem is nonsense unless the vector field is conservative. Therefore, it is reasonable to look for the function, ϕ satisfying $\nabla\phi = \mathbf{F}$. Such a function satisfies

$$\phi_x = \cos x - y(\sin xz)\, z$$

and so, assuming ϕ exists,

$$\phi(x,y,z) = \sin x + y\cos(xz) + \psi(y,z).$$

I have to add in the most general thing possible, $\psi(y,z)$ to ensure possible solutions are not being thrown out. It wouldn't be good at this point to add in a constant since the answer could involve a function of either or both of the other variables. Now from what was just obtained,

$$\phi_y = \cos(xz) + \psi_y = \cos xz$$

and so it is possible to take $\psi_y = 0$. Consequently, ϕ, if it exists is of the form

$$\phi(x,y,z) = \sin x + y\cos(xz) + \psi(z).$$

Now differentiating this with respect to z gives

$$\phi_z = -yx\sin(xz) + \psi_z = -yx\sin(xz)$$

and this shows ψ does not depend on z either. Therefore, it suffices to take $\psi = 0$ and

$$\phi(x,y,z) = \sin(x) + y\cos(xz).$$

Therefore, the desired line integral equals

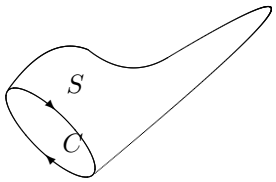
$$\sin\left(\frac{\pi}{2}\right) + 3\cos(\pi) - (\sin(\pi) + \cos(\pi)) = -1.$$

The above process for finding ϕ will not lead you astray in the case where there does not exist a scalar potential. As an example, consider the following.

Example 18.2.9 *Let* $\mathbf{F}(x,y,z) = \left(x, y^2x, z\right)$. *Find a scalar potential for* $\mathbf{F}$ *if it exists.*

If ϕ exists, then $\phi_x = x$ and so $\phi = \frac{x^2}{2} + \psi(y,z)$. Then $\phi_y = \psi_y(y,z) = xy^2$ but this is impossible because the left side depends only on y and z while the right side depends also on x. Therefore, this vector field is not conservative and there does not exist a scalar potential.

Definition 18.2.10 *A set of points in three dimensional space, V is simply connected if every piecewise smooth closed curve, C is the edge of a surface, S which is contained entirely within V in such a way that Stokes theorem holds for the surface, S and its edge, C.*



This is like a sock. The surface is the sock and the curve, C goes around the opening of the sock.

As an application of Stoke's theorem, here is a useful theorem which gives a way to check whether a vector field is conservative.

Theorem 18.2.11 *For a three dimensional simply connected open set, V and $\mathbf{F}$ a C^1 vector field defined in V, $\mathbf{F}$ is conservative if $\nabla \times \mathbf{F} = \mathbf{0}$ in V.*

Proof: If $\nabla \times \mathbf{F} = \mathbf{0}$ then taking an arbitrary closed curve, C, and letting S be a surface bounded by C which is contained in V, Stoke's theorem implies

$$0 = \int_S \nabla \times \mathbf{F} \cdot \mathbf{n}\, dA = \int_C \mathbf{F} \cdot d\mathbf{R}.$$

Thus $\mathbf{F}$ is conservative.

Example 18.2.12 *Determine whether the vector field,*

$$\left(4x^3 + 2\left(\cos\left(x^2+z^2\right)\right)x, 1, 2\left(\cos\left(x^2+z^2\right)\right)z\right)$$

is conservative.

Since this vector field is defined on all of $\mathbb{R}^3$, it only remains to take its curl and see if it is the zero vector.

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ 4x^3 + 2\left(\cos\left(x^2+z^2\right)\right)x & 1 & 2\left(\cos\left(x^2+z^2\right)\right)z \end{vmatrix}.$$

This is obviously equal to zero. Therefore, the given vector field is conservative. Can you find a potential function for it? Let ϕ be the potential function. Then $\phi_z = 2\left(\cos\left(x^2+z^2\right)\right)z$ and so $\phi(x,y,z) = \sin\left(x^2+z^2\right) + g(x,y)$. Now taking the derivative of ϕ with respect to y, you see $g_y = 1$ so $g(x,y) = y + h(x)$. Hence $\phi(x,y,z) = y + g(x) + \sin\left(x^2+z^2\right)$. Taking the derivative with respect to x, you get $4x^3 + 2\left(\cos\left(x^2+z^2\right)\right)x = g'(x) + 2x\cos\left(x^2+z^2\right)$ and so it suffices to take $g(x) = x^4$. Hence $\phi(x,y,z) = y + x^4 + \sin\left(x^2+z^2\right)$.

18.2.4 Some Terminology

If $\mathbf{F} = (P,Q,R)$ is a vector field. Then the statement that $\mathbf{F}$ is conservative is the same as saying the differential form $Pdx + Qdy + Rdz$ is exact. Some people like to say things in terms of vector fields and some say it in terms of differential forms. In Example 18.2.12, the differential form $\left(4x^3 + 2\left(\cos\left(x^2+z^2\right)\right)x\right)dx + dy + \left(2\left(\cos\left(x^2+z^2\right)\right)z\right)dz$ is exact.

18.2.5 Maxwell's Equations And The Wave Equation

Many of the ideas presented above are useful in analyzing Maxwell's equations. These equations are derived in advanced physics courses. They are

$$\nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = \mathbf{0} \tag{18.5}$$

$$\nabla \cdot \mathbf{E} = 4\pi\rho \tag{18.6}$$

$$\nabla \times \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{f} \tag{18.7}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{18.8}$$

and it is assumed these hold on all of $\mathbb{R}^3$ to eliminate technical considerations having to do with whether something is simply connected.

In these equations, $\mathbf{E}$ is the electrostatic field and $\mathbf{B}$ is the magnetic field while ρ and $\mathbf{f}$ are sources. By 18.8 $\mathbf{B}$ has a vector potential, $\mathbf{A}_1$ such that $\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}_1$. Now go to 18.5 and write

$$\nabla \times \mathbf{E} + \frac{1}{c}\nabla \times \frac{\partial \mathbf{A}_1}{\partial t} = \mathbf{0}$$

showing that

$$\nabla \times \left(\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}_1}{\partial t} \right) = \mathbf{0}$$

It follows $\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}_1}{\partial t}$ has a scalar potential, ψ_1 satisfying

$$\nabla \psi_1 = \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}_1}{\partial t}. \tag{18.9}$$

Now suppose ϕ is a time dependent scalar field satisfying

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = \frac{1}{c} \frac{\partial \psi_1}{\partial t} - \nabla \cdot \mathbf{A}_1. \tag{18.10}$$

Next define

$$\mathbf{A} \equiv \mathbf{A}_1 + \nabla \phi, \quad \psi \equiv \psi_1 + \frac{1}{c} \frac{\partial \phi}{\partial t}. \tag{18.11}$$

Therefore, in terms of the new variables, 18.10 becomes

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = \frac{1}{c} \left(\frac{\partial \psi}{\partial t} - \frac{1}{c} \frac{\partial^2 \phi}{\partial t^2} \right) - \nabla \cdot A + \nabla^2 \phi$$

which yields

$$0 = \frac{\partial \psi}{\partial t} - c \nabla \cdot A. \tag{18.12}$$

Then it follows from Theorem 17.1.3 on Page 370 that $\mathbf{A}$ is also a vector potential for $\mathbf{B}$. That is

$$\nabla \times \mathbf{A} = \mathbf{B}. \tag{18.13}$$

From 18.9

$$\nabla \left(\psi - \frac{1}{c} \frac{\partial \phi}{\partial t} \right) = \mathbf{E} + \frac{1}{c} \left(\frac{\partial A}{\partial t} - \nabla \frac{\partial \phi}{\partial t} \right)$$

and so

$$\nabla \psi = \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}. \tag{18.14}$$

Using 18.7 and 18.14,

$$\nabla \times (\nabla \times \mathbf{A}) - \frac{1}{c} \frac{\partial}{\partial t} \left(\nabla \psi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right) = \frac{4\pi}{c} \mathbf{f}. \tag{18.15}$$

Now from Theorem 17.1.3 on Page 370 this implies

$$\nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} - \nabla \left(\frac{1}{c} \frac{\partial \psi}{\partial t} \right) + \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = \frac{4\pi}{c} \mathbf{f}$$

and using 18.12, this gives

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} = \frac{4\pi}{c} \mathbf{f}. \tag{18.16}$$

Also from 18.14, 18.6, and 18.12,

$$\begin{aligned} \nabla^2 \psi &= \nabla \cdot \mathbf{E} + \frac{1}{c} \frac{\partial}{\partial t} (\nabla \cdot \mathbf{A}) \\ &= 4\pi \rho + \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} \end{aligned}$$

and so

$$\frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi = -4\pi \rho. \tag{18.17}$$

This is very interesting. If a solution to the wave equations, 18.17, and 18.16 can be found along with a solution to 18.12, then letting the magnetic field be given by 18.13 and letting $\mathbf{E}$ be given by 18.14 the result is a solution to Maxwells equations. This is significant because wave equations are easier to think of than Maxwell's equations. Note the above argument also showed that it is always possible, by solving another wave equation, to get 18.12 to hold.

18.3 Exercises

1. Determine whether the vector field,
$$\left(2xy^3 \sin z^4, 3x^2y^2 \sin z^4 + 1, 4x^2y^3 \left(\cos z^4\right) z^3 + 1\right)$$
is conservative. If it is conservative, find a potential function.

2. Determine whether the vector field,
$$\left(2xy^3 \sin z + y^2 + z, 3x^2y^2 \sin z + 2xy, x^2y^3 \cos z + x\right)$$
is conservative. If it is conservative, find a potential function.

3. Determine whether the vector field, $\left(2xy^3 \sin z + z, 3x^2y^2 \sin z + 2xy, x^2y^3 \cos z + x\right)$ is conservative. If it is conservative, find a potential function.

4. Find scalar potentials for the following vector fields if it is possible to do so. If it is not possible to do so, explain why.
   (a) $\left(y^2, 2xy + \sin z, 2z + y \cos z\right)$
   (b) $\left(2z \left(\cos\left(x^2 + y^2\right)\right) x, 2z \left(\cos\left(x^2 + y^2\right)\right) y, \sin\left(x^2 + y^2\right) + 2z\right)$
   (c) $\left(f(x), g(y), h(z)\right)$
   (d) $\left(xy, z^2, y^3\right)$
   (e) $\left(z + 2\frac{x}{x^2+y^2+1}, 2\frac{y}{x^2+y^2+1}, x + 3z^2\right)$

5. If a vector field is not conservative on the set U, is it possible the same vector field could be conservative on some subset of U? Explain and give examples if it is possible. If it is not possible also explain why.

6. Prove that if a vector field, $\mathbf{F}$ has a scalar potential, then it has infinitely many scalar potentials.

7. Here is a vector field: $\mathbf{F} \equiv \left(2xy, x^2 - 5y^4, 3z^2\right)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$ where C is a curve which goes from $(1, 2, 3)$ to $(4, -2, 1)$.

8. Here is a vector field: $\mathbf{F} \equiv \left(2xy, x^2 - 5y^4, 3\left(\cos z^3\right) z^2\right)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$ where C is a curve which goes from $(1, 0, 1)$ to $(-4, -2, 1)$.

9. Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set, $\{(x, y) : 2 \leq x \leq 4, 0 \leq y \leq x\}$ and $\mathbf{F}(x, y) = (x \sin y, y \sin x)$.

10. Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set, $\left\{(x, y) : 2 \leq x \leq 3, 0 \leq y \leq x^2\right\}$ and $\mathbf{F}(x, y) = (x \cos y, y + x)$.

11. Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set, $\{(x, y) : 1 \leq x \leq 2, x \leq y \leq 3\}$ and $\mathbf{F}(x, y) = (x \sin y, y \sin x)$.

12. Find $\int_{\partial U} \mathbf{F}\cdot d\mathbf{R}$ where U is the set, $\left\{(x,y) : x^2+y^2 \leq 2\right\}$ and $\mathbf{F}(x,y) = \left(-y^3, x^3\right)$.

13. Show that for many open sets in $\mathbb{R}^2$, Area of $U = \int_{\partial U} x dy$, and Area of $U = \int_{\partial U} -y dx$ and Area of $U = \frac{1}{2}\int_{\partial U} -y dx + x dy$. **Hint:** Use Green's theorem.

14. Two smooth oriented surfaces, S_1 and S_2 intersect in a piecewise smooth oriented closed curve, C. Let $\mathbf{F}$ be a C^1 vector field defined on $\mathbb{R}^3$. Explain why $\int_{S_1} \text{curl}(\mathbf{F})\cdot \mathbf{n}\, dS = \int_{S_2} \text{curl}(\mathbf{F})\cdot \mathbf{n}\, dS$. Here $\mathbf{n}$ is the normal to the surface which corresponds to the given orientation of the curve, C.

15. Show that $\text{curl}(\psi\nabla\phi) = \nabla\psi\times\nabla\phi$ and explain why $\int_S \nabla\psi\times\nabla\phi\cdot\mathbf{n}\, dS = \int_{\partial S}(\psi\nabla\phi)\cdot d\mathbf{r}$.

16. Find a simple formula for $\text{div}(\nabla(u^\alpha))$ where $\alpha \in \mathbb{R}$.

17. Parametric equations for one arch of a cycloid are given by $x = a(t-\sin t)$ and $y = a(1-\cos t)$ where here $t \in [0, 2\pi]$. Sketch a rough graph of this arch of a cycloid and then find the area between this arch and the x axis. **Hint:** This is very easy using Green's theorem and the vector field, $\mathbf{F} = (-y, x)$.

18. Let $\mathbf{r}(t) = \left(\cos^3(t), \sin^3(t)\right)$ where $t \in [0, 2\pi]$. Sketch this curve and find the area enclosed by it using Green's theorem.

19. Consider the vector field, $\left(\frac{-y}{(x^2+y^2)}, \frac{x}{(x^2+y^2)}, 0\right) = \mathbf{F}$. Show that $\nabla\times\mathbf{F} = \mathbf{0}$ but that for the closed curve, whose parameterization is $\mathbf{R}(t) = (\cos t, \sin t, 0)$ for $t \in [0, 2\pi]$, $\int_C \mathbf{F}\cdot d\mathbf{R} \neq 0$. Therefore, the vector field is not conservative. Does this contradict Theorem 18.2.11? Explain.

20. Let $\mathbf{x}$ be a point of $\mathbb{R}^3$ and let $\mathbf{n}$ be a unit vector. Let D_r be the circular disk of radius r containing $\mathbf{x}$ which is perpendicular to $\mathbf{n}$. Placing the tail of $\mathbf{n}$ at $\mathbf{x}$ and viewing D_r from the point of $\mathbf{n}$, orient ∂D_r in the counter clockwise direction. Now suppose $\mathbf{F}$ is a vector field defined near $\mathbf{x}$. Show $\text{curl}(\mathbf{F})\cdot\mathbf{n} = \lim_{r\to 0}\frac{1}{\pi r^2}\int_{\partial D_r}\mathbf{F}\cdot d\mathbf{R}$. This last integral is sometimes called the circulation density of $\mathbf{F}$. Explain how this shows that $\text{curl}(\mathbf{F})\cdot\mathbf{n}$ measures the tendency for the vector field to "curl" around the point, the vector $\mathbf{n}$ at the point $\mathbf{x}$.

21. The cylinder $x^2+y^2 = 4$ is intersected with the plane $x+y+z = 2$. This yields a closed curve, C. Orient this curve in the counter clockwise direction when viewed from a point high on the z axis. Let $\mathbf{F} = \left(x^2y, z+y, x^2\right)$. Find $\int_C \mathbf{F}\cdot d\mathbf{R}$.

22. The cylinder $x^2+4y^2 = 4$ is intersected with the plane $x+3y+2z = 1$. This yields a closed curve, C. Orient this curve in the counter clockwise direction when viewed from a point high on the z axis. Let $\mathbf{F} = \left(y, z+y, x^2\right)$. Find $\int_C \mathbf{F}\cdot d\mathbf{R}$.

23. The cylinder $x^2+y^2 = 4$ is intersected with the plane $x+3y+2z = 1$. This yields a closed curve, C. Orient this curve in the clockwise direction when viewed from a point high on the z axis. Let $\mathbf{F} = (y, z+y, x)$. Find $\int_C \mathbf{F}\cdot d\mathbf{R}$.

24. Let $\mathbf{F} = \left(xz, z^2(y+\sin x), z^3y\right)$. Find the surface integral, $\int_S \text{curl}(\mathbf{F})\cdot\mathbf{n}dA$ where S is the surface, $z = 4-\left(x^2+y^2\right)$, $z \geq 0$.

25. Let $\mathbf{F} = \left(xz, \left(y^3+x\right), z^3y\right)$. Find the surface integral, $\int_S \text{curl}(\mathbf{F})\cdot\mathbf{n}dA$ where S is the surface, $z = 16-\left(x^2+y^2\right)$, $z \geq 0$.

26. The cylinder $z = y^2$ intersects the surface $z = 8 - x^2 - 4y^2$ in a curve, C which is oriented in the counter clockwise direction when viewed high on the z axis. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$ if $\mathbf{F} = \left(\frac{z^2}{2}, xy, xz\right)$. **Hint:** This is not too hard if you show you can use Stokes theorem on a domain in the xy plane.

27. Suppose solutions have been found to 18.17, 18.16, and 18.12. Then define $\mathbf{E}$ and $\mathbf{B}$ using 18.14 and 18.13. Verify Maxwell's equations hold for $\mathbf{E}$ and $\mathbf{B}$.

28. Suppose now you have found solutions to 18.17 and 18.16, ψ_1 and A_1. Then go show again that if ϕ satisfies 18.10 and $\psi \equiv \psi_1 + \frac{1}{c}\frac{\partial \phi}{\partial t}$, while $\mathbf{A} \equiv \mathbf{A}_1 + \nabla \phi$, then 18.12 holds for $\mathbf{A}$ and ψ.

29. Why consider Maxwell's equations? Why not just consider 18.17, 18.16, and 18.12?

30. Tell which open sets are simply connected.

    (a) The inside of a car radiator.

    (b) A donut.

    (c) The solid part of a cannon ball which contains a void on the interior.

    (d) The inside of a donut which has had a large bite taken out of it.

    (e) All of $\mathbb{R}^3$ except the z axis.

    (f) All of $\mathbb{R}^3$except the xy plane.

31. Let P be a polygon with vertices $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n), (x_1, y_1)$ encountered as you move over the boundary of the polygon in the counter clockwise direction. Using Problem 13, find a nice formula for the area of the polygon in terms of the vertices.

The Mathematical Theory Of Determinants*

A.1 The Function sgn_n

It is easiest to give a different definition of the determinant which is clearly well defined and then prove the earlier one in terms of Laplace expansion. Let $(i_1, \cdots, i_n)$ be an ordered list of numbers from $\{1, \cdots, n\}$. This means the order is important so $(1,2,3)$ and $(2,1,3)$ are different. There will be some repetition between this section and the earlier section on determinants. The main purpose is to give all the missing proofs. Two books which give a good introduction to determinants are Apostol [2] and Rudin [23]. A recent book which also has a good introduction is Baker [4].

The following Lemma will be essential in the definition of the determinant.

Lemma A.1.1 *There exists a unique function,* sgn_n *which maps each list of numbers from* $\{1, \cdots, n\}$ *to one of the three numbers,* $0, 1,$ *or* -1 *which also has the following properties.*

$$\mathrm{sgn}_n(1, \cdots, n) = 1 \tag{1.1}$$

$$\mathrm{sgn}_n(i_1, \cdots, p, \cdots, q, \cdots, i_n) = -\mathrm{sgn}_n(i_1, \cdots, q, \cdots, p, \cdots, i_n) \tag{1.2}$$

In words, the second property states that if two of the numbers are switched, the value of the function is multiplied by -1*. Also, in the case where* $n > 1$ *and* $\{i_1, \cdots, i_n\} = \{1, \cdots, n\}$ *so that every number from* $\{1, \cdots, n\}$ *appears in the ordered list,* $(i_1, \cdots, i_n)$*,*

$$\mathrm{sgn}_n(i_1, \cdots, i_{\theta-1}, n, i_{\theta+1}, \cdots, i_n) \equiv$$

$$(-1)^{n-\theta}\,\mathrm{sgn}_{n-1}(i_1, \cdots, i_{\theta-1}, i_{\theta+1}, \cdots, i_n) \tag{1.3}$$

where $n = i_\theta$ *in the ordered list,* $(i_1, \cdots, i_n)$*.*

Proof: To begin with, it is necessary to show the existence of such a function. This is clearly true if $n = 1$. Define $\operatorname{sgn}_1(1) \equiv 1$ and observe that it works. No switching is possible. In the case where $n = 2$, it is also clearly true. Let $\operatorname{sgn}_2(1,2) = 1$ and $\operatorname{sgn}_2(2,1) = -1$ while $\operatorname{sgn}_2(2,2) = \operatorname{sgn}_2(1,1) = 0$ and verify it works. Assuming such a function exists for n, sgn_{n+1} will be defined in terms of sgn_n. If there are any repeated numbers in $(i_1, \cdots, i_{n+1})$, $\operatorname{sgn}_{n+1}(i_1, \cdots, i_{n+1}) \equiv 0$. If there are no repeats, then $n+1$ appears somewhere in the ordered list. Let θ be the position of the number $n+1$ in the list. Thus, the list is of the form $(i_1, \cdots, i_{\theta-1}, n+1, i_{\theta+1}, \cdots, i_{n+1})$. From 1.3 it must be that

$$\operatorname{sgn}_{n+1}(i_1, \cdots, i_{\theta-1}, n+1, i_{\theta+1}, \cdots, i_{n+1}) \equiv$$
$$(-1)^{n+1-\theta} \operatorname{sgn}_n(i_1, \cdots, i_{\theta-1}, i_{\theta+1}, \cdots, i_{n+1}).$$

It is necessary to verify this satisfies 1.1 and 1.2 with n replaced with $n+1$. The first of these is obviously true because

$$\operatorname{sgn}_{n+1}(1, \cdots, n, n+1) \equiv (-1)^{n+1-(n+1)} \operatorname{sgn}_n(1, \cdots, n) = 1.$$

If there are repeated numbers in $(i_1, \cdots, i_{n+1})$, then it is obvious 1.2 holds because both sides would equal zero from the above definition. It remains to verify 1.2 in the case where there are no numbers repeated in $(i_1, \cdots, i_{n+1})$. Consider

$$\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{p}, \cdots, \overset{s}{q}, \cdots, i_{n+1}\right),$$

where the r above the p indicates the number, p is in the r^{th} position and the s above the q indicates that the number, q is in the s^{th} position. Suppose first that $r < \theta < s$. Then

$$\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{p}, \cdots, \overset{\theta}{n+1}, \cdots, \overset{s}{q}, \cdots, i_{n+1}\right) \equiv$$
$$(-1)^{n+1-\theta} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{p}, \cdots, \overset{s-1}{q}, \cdots, i_{n+1}\right)$$

while

$$\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{q}, \cdots, \overset{\theta}{n+1}, \cdots, \overset{s}{p}, \cdots, i_{n+1}\right) =$$
$$(-1)^{n+1-\theta} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{q}, \cdots, \overset{s-1}{p}, \cdots, i_{n+1}\right)$$

and so, by induction, a switch of p and q introduces a minus sign in the result. Similarly, if $\theta > s$ or if $\theta < r$ it also follows that 1.2 holds. The interesting case is when $\theta = r$ or $\theta = s$. Consider the case where $\theta = r$ and note the other case is entirely similar.

$$\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{n+1}, \cdots, \overset{s}{q}, \cdots, i_{n+1}\right) =$$
$$(-1)^{n+1-r} \operatorname{sgn}_n\left(i_1, \cdots, \overset{s-1}{q}, \cdots, i_{n+1}\right) \tag{1.4}$$

while

$$\operatorname{sgn}_{n+1}\left(i_1, \cdots, \overset{r}{q}, \cdots, \overset{s}{n+1}, \cdots, i_{n+1}\right) =$$
$$(-1)^{n+1-s} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{q}, \cdots, i_{n+1}\right). \tag{1.5}$$

By making $s-1-r$ switches, move the q which is in the $s-1^{th}$ position in 1.4 to the r^{th} position in 1.5. By induction, each of these switches introduces a factor of -1 and so

$$\operatorname{sgn}_n\left(i_1, \cdots, \overset{s-1}{q}, \cdots, i_{n+1}\right) = (-1)^{s-1-r} \operatorname{sgn}_n\left(i_1, \cdots, \overset{r}{q}, \cdots, i_{n+1}\right).$$

Therefore,

$$\begin{aligned}\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{n+1},\cdots,\overset{s}{q},\cdots,i_{n+1}\right) &= (-1)^{n+1-r}\operatorname{sgn}_n\left(i_1,\cdots,\overset{s-1}{q},\cdots,i_{n+1}\right)\\ &= (-1)^{n+1-r}(-1)^{s-1-r}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,i_{n+1}\right)\\ = (-1)^{n+s}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,i_{n+1}\right) &= (-1)^{2s-1}(-1)^{n+1-s}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,i_{n+1}\right)\\ &= -\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{q},\cdots,\overset{s}{n+1},\cdots,i_{n+1}\right).\end{aligned}$$

This proves the existence of the desired function.

To see this function is unique, note that you can obtain any ordered list of distinct numbers from a sequence of switches. If there exist two functions, f and g both satisfying 1.1 and 1.2, you could start with $f(1,\cdots,n) = g(1,\cdots,n)$ and applying the same sequence of switches, eventually arrive at $f(i_1,\cdots,i_n) = g(i_1,\cdots,i_n)$. If any numbers are repeated, then 1.2 gives both functions are equal to zero for that ordered list. This proves the lemma.

A.2 The Determinant

A.2.1 The Definition

In what follows sgn will often be used rather than sgn_n because the context supplies the appropriate n.

Definition A.2.1 *Let f be a real valued function which has the set of ordered lists of numbers from $\{1,\cdots,n\}$ as its domain. Define*

$$\sum_{(k_1,\cdots,k_n)} f(k_1\cdots k_n)$$

to be the sum of all the $f(k_1\cdots k_n)$ for all possible choices of ordered lists $(k_1,\cdots,k_n)$ of numbers of $\{1,\cdots,n\}$. For example,

$$\sum_{(k_1,k_2)} f(k_1,k_2) = f(1,2)+f(2,1)+f(1,1)+f(2,2).$$

Definition A.2.2 *Let $(a_{ij}) = A$ denote an $n\times n$ matrix. The determinant of A, denoted by $\det(A)$ is defined by*

$$\det(A) \equiv \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, a_{1k_1}\cdots a_{nk_n}$$

where the sum is taken over all ordered lists of numbers from $\{1,\cdots,n\}$. Note it suffices to take the sum over only those ordered lists in which there are no repeats because if there are, $\operatorname{sgn}(k_1,\cdots,k_n) = 0$ and so that term contributes 0 to the sum.

Let A be an $n\times n$ matrix, $A = (a_{ij})$ and let $(r_1,\cdots,r_n)$ denote an ordered list of n numbers from $\{1,\cdots,n\}$. Let $A(r_1,\cdots,r_n)$ denote the matrix whose k^{th} row is the r_k row of the matrix, A. Thus

$$\det(A(r_1,\cdots,r_n)) = \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, a_{r_1k_1}\cdots a_{r_nk_n} \tag{1.6}$$

and

$$A(1,\cdots,n) = A.$$

A.2.2 Permuting Rows Or Columns

Proposition A.2.3 *Let*

$$(r_1, \cdots, r_n)$$

be an ordered list of numbers from $\{1, \cdots, n\}$*. Then*

$$\operatorname{sgn}(r_1, \cdots, r_n) \det(A)$$

$$= \sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_n) a_{r_1 k_1} \cdots a_{r_n k_n} \tag{1.7}$$

$$= \det(A(r_1, \cdots, r_n)). \tag{1.8}$$

Proof: Let $(1, \cdots, n) = (1, \cdots, r, \cdots s, \cdots, n)$ so $r < s$.

$$\det(A(1, \cdots, r, \cdots, s, \cdots, n)) = \tag{1.9}$$

$$\sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_r, \cdots, k_s, \cdots, k_n) a_{1k_1} \cdots a_{rk_r} \cdots a_{sk_s} \cdots a_{nk_n},$$

and renaming the variables, calling k_s, k_r and k_r, k_s, this equals

$$= \sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_s, \cdots, k_r, \cdots, k_n) a_{1k_1} \cdots a_{rk_s} \cdots a_{sk_r} \cdots a_{nk_n}$$

$$= \sum_{(k_1, \cdots, k_n)} -\operatorname{sgn}\left(k_1, \cdots, \overbrace{k_r, \cdots, k_s}^{\text{These got switched}}, \cdots, k_n\right) a_{1k_1} \cdots a_{sk_r} \cdots a_{rk_s} \cdots a_{nk_n}$$

$$= -\det(A(1, \cdots, s, \cdots, r, \cdots, n)). \tag{1.10}$$

Consequently,

$$\det(A(1, \cdots, s, \cdots, r, \cdots, n)) =$$

$$-\det(A(1, \cdots, r, \cdots, s, \cdots, n)) = -\det(A)$$

Now letting $A(1, \cdots, s, \cdots, r, \cdots, n)$ play the role of A, and continuing in this way, switching pairs of numbers,

$$\det(A(r_1, \cdots, r_n)) = (-1)^p \det(A)$$

where it took p switches to obtain$(r_1, \cdots, r_n)$ from $(1, \cdots, n)$. By Lemma A.1.1, this implies

$$\det(A(r_1, \cdots, r_n)) = (-1)^p \det(A) = \operatorname{sgn}(r_1, \cdots, r_n) \det(A)$$

and proves the proposition in the case when there are no repeated numbers in the ordered list, $(r_1, \cdots, r_n)$. However, if there is a repeat, say the r^{th} row equals the s^{th} row, then the reasoning of 1.9 -1.10 shows that $A(r_1, \cdots, r_n) = 0$ and also $\operatorname{sgn}(r_1, \cdots, r_n) = 0$ so the formula holds in this case also.

Observation A.2.4 *There are* $n!$ *ordered lists of distinct numbers from* $\{1, \cdots, n\}$.

To see this, consider n slots placed in order. There are n choices for the first slot. For each of these choices, there are $n-1$ choices for the second. Thus there are $n(n-1)$ ways to fill the first two slots. Then for each of these ways there are $n-2$ choices left for the third slot. Continuing this way, there are $n!$ ordered lists of distinct numbers from $\{1, \cdots, n\}$ as stated in the observation.

A.2.3 A Symmetric Definition

With the above, it is possible to give a more symmetric description of the determinant from which it will follow that $\det(A) = \det\left(A^T\right)$.

Corollary A.2.5 *The following formula for* $\det(A)$ *is valid.*

$$\det(A) = \frac{1}{n!}\cdot$$

$$\sum_{(r_1,\cdots,r_n)} \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(r_1,\cdots,r_n) \operatorname{sgn}(k_1,\cdots,k_n) a_{r_1k_1}\cdots a_{r_nk_n}. \tag{1.11}$$

And also $\det\left(A^T\right) = \det(A)$ *where* A^T *is the transpose of* A. *(Recall that for* $A^T = \left(a_{ij}^T\right)$, $a_{ij}^T = a_{ji}$.*)*

Proof: From Proposition A.2.3, if the r_i are distinct,

$$\det(A) = \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(r_1,\cdots,r_n) \operatorname{sgn}(k_1,\cdots,k_n) a_{r_1k_1}\cdots a_{r_nk_n}.$$

Summing over all ordered lists, $(r_1,\cdots,r_n)$ where the r_i are distinct, (If the r_i are not distinct, $\operatorname{sgn}(r_1,\cdots,r_n) = 0$ and so there is no contribution to the sum.)

$$n!\det(A) =$$

$$\sum_{(r_1,\cdots,r_n)} \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(r_1,\cdots,r_n) \operatorname{sgn}(k_1,\cdots,k_n) a_{r_1k_1}\cdots a_{r_nk_n}.$$

This proves the corollary since the formula gives the same number for A as it does for A^T.

A.2.4 The Alternating Property Of The Determinant

Corollary A.2.6 *If two rows or two columns in an* $n\times n$ *matrix,* A, *are switched, the determinant of the resulting matrix equals* (-1) *times the determinant of the original matrix. If* A *is an* $n\times n$ *matrix in which two rows are equal or two columns are equal then* $\det(A) = 0$. *Suppose the* i^{th} *row of* A *equals* $(xa_1+yb_1,\cdots,xa_n+yb_n)$. *Then*

$$\det(A) = x\det(A_1) + y\det(A_2)$$

where the i^{th} *row of* A_1 *is* $(a_1,\cdots,a_n)$ *and the* i^{th} *row of* A_2 *is* $(b_1,\cdots,b_n)$, *all other rows of* A_1 *and* A_2 *coinciding with those of* A. *In other words,* det *is a linear function of each row* A. *The same is true with the word "row" replaced with the word "column".*

Proof: By Proposition A.2.3 when two rows are switched, the determinant of the resulting matrix is (-1) times the determinant of the original matrix. By Corollary A.2.5 the same holds for columns because the columns of the matrix equal the rows of the transposed matrix. Thus if A_1 is the matrix obtained from A by switching two columns,

$$\det(A) = \det\left(A^T\right) = -\det\left(A_1^T\right) = -\det(A_1).$$

If A has two equal columns or two equal rows, then switching them results in the same matrix. Therefore, $\det(A) = -\det(A)$ and so $\det(A) = 0$.

It remains to verify the last assertion.

$$\det(A) \equiv \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n) a_{1k_1}\cdots(xa_{k_i}+yb_{k_i})\cdots a_{nk_n}$$

$$= x \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, a_{1k_1}\cdots a_{k_i}\cdots a_{nk_n}$$

$$+y \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, a_{1k_1}\cdots b_{k_i}\cdots a_{nk_n}$$

$$\equiv x\det(A_1) + y\det(A_2).$$

The same is true of columns because $\det\left(A^T\right) = \det(A)$ and the rows of A^T are the columns of A.

A.2.5 Linear Combinations And Determinants

Definition A.2.7 *A vector,* $\mathbf{w}$*, is a linear combination of the vectors* $\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}$ *if there exists scalars,* $c_1,\cdots c_r$ *such that* $\mathbf{w} = \sum_{k=1}^r c_k\mathbf{v}_k$*. This is the same as saying* $\mathbf{w} \in \operatorname{span}\{\mathbf{v}_1,\cdots,\mathbf{v}_r\}$.

The following corollary is also of great use.

Corollary A.2.8 *Suppose* A *is an* $n\times n$ *matrix and some column (row) is a linear combination of* r *other columns (rows). Then* $\det(A) = 0$.

Proof: Let $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$ be the columns of A and suppose the condition that one column is a linear combination of r of the others is satisfied. Then by using Corollary A.2.6 you may rearrange the columns to have the n^{th} column a linear combination of the first r columns. Thus $\mathbf{a}_n = \sum_{k=1}^r c_k\mathbf{a}_k$ and so

$$\det(A) = \det\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r & \cdots & \mathbf{a}_{n-1} & \sum_{k=1}^r c_k\mathbf{a}_k \end{pmatrix}.$$

By Corollary A.2.6

$$\det(A) = \sum_{k=1}^r c_k \det\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r & \cdots & \mathbf{a}_{n-1} & \mathbf{a}_k \end{pmatrix} = 0.$$

The case for rows follows from the fact that $\det(A) = \det\left(A^T\right)$. This proves the corollary.

A.2.6 The Determinant Of A Product

Recall the following definition of matrix multiplication.

Definition A.2.9 *If* A *and* B *are* $n\times n$ *matrices,* $A = (a_{ij})$ *and* $B = (b_{ij})$*,* $AB = (c_{ij})$ *where*

$$c_{ij} \equiv \sum_{k=1}^n a_{ik}b_{kj}.$$

One of the most important rules about determinants is that the determinant of a product equals the product of the determinants.

Theorem A.2.10 *Let* A *and* B *be* $n\times n$ *matrices. Then*

$$\det(AB) = \det(A)\det(B).$$

Proof: Let c_{ij} be the ij^{th} entry of AB. Then by Proposition A.2.3,

$$\det(AB) =$$

$$\begin{aligned}
&\sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, c_{1k_1}\cdots c_{nk_n} \\
&= \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n) \left(\sum_{r_1} a_{1r_1}b_{r_1k_1}\right)\cdots\left(\sum_{r_n} a_{nr_n}b_{r_nk_n}\right) \\
&= \sum_{(r_1\cdots,r_n)} \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n)\, b_{r_1k_1}\cdots b_{r_nk_n}\left(a_{1r_1}\cdots a_{nr_n}\right) \\
&= \sum_{(r_1\cdots,r_n)} \operatorname{sgn}(r_1\cdots r_n)\, a_{1r_1}\cdots a_{nr_n}\det(B) = \det(A)\det(B).
\end{aligned}$$

This proves the theorem.

A.2.7 Cofactor Expansions

Lemma A.2.11 *Suppose a matrix is of the form*

$$M = \begin{pmatrix} A & * \\ \mathbf{0} & a \end{pmatrix} \tag{1.12}$$

or

$$M = \begin{pmatrix} A & \mathbf{0} \\ * & a \end{pmatrix} \tag{1.13}$$

where a is a number and A is an $(n-1)\times(n-1)$ matrix and $*$ denotes either a column or a row having length $n-1$ and the $\mathbf{0}$ denotes either a column or a row of length $n-1$ consisting entirely of zeros. Then

$$\det(M) = a\det(A).$$

Proof: Denote M by (m_{ij}). Thus in the first case, $m_{nn} = a$ and $m_{ni} = 0$ if $i \neq n$ while in the second case, $m_{nn} = a$ and $m_{in} = 0$ if $i \neq n$. From the definition of the determinant,

$$\det(M) \equiv \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}_n(k_1,\cdots,k_n)\, m_{1k_1}\cdots m_{nk_n}$$

Letting θ denote the position of n in the ordered list, $(k_1,\cdots,k_n)$ then using the earlier conventions used to prove Lemma A.1.1, $\det(M)$ equals

$$\sum_{(k_1,\cdots,k_n)} (-1)^{n-\theta} \operatorname{sgn}_{n-1}\left(k_1,\cdots,k_{\theta-1},\overset{\theta}{k_{\theta+1}},\cdots,\overset{n-1}{k_n}\right) m_{1k_1}\cdots m_{nk_n}$$

Now suppose 1.13. Then if $k_n \neq n$, the term involving m_{nk_n} in the above expression equals zero. Therefore, the only terms which survive are those for which $\theta = n$ or in other words, those for which $k_n = n$. Therefore, the above expression reduces to

$$a \sum_{(k_1,\cdots,k_{n-1})} \operatorname{sgn}_{n-1}(k_1,\cdots k_{n-1})\, m_{1k_1}\cdots m_{(n-1)k_{n-1}} = a\det(A).$$

To get the assertion in the situation of 1.12 use Corollary A.2.5 and 1.13 to write

$$\det(M) = \det\left(M^T\right) = \det\left(\begin{pmatrix} A^T & \mathbf{0} \\ * & a \end{pmatrix}\right) = a\det\left(A^T\right) = a\det(A).$$

This proves the lemma.

In terms of the theory of determinants, arguably the most important idea is that of Laplace expansion along a row or a column. This will follow from the above definition of a determinant.

Definition A.2.12 *Let $A = (a_{ij})$ be an $n \times n$ matrix. Then a new matrix called the cofactor matrix,* $\operatorname{cof}(A)$ *is defined by* $\operatorname{cof}(A) = (c_{ij})$ *where to obtain c_{ij} delete the i^{th} row and the j^{th} column of A, take the determinant of the $(n-1)\times(n-1)$ matrix which results, (This is called the ij^{th} minor of A.) and then multiply this number by $(-1)^{i+j}$. To make the formulas easier to remember,* $\operatorname{cof}(A)_{ij}$ *will denote the ij^{th} entry of the cofactor matrix.*

The following is the main result. Earlier this was given as a definition and the outrageous totally unjustified assertion was made that the same number would be obtained by expanding the determinant along any row or column. The following theorem proves this assertion.

Theorem A.2.13 *Let A be an $n \times n$ matrix where $n \geq 2$. Then*

$$\det(A) = \sum_{j=1}^{n} a_{ij} \operatorname{cof}(A)_{ij} = \sum_{i=1}^{n} a_{ij} \operatorname{cof}(A)_{ij}.$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column.

Proof: Let $(a_{i1}, \cdots, a_{in})$ be the i^{th} row of A. Let B_j be the matrix obtained from A by leaving every row the same except the i^{th} row which in B_j equals $(0, \cdots, 0, a_{ij}, 0, \cdots, 0)$. Then by Corollary A.2.6,

$$\det(A) = \sum_{j=1}^{n} \det(B_j)$$

Denote by A^{ij} the $(n-1)\times(n-1)$ matrix obtained by deleting the i^{th} row and the j^{th} column of A. Thus $\operatorname{cof}(A)_{ij} \equiv (-1)^{i+j} \det\left(A^{ij}\right)$. At this point, recall that from Proposition A.2.3, when two rows or two columns in a matrix, M, are switched, this results in multiplying the determinant of the old matrix by -1 to get the determinant of the new matrix. Therefore, by Lemma A.2.11,

$$\begin{aligned} \det(B_j) &= (-1)^{n-j}(-1)^{n-i} \det\left(\begin{pmatrix} A^{ij} & * \\ \mathbf{0} & a_{ij} \end{pmatrix}\right) \\ &= (-1)^{i+j} \det\left(\begin{pmatrix} A^{ij} & * \\ \mathbf{0} & a_{ij} \end{pmatrix}\right) = a_{ij} \operatorname{cof}(A)_{ij}. \end{aligned}$$

Therefore,

$$\det(A) = \sum_{j=1}^{n} a_{ij} \operatorname{cof}(A)_{ij}$$

which is the formula for expanding $\det(A)$ along the i^{th} row. Also,

$$\begin{aligned} \det(A) &= \det\left(A^T\right) = \sum_{j=1}^{n} a_{ij}^T \operatorname{cof}\left(A^T\right)_{ij} \\ &= \sum_{j=1}^{n} a_{ji} \operatorname{cof}(A)_{ji} \end{aligned}$$

which is the formula for expanding $\det(A)$ along the i^{th} column. This proves the theorem.Note that this gives an easy way to write a formula for the inverse of an $n \times n$ matrix. Recall the definition of the inverse of a matrix in Definition 2.1.28 on Page 39.

A.2.8 Formula For The Inverse

Theorem A.2.14 *A^{-1} exists if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$, then $A^{-1} = \left(a_{ij}^{-1}\right)$ where*

$$a_{ij}^{-1} = \det(A)^{-1} \operatorname{cof}(A)_{ji}$$

for $\operatorname{cof}(A)_{ij}$ the ij^{th} cofactor of A.

Proof: By Theorem A.2.13 and letting $(a_{ir}) = A$, if $\det(A) \neq 0$,

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ir} \det(A)^{-1} = \det(A) \det(A)^{-1} = 1.$$

Now consider

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

when $k \neq r$. Replace the k^{th} column with the r^{th} column to obtain a matrix, B_k whose determinant equals zero by Corollary A.2.6. However, expanding this matrix along the k^{th} column yields

$$0 = \det(B_k) \det(A)^{-1} = \sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

Summarizing,

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1} = \delta_{rk}.$$

Using the other formula in Theorem A.2.13, and similar reasoning,

$$\sum_{j=1}^{n} a_{rj} \operatorname{cof}(A)_{kj} \det(A)^{-1} = \delta_{rk}$$

This proves that if $\det(A) \neq 0$, then A^{-1} exists with $A^{-1} = \left(a_{ij}^{-1}\right)$, where

$$a_{ij}^{-1} = \operatorname{cof}(A)_{ji} \det(A)^{-1}.$$

Now suppose A^{-1} exists. Then by Theorem A.2.10,

$$1 = \det(I) = \det\left(AA^{-1}\right) = \det(A) \det\left(A^{-1}\right)$$

so $\det(A) \neq 0$. This proves the theorem.

The next corollary points out that if an $n \times n$ matrix, A has a right or a left inverse, then it has an inverse.

Corollary A.2.15 *Let A be an $n \times n$ matrix and suppose there exists an $n \times n$ matrix, B such that $BA = I$. Then A^{-1} exists and $A^{-1} = B$. Also, if there exists C an $n \times n$ matrix such that $AC = I$, then A^{-1} exists and $A^{-1} = C$.*

Proof: Since $BA = I$, Theorem A.2.10 implies

$$\det B \det A = 1$$

and so $\det A \neq 0$. Therefore from Theorem A.2.14, A^{-1} exists. Therefore,

$$A^{-1} = (BA) A^{-1} = B\left(AA^{-1}\right) = BI = B.$$

The case where $CA = I$ is handled similarly.

The conclusion of this corollary is that left inverses, right inverses and inverses are all the same in the context of $n \times n$ matrices.

Theorem A.2.14 says that to find the inverse, take the transpose of the cofactor matrix and divide by the determinant. The transpose of the cofactor matrix is called the adjugate or sometimes the classical adjoint of the matrix A. It is an abomination to call it the adjoint although you do sometimes see it referred to in this way. In words, A^{-1} is equal to one over the determinant of A times the adjugate matrix of A.

A.2.9 Cramer's Rule

In case you are solving a system of equations, $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, it follows that if A^{-1} exists,

$$\mathbf{x} = \left(A^{-1}A\right)\mathbf{x} = A^{-1}\left(A\mathbf{x}\right) = A^{-1}\mathbf{y}$$

thus solving the system. Now in the case that A^{-1} exists, there is a formula for A^{-1} given above. Using this formula,

$$x_i = \sum_{j=1}^{n} a_{ij}^{-1} y_j = \sum_{j=1}^{n} \frac{1}{\det(A)} \operatorname{cof}(A)_{ji}\, y_j.$$

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det \begin{pmatrix} * & \cdots & y_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & y_n & \cdots & * \end{pmatrix},$$

where here the i^{th} column of A is replaced with the column vector, $(y_1 \cdots, y_n)^T$, and the determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

A.2.10 Upper Triangular Matrices

Definition A.2.16 *A matrix M, is upper triangular if $M_{ij} = 0$ whenever $i > j$. Thus such a matrix equals zero below the main diagonal, the entries of the form M_{ii} as shown.*

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

With this definition, here is a simple corollary of Theorem A.2.13.

Corollary A.2.17 *Let M be an upper (lower) triangular matrix. Then $\det(M)$ is obtained by taking the product of the entries on the main diagonal.*

A.2.11 The Determinant Rank

Definition A.2.18 *A submatrix of a matrix A is the rectangular array of numbers obtained by deleting some rows and columns of A. Let A be an $m \times n$ matrix. The* ***determinant rank*** *of the matrix equals r where r is the largest number such that some $r \times r$ submatrix of A has a non zero determinant.*

Theorem A.2.19 *If A, an $m \times n$ matrix has determinant rank, r, then there exist r rows (columns) of the matrix such that every other row (column) is a linear combination of these r rows (columns).*

Proof: Suppose the determinant rank of $A = (a_{ij})$ equals r. Thus some $r \times r$ submatrix has non zero determinant and there is no larger square submatrix which has non zero determinant. Suppose such a submatrix is determined by the r columns whose indices are

$$j_1 < \cdots < j_r$$

and the r rows whose indices are

$$i_1 < \cdots < i_r$$

I want to show that every row is a linear combination of these rows. Consider the l^{th} row and let p be an index between 1 and n. Form the following $(r+1) \times (r+1)$ matrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} & a_{i_1 p} \\ \vdots & & \vdots & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} & a_{i_r p} \\ a_{l j_1} & \cdots & a_{l j_r} & a_{l p} \end{pmatrix}$$

Of course you can assume $l \notin \{i_1, \cdots, i_r\}$ because there is nothing to prove if the l^{th} row is one of the chosen ones. The above matrix has determinant 0. This is because if $p \notin \{j_1, \cdots, j_r\}$ then the above would be a submatrix of A which is too large to have non zero determinant. On the other hand, if $p \in \{j_1, \cdots, j_r\}$ then the above matrix has two columns which are equal so its determinant is still 0.

Expand the determinant of the above matrix along the last column. Let C_k denote the cofactor associated with the entry $a_{i_k p}$. This is not dependent on the choice of p. Remember, you delete the column and the row the entry is in and take the determinant of what is left and multiply by -1 raised to an appropriate power. Let C denote the cofactor associated with a_{lp}. This is given to be nonzero, it being the determinant of the matrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} \\ \vdots & & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} \end{pmatrix}$$

Thus

$$0 = a_{lp} C + \sum_{k=1}^{r} C_k a_{i_k p}$$

which implies

$$a_{lp} = \sum_{k=1}^{r} \frac{-C_k}{C} a_{i_k p} \equiv \sum_{k=1}^{r} m_k a_{i_k p}$$

Since this is true for every p and since m_k does not depend on p, this has shown the l^{th} row is a linear combination of the $i_1, i_2, \cdots, i_r$ rows. The determinant rank does not change when you replace A with A^T. Therefore, the same conclusion holds for the columns. This proves the theorem.

A.2.12 Telling Whether A Is One To One Or Onto

The following theorem is of fundamental importance and ties together many of the ideas presented above.

Theorem A.2.20 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. $\det(A) = 0$.
2. A, A^T *are not one to one.*
3. A *is not onto.*

Proof: Suppose $\det(A) = 0$. Then the determinant rank of $A = r < n$. Therefore, there exist r columns such that every other column is a linear combination of these columns by Theorem A.2.19. In particular, it follows that for some m, the m^{th} column is a linear combination of all the others. Thus letting $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m & \cdots & \mathbf{a}_n \end{pmatrix}$ where the columns are denoted by $\mathbf{a}_i$, there exists scalars, α_i such that

$$\mathbf{a}_m = \sum_{k \neq m} \alpha_k \mathbf{a}_k.$$

Now consider the column vector, $\mathbf{x} \equiv \begin{pmatrix} \alpha_1 & \cdots & -1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix}^T$. Then

$$A\mathbf{x} = -\mathbf{a}_m + \sum_{k \neq m} \alpha_k \mathbf{a}_k = \mathbf{0}.$$

Since also $A\mathbf{0} = \mathbf{0}$, it follows A is not one to one. Similarly, A^T is not one to one by the same argument applied to A^T. This verifies that 1.) implies 2.).

Now suppose 2.). Then since A^T is not one to one, it follows there exists $\mathbf{x} \neq \mathbf{0}$ such that

$$A^T\mathbf{x} = \mathbf{0}.$$

Taking the transpose of both sides yields

$$\mathbf{x}^T A = \mathbf{0}^T$$

where the $\mathbf{0}^T$ is a $1 \times n$ matrix or row vector. Now if $A\mathbf{y} = \mathbf{x}$, then

$$|\mathbf{x}|^2 = \mathbf{x}^T (A\mathbf{y}) = \left(\mathbf{x}^T A\right) \mathbf{y} = \mathbf{0}^T \mathbf{y} = 0$$

contrary to $\mathbf{x} \neq \mathbf{0}$. Consequently there can be no $\mathbf{y}$ such that $A\mathbf{y} = \mathbf{x}$ and so A is not onto. This shows that 2.) implies 3.).

Finally, suppose 3.). If 1.) does not hold, then $\det(A) \neq 0$ but then from Theorem A.2.14 A^{-1} exists and so for every $\mathbf{y} \in \mathbb{F}^n$ there exists a unique $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{y}$. In fact $\mathbf{x} = A^{-1}\mathbf{y}$. Thus A would be onto contrary to 3.). This shows 3.) implies 1.) and proves the theorem.

Corollary A.2.21 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. $det(A) \neq 0$.
2. A *and* A^T *are one to one.*
3. A *is onto.*

Proof: This follows immediately from the above theorem.

A.2.13 Schur's Theorem

Consider the following system of equations for $x_1, x_2, \cdots, x_n$

$$\sum_{j=1}^{n} a_{ij}x_j = 0,\ i = 1, 2, \cdots, m \tag{1.14}$$

where $m < n$. Then the following theorem is a fundamental observation.

Theorem A.2.22 *Let the system of equations be as just described in 1.14 where $m < n$. Then letting*

$$\mathbf{x}^T \equiv (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n,$$

there exists $\mathbf{x} \neq \mathbf{0}$ such that the components satisfy each of the equations of 1.14.

Proof: The above system is of the form

$$A\mathbf{x} = \mathbf{0}$$

where A is an $m \times n$ matrix with $m < n$. Therefore, if you form the matrix

$$\begin{pmatrix} A \\ 0 \end{pmatrix},$$

an $n \times n$ matrix having $n - m$ rows of zeros on the bottom, it follows this matrix has determinant equal to 0. Therefore, from Theorem A.2.19, there exists $\mathbf{x} \neq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{0}$. This proves the theorem.

Definition A.2.23 *A set of vectors in $\mathbb{R}^n$ $\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$ is called an **orthonormal** set of vectors if*

$$\mathbf{x}_i \cdot \mathbf{x}_j = \delta_{ij} \equiv \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Theorem A.2.24 *Let $\mathbf{v}_1$ be a unit vector ($|\mathbf{v}_1| = 1$) in $\mathbb{R}^n$, $n > 1$. Then there exist vectors $\{\mathbf{v}_2, \cdots, \mathbf{v}_n\}$ such that*

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$$

is an orthonormal set of vectors.

Proof: The equation for $\mathbf{x}$

$$\mathbf{v}_1 \cdot \mathbf{x} = 0$$

has a nonzero solution $\mathbf{x}$ by Theorem A.2.22. Pick such a solution and divide by its magnitude to get $\mathbf{v}_2$ a unit vector such that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. Now suppose $\mathbf{v}_1, \cdots, \mathbf{v}_k$ have been chosen such that $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ us an orthonormal set of vectors. Then consider the equations

$$\mathbf{v}_j \cdot \mathbf{x} = 0 \quad j = 1, 2, \cdots, k$$

This amounts to the situation of Theorem A.2.22 in which there are more variables than equations. Therefore, by this theorem, there exists a nonzero $\mathbf{x}$ solving all these equations. Divide by its magnitude and this gives $\mathbf{v}_{k+1}$. This proves the theorem.

Definition A.2.25 *If U is an $n \times n$ matrix whose columns form an orthonormal set of vectors, then Q is called an **orthogonal matrix**. Note that from the way we multiply matrices,*

$$U^T U = UU^T = I.$$

Thus $U^{-1} = U^T$.

Note the product of orthogonal matrices is orthogonal because

$$(U_1U_2)^T (U_1U_2) = U_2^T U_1^T U_1U_2 = I.$$

Two matrices A and B are similar if there is some invertible matrix S such that $A = S^{-1}BS$. Note that similar matrices have the same characteristic equation because by Theorem A.2.10 which says the determinant of a product is the product of the determinants,

$$\det(\lambda I - A) = \det\left(\lambda I - S^{-1}BS\right) = \det\left(S^{-1}(\lambda I - B)S\right)$$

$$= \det\left(S^{-1}\right)\det(\lambda I - B)\det(S) = \det\left(S^{-1}S\right)\det(\lambda I - B) = \det(\lambda I - B)$$

With this preparation, here is a case of Schur's theorem.

Theorem A.2.26 *Let A be a real $n \times n$ matrix which has all real eigenvalues. Then there exists an orthogonal matrix, U such that*

$$U^T AU = T, \tag{1.15}$$

where T is an upper triangular matrix having the eigenvalues of A on the main diagonal listed according to multiplicity as zeros of the characteristic equation.

Proof: The theorem is clearly true if A is a 1×1 matrix. Just let $U = 1$ the 1×1 matrix which has 1 down the main diagonal and zeros elsewhere. Suppose it is true for $(n-1) \times (n-1)$ matrices and let A be an $n \times n$ matrix. Then let $\mathbf{v}_1$ be a unit eigenvector for A . Then there exists λ_1 such that

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1,\ |\mathbf{v}_1| = 1.$$

By Theorem A.2.24 there exists $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$, an orthonormal set in $\mathbb{R}^n$. Let U_0 be a matrix whose i^{th} column is $\mathbf{v}_i$. Then from the above, it follows U_0 is orthogonal. Then from the way you multiply matrices $U_0^T AU_0$ is of the form

$$\begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

where A_1 is an $n-1 \times n-1$ matrix. The above matrix is similar to A so it has the same eigenvalues and indeed the same characteristic equation. Also the eigenvalues of A_1 are all real because each of these eigenvalues is an eigenvalue of the above matrix and is therefore an eigenvalue of A. Now by induction there exists an $(n-1) \times (n-1)$ orthogonal matrix $\widetilde{U}_1$ such that

$$\widetilde{U}_1^* A_1 \widetilde{U}_1 = T_{n-1},$$

an upper trianguar matrix. Consider

$$U_1 \equiv \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix}$$

This is a orthogonal matrix and

$$\begin{aligned} U_1^T U_0^T AU_0U_1 &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1^* \end{pmatrix} \begin{pmatrix} \lambda_1 & * \\ \mathbf{0} & A_1 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & * \\ \mathbf{0} & T_{n-1} \end{pmatrix} \equiv T \end{aligned}$$

where T is upper triangular. Then let $U = U_0U_1$. Since $(U_0U_1)^T = U_1^TU_0^T$, it follows A is similar to T and that U_0U_1 is orthogonal. Hence A and T have the same characteristic polynomials and since the eigenvalues of T are the diagonal entries listed according to algebraic multiplicity, this proves the theorem.

A.2.14 Symmetric Matrices

Recall a real matrix A is symmetric if $A = A^T$.

Lemma A.2.27 *A real symmetric matrix has all real eigenvalues.*

Proof: Recall the eigenvalues are solutions λ to

$$\det(\lambda I - A) = 0$$

and so by Theorem A.2.20, there exists $\mathbf{x}$ a vector such that

$$A\mathbf{x} = \lambda\mathbf{x},\ \mathbf{x} \neq \mathbf{0}$$

Of course if A is real, it is still possible that the eigenvalue could be complex and if this is the case, then the vector $\mathbf{x}$ will also end up being complex. I wish to show the eigenvalues are all real. Suppose then that λ is an eigenvalue and let $\mathbf{x}$ be the corresponding eigenvector described above. Then letting $\overline{\mathbf{x}}$ denote the complex conjugate of $\mathbf{x}$,

$$\lambda\mathbf{x}^T\overline{\mathbf{x}} = (A\mathbf{x})^T\overline{\mathbf{x}} = \mathbf{x}^TA^T\overline{\mathbf{x}} = \mathbf{x}^TA\overline{\mathbf{x}} = \mathbf{x}^T\overline{A\mathbf{x}} = \mathbf{x}^T\overline{\mathbf{x}}\overline{\lambda}$$

and so, cancelling $\mathbf{x}^T\overline{\mathbf{x}}$, it follows $\lambda = \overline{\lambda}$ showing λ is real. This proves the lemma.

Theorem A.2.28 *Let A be a real symmetric matrix. Then there exists a diagonal matrix D consisting of the eigenvalues of A down the main diagonal and an orthogonal matrix U such that*

$$U^TAU = D.$$

Proof: Since A has all real eigenvalues, it follows from Theorem A.2.26, there exists an orthogonal matrix U such that

$$U^TAU = T$$

where T is upper triangular. Now

$$T^T = U^TA^TU = U^TAU = T$$

and so in fact T is a diagonal matrix having the eigenvalues of A down the diagonal. This proves the theorem.

Theorem A.2.29 *Let A be a real symmetric matrix which has all positive eigenvalues $0 < \lambda_1 \leq \lambda_2 \cdots \leq \lambda_n$. Then*

$$(A\mathbf{x}\cdot\mathbf{x}) \equiv \mathbf{x}^TA\mathbf{x} \geq \lambda_1|\mathbf{x}|^2$$

Proof: Let U be the orthogonal matrix of Theorem A.2.28. Then

$$\begin{aligned}
(A\mathbf{x}\cdot\mathbf{x}) &= \mathbf{x}^TA\mathbf{x} = \left(\mathbf{x}^TU\right)D\left(U^T\mathbf{x}\right) \\
&= \left(U^T\mathbf{x}\right)D\left(U^T\mathbf{x}\right) = \sum_i \lambda_i\left|\left(U^T\mathbf{x}\right)_i\right|^2 \\
&\geq \lambda_1\sum_i\left|\left(U^T\mathbf{x}\right)_i\right|^2 = \lambda_1\left(U^T\mathbf{x}\cdot U^T\mathbf{x}\right) \\
&= \lambda_1\left(U^T\mathbf{x}\right)^TU^T\mathbf{x} = \lambda_1\mathbf{x}^TUU^T\mathbf{x} = \lambda_1\mathbf{x}^TI\mathbf{x} = \lambda_1|\mathbf{x}|^2.
\end{aligned}$$

A.3 Exercises

1. Let $m < n$ and let A be an $m \times n$ matrix. Show that A is **not** one to one. **Hint:** Consider the $n \times n$ matrix, A_1 which is of the form

$$A_1 \equiv \begin{pmatrix} A \\ 0 \end{pmatrix}$$

where the 0 denotes an $(n-m) \times n$ matrix of zeros. Thus $\det A_1 = 0$ and so A_1 is not one to one. Now observe that $A_1\mathbf{x}$ is the vector,

$$A_1\mathbf{x} = \begin{pmatrix} A\mathbf{x} \\ \mathbf{0} \end{pmatrix}$$

which equals zero if and only if $A\mathbf{x} = \mathbf{0}$.

2. Show that matrix multiplication is associative. That is, $(AB)\,C = A\,(BC)$.

3. Show the inverse of a matrix, if it exists, is unique. Thus if $AB = BA = I$, then $B = A^{-1}$.

4. In the proof of Theorem A.2.14 it was claimed that $\det(I) = 1$. Here $I = (\delta_{ij})$. Prove this assertion. Also prove Corollary A.2.17.

5. Let $\mathbf{v}_1, \cdots, \mathbf{v}_n$ be vectors in $\mathbb{F}^n$ and let $M(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ denote the matrix whose i^{th} column equals $\mathbf{v}_i$. Define

$$d(\mathbf{v}_1, \cdots, \mathbf{v}_n) \equiv \det(M(\mathbf{v}_1, \cdots, \mathbf{v}_n)).$$

Prove that d is linear in each variable, (multilinear), that

$$d(\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_n) = -d(\mathbf{v}_1, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_n), \tag{1.16}$$

and

$$d(\mathbf{e}_1, \cdots, \mathbf{e}_n) = 1 \tag{1.17}$$

where here $\mathbf{e}_j$ is the vector in $\mathbb{F}^n$ which has a zero in every position except the j^{th} position in which it has a one.

6. Suppose $f : \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ satisfies 1.16 and 1.17 and is linear in each variable. Show that $f = d$.

7. Show that if you replace a row (column) of an $n \times n$ matrix A with itself added to some multiple of another row (column) then the new matrix has the same determinant as the original one.

8. If $A = (a_{ij})$, show $\det(A) = \sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_n)\, a_{k_1 1} \cdots a_{k_n n}$.

9. Use the result of Problem 7 to evaluate by hand the determinant

$$\det \begin{pmatrix} 1 & 2 & 3 & 2 \\ -6 & 3 & 2 & 3 \\ 5 & 2 & 2 & 3 \\ 3 & 4 & 6 & 4 \end{pmatrix}.$$

10. Find the inverse if it exists of the matrix,

$$\begin{pmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{pmatrix}.$$

11. Let $Ly = y^{(n)} + a_{n-1}(x) y^{(n-1)} + \cdots + a_1(x) y' + a_0(x) y$ where the a_i are given continuous functions defined on a closed interval, (a,b) and y is some function which has n derivatives so it makes sense to write Ly. Suppose $Ly_k = 0$ for $k = 1, 2, \cdots, n$. The Wronskian of these functions, y_i is defined as

$$W(y_1, \cdots, y_n)(x) \equiv \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{pmatrix}$$

Show that for $W(x) = W(y_1, \cdots, y_n)(x)$ to save space,

$$W'(x) = \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n)}(x) & \cdots & y_n^{(n)}(x) \end{pmatrix}.$$

Now use the differential equation, $Ly = 0$ which is satisfied by each of these functions, y_i and properties of determinants presented above to verify that $W' + a_{n-1}(x) W = 0$. Give an explicit solution of this linear differential equation, Abel's formula, and use your answer to verify that the Wronskian of these solutions to the equation, $Ly = 0$ either vanishes identically on (a,b) or never.

12. Two $n \times n$ matrices, A and B, are similar if $B = S^{-1}AS$ for some invertible $n \times n$ matrix, S. Show that if two matrices are similar, they have the same characteristic polynomials.

13. Suppose the characteristic polynomial of an $n \times n$ matrix, A is of the form

$$t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$$

and that $a_0 \neq 0$. Find a formula A^{-1} in terms of powers of the matrix, A. Show that A^{-1} exists if and only if $a_0 \neq 0$.

14. In constitutive modelling of the stress and strain tensors, one sometimes considers sums of the form $\sum_{k=0}^{\infty} a_k A^k$ where A is a 3×3 matrix. Show using the Cayley Hamilton theorem that if such a thing makes any sense, you can always obtain it as a finite sum having no more than n terms.

Implicit Function Theorem*

The implicit function theorem is one of the greatest theorems in mathematics. There are many versions of this theorem which are of far greater generality than the one given here. The proof given here is like one found in one of Caratheodory's books on the calculus of variations. It is not as elegant as some of the others which are based on a contraction mapping principle but it may be more accessible. However, it is an advanced topic. Don't waste your time with it unless you have first read and understood the material on rank and determinants found in the chapter on the mathematical theory of determinants. You will also need to use the extreme value theorem for a function of n variables and the chain rule as well as everything about matrix multiplication.

Definition B.0.1 *Suppose U is an open set in $\mathbb{R}^n \times \mathbb{R}^m$ and $(\mathbf{x}, \mathbf{y})$ will denote a typical point of $\mathbb{R}^n \times \mathbb{R}^m$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Let $\mathbf{f} : U \to \mathbb{R}^p$ be in $C^1(U)$. Then define*

$$D_1\mathbf{f}(\mathbf{x},\mathbf{y}) \equiv \begin{pmatrix} f_{1,x_1}(\mathbf{x},\mathbf{y}) & \cdots & f_{1,x_n}(\mathbf{x},\mathbf{y}) \\ \vdots & & \vdots \\ f_{p,x_1}(\mathbf{x},\mathbf{y}) & \cdots & f_{p,x_n}(\mathbf{x},\mathbf{y}) \end{pmatrix},$$

$$D_2\mathbf{f}(\mathbf{x},\mathbf{y}) \equiv \begin{pmatrix} f_{1,y_1}(\mathbf{x},\mathbf{y}) & \cdots & f_{1,y_m}(\mathbf{x},\mathbf{y}) \\ \vdots & & \vdots \\ f_{p,y_1}(\mathbf{x},\mathbf{y}) & \cdots & f_{p,y_m}(\mathbf{x},\mathbf{y}) \end{pmatrix}.$$

Thus $D\mathbf{f}(\mathbf{x},\mathbf{y})$ is a $p \times (n+m)$ matrix of the form

$$D\mathbf{f}(\mathbf{x},\mathbf{y}) = \left(\begin{array}{c|c} D_1\mathbf{f}(\mathbf{x},\mathbf{y}) & D_2\mathbf{f}(\mathbf{x},\mathbf{y}) \end{array} \right).$$

Note that $D_1\mathbf{f}(\mathbf{x},\mathbf{y})$ is an $p \times n$ matrix and $D_2\mathbf{f}(\mathbf{x},\mathbf{y})$ is a $p \times m$ matrix.

Theorem B.0.2 *(implicit function theorem) Suppose U is an open set in $\mathbb{R}^n \times \mathbb{R}^m$. Let $\mathbf{f} : U \to \mathbb{R}^n$ be in $C^1(U)$ and suppose*

$$\mathbf{f}(\mathbf{x}_0,\mathbf{y}_0) = \mathbf{0},\ D_1\mathbf{f}(\mathbf{x}_0,\mathbf{y}_0)^{-1} \text{ exists.} \tag{2.1}$$

Then there exist positive constants, δ, η, such that for every $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ there exists a unique $\mathbf{x}(\mathbf{y}) \in B(\mathbf{x}_0, \delta)$ such that

$$\mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) = \mathbf{0}. \tag{2.2}$$

Furthermore, the mapping, $\mathbf{y} \to \mathbf{x}(\mathbf{y})$ is in $C^1(B(\mathbf{y}_0, \eta))$.

Proof: Let

$$\mathbf{f}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} f_1(\mathbf{x}, \mathbf{y}) \\ f_2(\mathbf{x}, \mathbf{y}) \\ \vdots \\ f_n(\mathbf{x}, \mathbf{y}) \end{pmatrix}.$$

Define for $(\mathbf{x}^1, \cdots, \mathbf{x}^n) \in \overline{B(\mathbf{x}_0, \delta)}^n$ and $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ the following matrix.

$$J(\mathbf{x}^1, \cdots, \mathbf{x}^n, \mathbf{y}) \equiv \begin{pmatrix} f_{1,x_1}(\mathbf{x}^1, \mathbf{y}) & \cdots & f_{1,x_n}(\mathbf{x}^1, \mathbf{y}) \\ \vdots & & \vdots \\ f_{n,x_1}(\mathbf{x}^n, \mathbf{y}) & \cdots & f_{n,x_n}(\mathbf{x}^n, \mathbf{y}) \end{pmatrix}.$$

Then by the assumption of continuity of all the partial derivatives and the extreme value theorem, there exists $r > 0$ and $\delta_0, \eta_0 > 0$ such that if $\delta \leq \delta_0$ and $\eta \leq \eta_0$, it follows that for all $(\mathbf{x}^1, \cdots, \mathbf{x}^n) \in \overline{B(\mathbf{x}_0, \delta)}^n$ and $\mathbf{y} \in \overline{B(\mathbf{y}_0, \eta)}$,

$$\det\left(J(\mathbf{x}^1, \cdots, \mathbf{x}^n, \mathbf{y})\right) > r > 0. \tag{2.3}$$

and $\overline{B(\mathbf{x}_0, \delta_0)} \times \overline{B(\mathbf{y}_0, \eta_0)} \subseteq U$. By continuity of all the partial derivatives and the extreme value theorem, it can also be assumed there exists a constant, K such that for all $(\mathbf{x}, \mathbf{y}) \in \overline{B(\mathbf{x}_0, \delta_0)} \times \overline{B(\mathbf{y}_0, \eta_0)}$ and $i = 1, 2, \cdots, n$, the i^{th} row of $D_2\mathbf{f}(\mathbf{x}, \mathbf{y})$, given by $D_2 f_i(\mathbf{x}, \mathbf{y})$ satisfies

$$|D_2 f_i(\mathbf{x}, \mathbf{y})| < K, \tag{2.4}$$

and for all $(\mathbf{x}^1, \cdots, \mathbf{x}^n) \in \overline{B(\mathbf{x}_0, \delta_0)}^n$ and $\mathbf{y} \in \overline{B(\mathbf{y}_0, \eta_0)}$ the i^{th} row of the matrix,

$$J(\mathbf{x}^1, \cdots, \mathbf{x}^n, \mathbf{y})^{-1}$$

which equals $\mathbf{e}_i^T \left(J(\mathbf{x}^1, \cdots, \mathbf{x}^n, \mathbf{y})^{-1}\right)$ satisfies

$$\left|\mathbf{e}_i^T \left(J(\mathbf{x}^1, \cdots, \mathbf{x}^n, \mathbf{y})^{-1}\right)\right| < K. \tag{2.5}$$

(Recall that $\mathbf{e}_i$ is the column vector consisting of all zeros except for a 1 in the i^{th} position.)

To begin with it is shown that for a given $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ there is at most one $\mathbf{x} \in B(\mathbf{x}_0, \delta)$ such that $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$.

Pick $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ and suppose there exist $\mathbf{x}, \mathbf{z} \in \overline{B(\mathbf{x}_0, \delta)}$ such that $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{f}(\mathbf{z}, \mathbf{y}) = \mathbf{0}$. Consider f_i and let

$$h(t) \equiv f_i(\mathbf{x} + t(\mathbf{z} - \mathbf{x}), \mathbf{y}).$$

Then $h(1) = h(0)$ and so by the mean value theorem, $h'(t_i) = 0$ for some $t_i \in (0, 1)$. Therefore, from the chain rule and for this value of t_i,

$$h'(t_i) = Df_i(\mathbf{x} + t_i(\mathbf{z} - \mathbf{x}), \mathbf{y})(\mathbf{z} - \mathbf{x}) = 0. \tag{2.6}$$

Then denote by $\mathbf{x}^i$ the vector, $\mathbf{x}+t_i(\mathbf{z}-\mathbf{x})$. It follows from 2.6 that

$$J\left(\mathbf{x}^1,\cdots,\mathbf{x}^n,\mathbf{y}\right)(\mathbf{z}-\mathbf{x})=\mathbf{0}$$

and so from 2.3 $\mathbf{z}-\mathbf{x}=\mathbf{0}$. (The matrix, in the above is invertible since its determinant is nonzero.) Now it will be shown that if η is chosen sufficiently small, then for all $\mathbf{y}\in B(\mathbf{y}_0,\eta)$, there exists a unique $\mathbf{x}(\mathbf{y})\in B(\mathbf{x}_0,\delta)$ such that $\mathbf{f}(\mathbf{x}(\mathbf{y}),\mathbf{y})=\mathbf{0}$.

Claim: If η is small enough, then the function, $h_{\mathbf{y}}(\mathbf{x})\equiv|\mathbf{f}(\mathbf{x},\mathbf{y})|^2$ achieves its minimum value on $\overline{B(\mathbf{x}_0,\delta)}$ at a point of $B(\mathbf{x}_0,\delta)$. (The existence of a point in $\overline{B(\mathbf{x}_0,\delta)}$ at which $h_{\mathbf{y}}$ achieves its minimum follows from the extreme value theorem.)

Proof of claim: Suppose this is not the case. Then there exists a sequence $\eta_k\to 0$ and for some $\mathbf{y}_k$ having $|\mathbf{y}_k-\mathbf{y}_0|<\eta_k$, the minimum of $h_{\mathbf{y}_k}$on $\overline{B(\mathbf{x}_0,\delta)}$ occurs on a point of $\overline{B(\mathbf{x}_0,\delta)}$, $\mathbf{x}_k$ such that $|\mathbf{x}_0-\mathbf{x}_k|=\delta$. Now taking a subsequence, still denoted by k, it can be assumed that $\mathbf{x}_k\to\mathbf{x}$ with $|\mathbf{x}-\mathbf{x}_0|=\delta$ and $\mathbf{y}_k\to\mathbf{y}_0$. This follows from the fact that $\left\{\mathbf{x}\in\overline{B(\mathbf{x}_0,\delta)}:|\mathbf{x}-\mathbf{x}_0|=\delta\right\}$ is a closed and bounded set and is therefore sequentially compact. Let $\varepsilon>0$. Then for k large enough, the continuity of $\mathbf{y}\to h_{\mathbf{y}}(\mathbf{x}_0)$ implies $h_{\mathbf{y}_k}(\mathbf{x}_0)<\varepsilon$ because $h_{\mathbf{y}_0}(\mathbf{x}_0)=0$ since $\mathbf{f}(\mathbf{x}_0,\mathbf{y}_0)=\mathbf{0}$. Therefore, from the definition of $\mathbf{x}_k$, it is also the case that $h_{\mathbf{y}_k}(\mathbf{x}_k)<\varepsilon$. Passing to the limit yields $h_{\mathbf{y}_0}(\mathbf{x})\le\varepsilon$. Since $\varepsilon>0$ is arbitrary, it follows that $h_{\mathbf{y}_0}(\mathbf{x})=0$ which contradicts the first part of the argument in which it was shown that for $\mathbf{y}\in B(\mathbf{y}_0,\eta)$ there is at most one point, $\mathbf{x}$ of $\overline{B(\mathbf{x}_0,\delta)}$ where $\mathbf{f}(\mathbf{x},\mathbf{y})=\mathbf{0}$. Here two have been obtained, $\mathbf{x}_0$ and $\mathbf{x}$. This proves the claim.

Choose $\eta<\eta_0$ and also small enough that the above claim holds and let $\mathbf{x}(\mathbf{y})$ denote a point of $B(\mathbf{x}_0,\delta)$ at which the minimum of $h_{\mathbf{y}}$ on $\overline{B(\mathbf{x}_0,\delta)}$ is achieved. Since $\mathbf{x}(\mathbf{y})$ is an interior point, you can consider $h_{\mathbf{y}}(\mathbf{x}(\mathbf{y})+t\mathbf{v})$ for $|t|$ small and conclude this function of t has a zero derivative at $t=0$. Now

$$h_{\mathbf{y}}(\mathbf{x}(\mathbf{y})+t\mathbf{v})=\sum_{i=1}^{n}f_i^2(\mathbf{x}(\mathbf{y})+t\mathbf{v},\mathbf{y})$$

and so from the chain rule,

$$\frac{d}{dt}h_{\mathbf{y}}(\mathbf{x}(\mathbf{y})+t\mathbf{v})=\sum_{i=1}^{n}2f_i(\mathbf{x}(\mathbf{y})+t\mathbf{v},\mathbf{y})\frac{\partial f_i(\mathbf{x}(\mathbf{y})+t\mathbf{v},\mathbf{y})}{\partial x_j}v_j.$$

Therefore, letting $t=0$, it is required that for every $\mathbf{v}$,

$$\sum_{i=1}^{n}2f_i(\mathbf{x}(\mathbf{y}),\mathbf{y})\frac{\partial f_i(\mathbf{x}(\mathbf{y}),\mathbf{y})}{\partial x_j}v_j=0.$$

In terms of matrices this reduces to

$$0=2\mathbf{f}(\mathbf{x}(\mathbf{y}),\mathbf{y})^T D_1\mathbf{f}(\mathbf{x}(\mathbf{y}),\mathbf{y})\mathbf{v}$$

for every vector $\mathbf{v}$. Therefore,

$$\mathbf{0}=\mathbf{f}(\mathbf{x}(\mathbf{y}),\mathbf{y})^T D_1\mathbf{f}(\mathbf{x}(\mathbf{y}),\mathbf{y})$$

From 2.3, it follows $\mathbf{f}(\mathbf{x}(\mathbf{y}),\mathbf{y})=\mathbf{0}$. This proves the existence of the function $\mathbf{y}\to\mathbf{x}(\mathbf{y})$ such that $\mathbf{f}(\mathbf{x}(\mathbf{y}),\mathbf{y})=\mathbf{0}$ for all $\mathbf{y}\in B(\mathbf{y}_0,\eta)$.

It remains to verify this function is a C^1 function. To do this, let $\mathbf{y}_1$ and $\mathbf{y}_2$ be points of $B(\mathbf{y}_0,\eta)$. Then as before, consider the i^{th} component of $\mathbf{f}$ and consider the same argument using the mean value theorem to write

$$\begin{aligned}0&=f_i(\mathbf{x}(\mathbf{y}_1),\mathbf{y}_1)-f_i(\mathbf{x}(\mathbf{y}_2),\mathbf{y}_2)\\&=f_i(\mathbf{x}(\mathbf{y}_1),\mathbf{y}_1)-f_i(\mathbf{x}(\mathbf{y}_2),\mathbf{y}_1)+f_i(\mathbf{x}(\mathbf{y}_2),\mathbf{y}_1)-f_i(\mathbf{x}(\mathbf{y}_2),\mathbf{y}_2)\\&=D_1f_i\left(\mathbf{x}^i,\mathbf{y}_1\right)(\mathbf{x}(\mathbf{y}_1)-\mathbf{x}(\mathbf{y}_2))+D_2f_i\left(\mathbf{x}(\mathbf{y}_2),\mathbf{y}^i\right)(\mathbf{y}_1-\mathbf{y}_2).\end{aligned}\tag{2.7}$$

where $\mathbf{y}^i$ is a point on the line segment joining $\mathbf{y}_1$ and $\mathbf{y}_2$. Thus from 2.4 and the Cauchy Schwarz inequality,

$$\left|D_2 f_i\left(\mathbf{x}\left(\mathbf{y}_2\right),\mathbf{y}^i\right)\left(\mathbf{y}_1-\mathbf{y}_2\right)\right| \leq K\left|\mathbf{y}_1-\mathbf{y}_2\right|.$$

Therefore, letting $M\left(\mathbf{y}^1,\cdots,\mathbf{y}^n\right)\equiv M$ denote the matrix having the i^{th} row equal to $D_2 f_i\left(\mathbf{x}\left(\mathbf{y}_2\right),\mathbf{y}^i\right)$, it follows

$$\left|M\left(\mathbf{y}_1-\mathbf{y}_2\right)\right| \leq \left(\sum_i K^2\left|\mathbf{y}_1-\mathbf{y}_2\right|^2\right)^{1/2} = \sqrt{m}K\left|\mathbf{y}_1-\mathbf{y}_2\right|. \tag{2.8}$$

Also, from 2.7,

$$J\left(\mathbf{x}^1,\cdots,\mathbf{x}^n,\mathbf{y}_1\right)\left(\mathbf{x}\left(\mathbf{y}_1\right)-\mathbf{x}\left(\mathbf{y}_2\right)\right) = -M\left(\mathbf{y}_1-\mathbf{y}_2\right) \tag{2.9}$$

and so from 2.8 and 2.10,

$$\begin{aligned}
\left|\mathbf{x}\left(\mathbf{y}_1\right)-\mathbf{x}\left(\mathbf{y}_2\right)\right| &= \left|J\left(\mathbf{x}^1,\cdots,\mathbf{x}^n,\mathbf{y}_1\right)^{-1} M\left(\mathbf{y}_1-\mathbf{y}_2\right)\right| &(2.10)\\
&= \left(\sum_{i=1}^n \left|\mathbf{e}_i^T J\left(\mathbf{x}^1,\cdots,\mathbf{x}^n,\mathbf{y}_1\right)^{-1} M\left(\mathbf{y}_1-\mathbf{y}_2\right)\right|^2\right)^{1/2}\\
&\leq \left(\sum_{i=1}^n K^2\left|M\left(\mathbf{y}_1-\mathbf{y}_2\right)\right|^2\right)^{1/2} \leq \left(\sum_{i=1}^n K^2\left(\sqrt{m}K\left|\mathbf{y}_1-\mathbf{y}_2\right|\right)^2\right)^{1/2}\\
&= K^2\sqrt{mn}\left|\mathbf{y}_1-\mathbf{y}_2\right| &(2.11)
\end{aligned}$$

It follows as in the proof of the chain rule that

$$\mathbf{o}\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right)-\mathbf{x}\left(\mathbf{y}\right)\right) = \mathbf{o}\left(\mathbf{v}\right). \tag{2.12}$$

Now let $\mathbf{y}\in B\left(\mathbf{y}_0,\eta\right)$ and let $|\mathbf{v}|$ be sufficiently small that $\mathbf{y}+\mathbf{v}\in B\left(\mathbf{y}_0,\eta\right)$. Then

$$\begin{aligned}
\mathbf{0} &= \mathbf{f}\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right),\mathbf{y}+\mathbf{v}\right)-\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)\\
&= \mathbf{f}\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right),\mathbf{y}+\mathbf{v}\right)-\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right),\mathbf{y}\right)+\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right),\mathbf{y}\right)-\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)
\end{aligned}$$

$$= D_2\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right),\mathbf{y}\right)\mathbf{v}+D_1\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right)-\mathbf{x}\left(\mathbf{y}\right)\right)+\mathbf{o}\left(\left|\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right)-\mathbf{x}\left(\mathbf{y}\right)\right|\right)$$

$$\begin{aligned}
&= D_2\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)\mathbf{v}+D_1\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right)-\mathbf{x}\left(\mathbf{y}\right)\right)+\\
&\quad \mathbf{o}\left(\left|\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right)-\mathbf{x}\left(\mathbf{y}\right)\right|\right)+\left(D_2\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right),\mathbf{y}\right)\mathbf{v}-D_2\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)\mathbf{v}\right)\\
&= D_2\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)\mathbf{v}+D_1\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)\left(\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right)-\mathbf{x}\left(\mathbf{y}\right)\right)+\mathbf{o}\left(\mathbf{v}\right).
\end{aligned}$$

Therefore,

$$\mathbf{x}\left(\mathbf{y}+\mathbf{v}\right)-\mathbf{x}\left(\mathbf{y}\right) = -D_1\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)^{-1} D_2\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)\mathbf{v}+\mathbf{o}\left(\mathbf{v}\right)$$

which shows that $D\mathbf{x}\left(\mathbf{y}\right) = -D_1\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)^{-1} D_2\mathbf{f}\left(\mathbf{x}\left(\mathbf{y}\right),\mathbf{y}\right)$ and $\mathbf{y}\to D\mathbf{x}\left(\mathbf{y}\right)$ is continuous. This proves the theorem.

B.1 The Method Of Lagrange Multipliers

As an application of the implicit function theorem, consider the method of Lagrange multipliers. Recall the problem is to maximize or minimize a function subject to equality constraints. Let $f: U\to\mathbb{R}$ be a C^1 function where $U\subseteq\mathbb{R}^n$ and let

$$g_i\left(\mathbf{x}\right) = 0,\ i=1,\cdots,m \tag{2.13}$$

be a collection of equality constraints with $m < n$. Now consider the system of nonlinear equations

$$\begin{aligned} f(\mathbf{x}) &= a \\ g_i(\mathbf{x}) &= 0,\ i = 1, \cdots, m. \end{aligned}$$

Recall $\mathbf{x}_0$ is a local maximum if $f(\mathbf{x}_0) \geq f(\mathbf{x})$ for all $\mathbf{x}$ near $\mathbf{x}_0$ which also satisfies the constraints 2.13. A local minimum is defined similarly. Let $\mathbf{F} : U \times \mathbb{R} \to \mathbb{R}^{m+1}$ be defined by

$$\mathbf{F}(\mathbf{x},a) \equiv \begin{pmatrix} f(\mathbf{x}) - a \\ g_1(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{pmatrix}. \tag{2.14}$$

Now consider the $m+1 \times n$ matrix,

$$\begin{pmatrix} f_{x_1}(\mathbf{x}_0) & \cdots & f_{x_n}(\mathbf{x}_0) \\ g_{1x_1}(\mathbf{x}_0) & \cdots & g_{1x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ g_{mx_1}(\mathbf{x}_0) & \cdots & g_{mx_n}(\mathbf{x}_0) \end{pmatrix}.$$

If this matrix has rank $m+1$ then some $m+1 \times m+1$ submatrix has nonzero determinant. It follows from the implicit function theorem there exists $m+1$ variables, $x_{i_1}, \cdots, x_{i_{m+1}}$ such that the system

$$\mathbf{F}(\mathbf{x},a) = \mathbf{0} \tag{2.15}$$

specifies these $m+1$ variables as a function of the remaining $n-(m+1)$ variables and a in an open set of $\mathbb{R}^{n-m}$. Thus there is a solution $(\mathbf{x},a)$ to 2.15 for some $\mathbf{x}$ close to $\mathbf{x}_0$ whenever a is in some open interval. Therefore, $\mathbf{x}_0$ cannot be either a local minimum or a local maximum. It follows that if $\mathbf{x}_0$ is either a local maximum or a local minimum, then the above matrix must have rank less than $m+1$ which requires the rows to be linearly dependent. Thus, there exist m scalars,

$$\lambda_1, \cdots, \lambda_m,$$

and a scalar μ, not all zero such that

$$\mu \begin{pmatrix} f_{x_1}(\mathbf{x}_0) \\ \vdots \\ f_{x_n}(\mathbf{x}_0) \end{pmatrix} = \lambda_1 \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix} + \cdots + \lambda_m \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix}. \tag{2.16}$$

If the column vectors

$$\begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix}, \cdots \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix} \tag{2.17}$$

are linearly independent, then, $\mu \neq 0$ and dividing by μ yields an expression of the form

$$\begin{pmatrix} f_{x_1}(\mathbf{x}_0) \\ \vdots \\ f_{x_n}(\mathbf{x}_0) \end{pmatrix} = \lambda_1 \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix} + \cdots + \lambda_m \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix} \tag{2.18}$$

at every point $\mathbf{x}_0$ which is either a local maximum or a local minimum. This proves the following theorem.

Theorem B.1.1 *Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be a C^1 function. Then if $\mathbf{x}_0 \in U$ is either a local maximum or local minimum of f subject to the constraints 2.13, then 2.16 must hold for some scalars $\mu, \lambda_1, \cdots, \lambda_m$ not all equal to zero. If the vectors in 2.17 are linearly independent, it follows that an equation of the form 2.18 holds.*

B.2 The Local Structure Of C^1 Mappings

Definition B.2.1 *Let U be an open set in $\mathbb{R}^n$ and let $\mathbf{h} : U \to \mathbb{R}^n$. Then $\mathbf{h}$ is called primitive if it is of the form*

$$\mathbf{h}(\mathbf{x}) = \begin{pmatrix} x_1 & \cdots & \alpha(\mathbf{x}) & \cdots & x_n \end{pmatrix}^T.$$

Thus, $\mathbf{h}$ is primitive if it only changes one of the variables. A function, $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is called a flip if

$$\mathbf{F}(x_1, \cdots, x_k, \cdots, x_l, \cdots, x_n) = (x_1, \cdots, x_l, \cdots, x_k, \cdots, x_n)^T.$$

Thus a function is a flip if it interchanges two coordinates. Also, for $m = 1, 2, \cdots, n$,

$$P_m(\mathbf{x}) \equiv \begin{pmatrix} x_1 & x_2 & \cdots & x_m & 0 & \cdots & 0 \end{pmatrix}^T$$

It turns out that if $\mathbf{h}(\mathbf{0}) = \mathbf{0}$,$D\mathbf{h}(0)^{-1}$ exists, and $\mathbf{h}$ is C^1 on U, then $\mathbf{h}$ can be written as a composition of primitive functions and flips. This is a very interesting application of the inverse function theorem.

Theorem B.2.2 *Let $\mathbf{h} : U \to \mathbb{R}^n$ be a C^1 function with $\mathbf{h}(\mathbf{0}) = \mathbf{0}$,$D\mathbf{h}(\mathbf{0})^{-1}$ exists. Then there an open set, $V \subseteq U$ containing $\mathbf{0}$, flips, $\mathbf{F}_1, \cdots, \mathbf{F}_{n-1}$, and primitive functions, $\mathbf{G}_n, \mathbf{G}_{n-1}, \cdots, \mathbf{G}_1$ such that for $\mathbf{x} \in V$,*

$$\mathbf{h}(\mathbf{x}) = \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{x}).$$

Proof: Let

$$\mathbf{h}_1(\mathbf{x}) \equiv \mathbf{h}(\mathbf{x}) = \begin{pmatrix} \alpha_1(\mathbf{x}) & \cdots & \alpha_n(\mathbf{x}) \end{pmatrix}^T$$

$$D\mathbf{h}(\mathbf{0})\mathbf{e}_1 = \begin{pmatrix} \alpha_{1,1}(\mathbf{0}) & \cdots & \alpha_{n,1}(\mathbf{0}) \end{pmatrix}^T$$

where $\alpha_{k,1}$ denotes $\frac{\partial \alpha_k}{\partial x_1}$. Since $D\mathbf{h}(\mathbf{0})$ is one to one, the right side of this expression cannot be zero. Hence there exists some k such that $\alpha_{k,1}(\mathbf{0}) \neq 0$. Now define

$$\mathbf{G}_1(\mathbf{x}) \equiv \begin{pmatrix} \alpha_k(\mathbf{x}) & x_2 & \cdots & x_n \end{pmatrix}^T$$

Then the matrix of $D\mathbf{G}(\mathbf{0})$ is of the form

$$\begin{pmatrix} \alpha_{k,1}(\mathbf{0}) & \cdots & \cdots & \alpha_{k,n}(0) \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

and its determinant equals $\alpha_{k,1}(\mathbf{0}) \neq 0$. Therefore, by the inverse function theorem, there exists an open set, U_1 containing $\mathbf{0}$ and an open set, V_2 containing $\mathbf{0}$ such that $\mathbf{G}_1(U_1) = V_2$ and $\mathbf{G}_1$ is one to one and onto such that it and its inverse are both C^1. Let $\mathbf{F}_1$ denote the flip which interchanges x_k with x_1. Now define

$$\mathbf{h}_2(\mathbf{y}) \equiv \mathbf{F}_1 \circ \mathbf{h}_1 \circ \mathbf{G}_1^{-1}(\mathbf{y})$$

Thus

$$\begin{aligned}\mathbf{h}_2(\mathbf{G}_1(\mathbf{x})) &\equiv \mathbf{F}_1 \circ \mathbf{h}_1(\mathbf{x}) \\ &= \begin{pmatrix} \alpha_k(\mathbf{x}) & \cdots & \alpha_1(\mathbf{x}) & \cdots & \alpha_n(\mathbf{x}) \end{pmatrix}^T\end{aligned} \tag{2.19}$$

Therefore,

$$P_1\mathbf{h}_2(\mathbf{G}_1(\mathbf{x})) = \begin{pmatrix} \alpha_k(\mathbf{x}) & 0 & \cdots & 0 \end{pmatrix}^T.$$

Also

$$P_1(\mathbf{G}_1(\mathbf{x})) = \begin{pmatrix} \alpha_k(\mathbf{x}) & x_2 & \cdots & x_n \end{pmatrix}^T$$

so $P_1\mathbf{h}_2(\mathbf{y}) = P_1(\mathbf{y})$ for all $\mathbf{y} \in V_2$. Also, $\mathbf{h}_2(\mathbf{0}) = \mathbf{0}$ and $D\mathbf{h}_2(0)^{-1}$ exists because of the definition of $\mathbf{h}_2$ above and the chain rule. Also, since $\mathbf{F}_1^2 = identity$, it follows from 2.19 that

$$\mathbf{h}(\mathbf{x}) = \mathbf{h}_1(\mathbf{x}) = \mathbf{F}_1 \circ \mathbf{h}_2 \circ \mathbf{G}_1(\mathbf{x}). \tag{2.20}$$

Suppose then that for $m \geq 2$,

$$P_{m-1}\mathbf{h}_m(\mathbf{x}) = P_{m-1}(\mathbf{x}) \tag{2.21}$$

for all $\mathbf{x} \in U_m$, an open subset of U containing $\mathbf{0}$ and $\mathbf{h}_m(\mathbf{0}) = \mathbf{0}$,$D\mathbf{h}_m(\mathbf{0})^{-1}$ exists. From 2.21, $\mathbf{h}_m(\mathbf{x})$ must be of the form

$$\mathbf{h}_m(\mathbf{x}) = \begin{pmatrix} x_1 & \cdots & x_{m-1} & \alpha_1(\mathbf{x}) & \cdots & \alpha_n(\mathbf{x}) \end{pmatrix}^T$$

where these α_k are different than the ones used earlier. Then

$$D\mathbf{h}_m(\mathbf{0})\mathbf{e}_m = \begin{pmatrix} 0 & \cdots & 0 & \alpha_{1,m}(\mathbf{0}) & \cdots & \alpha_{n,m}(\mathbf{0}) \end{pmatrix}^T \neq \mathbf{0}$$

because $D\mathbf{h}_m(\mathbf{0})^{-1}$ exists. Therefore, there exists a k such that $\alpha_{k,m}(\mathbf{0}) \neq 0$, not the same k as before. Define

$$\mathbf{G}_{m+1}(\mathbf{x}) \equiv \begin{pmatrix} x_1 & \cdots & x_{m-1} & \alpha_k(\mathbf{x}) & x_{m+1} & \cdots & x_n \end{pmatrix}^T \tag{2.22}$$

Then $G_{m+1}(\mathbf{0}) = \mathbf{0}$ and $D\mathbf{G}_{m+1}(\mathbf{0})^{-1}$ exists similar to the above. In fact $\det(D\mathbf{G}_{m+1}(\mathbf{0})) = \alpha_{k,m}(\mathbf{0})$. Therefore, by the inverse function theorem, there exists an open set, V_{m+1} containing $\mathbf{0}$ such that $V_{m+1} = \mathbf{G}_{m+1}(U_m)$ with $\mathbf{G}_{m+1}$ and its inverse being one to one continuous and onto. Let $\mathbf{F}_m$ be the flip which flips x_m and x_k. Then define $\mathbf{h}_{m+1}$ on V_{m+1} by

$$\mathbf{h}_{m+1}(\mathbf{y}) = \mathbf{F}_m \circ \mathbf{h}_m \circ \mathbf{G}_{m+1}^{-1}(\mathbf{y}).$$

Thus for $\mathbf{x} \in U_m$,

$$\mathbf{h}_{m+1}(\mathbf{G}_{m+1}(\mathbf{x})) = (\mathbf{F}_m \circ \mathbf{h}_m)(\mathbf{x}). \tag{2.23}$$

and consequently,

$$\mathbf{F}_m \circ \mathbf{h}_{m+1} \circ \mathbf{G}_{m+1}(\mathbf{x}) = \mathbf{h}_m(\mathbf{x}) \tag{2.24}$$

It follows

$$\begin{aligned}P_m\mathbf{h}_{m+1}(\mathbf{G}_{m+1}(\mathbf{x})) &= P_m(\mathbf{F}_m \circ \mathbf{h}_m)(\mathbf{x}) \\ &= \begin{pmatrix} x_1 & \cdots & x_{m-1} & \alpha_k(\mathbf{x}) & 0 & \cdots & 0 \end{pmatrix}^T\end{aligned}$$

and

$$P_m(\mathbf{G}_{m+1}(\mathbf{x})) = \begin{pmatrix} x_1 & \cdots & x_{m-1} & \alpha_k(\mathbf{x}) & 0 & \cdots & 0 \end{pmatrix}^T.$$

Therefore, for $\mathbf{y} \in V_{m+1}$,

$$P_m\mathbf{h}_{m+1}(\mathbf{y}) = P_m(\mathbf{y}).$$

As before, $\mathbf{h}_{m+1}(\mathbf{0}) = \mathbf{0}$ and $D\mathbf{h}_{m+1}(\mathbf{0})^{-1}$ exists. Therefore, we can apply 2.24 repeatedly, obtaining the following:

$$\begin{aligned}\mathbf{h}(\mathbf{x}) &= \mathbf{F}_1 \circ \mathbf{h}_2 \circ \mathbf{G}_1(\mathbf{x}) \\ &= \mathbf{F}_1 \circ \mathbf{F}_2 \circ \mathbf{h}_3 \circ \mathbf{G}_2 \circ \mathbf{G}_1(\mathbf{x}) \\ &\vdots \\ &= \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{h}_n \circ \mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{x})\end{aligned}$$

where

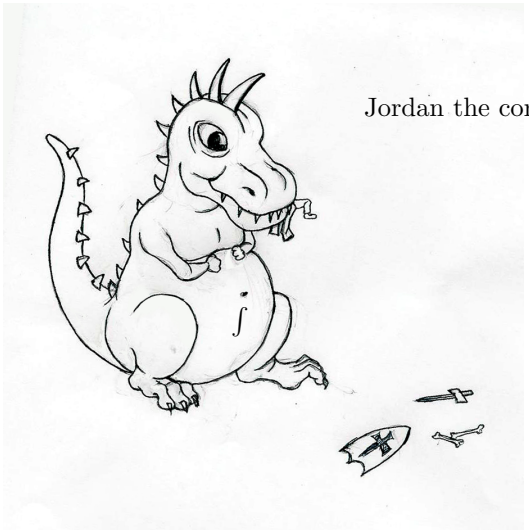
$$P_{n-1}\mathbf{h}_n(\mathbf{x}) = P_{n-1}(\mathbf{x}) = \begin{pmatrix} x_1 & \cdots & x_{n-1} & 0 \end{pmatrix}^T$$

and so $\mathbf{h}_n(\mathbf{x})$ is of the form

$$\mathbf{h}_n(\mathbf{x}) = \begin{pmatrix} x_1 & \cdots & x_{n-1} & \alpha(\mathbf{x}) \end{pmatrix}^T.$$

Therefore, define the primitive function, $\mathbf{G}_n(\mathbf{x})$ to equal $\mathbf{h}_n(\mathbf{x})$. This proves the theorem.

The Theory Of The Riemann Integral*



Jordan the contented dragon

The definition of the Riemann integral of a function of n variables uses the following definition.

Definition C.0.3 *For $i = 1, \cdots, n$, let $\left\{\alpha_k^i\right\}_{k=-\infty}^{\infty}$ be points on $\mathbb{R}$ which satisfy*

$$\lim_{k\to\infty} \alpha_k^i = \infty,\ \lim_{k\to-\infty} \alpha_k^i = -\infty,\ \alpha_k^i < \alpha_{k+1}^i. \tag{3.1}$$

For such sequences, define a grid on $\mathbb{R}^n$ denoted by $\mathcal{G}$ or $\mathcal{F}$ as the collection of boxes of the form

$$Q = \prod_{i=1}^{n} \left[\alpha_{j_i}^i, \alpha_{j_i+1}^i\right]. \tag{3.2}$$

If $\mathcal{G}$ is a grid, $\mathcal{F}$ is called a refinement of $\mathcal{G}$ if every box of $\mathcal{G}$ is the union of boxes of $\mathcal{F}$.

Lemma C.0.4 *If $\mathcal{G}$ and $\mathcal{F}$ are two grids, they have a common refinement, denoted here by $\mathcal{G} \vee \mathcal{F}$.*

Proof: Let $\left\{\alpha_k^i\right\}_{k=-\infty}^{\infty}$ be the sequences used to construct $\mathcal{G}$ and let $\left\{\beta_k^i\right\}_{k=-\infty}^{\infty}$ be the sequence used to construct $\mathcal{F}$. Now let $\left\{\gamma_k^i\right\}_{k=-\infty}^{\infty}$ denote the union of $\left\{\alpha_k^i\right\}_{k=-\infty}^{\infty}$

and $\left\{\beta_k^i\right\}_{k=-\infty}^{\infty}$. It is necessary to show that for each i these points can be arranged in order. To do so, let $\gamma_0^i \equiv \alpha_0^i$. Now if

$$\gamma_{-j}^i, \cdots, \gamma_0^i, \cdots, \gamma_j^i$$

have been chosen such that they are in order and all distinct, let γ_{j+1}^i be the first element of

$$\left\{\alpha_k^i\right\}_{k=-\infty}^{\infty} \cup \left\{\beta_k^i\right\}_{k=-\infty}^{\infty} \tag{3.3}$$

which is larger than γ_j^i and let $\gamma_{-(j+1)}^i$ be the last element of 3.3 which is strictly smaller than γ_{-j}^i. The assumption 3.1 insures such a first and last element exists. Now let the grid $\mathcal{G} \vee \mathcal{F}$ consist of boxes of the form

$$Q \equiv \prod_{i=1}^{n} \left[\gamma_{j_i}^i, \gamma_{j_i+1}^i\right].$$

The Riemann integral is only defined for functions, f which are bounded and are equal to zero off some bounded set, D. In what follows f will always be such a function.

Definition C.0.5 *Let f be a bounded function which equals zero off a bounded set, D, and let $\mathcal{G}$ be a grid. For $Q \in \mathcal{G}$, define*

$$M_Q(f) \equiv \sup\{f(\mathbf{x}) : \mathbf{x} \in Q\},\ m_Q(f) \equiv \inf\{f(\mathbf{x}) : \mathbf{x} \in Q\}. \tag{3.4}$$

Also define for Q a box, the volume of Q, denoted by $v(Q)$ by

$$v(Q) \equiv \prod_{i=1}^{n} (b_i - a_i),\ Q \equiv \prod_{i=1}^{n} [a_i, b_i].$$

Now define upper sums, $\mathcal{U}_{\mathcal{G}}(f)$ and lower sums, $\mathcal{L}_{\mathcal{G}}(f)$ with respect to the indicated grid, by the formulas

$$\mathcal{U}_{\mathcal{G}}(f) \equiv \sum_{Q \in \mathcal{G}} M_Q(f)\, v(Q),\ \mathcal{L}_{\mathcal{G}}(f) \equiv \sum_{Q \in \mathcal{G}} m_Q(f)\, v(Q).$$

A function of n variables is Riemann integrable when there is a unique number between all the upper and lower sums. This number is the value of the integral.

Note that in this definition, $M_Q(f) = m_Q(f) = 0$ for all but finitely many $Q \in \mathcal{G}$ so there are no convergence questions to be considered here.

Lemma C.0.6 *If $\mathcal{F}$ is a refinement of $\mathcal{G}$ then*

$$\mathcal{U}_{\mathcal{G}}(f) \geq \mathcal{U}_{\mathcal{F}}(f),\ \mathcal{L}_{\mathcal{G}}(f) \leq \mathcal{L}_{\mathcal{F}}(f).$$

Also if $\mathcal{F}$ and $\mathcal{G}$ are two grids,

$$\mathcal{L}_{\mathcal{G}}(f) \leq \mathcal{U}_{\mathcal{F}}(f).$$

Proof: For $P \in \mathcal{G}$ let $\widehat{P}$ denote the set,

$$\{Q \in \mathcal{F} : Q \subseteq P\}.$$

Then $P = \cup \widehat{P}$ and

$$\mathcal{L}_{\mathcal{F}}(f) \equiv \sum_{Q \in \mathcal{F}} m_Q(f)\, v(Q) = \sum_{P \in \mathcal{G}} \sum_{Q \in \widehat{P}} m_Q(f)\, v(Q)$$

$$\geq \sum_{P\in\mathcal{G}} m_P(f) \sum_{Q\in\widehat{P}} v(Q) = \sum_{P\in\mathcal{G}} m_P(f)\, v(P) \equiv \mathcal{L}_{\mathcal{G}}(f).$$

Similarly, the other inequality for the upper sums is valid.

To verify the last assertion of the lemma, use Lemma C.0.4 to write

$$\mathcal{L}_{\mathcal{G}}(f) \leq \mathcal{L}_{\mathcal{G}\vee\mathcal{F}}(f) \leq \mathcal{U}_{\mathcal{G}\vee\mathcal{F}}(f) \leq \mathcal{U}_{\mathcal{F}}(f).$$

This proves the lemma.

This lemma makes it possible to define the Riemann integral.

Definition C.0.7 *Define an upper and a lower integral as follows.*

$$\overline{I}(f) \equiv \inf\{\mathcal{U}_{\mathcal{G}}(f) : \mathcal{G} \text{ is a grid}\},$$

$$\underline{I}(f) \equiv \sup\{\mathcal{L}_{\mathcal{G}}(f) : \mathcal{G} \text{ is a grid}\}.$$

Lemma C.0.8 $\overline{I}(f) \geq \underline{I}(f)$.

Proof: From Lemma C.0.6 it follows for any two grids $\mathcal{G}$ and $\mathcal{F}$,

$$\mathcal{L}_{\mathcal{G}}(f) \leq \mathcal{U}_{\mathcal{F}}(f).$$

Therefore, taking the supremum for all grids on the left in this inequality,

$$\underline{I}(f) \leq \mathcal{U}_{\mathcal{F}}(f)$$

for all grids $\mathcal{F}$. Taking the infimum in this inequality, yields the conclusion of the lemma.

Definition C.0.9 *A bounded function, f which equals zero off a bounded set, D, is said to be Riemann integrable, written as $f \in \mathcal{R}(\mathbb{R}^n)$ exactly when $\underline{I}(f) = \overline{I}(f)$. In this case define*

$$\int f\, dV \equiv \int f\, dx = \overline{I}(f) = \underline{I}(f).$$

As in the case of integration of functions of one variable, one obtains the Riemann criterion which is stated as the following theorem.

Theorem C.0.10 *(Riemann criterion) $f \in \mathcal{R}(\mathbb{R}^n)$ if and only if for all $\varepsilon > 0$ there exists a grid $\mathcal{G}$ such that*

$$\mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon.$$

Proof: If $f \in \mathcal{R}(\mathbb{R}^n)$, then $\overline{I}(f) = \underline{I}(f)$ and so there exist grids $\mathcal{G}$ and $\mathcal{F}$ such that

$$\mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{F}}(f) \leq \overline{I}(f) + \frac{\varepsilon}{2} - \left(\underline{I}(f) - \frac{\varepsilon}{2}\right) = \varepsilon.$$

Then letting $\mathcal{H} = \mathcal{G}\vee\mathcal{F}$, Lemma C.0.6 implies

$$\mathcal{U}_{\mathcal{H}}(f) - \mathcal{L}_{\mathcal{H}}(f) \leq \mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{F}}(f) < \varepsilon.$$

Conversely, if for all $\varepsilon > 0$ there exists $\mathcal{G}$ such that

$$\mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon,$$

then

$$\overline{I}(f) - \underline{I}(f) \leq \mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this proves the theorem.

C.1 Basic Properties

It is important to know that certain combinations of Riemann integrable functions are Riemann integrable. The following theorem will include all the important cases.

Theorem C.1.1 *Let $f, g \in \mathcal{R}(\mathbb{R}^n)$ and let $\phi : K \to \mathbb{R}$ be continuous where K is a compact set in $\mathbb{R}^2$ containing $f(\mathbb{R}^n) \times g(\mathbb{R}^n)$. Also suppose that $\phi(0,0) = 0$. Then defining*

$$h(\mathbf{x}) \equiv \phi(f(\mathbf{x}), g(\mathbf{x})),$$

it follows that h is also in $\mathcal{R}(\mathbb{R}^n)$.

Proof: Let $\varepsilon > 0$ and let $\delta_1 > 0$ be such that if $(y_i, z_i), i = 1, 2$ are points in K, such that $|z_1 - z_2| \leq \delta_1$ and $|y_1 - y_2| \leq \delta_1$, then

$$|\phi(y_1, z_1) - \phi(y_2, z_2)| < \varepsilon.$$

Let $0 < \delta < \min(\delta_1, \varepsilon, 1)$. Let $\mathcal{G}$ be a grid with the property that for $Q \in \mathcal{G}$, the diameter of Q is less than δ and also for $k = f, g$,

$$\mathcal{U}_{\mathcal{G}}(k) - \mathcal{L}_{\mathcal{G}}(k) < \delta^2. \tag{3.5}$$

Then defining for $k = f, g$,

$$\mathcal{P}_k \equiv \{Q \in \mathcal{G} : M_Q(k) - m_Q(k) > \delta\},$$

it follows

$$\delta^2 > \sum_{Q \in \mathcal{G}} (M_Q(k) - m_Q(k)) v(Q) \geq$$

$$\sum_{\mathcal{P}_k} (M_Q(k) - m_Q(k)) v(Q) \geq \delta \sum_{\mathcal{P}_k} v(Q)$$

and so for $k = f, g$,

$$\varepsilon > \delta > \sum_{\mathcal{P}_k} v(Q). \tag{3.6}$$

Suppose for $k = f, g$,

$$M_Q(k) - m_Q(k) \leq \delta.$$

Then if $\mathbf{x}_1, \mathbf{x}_2 \in Q$,

$$|f(\mathbf{x}_1) - f(\mathbf{x}_2)| < \delta, \text{ and } |g(\mathbf{x}_1) - g(\mathbf{x}_2)| < \delta.$$

Therefore,

$$|h(\mathbf{x}_1) - h(\mathbf{x}_2)| \equiv |\phi(f(\mathbf{x}_1), g(\mathbf{x}_1)) - \phi(f(\mathbf{x}_2), g(\mathbf{x}_2))| < \varepsilon$$

and it follows that

$$|M_Q(h) - m_Q(h)| \leq \varepsilon.$$

Now let

$$\mathcal{S} \equiv \{Q \in \mathcal{G} : 0 < M_Q(k) - m_Q(k) \leq \delta,\ k = f, g\}.$$

Thus the union of the boxes in $\mathcal{S}$ is contained in some large box, R, which depends only on f and g and also, from the assumption that $\phi(0,0) = 0$, $M_Q(h) - m_Q(h) = 0$ unless $Q \subseteq R$. Then

$$\mathcal{U}_{\mathcal{G}}(h) - \mathcal{L}_{\mathcal{G}}(h) \leq \sum_{Q \in \mathcal{P}_f} (M_Q(h) - m_Q(h)) v(Q) +$$

$$\sum_{Q\in\mathcal{P}_g}\left(M_Q(h)-m_Q(h)\right)v(Q)+\sum_{Q\in\mathcal{S}}\delta v(Q).$$

Now since K is compact, it follows $\phi(K)$ is bounded and so there exists a constant, C, depending only on h and ϕ such that $M_Q(h)-m_Q(h)<C$. Therefore, the above inequality implies

$$\mathcal{U}_{\mathcal{G}}(h)-\mathcal{L}_{\mathcal{G}}(h)\leq C\sum_{Q\in\mathcal{P}_f}v(Q)+C\sum_{Q\in\mathcal{P}_g}v(Q)+\sum_{Q\in\mathcal{S}}\delta v(Q),$$

which by 3.6 implies

$$\mathcal{U}_{\mathcal{G}}(h)-\mathcal{L}_{\mathcal{G}}(h)\leq 2C\varepsilon+\delta v(R)\leq 2C\varepsilon+\varepsilon v(R).$$

Since ε is arbitrary, the Riemann criterion is satisfied and so $h\in\mathcal{R}(\mathbb{R}^n)$.

Corollary C.1.2 *Let $f,g\in\mathcal{R}(\mathbb{R}^n)$ and let $a,b\in\mathbb{R}$. Then $af+bg$, fg, and $|f|$ are all in $\mathcal{R}(\mathbb{R}^n)$. Also,*

$$\int_{\mathbb{R}^n}(af+bg)\,dx=a\int_{\mathbb{R}^n}f\,dx+b\int_{\mathbb{R}^n}g\,dx, \tag{3.7}$$

and

$$\int|f|\,dx\geq\left|\int f\,dx\right|. \tag{3.8}$$

Proof: Each of the combinations of functions described above is Riemann integrable by Theorem C.1.1. For example, to see $af+bg\in\mathcal{R}(\mathbb{R}^n)$ consider $\phi(y,z)\equiv ay+bz$. This is clearly a continuous function of (y,z) such that $\phi(0,0)=0$. To obtain $|f|\in\mathcal{R}(\mathbb{R}^n)$, let $\phi(y,z)\equiv|y|$. It remains to verify the formulas. To do so, let $\mathcal{G}$ be a grid with the property that for $k=f,g$, $|f|$ and $af+bg$,

$$\mathcal{U}_{\mathcal{G}}(k)-\mathcal{L}_{\mathcal{G}}(k)<\varepsilon. \tag{3.9}$$

Consider 3.7. For each $Q\in\mathcal{G}$ pick a point in Q, $\mathbf{x}_Q$. Then

$$\sum_{Q\in\mathcal{G}}k(\mathbf{x}_Q)\,v(Q)\in[\mathcal{L}_{\mathcal{G}}(k),\mathcal{U}_{\mathcal{G}}(k)]$$

and so

$$\left|\int k\,dx-\sum_{Q\in\mathcal{G}}k(\mathbf{x}_Q)\,v(Q)\right|<\varepsilon.$$

Consequently, since

$$\sum_{Q\in\mathcal{G}}(af+bg)(\mathbf{x}_Q)\,v(Q)$$

$$=a\sum_{Q\in\mathcal{G}}f(\mathbf{x}_Q)\,v(Q)+b\sum_{Q\in\mathcal{G}}g(\mathbf{x}_Q)\,v(Q),$$

it follows

$$\left|\int(af+bg)\,dx-a\int f\,dx-b\int g\,dx\right|\leq$$

$$\left|\int(af+bg)\,dx-\sum_{Q\in\mathcal{G}}(af+bg)(\mathbf{x}_Q)\,v(Q)\right|+$$

$$\left| a\sum_{Q\in\mathcal{G}} f(\mathbf{x}_Q)\, v(Q) - a\int f\,dx \right| + \left| b\sum_{Q\in\mathcal{G}} g(\mathbf{x}_Q)\, v(Q) - b\int g\,dx \right|$$
$$\leq \varepsilon + |a|\,\varepsilon + |b|\,\varepsilon.$$

Since ε is arbitrary, this establishes Formula 3.7 and shows the integral is linear.

It remains to establish the inequality 3.8. By 3.9, and the triangle inequality for sums,

$$\int |f|\ dx + \varepsilon \geq \sum_{Q\in\mathcal{G}} |f(\mathbf{x}_Q)|\, v(Q) \geq$$
$$\geq \left| \sum_{Q\in\mathcal{G}} f(\mathbf{x}_Q)\, v(Q) \right| \geq \left| \int f\,dx \right| - \varepsilon.$$

Then since ε is arbitrary, this establishes the desired inequality. This proves the corollary.

C.2 Which Functions Are Integrable?

Which functions are in $\mathcal{R}(\mathbb{R}^n)$? As in the case of integrals of functions of one variable, this is an important question. It turns out the Riemann integrable functions are characterized by being continuous except on a very small set. This has to do with Jordan content.

Definition C.2.1 *A bounded set, E, has Jordan content 0 or content 0 if for every $\varepsilon > 0$ there exists a grid, $\mathcal{G}$ such that*

$$\sum_{Q\cap E\neq\emptyset} v(Q) < \varepsilon.$$

This symbol says to sum the volumes of all boxes from $\mathcal{G}$ which have nonempty intersection with E.

Next it is necessary to define the oscillation of a function.

Definition C.2.2 *Let f be a function defined on $\mathbb{R}^n$ and let*

$$\omega_{f,r}(\mathbf{x}) \equiv \sup\{|f(\mathbf{z}) - f(\mathbf{y})| : \mathbf{z},\mathbf{y} \in B(\mathbf{x},r)\}.$$

This is called the oscillation of f on $B(\mathbf{x},r)$. Note that this function of r is decreasing in r. Define the oscillation of f as

$$\omega_f(\mathbf{x}) \equiv \lim_{r\to 0+} \omega_{f,r}(\mathbf{x}).$$

Note that as r decreases, the function, $\omega_{f,r}(\mathbf{x})$ decreases. It is also bounded below by 0 and so the limit must exist and equals $\inf\{\omega_{f,r}(\mathbf{x}) : r > 0\}$. (Why?) Then the following simple lemma whose proof follows directly from the definition of continuity gives the reason for this definition.

Lemma C.2.3 *A function f is continuous at $\mathbf{x}$ if and only if $\omega_f(\mathbf{x}) = 0$.*

This concept of oscillation gives a way to define how discontinuous a function is at a point. The discussion will depend on the following fundamental lemma which gives the existence of something called the Lebesgue number.

Definition C.2.4 *Let $\mathfrak{C}$ be a set whose elements are sets of $\mathbb{R}^n$ and let $K \subseteq \mathbb{R}^n$. The set, $\mathfrak{C}$ is called a cover of K if every point of K is contained in some set of $\mathfrak{C}$. If the elements of $\mathfrak{C}$ are open sets, it is called an open cover.*

Lemma C.2.5 *Let K be sequentially compact and let $\mathfrak{C}$ be an open cover of K. Then there exists $r > 0$ such that whenever $\mathbf{x} \in K$, $B(\mathbf{x}, r)$ is contained in some set of $\mathfrak{C}$.*

Proof: Suppose this is not so. Then letting $r_n = 1/n$, there exists $\mathbf{x}_n \in K$ such that $B(\mathbf{x}_n, r_n)$ is not contained in any set of $\mathfrak{C}$. Since K is sequentially compact, there is a subsequence, $\mathbf{x}_{n_k}$ which converges to a point, $\mathbf{x} \in K$. But there exists $\delta > 0$ such that $B(\mathbf{x}, \delta) \subseteq U$ for some $U \in \mathfrak{C}$. Let k be so large that $1/k < \delta/2$ and $|\mathbf{x}_{n_k} - \mathbf{x}| < \delta/2$ also. Then if $\mathbf{z} \in B(\mathbf{x}_{n_k}, r_{n_k})$, it follows

$$|\mathbf{z} - \mathbf{x}| \leq |\mathbf{z} - \mathbf{x}_{n_k}| + |\mathbf{x}_{n_k} - \mathbf{x}| < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

and so $B(\mathbf{x}_{n_k}, r_{n_k}) \subseteq U$ contrary to supposition. Therefore, the desired number exists after all.

Theorem C.2.6 *Let f be a bounded function which equals zero off a bounded set and let W denote the set of points where f fails to be continuous. Then $f \in \mathcal{R}(\mathbb{R}^n)$ if W has content zero. That is, for all $\varepsilon > 0$ there exists a grid, $\mathcal{G}$ such that*

$$\sum_{Q \in \mathcal{G}_W} v(Q) < \varepsilon \tag{3.10}$$

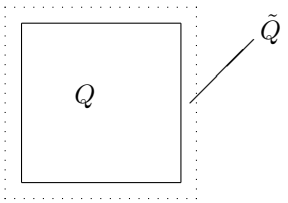
where

$$\mathcal{G}_W \equiv \{Q \in \mathcal{G} : Q \cap W \neq \emptyset\}.$$

Proof: Let W have content zero. Also let $|f(\mathbf{x})| < C/2$ for all $\mathbf{x} \in \mathbb{R}^n$, let $\varepsilon > 0$ be given, and let $\mathcal{G}$ be a grid which satisfies 3.10. Since f equals zero off some bounded set, there exists R such that f equals zero off of $B(\mathbf{0}, \frac{R}{2})$. Thus $W \subseteq B(\mathbf{0}, \frac{R}{2})$. Also note that if $\mathcal{G}$ is a grid for which 3.10 holds, then this inequality continues to hold if $\mathcal{G}$ is replaced with a refined grid. Therefore, you may assume the diameter of every box in $\mathcal{G}$ which intersects $B(\mathbf{0}, R)$ is less than $\frac{R}{3}$ and so all boxes of $\mathcal{G}$ which intersect the set where f is nonzero are contained in $B(\mathbf{0}, R)$. Since W is bounded, $\mathcal{G}_W$ contains only finitely many boxes. Letting

$$Q \equiv \prod_{i=1}^{n} [a_i, b_i]$$

be one of these boxes, enlarge the box slightly as indicated in the following picture.



The enlarged box is an open set of the form,

$$\widetilde{Q} \equiv \prod_{i=1}^{n} (a_i - \eta_i, b_i + \eta_i)$$

where η_i is chosen small enough that if

$$\prod_{i=1}^{n} (\, b_i + \eta_i - (a_i - \eta_i)) \equiv v\left(\widetilde{Q}\right),$$

and $\widetilde{\mathcal{G}_W}$ denotes those $\widetilde{Q}$ for $Q \in \mathcal{G}$ which have nonempty intersection with W, then

$$\sum_{\widetilde{Q} \in \widetilde{\mathcal{G}_W}} v\left(\widetilde{\widetilde{Q}}\right) < \varepsilon \tag{3.11}$$

where $\widetilde{\widetilde{Q}}$ is the box,

$$\prod_{i=1}^{n} (\ b_i + 2\eta_i - (a_i - 2\eta_i))$$

For each $\mathbf{x} \in \mathbb{R}^n$, let $r_{\mathbf{x}} < \min(\eta_1/2, \cdots, \eta_n/2)$ be such that

$$\omega_{f,r_{\mathbf{x}}}(\mathbf{x}) < \varepsilon + \omega_f(\mathbf{x}). \tag{3.12}$$

Now let $\mathfrak{C}$ denote all intersections of the form $\widetilde{Q} \cap B(\mathbf{x},r_{\mathbf{x}})$ such that $\mathbf{x} \in \overline{B(\mathbf{0},R)}$ so that $\mathfrak{C}$ is an open cover of the compact set, $\overline{B(\mathbf{0},R)}$. Let δ be a Lebesgue number for this open cover of $\overline{B(\mathbf{0},R)}$ and let $\mathcal{F}$ be a refinement of $\mathcal{G}$ such that every box in $\mathcal{F}$ has diameter less than δ. Now let $\mathcal{F}_1$ consist of those boxes of $\mathcal{F}$ which have nonempty intersection with $B(\mathbf{0},R/2)$. Thus all boxes of $\mathcal{F}_1$ are contained in $B(\mathbf{0},R)$ and each one is contained in some set of $\mathfrak{C}$. Let $\mathfrak{C}_W$ be those open sets of $\mathfrak{C}$, $\widetilde{Q} \cap B(\mathbf{x},r_{\mathbf{x}})$, for which $\mathbf{x} \in W$. Thus each of these sets is contained in some $\widetilde{\widetilde{Q}}$ where $Q \in \mathcal{G}_W$. Let $\mathcal{F}_W$ be those sets of $\mathcal{F}_1$ which are subsets of some set of $\mathfrak{C}_W$. Thus

$$\sum_{Q \in \mathcal{F}_W} v(Q) < \varepsilon. \tag{3.13}$$

because each Q in $\mathcal{F}_W$ is contained in a set, $\widetilde{\widetilde{Q}}$ described above and the sum of the volumes of these is less than ε by 3.11. Then

$$\mathcal{U}_{\mathcal{F}}(f) - \mathcal{L}_{\mathcal{F}}(f) = \sum_{Q \in \mathcal{F}_W} (M_Q(f) - m_Q(f))\, v(Q)$$

$$+ \sum_{Q \in \mathcal{F}_1 \setminus \mathcal{F}_W} (M_Q(f) - m_Q(f))\, v(Q).$$

If $Q \in \mathcal{F}_1 \setminus \mathcal{F}_W$, then Q must be a subset of some set of $\mathfrak{C} \setminus \mathfrak{C}_W$ since it is not in any set of $\mathfrak{C}_W$. Say $Q \subseteq \widetilde{Q_1} \cap B(\mathbf{x},r_{\mathbf{x}})$ where $\mathbf{x} \notin W$. Therefore, from 3.12 and the observation that $\mathbf{x} \notin W$, it follows $\omega_f(\mathbf{x}) = 0$ and so

$$M_Q(f) - m_Q(f) \leq \varepsilon.$$

Therefore, from 3.13 and the estimate on f,

$$\mathcal{U}_{\mathcal{F}}(f) - \mathcal{L}_{\mathcal{F}}(f) \leq \sum_{Q \in \mathcal{F}_W} Cv(Q) + \sum_{Q \in \mathcal{F}_1 \setminus \mathcal{F}_W} \varepsilon v(Q)$$

$$\leq C\varepsilon + \varepsilon (2R)^n,$$

the estimate of the second sum coming from the fact

$$B(\mathbf{0},R) \subseteq \prod_{i=1}^{n} [-R,R].$$

Since ε is arbitrary, this proves the theorem.[1]

[1] In fact one cannot do any better. It can be shown that if a function is Riemann integrable, then it must be the case that for all $\varepsilon > 0$, 3.10 is satisfied for some grid, $\mathcal{G}$. This along with what was just shown is known as Lebesgue's theorem after Lebesgue who discovered it in the early years of the twentieth century. Actually, he also invented a far superior integral which has been the integral of serious mathematicians since that time.

Definition C.2.7 *A bounded set, E is a Jordan set in $\mathbb{R}^n$ or a contented set in $\mathbb{R}^n$ if $\mathcal{X}_E \in \mathcal{R}(\mathbb{R}^n)$. The symbol $\mathcal{X}_E$ means*

$$\mathcal{X}_E(\mathbf{x}) = \begin{cases} 1 \text{ if } \mathbf{x} \in E \\ 0 \text{ if } \mathbf{x} \notin E \end{cases}$$

It is called the indicator function because it indicates whether $\mathbf{x}$ is in E according to whether it equals 1. For a function $f \in \mathcal{R}(\mathbb{R}^n)$ and E a contented set, $f\mathcal{X}_E \in \mathcal{R}(\mathbb{R}^n)$ by Corollary C.1.2. Then

$$\int_E f dV \equiv \int f\mathcal{X}_E dV.$$

So what are examples of contented sets?

Theorem C.2.8 *Suppose E is a bounded contented set in $\mathbb{R}^n$ and $f, g : E \to \mathbb{R}$ are two functions satisfying $f(\mathbf{x}) \geq g(\mathbf{x})$ for all $\mathbf{x} \in E$ and $f\mathcal{X}_E$ and $g\mathcal{X}_E$ are both in $\mathcal{R}(\mathbb{R}^n)$. Now define*

$$P \equiv \{(\mathbf{x}, x_{n+1}) : \mathbf{x} \in E \text{ and } g(\mathbf{x}) \leq x_{n+1} \leq f(\mathbf{x})\}.$$

Then P is a contented set in $\mathbb{R}^{n+1}$.

Proof: Let $\mathcal{G}$ be a grid such that for $k = f\mathcal{X}_E, g\mathcal{X}_E$,

$$\mathcal{U}_{\mathcal{G}}(k) - \mathcal{L}_{\mathcal{G}}(k) < \varepsilon/4. \tag{3.14}$$

Also let $K \geq \sum_{j=1}^m v_n(Q_j)$ where the Q_j are the boxes which intersect E. Let $\{a_i\}_{i=-\infty}^{\infty}$ be a sequence on $\mathbb{R}$, $a_i < a_{i+1}$ for all i, which includes

$$M_{Q_j}(f\mathcal{X}_E) + \frac{\varepsilon}{4mK}, M_{Q_j}(f\mathcal{X}_E), M_{Q_j}(g\mathcal{X}_E),$$
$$m_{Q_j}(f\mathcal{X}_E), m_{Q_j}(g\mathcal{X}_E), m_{Q_j}(g\mathcal{X}_E) - \frac{\varepsilon}{4mK}$$

for all $j = 1, \cdots, m$. Now define a grid on $\mathbb{R}^{n+1}$ as follows.

$$\mathcal{G}' \equiv \{Q \times [a_i, a_{i+1}] : Q \in \mathcal{G},\ i \in \mathbb{Z}\}$$

In words, this grid consists of all possible boxes of the form $Q \times [a_i, a_{i+1}]$ where $Q \in \mathcal{G}$ and a_i is a term of the sequence just described. It is necessary to verify that for $P \in \mathcal{G}'$, $\mathcal{X}_P \in \mathcal{R}\left(\mathbb{R}^{n+1}\right)$. This is done by showing that $\mathcal{U}_{\mathcal{G}'}(\mathcal{X}_P) - \mathcal{L}_{\mathcal{G}'}(\mathcal{X}_P) < \varepsilon$ and then noting that $\varepsilon > 0$ was arbitrary. For $\mathcal{G}'$ just described, denote by Q' a box in $\mathcal{G}'$. Thus $Q' = Q \times [a_i, a_{i+1}]$ for some i.

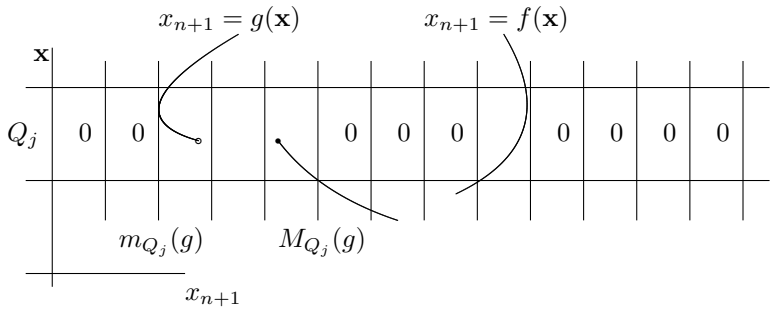
$$\begin{aligned} \mathcal{U}_{\mathcal{G}'}(\mathcal{X}_P) - \mathcal{L}_{\mathcal{G}'}(\mathcal{X}_P) &\equiv \sum_{Q' \in \mathcal{G}'} \left(M_{Q'}(\mathcal{X}_P) - m_{Q'}(\mathcal{X}_P)\right) v_{n+1}(Q') \\ &= \sum_{i=-\infty}^{\infty} \sum_{j=1}^{m} \left(M_{Q_j'}(\mathcal{X}_P) - m_{Q_j'}(\mathcal{X}_P)\right) v_n(Q_j)(a_{i+1} - a_i) \end{aligned}$$

and all sums are bounded because the functions, f and g are given to be bounded. Therefore, there are no limit considerations needed here. Thus

$$\mathcal{U}_{\mathcal{G}'}(\mathcal{X}_P) - \mathcal{L}_{\mathcal{G}'}(\mathcal{X}_P) =$$

$$\sum_{j=1}^{m} v_n(Q_j) \sum_{i=-\infty}^{\infty} \left(M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P)\right)(a_{i+1} - a_i).$$

Consider the inside sum with the aid of the following picture.



In this picture, the little rectangles represent the boxes $Q_j \times [a_i, a_{i+1}]$ for fixed j. The part of P having $\mathbf{x}$ contained in Q_j is between the two surfaces, $x_{n+1} = g(\mathbf{x})$ and $x_{n+1} = f(\mathbf{x})$ and there is a zero placed in those boxes for which

$$M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) = 0.$$

You see, $\mathcal{X}_P$ has either the value of 1 or the value of 0 depending on whether $(\mathbf{x}, y)$ is contained in P. For the boxes shown with 0 in them, either all of the box is contained in P or none of the box is contained in P. Either way,

$$M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) = 0$$

on these boxes. However, on the boxes intersected by the surfaces, the value of

$$M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P)$$

is 1 because there are points in this box which are not in P as well as points which are in P. Because of the construction of $\mathcal{G}'$ which included all values of

$$M_{Q_j}(f\mathcal{X}_E) + \frac{\varepsilon}{4mK}, M_{Q_j}(f\mathcal{X}_E),$$
$$M_{Q_j}(g\mathcal{X}_E), m_{Q_j}(f\mathcal{X}_E), m_{Q_j}(g\mathcal{X}_E)$$

for all $j = 1, \cdots, m$,

$$\sum_{i=-\infty}^{\infty} \left(M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P)\right)(a_{i+1} - a_i) \leq$$

$$\sum_{\{i : m_{Q_j}(g\mathcal{X}_E) \leq a_i < M_{Q_j}(g\mathcal{X}_E)\}} 1\,(a_{i+1} - a_i) + \sum_{\{i : m_{Q_j}(f\mathcal{X}_E) \leq a_i < M_{Q_j}(f\mathcal{X}_E)\}} 1\,(a_{i+1} - a_i) \tag{3.15}$$

The first of the sums in 3.15 contains all possible terms for which

$$M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P)$$

might be 1 due to the graph of the bottom surface $g\mathcal{X}_E$ while the second sum contains all possible terms for which the expression might be 1 due to the graph of the top surface $f\mathcal{X}_E$.

$$\leq \left(M_{Q_j}(g\mathcal{X}_E) + \frac{\varepsilon}{4mK} - m_{Q_j}(g\mathcal{X}_E)\right) + \left(M_{Q_j}(f\mathcal{X}_E) + \frac{\varepsilon}{4mK} - m_{Q_j}(f\mathcal{X}_E)\right)$$

$$= \left(M_{Q_j}(g\mathcal{X}_E) - m_{Q_j}(g\mathcal{X}_E)\right) + \left(M_{Q_j}(f\mathcal{X}_E) - m_{Q_j}(f\mathcal{X}_E)\right) + \frac{\varepsilon}{2m}\left(\sum_{j=1}^{m} v(Q_j)\right)^{-1}.$$

Therefore, by 3.14,

$$\mathcal{U}_{\mathcal{G}'}\left(\mathcal{X}_P\right)-\mathcal{L}_{\mathcal{G}'}\left(\mathcal{X}_P\right)\leq$$

$$\begin{aligned}
&\sum_{j=1}^{m} v_n\left(Q_j\right)\left[\left(M_{Q_j}\left(g\mathcal{X}_E\right)-m_{Q_j}\left(g\mathcal{X}_E\right)\right)+\left(M_{Q_j}\left(f\mathcal{X}_E\right)-m_{Q_j}\left(f\mathcal{X}_E\right)\right)\right]\\
&+\sum_{j=1}^{m} v\left(Q_j\right)\frac{\varepsilon}{2m}\left(\sum_{j=1}^{m} v\left(Q_j\right)\right)^{-1}\\
=\;&\mathcal{U}_{\mathcal{G}}\left(f\right)-\mathcal{L}_{\mathcal{G}}\left(f\right)+\mathcal{U}_{\mathcal{G}}\left(g\right)-\mathcal{L}_{\mathcal{G}}\left(g\right)+\frac{\varepsilon}{2}\\
<\;&\frac{\varepsilon}{4}+\frac{\varepsilon}{4}+\frac{\varepsilon}{2}=\varepsilon.
\end{aligned}$$

Since $\varepsilon>0$ is arbitrary, this proves the theorem.

Corollary C.2.9 *Suppose f and g are continuous functions defined on E, a contented set in $\mathbb{R}^n$ and that $g\left(\mathbf{x}\right)\leq f\left(\mathbf{x}\right)$ for all $\mathbf{x}\in E$. Then*

$$P\equiv\left\{\left(\mathbf{x},x_{n+1}\right):\mathbf{x}\in E\text{ and }g\left(\mathbf{x}\right)\leq x_{n+1}\leq f\left(\mathbf{x}\right)\right\}$$

is a contented set in $\mathbb{R}^n$.

Proof: Since E is contented, meaning $\mathcal{X}_E$ is integrable, it follows from Theorem C.2.6 the set of discontinuities of $\mathcal{X}_E$ has Jordan content 0. But the set of discontinuities of $\mathcal{X}_E$ is ∂E defined as those points $\mathbf{x}$ such that $B\left(\mathbf{x},r\right)$ contains points of E and points of E^C for every $r>0$. Extend f and g to equal 0 off E. Then the set of discontinuities of these extended functions, still denoted as f,g is ∂E which has Jordan content 0. This reduces to the situation of Theorem C.2.8. This proves the corollary.

As an example of how this can be applied, it is obvious a closed interval is a contented set in $\mathbb{R}$. Therefore, if f,g are two continuous functions with $f\left(x\right)\geq g\left(x\right)$ for $x\in\left[a,b\right]$, it follows from the above theorem or its corollary that the set,

$$P_1\equiv\left\{\left(x,y\right):g\left(x\right)\leq y\leq f\left(x\right)\right\}$$

is a contented set in $\mathbb{R}^2$. Now using the theorem and corollary again, suppose $f_1\left(x,y\right)\geq g_1\left(x,y\right)$ for $\left(x,y\right)\in P_1$ and f,g are continuous. Then the set

$$P_2\equiv\left\{\left(x,y,z\right):g_1\left(x,y\right)\leq z\leq f_1\left(x,y\right)\right\}$$

is a contented set in $\mathbb{R}^3$. Clearly you can continue this way obtaining examples of contented sets.

Note that as a special case, it follows that every box is a contented set. Therefore, if B_i is a box, functions of the form

$$\sum_{i=1}^{m} a_i\mathcal{X}_{B_i}$$

are integrable. These functions are called step functions.

The following theorem is analogous to the fact that in one dimension, when you integrate over a point, the answer is 0.

Theorem C.2.10 *If a bounded set, E, has Jordan content 0, then E is a Jordan (contented) set and if f is any bounded function defined on E, then $f\mathcal{X}_E\in\mathcal{R}\left(\mathbb{R}^n\right)$ and*

$$\int_E f\,dV=0.$$

Proof: Let $\mathcal{G}$ be a grid with

$$\sum_{Q\cap E\neq\emptyset} v(Q) < \frac{\varepsilon}{1+(M-m)}.$$

Then

$$\mathcal{U}_{\mathcal{G}}(f\mathcal{X}_E) \leq \sum_{Q\cap E\neq\emptyset} Mv(Q) \leq \frac{\varepsilon M}{1+(M-m)}$$

and

$$\mathcal{L}_{\mathcal{G}}(f\mathcal{X}_E) \geq \sum_{Q\cap E\neq\emptyset} mv(Q) \geq \frac{\varepsilon m}{1+(M-m)}$$

and so

$$\begin{aligned}\mathcal{U}_{\mathcal{G}}(f\mathcal{X}_E) - \mathcal{L}_{\mathcal{G}}(f\mathcal{X}_E) &\leq \sum_{Q\cap E\neq\emptyset} Mv(Q) - \sum_{Q\cap E\neq\emptyset} mv(Q)\\ &= (M-m)\sum_{Q\cap E\neq\emptyset} v(Q) < \frac{\varepsilon(M-m)}{1+(M-m)} < \varepsilon.\end{aligned}$$

This shows $f\mathcal{X}_E \in \mathcal{R}(\mathbb{R}^n)$. Now also,

$$m\varepsilon \leq \int f\mathcal{X}_E\, dV \leq M\varepsilon$$

and since ε is arbitrary, this shows

$$\int_E f\, dV \equiv \int f\mathcal{X}_E\, dV = 0$$

Why is E contented? Let $\mathcal{G}$ be a grid for which

$$\sum_{Q\cap E\neq\emptyset} v(Q) < \varepsilon$$

Then for this grid,

$$\mathcal{U}_{\mathcal{G}}(\mathcal{X}_E) - \mathcal{L}_{\mathcal{G}}(\mathcal{X}_E) \leq \sum_{Q\cap E\neq\emptyset} v(Q) < \varepsilon$$

and this proves the theorem.

Corollary C.2.11 *If $f\mathcal{X}_{E_i} \in \mathcal{R}(\mathbb{R}^n)$ for $i = 1, 2, \cdots, r$ and for all $i \neq j$, $E_i \cap E_j$ is either the empty set or a set of Jordan content 0, then letting $F \equiv \cup_{i=1}^r E_i$, it follows $f\mathcal{X}_F \in \mathcal{R}(\mathbb{R}^n)$ and*

$$\int f\mathcal{X}_F\, dV \equiv \int_F f\, dV = \sum_{i=1}^r \int_{E_i} f\, dV.$$

Proof: This is true if $r = 1$. Suppose it is true for r. It will be shown that it is true for $r+1$. Let $F_r = \cup_{i=1}^r E_i$ and let F_{r+1} be defined similarly. By the induction hypothesis, $f\mathcal{X}_{F_r} \in \mathcal{R}(\mathbb{R}^n)$. Also, since F_r is a finite union of the E_i, it follows that $F_r \cap E_{r+1}$ is either empty or a set of Jordan content 0.

$$-f\mathcal{X}_{F_r\cap E_{r+1}} + f\mathcal{X}_{F_r} + f\mathcal{X}_{E_{r+1}} = f\mathcal{X}_{F_{r+1}}$$

and by Theorem C.2.10 each function on the left is in $\mathcal{R}(\mathbb{R}^n)$ and the first one on the left has integral equal to zero. Therefore,

$$\int f\mathcal{X}_{F_{r+1}}\, dV = \int f\mathcal{X}_{F_r}\, dV + \int f\mathcal{X}_{E_{r+1}}\, dV$$

which by induction equals

$$\sum_{i=1}^{r}\int_{E_i} f\,dV + \int_{E_{r+1}} f\,dV = \sum_{i=1}^{r+1}\int_{E_i} f\,dV$$

and this proves the corollary.

In particular, for

$$Q = \prod_{i=1}^{n}[a_i,b_i]\,,\ Q' = \prod_{i=1}^{n}(a_i,b_i]$$

both are contented sets and

$$\int \mathcal{X}_Q dV = \int_{Q'} \mathcal{X}_{Q'} dV = v(Q)\,. \tag{3.16}$$

This is because

$$Q \setminus Q' = \cup_{i=1}^{n} a_i \times \prod_{j\neq i}(a_j,b_j]$$

a finite union of sets of content 0. It is obvious $\int \mathcal{X}_Q dV = v(Q)$ because you can use a grid which has Q as one of the boxes and then the upper and lower sums are the same and equal to $v(Q)$. Therefore, the claim about the equality of the two integrals in 3.16 follows right away from Corollary C.2.11. That $\mathcal{X}_{Q'}$ is integrable follows from

$$\mathcal{X}_{Q'} = \mathcal{X}_Q - \mathcal{X}_{Q\setminus Q'}$$

and each of the two functions on the right is integrable thanks to Theorem C.2.10.

In fact, here is an interesting version of the Riemann criterion which depends on these half open boxes.

Lemma C.2.12 *Suppose f is a bounded function which equals zero off some bounded set. Then $f \in \mathcal{R}(\mathbb{R}^n)$ if and only if for all $\varepsilon > 0$ there exists a grid, $\mathcal{G}$ such that*

$$\sum_{Q\in\mathcal{G}} \left(M_{Q'}(f) - m_{Q'}(f)\right) v(Q) < \varepsilon. \tag{3.17}$$

Proof: Since $Q' \subseteq Q$,

$$M_{Q'}(f) - m_{Q'}(f) \leq M_Q(f) - m_Q(f)$$

and therefore, the only if part of the equivalence is obvious.

Conversely, let $\mathcal{G}$ be a grid such that 3.17 holds with ε replaced with $\frac{\varepsilon}{2}$. It is necessary to show there is a grid such that 3.17 holds with no primes on the Q. Let $\mathcal{F}$ be a refinement of $\mathcal{G}$ obtained by adding the points $\alpha_k^i + \eta_k$ where $\eta_k \leq \eta$ and is also chosen so small that for each $i = 1,\cdots,n$,

$$\alpha_k^i + \eta_k < \alpha_{k+1}^i.$$

You only need to have $\eta_k > 0$ for the finitely many boxes of $\mathcal{G}$ which intersect the bounded set where f is not zero. Then for

$$Q \equiv \prod_{i=1}^{n}\left[\alpha_{k_i}^i, \alpha_{k_i+1}^i\right] \in \mathcal{G},$$

Let

$$\widehat{Q} \equiv \prod_{i=1}^{n}\left[\alpha_{k_i}^i + \eta_{k_i}, \alpha_{k_i+1}^i\right]$$

and denote by $\widehat{\mathcal{G}}$ the collection of these smaller boxes. For each set, Q in $\mathcal{G}$ there is the smaller set, $\widehat{Q}$ along with n boxes, $B_k, k = 1, \cdots, n$, one of whose sides is of length η_k and the remainder of whose sides are shorter than the diameter of Q such that the set, Q is the union of $\widehat{Q}$ and these sets, B_k. Now suppose f equals zero off the ball $B\left(\mathbf{0}, \frac{R}{2}\right)$. Then without loss of generality, you may assume the diameter of every box in $\mathcal{G}$ which has nonempty intersection with $B(\mathbf{0},R)$ is smaller than $\frac{R}{3}$. (If this is not so, simply refine $\mathcal{G}$ to make it so, such a refinement leaving 3.17 valid because refinements do not increase the difference between upper and lower sums in this context either.) Suppose there are P sets of $\mathcal{G}$ contained in $B(\mathbf{0},R)$ (So these are the only sets of $\mathcal{G}$ which could have nonempty intersection with the set where f is nonzero.) and suppose that for all $\mathbf{x}$, $|f(\mathbf{x})| < C/2$. Then

$$\sum_{Q\in\mathcal{F}} \left(M_Q(f) - m_Q(f)\right) v(Q) \leq \sum_{\widehat{Q}\in\widehat{\mathcal{G}}} \left(M_{\widehat{Q}}(f) - m_{\widehat{Q}}(f)\right) v(Q)$$

$$+ \sum_{Q\in\mathcal{F}\setminus\widehat{\mathcal{G}}} \left(M_Q(f) - m_Q(f)\right) v(Q)$$

The first term on the right of the inequality in the above is no larger than $\varepsilon/2$ because $M_{\widehat{Q}}(f) - m_{\widehat{Q}}(f) \leq M_{Q'}(f) - m_{Q'}(f)$ for each Q. Therefore, the above is dominated by

$$\leq \varepsilon/2 + CPnR^{n-1}\eta < \varepsilon$$

whenever η is small enough. Since ε is arbitrary, $f \in \mathcal{R}(\mathbb{R}^n)$ as claimed.

C.3 Iterated Integrals

To evaluate an n dimensional Riemann integral, one uses iterated integrals. Formally, an iterated integral is defined as follows. For f a function defined on $\mathbb{R}^{n+m}$,

$$\mathbf{y} \to f(\mathbf{x},\mathbf{y})$$

is a function of $\mathbf{y}$ for each $\mathbf{x} \in \mathbb{R}^n$. Therefore, it might be possible to integrate this function of $\mathbf{y}$ and write

$$\int_{\mathbb{R}^m} f(\mathbf{x},\mathbf{y})\, dV_y.$$

Now the result is clearly a function of $\mathbf{x}$ and so, it might be possible to integrate this and write

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} f(\mathbf{x},\mathbf{y})\, dV_y\, dV_x.$$

This symbol is called an iterated integral, because it involves the iteration of two lower dimensional integrations. Under what conditions are the two iterated integrals equal to the integral

$$\int_{\mathbb{R}^{n+m}} f(\mathbf{z})\, dV?$$

Definition C.3.1 *Let $\mathcal{G}$ be a grid on $\mathbb{R}^{n+m}$ defined by the $n+m$ sequences,*

$$\left\{\alpha_k^i\right\}_{k=-\infty}^{\infty} \quad i = 1, \cdots, n+m.$$

Let $\mathcal{G}_n$ be the grid on $\mathbb{R}^n$ obtained by considering only the first n of these sequences and let $\mathcal{G}_m$ be the grid on $\mathbb{R}^m$ obtained by considering only the last m of the sequences. Thus a typical box in $\mathcal{G}_m$ would be

$$\prod_{i=n+1}^{n+m} \left[\alpha_{k_i}^i, \alpha_{k_i+1}^i\right], \; k_i \geq n+1$$

and a box in $\mathcal{G}_n$ would be of the form

$$\prod_{i=1}^{n}\left[\alpha_{k_i}^i,\alpha_{k_i+1}^i\right],\ k_i\le n.$$

Lemma C.3.2 *Let $\mathcal{G}$, $\mathcal{G}_n$, and $\mathcal{G}_m$ be the grids defined above. Then*

$$\mathcal{G}=\{R\times P: R\in\mathcal{G}_n \text{ and } P\in\mathcal{G}_m\}.$$

Proof: If $Q\in\mathcal{G}$, then Q is clearly of this form. On the other hand, if $R\times P$ is one of the sets described above, then from the above description of R and P, it follows $R\times P$ is one of the sets of $\mathcal{G}$. This proves the lemma.

Now let $\mathcal{G}$ be a grid on $\mathbb{R}^{n+m}$ and suppose

$$\phi(\mathbf{z})=\sum_{Q\in\mathcal{G}}\phi_Q\mathcal{X}_{Q'}(\mathbf{z}) \tag{3.18}$$

where ϕ_Q equals zero for all but finitely many Q. Thus ϕ is a step function. Recall that for

$$Q=\prod_{i=1}^{n+m}[a_i,b_i],\ Q'\equiv\prod_{i=1}^{n+m}(a_i,b_i]$$

The function

$$\phi=\sum_{Q\in\mathcal{G}}\phi_Q\mathcal{X}_{Q'}$$

is integrable because it is a finite sum of integrable functions, each function in the sum being integrable because the set of discontinuities has Jordan content 0. (why?) Letting $(\mathbf{x},\mathbf{y})=\mathbf{z}$,

$$\begin{aligned}\phi(\mathbf{z})=\phi(\mathbf{x},\mathbf{y})&=\sum_{R\in\mathcal{G}_n}\sum_{P\in\mathcal{G}_m}\phi_{R\times P}\mathcal{X}_{R'\times P'}(\mathbf{x},\mathbf{y})\\&=\sum_{R\in\mathcal{G}_n}\sum_{P\in\mathcal{G}_m}\phi_{R\times P}\mathcal{X}_{R'}(\mathbf{x})\,\mathcal{X}_{P'}(\mathbf{y}).\end{aligned} \tag{3.19}$$

For a function of two variables, h, denote by $h(\cdot,\mathbf{y})$ the function, $\mathbf{x}\to h(\mathbf{x},\mathbf{y})$ and $h(\mathbf{x},\cdot)$ the function $\mathbf{y}\to h(\mathbf{x},\mathbf{y})$. The following lemma is a preliminary version of Fubini's theorem.

Lemma C.3.3 *Let ϕ be a step function as described in 3.18. Then*

$$\phi(\mathbf{x},\cdot)\in\mathcal{R}(\mathbb{R}^m), \tag{3.20}$$

$$\int_{\mathbb{R}^m}\phi(\cdot,\mathbf{y})\,dV_y\in\mathcal{R}(\mathbb{R}^n), \tag{3.21}$$

and

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^m}\phi(\mathbf{x},\mathbf{y})\,dV_y\,dV_x=\int_{\mathbb{R}^{n+m}}\phi(\mathbf{z})\,dV. \tag{3.22}$$

Proof: To verify 3.20, note that $\phi(\mathbf{x},\cdot)$ is the step function

$$\phi(\mathbf{x},\mathbf{y})=\sum_{P\in\mathcal{G}_m}\phi_{R\times P}\mathcal{X}_{P'}(\mathbf{y}).$$

Where $\mathbf{x}\in R'$ and this is a finite sum of integrable functions because each has set of discontinuities with Jordan content 0. From the description in 3.19,

$$\int_{\mathbb{R}^m}\phi(\mathbf{x},\mathbf{y})\,dV_y=\sum_{R\in\mathcal{G}_n}\sum_{P\in\mathcal{G}_m}\phi_{R\times P}\mathcal{X}_{R'}(\mathbf{x})\,v(P)$$

$$= \sum_{R\in\mathcal{G}_n} \left(\sum_{P\in\mathcal{G}_m} \phi_{R\times P} v(P) \right) \mathcal{X}_{R'}(\mathbf{x}), \tag{3.23}$$

another step function. Therefore,

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \phi(\mathbf{x},\mathbf{y})\, dV_y\, dV_x = \sum_{R\in\mathcal{G}_n} \sum_{P\in\mathcal{G}_m} \phi_{R\times P} v(P) v(R)$$

$$= \sum_{Q\in\mathcal{G}} \phi_Q v(Q) = \int_{\mathbb{R}^{n+m}} \phi(\mathbf{z})\, dV.$$

and this proves the lemma.

From 3.23,

$$M_{R_1'}\left(\int_{\mathbb{R}^m} \phi(\cdot,\mathbf{y})\, dV_y \right) \equiv \sup\left\{ \sum_{R\in\mathcal{G}_n} \left(\sum_{P\in\mathcal{G}_m} \phi_{R\times P} v(P) \right) \mathcal{X}_{R'}(\mathbf{x}) : \mathbf{x}\in R_1' \right\}$$

$$= \sum_{P\in\mathcal{G}_m} \phi_{R_1\times P} v(P) \tag{3.24}$$

because $\int_{\mathbb{R}^m} \phi(\cdot,\mathbf{y})\, dV_y$ has the constant value given in 3.24 for $\mathbf{x}\in R_1'$. Similarly,

$$m_{R_1'}\left(\int_{\mathbb{R}^m} \phi(\cdot,\mathbf{y})\, dV_y \right) \equiv \inf\left\{ \sum_{R\in\mathcal{G}_n} \left(\sum_{P\in\mathcal{G}_m} \phi_{R\times P} v(P) \right) \mathcal{X}_{R'}(\mathbf{x}) : \mathbf{x}\in R_1' \right\}$$

$$= \sum_{P\in\mathcal{G}_m} \phi_{R_1\times P} v(P). \tag{3.25}$$

Theorem C.3.4 *(Fubini) Let $f\in\mathcal{R}(\mathbb{R}^{n+m})$ and suppose also that $f(\mathbf{x},\cdot)\in\mathcal{R}(\mathbb{R}^m)$ for each $\mathbf{x}$. Then*

$$\int_{\mathbb{R}^m} f(\cdot,\mathbf{y})\, dV_y \in \mathcal{R}(\mathbb{R}^n) \tag{3.26}$$

and

$$\int_{\mathbb{R}^{n+m}} f(\mathbf{z})\, dV = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} f(\mathbf{x},\mathbf{y})\, dV_y\, dV_x. \tag{3.27}$$

Proof: Let $\mathcal{G}$ be a grid such that $\mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon$ and let $\mathcal{G}_n$ and $\mathcal{G}_m$ be as defined above. Let

$$\phi(\mathbf{z}) \equiv \sum_{Q\in\mathcal{G}} M_{Q'}(f)\mathcal{X}_{Q'}(\mathbf{z}),\ \psi(\mathbf{z}) \equiv \sum_{Q\in\mathcal{G}} m_{Q'}(f)\mathcal{X}_{Q'}(\mathbf{z}).$$

Observe that $M_{Q'}(f) \le M_Q(f)$ and $m_{Q'}(f) \ge m_Q(f)$. Then

$$\mathcal{U}_{\mathcal{G}}(f) \ge \int \phi\, dV,\ \mathcal{L}_{\mathcal{G}}(f) \le \int \psi\, dV.$$

Also $f(\mathbf{z}) \in (\psi(\mathbf{z}), \phi(\mathbf{z}))$ for all $\mathbf{z}$. Thus from 3.24,

$$M_{R'}\left(\int_{\mathbb{R}^m} f(\cdot,\mathbf{y})\, dV_y \right) \le M_{R'}\left(\int_{\mathbb{R}^m} \phi(\cdot,\mathbf{y})\, dV_y \right) = \sum_{P\in\mathcal{G}_m} M_{R'\times P'}(f) v(P)$$

and from 3.25,

$$m_{R'}\left(\int_{\mathbb{R}^m} f(\cdot,\mathbf{y})\, dV_y \right) \ge m_{R'}\left(\int_{\mathbb{R}^m} \psi(\cdot,\mathbf{y})\, dV_y \right) = \sum_{P\in\mathcal{G}_m} m_{R'\times P'}(f) v(P).$$

Therefore,

$$\sum_{R\in\mathcal{G}_n}\left[M_{R'}\left(\int_{\mathbb{R}^m} f(\cdot,\mathbf{y})\,dV_y\right)-m_{R'}\left(\int_{\mathbb{R}^m} f(\cdot,\mathbf{y})\,dV_y\right)\right]v(R)\leq$$

$$\sum_{R\in\mathcal{G}_n}\sum_{P\in\mathcal{G}_m}\left[M_{R'\times P'}(f)-m_{R'\times P'}(f)\right]v(P)\,v(R)\leq\mathcal{U}_{\mathcal{G}}(f)-\mathcal{L}_{\mathcal{G}}(f)<\varepsilon.$$

This shows, from Lemma C.2.12 and the Riemann criterion, that $\int_{\mathbb{R}^m} f(\cdot,\mathbf{y})\,dV_y\in\mathcal{R}(\mathbb{R}^n)$. It remains to verify 3.27. First note

$$\int_{\mathbb{R}^{n+m}} f(\mathbf{z})\,dV\in[\mathcal{L}_{\mathcal{G}}(f),\mathcal{U}_{\mathcal{G}}(f)\,].$$

Next,

$$\mathcal{L}_{\mathcal{G}}(f)\leq\int_{\mathbb{R}^{n+m}}\psi\,dV=\int_{\mathbb{R}^n}\int_{\mathbb{R}^m}\psi\,dV_y\,dV_x\leq\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} f(\mathbf{x},\mathbf{y})\,dV_y\,dV_x$$

$$\leq\int_{\mathbb{R}^n}\int_{\mathbb{R}^m}\phi(\mathbf{x},\mathbf{y})\,dV_y\,dV_x=\int_{\mathbb{R}^{n+m}}\phi\,dV\leq\mathcal{U}_{\mathcal{G}}(f).$$

Therefore,

$$\left|\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} f(\mathbf{x},\mathbf{y})\,dV_y\,dV_x-\int_{\mathbb{R}^{n+m}} f(\mathbf{z})\,dV\right|\leq\varepsilon$$

and since $\varepsilon>0$ is arbitrary, this proves Fubini's theorem[2].

Corollary C.3.5 *Suppose E is a bounded contented set in $\mathbb{R}^n$ and let ϕ,ψ be continuous functions defined on E such that $\phi(\mathbf{x})\geq\psi(\mathbf{x})$. Also suppose f is a continuous bounded function defined on the set,*

$$P\equiv\{(\mathbf{x},y):\psi(\mathbf{x})\leq y\leq\phi(\mathbf{x})\},$$

It follows $f\mathcal{X}_P\in\mathcal{R}(\mathbb{R}^{n+1})$ and

$$\int_P f\,dV=\int_E\int_{\psi(\mathbf{x})}^{\phi(\mathbf{x})} f(\mathbf{x},y)\,dy\,dV_x.$$

Proof: Since f is continuous, there is no problem in writing $f(\mathbf{x},\cdot)\,\mathcal{X}_{[\psi(\mathbf{x}),\phi(\mathbf{x})]}(\cdot)\in\mathcal{R}(\mathbb{R}^1)$. Also, $f\mathcal{X}_P\in\mathcal{R}(\mathbb{R}^{n+1})$ because P is contented thanks to Corollary C.2.9. Therefore, by Fubini's theorem

$$\begin{aligned}\int_P f\,dV&=\int_{\mathbb{R}^n}\int_{\mathbb{R}} f\mathcal{X}_P\,dy\,dV_x\\&=\int_E\int_{\psi(\mathbf{x})}^{\phi(\mathbf{x})} f(\mathbf{x},y)\,dy\,dV_x\end{aligned}$$

proving the corollary.

Other versions of this corollary are immediate and should be obvious whenever encountered.

[2]Actually, Fubini's theorem usually refers to a much more profound result in the theory of Lebesgue integration.

C.4 The Change Of Variables Formula

First recall Theorem B.2.2 on Page 434 which is listed here for convenience.

Theorem C.4.1 *Let $\mathbf{h} : U \to \mathbb{R}^n$ be a C^1 function with $\mathbf{h}(\mathbf{0}) = \mathbf{0}$,$D\mathbf{h}(\mathbf{0})^{-1}$ exists. Then there exists an open set, $V \subseteq U$ containing $\mathbf{0}$, flips, $\mathbf{F}_1, \cdots, \mathbf{F}_{n-1}$, and primitive functions, $\mathbf{G}_n, \mathbf{G}_{n-1}, \cdots, \mathbf{G}_1$ such that for $\mathbf{x} \in V$,*

$$\mathbf{h}(\mathbf{x}) = \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{x}).$$

Also recall Theorem 8.14.12 on Page 194.

Theorem C.4.2 *Let $\phi : [a,b] \to [c,d]$ be one to one and suppose ϕ' exists and is continuous on $[a,b]$. Then if f is a continuous function defined on $[a,b]$,*

$$\int_c^d f(s)\, ds = \int_a^b f(\phi(t)) \left|\phi'(t)\right| dt$$

The following is a simple corollary to this theorem.

Corollary C.4.3 *Let $\phi : [a,b] \to [c,d]$ be one to one and suppose ϕ' exists and is continuous on $[a,b]$. Then if f is a continuous function defined on $[a,b]$,*

$$\int_{\mathbb{R}} \mathcal{X}_{[a,b]}\left(\phi^{-1}(x)\right) f(x)\, dx = \int_{\mathbb{R}} \mathcal{X}_{[a,b]}(t)\, f(\phi(t)) \left|\phi'(t)\right| dt$$

Lemma C.4.4 *Let $\mathbf{h} : V \to \mathbb{R}^n$ be a C^1 function and suppose H is a compact subset of V. Then there exists a constant, C independent of $\mathbf{x} \in H$ such that*

$$|D\mathbf{h}(\mathbf{x})\,\mathbf{v}| \leq C\,|\mathbf{v}|.$$

Proof: Consider the compact set, $H \times \partial B(\mathbf{0},1) \subseteq \mathbb{R}^{2n}$. Let $f : H \times \partial B(\mathbf{0},1) \to \mathbb{R}$ be given by $f(\mathbf{x},\mathbf{v}) = |D\mathbf{h}(\mathbf{x})\,\mathbf{v}|$. Then let C denote the maximum value of f. It follows that for $\mathbf{v} \in \mathbb{R}^n$,

$$\left|D\mathbf{h}(\mathbf{x})\,\frac{\mathbf{v}}{|\mathbf{v}|}\right| \leq C$$

and so the desired formula follows when you multiply both sides by $|\mathbf{v}|$.

Definition C.4.5 *Let A be an open set. Write $C^k(A;\mathbb{R}^n)$ to denote a C^k function whose domain is A and whose range is in $\mathbb{R}^n$. Let U be an open set in $\mathbb{R}^n$. Then $\mathbf{h} \in C^k\left(\overline{U};\mathbb{R}^n\right)$ if there exists an open set, $V \supseteq \overline{U}$ and a function, $\mathbf{g} \in C^1(V;\mathbb{R}^n)$ such that $\mathbf{g} = \mathbf{h}$ on $\overline{U}$. $f \in C^k\left(\overline{U}\right)$ means the same thing except that f has values in $\mathbb{R}$.*

Theorem C.4.6 *Let U be a bounded open set such that ∂U has zero content and let $\mathbf{h} \in C\left(\overline{U};\mathbb{R}^n\right)$ be one to one and $D\mathbf{h}(\mathbf{x})^{-1}$ exists for all $\mathbf{x} \in U$. Then $\mathbf{h}(\partial U) = \partial(\mathbf{h}(U))$ and $\partial(\mathbf{h}(U))$ has zero content.*

Proof: Let $\mathbf{x} \in \partial U$ and let $\mathbf{g} = \mathbf{h}$ where $\mathbf{g}$ is a C^1 function defined on an open set containing $\overline{U}$. By the inverse function theorem, $\mathbf{g}$ is locally one to one and an open mapping near $\mathbf{x}$. Thus $\mathbf{g}(\mathbf{x}) = \mathbf{h}(\mathbf{x})$ and is in an open set containing points of $\mathbf{g}(U)$ and points of $\mathbf{g}\left(U^C\right)$. These points of $\mathbf{g}\left(U^C\right)$ cannot equal any points of $\mathbf{h}(U)$ because $\mathbf{g}$ is one to one locally. Thus $\mathbf{h}(\mathbf{x}) \in \partial(\mathbf{h}(U))$ and so $\mathbf{h}(\partial U) \subseteq \partial(\mathbf{h}(U))$. Now suppose $\mathbf{y} \in \partial(\mathbf{h}(U))$. By the inverse function theorem $\mathbf{y}$ cannot be in the open set $\mathbf{h}(U)$. Since $\mathbf{y} \in \partial(\mathbf{h}(U))$, every ball centered at $\mathbf{y}$ contains points of $\mathbf{h}(U)$ and so $\mathbf{y} \in \overline{\mathbf{h}(U)} \setminus \mathbf{h}(U)$. Thus there exists a sequence, $\{\mathbf{x}_n\} \subseteq U$ such that $\mathbf{h}(\mathbf{x}_n) \to \mathbf{y}$. But then, by the inverse

function theorem, $\mathbf{x}_n \to \mathbf{h}^{-1}(\mathbf{y})$ and so $\mathbf{h}^{-1}(\mathbf{y}) \in \partial U$. Therefore, $\mathbf{y} \in \mathbf{h}(\partial U)$ and this proves the two sets are equal. It remains to verify the claim about content.

First let H denote a compact set whose interior contains $\overline{U}$ which is also in the interior of the domain of $\mathbf{g}$. Now since ∂U has content zero, it follows that for $\varepsilon > 0$ given, there exists a grid, $\mathcal{G}$ such that if $\mathcal{G}'$ are those boxes of $\mathcal{G}$ which have nonempty intersection with ∂U, then

$$\sum_{Q \in \mathcal{G}'} v(Q) < \varepsilon.$$

and by refining the grid if necessary, no box of $\mathcal{G}$ has nonempty intersection with both $\overline{U}$ and H^C. Refining this grid still more, you can also assume that for all boxes in $\mathcal{G}'$,

$$\frac{l_i}{l_j} < 2$$

where l_i is the length of the i^{th} side. (Thus the boxes are not too far from being cubes.)

Let C be the constant of Lemma C.4.4 applied to $\mathbf{g}$ on H.

Now consider one of these boxes, $Q \in \mathcal{G}'$. If $\mathbf{x}, \mathbf{y} \in Q$, it follows from the chain rule that

$$\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{x}) = \int_0^1 D\mathbf{g}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x})\,dt$$

By Lemma C.4.4 applied to H

$$\begin{aligned} |\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{x})| &\leq \int_0^1 |D\mathbf{g}(\mathbf{x}+t(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x})|\,dt \\ &\leq C\int_0^1 |\mathbf{x}-\mathbf{y}|\,dt \leq C\,\mathrm{diam}(Q) \\ &= C\left(\sum_{i=1}^n l_i^2\right)^{1/2} \leq C\sqrt{n}L \end{aligned}$$

where L is the length of the longest side of Q. Thus $\mathrm{diam}(\mathbf{g}(Q)) \leq C\sqrt{n}L$ and so $\mathbf{g}(Q)$ is contained in a cube having sides equal to $C\sqrt{n}L$ and volume equal to

$$C^n n^{n/2} L^n \leq C^n n^{n/2} 2^n l_1 l_2 \cdots l_n = C^n n^{n/2} 2^n v(Q).$$

Denoting by P_Q this cube, it follows

$$\mathbf{h}(\partial U) \subseteq \cup_{Q \in \mathcal{G}'} v(P_Q)$$

and

$$\sum_{Q \in \mathcal{G}'} v(P_Q) \leq C^n n^{n/2} 2^n \sum_{Q \in \mathcal{G}'} v(Q) < \varepsilon C^n n^{n/2} 2^n.$$

Since $\varepsilon > 0$ is arbitrary, this shows $\mathbf{h}(\partial U)$ has content zero as claimed.

Theorem C.4.7 *Suppose $f \in C(\overline{U})$ where U is a bounded open set with ∂U having content 0. Then $f\mathcal{X}_U \in \mathcal{R}(\mathbb{R}^n)$.*

Proof: Let H be a compact set whose interior contains $\overline{U}$ which is also contained in the domain of g where g is a continuous functions whose restriction to U equals f. Consider $g\mathcal{X}_U$, a function whose set of discontinuities has content 0. Then $g\mathcal{X}_U = f\mathcal{X}_U \in \mathcal{R}(\mathbb{R}^n)$ as claimed. This is by the big theorem which tells which functions are Riemann integrable.

The following lemma is obvious from the definition of the integral.

Lemma C.4.8 *Let U be a bounded open set and let $f\mathcal{X}_U \in \mathcal{R}(\mathbb{R}^n)$. Then*

$$\int f(\mathbf{x}+\mathbf{p})\,\mathcal{X}_{U-\mathbf{p}}(\mathbf{x})\,dx = \int f(\mathbf{x})\,\mathcal{X}_U(\mathbf{x})\,dx$$

A few more lemmas are needed.

Lemma C.4.9 *Let S be a nonempty subset of $\mathbb{R}^n$. Define*

$$f(\mathbf{x}) \equiv \operatorname{dist}(\mathbf{x},S) \equiv \inf\{|\mathbf{x}-\mathbf{y}| : \mathbf{y}\in S\}.$$

Then f is continuous.

Proof: Consider $|f(\mathbf{x})-f(\mathbf{x}_1)|$and suppose without loss of generality that $f(\mathbf{x}_1)\geq f(\mathbf{x})$. Then choose $\mathbf{y}\in S$ such that $f(\mathbf{x})+\varepsilon > |\mathbf{x}-\mathbf{y}|$. Then

$$\begin{aligned} |f(\mathbf{x}_1)-f(\mathbf{x})| &= f(\mathbf{x}_1)-f(\mathbf{x}) \leq f(\mathbf{x}_1)-|\mathbf{x}-\mathbf{y}|+\varepsilon \\ &\leq |\mathbf{x}_1-\mathbf{y}|-|\mathbf{x}-\mathbf{y}|+\varepsilon \\ &\leq |\mathbf{x}-\mathbf{x}_1|+|\mathbf{x}-\mathbf{y}|-|\mathbf{x}-\mathbf{y}|+\varepsilon \\ &= |\mathbf{x}-\mathbf{x}_1|+\varepsilon. \end{aligned}$$

Since ε is arbitrary, it follows that $|f(\mathbf{x}_1)-f(\mathbf{x})|\leq|\mathbf{x}-\mathbf{x}_1|$ and this proves the lemma.

Theorem C.4.10 *(Urysohn's lemma for $\mathbb{R}^n$) Let H be a closed subset of an open set, U. Then there exists a continuous function, $g:\mathbb{R}^n\to[0,1]$ such that $g(\mathbf{x})=1$ for all $\mathbf{x}\in H$ and $g(\mathbf{x})=0$ for all $\mathbf{x}\notin U$.*

Proof: If $\mathbf{x}\notin C$, a closed set, then $\operatorname{dist}(\mathbf{x},C)>0$ because if not, there would exist a sequence of points of C converging to $\mathbf{x}$ and it would follow that $\mathbf{x}\in C$. Therefore, $\operatorname{dist}(\mathbf{x},H)+\operatorname{dist}\left(\mathbf{x},U^C\right)>0$ for all $\mathbf{x}\in\mathbb{R}^n$. Now define a continuous function, g as

$$g(\mathbf{x}) \equiv \frac{\operatorname{dist}\left(\mathbf{x},U^C\right)}{\operatorname{dist}(\mathbf{x},H)+\operatorname{dist}(\mathbf{x},U^C)}.$$

It is easy to see this verifies the conclusions of the theorem and this proves the theorem.

Definition C.4.11 *Define* $\operatorname{spt}(f)$ *(support of f) to be the closure of the set $\{x : f(x)\neq 0\}$. If V is an open set, $C_c(V)$ will be the set of continuous functions f, defined on $\mathbb{R}^n$ having* $\operatorname{spt}(f)\subseteq V$.

Definition C.4.12 *If K is a compact subset of an open set, V, then $K\prec\phi\prec V$ if*

$$\phi\in C_c(V),\ \phi(K)=\{1\},\ \phi(\mathbb{R}^n)\subseteq[0,1].$$

Also for $\phi\in C_c(\mathbb{R}^n)$, $K\prec\phi$ if

$$\phi(\mathbb{R}^n)\subseteq[0,1] \text{ and } \phi(K)=1.$$

and $\phi\prec V$ if

$$\phi(\mathbb{R}^n)\subseteq[0,1] \text{ and } \operatorname{spt}(\phi)\subseteq V.$$

Theorem C.4.13 *(Partition of unity) Let K be a compact subset of $\mathbb{R}^n$ and suppose*

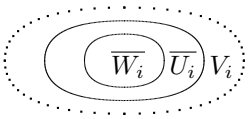
$$K\subseteq V=\cup_{i=1}^n V_i,\ V_i \text{ open.}$$

Then there exist $\psi_i\prec V_i$ with

$$\sum_{i=1}^n \psi_i(\mathbf{x})=1$$

for all $\mathbf{x}\in K$.

Proof: Let $K_1 = K \setminus \cup_{i=2}^n V_i$. Thus K_1 is compact because it is the intersection of a closed set with a compact set and $K_1 \subseteq V_1$. Let $K_1 \subseteq W_1 \subseteq \overline{W}_1 \subseteq V_1$ with $\overline{W}_1$compact. To obtain W_1, use Theorem C.4.10 to get f such that $K_1 \prec f \prec V_1$ and let $W_1 \equiv \{\mathbf{x} : f(\mathbf{x}) \neq 0\}$. Thus $W_1, V_2, \cdots V_n$ covers K and $\overline{W}_1 \subseteq V_1$. Let $K_2 = K \setminus (\cup_{i=3}^n V_i \cup W_1)$. Then K_2 is compact and $K_2 \subseteq V_2$. Let $K_2 \subseteq W_2 \subseteq \overline{W}_2 \subseteq V_2$ $\overline{W}_2$ compact. Continue this way finally obtaining $W_1, \cdots, W_n$, $K \subseteq W_1 \cup \cdots \cup W_n$, and $\overline{W}_i \subseteq V_i$ $\overline{W}_i$ compact. Now let $\overline{W}_i \subseteq U_i \subseteq \overline{U}_i \subseteq V_i$, $\overline{U}_i$ compact.



By Theorem C.4.10, there exist functions, ϕ_i, γ such that $\overline{U}_i \prec \phi_i \prec V_i$, $\cup_{i=1}^n \overline{W}_i \prec \gamma \prec \cup_{i=1}^n U_i$. Define

$$\psi_i(\mathbf{x}) = \begin{cases} \gamma(\mathbf{x})\phi_i(\mathbf{x}) / \sum_{j=1}^n \phi_j(\mathbf{x}) \text{ if } \sum_{j=1}^n \phi_j(\mathbf{x}) \neq 0, \\ 0 \text{ if } \sum_{j=1}^n \phi_j(\mathbf{x}) = 0. \end{cases}$$

If $\mathbf{x}$ is such that $\sum_{j=1}^n \phi_j(\mathbf{x}) = 0$, then $\mathbf{x} \notin \cup_{i=1}^n \overline{U}_i$. Consequently $\gamma(\mathbf{y}) = 0$ for all $\mathbf{y}$ near $\mathbf{x}$ and so $\psi_i(\mathbf{y}) = 0$ for all $\mathbf{y}$ near $\mathbf{x}$. Hence ψ_i is continuous at such $\mathbf{x}$. If $\sum_{j=1}^n \phi_j(\mathbf{x}) \neq 0$, this situation persists near $\mathbf{x}$ and so ψ_i is continuous at such points. Therefore ψ_i is continuous. If $\mathbf{x} \in K$, then $\gamma(\mathbf{x}) = 1$ and so $\sum_{j=1}^n \psi_j(\mathbf{x}) = 1$. Clearly $0 \leq \psi_i(\mathbf{x}) \leq 1$ and $\mathrm{spt}(\psi_j) \subseteq V_j$. This proves the theorem.

The next lemma contains the main ideas.

Lemma C.4.14 *Let U be a bounded open set with ∂U having content 0. Also let $\mathbf{h} \in C^1\left(\overline{U}; \mathbb{R}^n\right)$ be one to one on U with $D\mathbf{h}(\mathbf{x})^{-1}$ exists for all $\mathbf{x} \in U$. Let $f \in C\left(\overline{U}\right)$ be nonnegative. Then*

$$\int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f(\mathbf{z}) \, dV_n = \int \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) \left|\det D\mathbf{h}(\mathbf{x})\right| dV_n$$

Proof: Let $\varepsilon > 0$ be given. Then by Theorem C.4.7,

$$\mathbf{x} \to \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) \left|\det D\mathbf{h}(\mathbf{x})\right|$$

is Riemann integrable. Therefore, there exists a grid, $\mathcal{G}$ such that, letting

$$g(\mathbf{x}) = \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) \left|\det D\mathbf{h}(\mathbf{x})\right|,$$

$$\mathcal{L}_{\mathcal{G}}(g) + \varepsilon > \mathcal{U}_{\mathcal{G}}(g).$$

Let K denote the union of the boxes, Q of $\mathcal{G}$ for which $m_Q(g) > 0$. Thus K is a compact subset of U and it is only the terms from these boxes which contribute anything nonzero to the lower sum. By Theorem B.2.2 on Page 434 which is stated above, it follows that for $\mathbf{p} \in K$, there exists an open set contained in U which contains $\mathbf{p}$,$O_{\mathbf{p}}$ such that for $\mathbf{x} \in O_{\mathbf{p}} - \mathbf{p}$,

$$\mathbf{h}(\mathbf{x} + \mathbf{p}) - \mathbf{h}(\mathbf{p}) = \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \cdots \circ \mathbf{G}_1(\mathbf{x})$$

where the $\mathbf{G}_i$ are primitive functions and the $\mathbf{F}_j$ are flips. Finitely many of these open sets, $\{O_j\}_{j=1}^q$ cover K. Let the distinguished point for O_j be denoted by $\mathbf{p}_j$. Now refine $\mathcal{G}$ if necessary such that the diameter of every cell of the new $\mathcal{G}$ which intersects U is

smaller than a Lebesgue number for this open cover. Denote by $\mathcal{G}'$ those boxes of $\mathcal{G}$ whose union equals the set, K. Thus every box of $\mathcal{G}'$ is contained in one of these O_j. By Theorem C.4.13 there exists a partition of unity, $\{\psi_j\}$ on $\mathbf{h}(K)$ such that $\psi_j \prec \mathbf{h}(O_j)$. Then

$$\begin{aligned}\mathcal{L}_{\mathcal{G}}(g) &\leq \sum_{Q\in\mathcal{G}'}\int \mathcal{X}_Q(\mathbf{x})\, f(\mathbf{h}(\mathbf{x}))\,|\det D\mathbf{h}(\mathbf{x})|\,dx \\ &= \sum_{Q\in\mathcal{G}'}\sum_{j=1}^{q}\int \mathcal{X}_Q(\mathbf{x})\,(\psi_j f)(\mathbf{h}(\mathbf{x}))\,|\det D\mathbf{h}(\mathbf{x})|\,dx. \qquad (3.28)\end{aligned}$$

Consider the term $\int \mathcal{X}_Q(\mathbf{x})\,(\psi_j f)(\mathbf{h}(\mathbf{x}))\,|\det D\mathbf{h}(\mathbf{x})|\,dx$. By Lemma C.4.8 and Fubini's theorem this equals

$$\begin{aligned}&\int_{\mathbb{R}^{n-1}}\int_{\mathbb{R}} \mathcal{X}_{Q-\mathbf{p}_j}(\mathbf{x})\,(\psi_j f)(\mathbf{h}(\mathbf{p}_i)+\mathbf{F}_1\circ\cdots\circ\mathbf{F}_{n-1}\circ\mathbf{G}_n\circ\cdots\circ\mathbf{G}_1(\mathbf{x}))\cdot \\ &|D\mathbf{F}(\mathbf{G}_n\circ\cdots\circ\mathbf{G}_1(\mathbf{x}))|\,|D\mathbf{G}_n(\mathbf{G}_{n-1}\circ\cdots\circ\mathbf{G}_1(\mathbf{x}))|\,|D\mathbf{G}_{n-1}(\mathbf{G}_{n-2}\circ\cdots\circ\mathbf{G}_1(\mathbf{x}))|\cdot \\ &\cdots|D\mathbf{G}_2(\mathbf{G}_1(\mathbf{x}))|\,|D\mathbf{G}_1(\mathbf{x})|\,dx_1 dV_{n-1}. \qquad (3.29)\end{aligned}$$

Here dV_{n-1} is with respect to the variables, $x_2,\cdots,x_n$. Also $\mathbf{F}$ denotes $\mathbf{F}_1\circ\cdots\circ\mathbf{F}_{n-1}$. Now

$$\mathbf{G}_1(\mathbf{x}) = (\alpha(\mathbf{x}), x_2,\cdots,x_n)^T$$

and is one to one. Therefore, fixing $x_2,\cdots,x_n$, $x_1 \to \alpha(\mathbf{x})$ is one to one. Also, $D\mathbf{G}_1(\mathbf{x}) = \frac{\partial\alpha}{\partial x_1}(\mathbf{x})$. Fixing $x_2,\cdots,x_n$, change the variable,

$$y_1 = \alpha(x_1, x_2,\cdots,x_n).$$

Thus

$$\mathbf{x} = (x_1,x_2,\cdots,x_n)^T = \mathbf{G}_1^{-1}(y_1,x_2,\cdots,x_n) \equiv \mathbf{G}_1^{-1}(\mathbf{x}')$$

Then in 3.29 you can use Corollary C.4.3 to write 3.29 as

$$\begin{aligned}&\int_{\mathbb{R}^{n-1}}\int_{\mathbb{R}} \mathcal{X}_{Q-\mathbf{p}_j}(\mathbf{G}_1^{-1}(\mathbf{x}'))\,(\psi_j f)(\mathbf{h}(\mathbf{p}_i)+\mathbf{F}_1\circ\cdots\circ\mathbf{F}_{n-1}\circ\mathbf{G}_n\circ\cdots\circ\mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))\cdot \\ &|D\mathbf{F}(\mathbf{G}_n\circ\cdots\circ\mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))|\,|D\mathbf{G}_n(\mathbf{G}_{n-1}\circ\cdots\circ\mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))|\cdot \\ &|D\mathbf{G}_{n-1}(\mathbf{G}_{n-2}\circ\cdots\circ\mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))|\cdots|D\mathbf{G}_2(\mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))|\,dy_1 dV_{n-1} \qquad (3.30)\end{aligned}$$

which reduces to

$$\begin{aligned}&\int_{\mathbb{R}^n} \mathcal{X}_{Q-\mathbf{p}_j}(\mathbf{G}_1^{-1}(\mathbf{x}'))\,(\psi_j f)(\mathbf{h}(\mathbf{p}_i)+\mathbf{F}_1\circ\cdots\circ\mathbf{F}_{n-1}\circ\mathbf{G}_n\circ\cdots\circ\mathbf{G}_2(\mathbf{x}'))\cdot \\ &|D\mathbf{F}(\mathbf{G}_n\circ\cdots\circ\mathbf{G}_2(\mathbf{x}'))|\,|D\mathbf{G}_n(\mathbf{G}_{n-1}\circ\cdots\circ\mathbf{G}_2(\mathbf{x}'))|\,|D\mathbf{G}_{n-1}(\mathbf{G}_{n-2}\circ\cdots\circ\mathbf{G}_2(\mathbf{x}'))|\cdot \\ &\cdots|D\mathbf{G}_2(\mathbf{x}')|\,dV_n. \qquad (3.31)\end{aligned}$$

Now use Fubini's theorem again to make the inside integral taken with respect to x_2. Exactly the same process yields

$$\begin{aligned}&\int_{\mathbb{R}^{n-1}}\int_{\mathbb{R}} \mathcal{X}_{Q-\mathbf{p}_j}(\mathbf{G}_1^{-1}\circ\mathbf{G}_2^{-1}(\mathbf{x}''))\,(\psi_j f)(\mathbf{h}(\mathbf{p}_i)+\mathbf{F}_1\circ\cdots\circ\mathbf{F}_{n-1}\circ\mathbf{G}_n\circ\cdots\circ\mathbf{G}_3(\mathbf{x}''))\cdot \\ &|D\mathbf{F}(\mathbf{G}_n\circ\cdots\circ\mathbf{G}_3(\mathbf{x}''))|\,|D\mathbf{G}_n(\mathbf{G}_{n-1}\circ\cdots\circ\mathbf{G}_3(\mathbf{x}''))|\,|D\mathbf{G}_{n-1}(\mathbf{G}_{n-2}\circ\cdots\circ\mathbf{G}_3(\mathbf{x}''))|\cdot \\ &\cdots dy_2 dV_{n-1}. \qquad (3.32)\end{aligned}$$

Now $\mathbf{F}$ is just a composition of flips and so $|D\mathbf{F}(\mathbf{G}_n\circ\cdots\circ\mathbf{G}_3(\mathbf{x}''))| = 1$ and so this term can be replaced with 1. Continuing this process, eventually yields an expression of the form

$$\int_{\mathbb{R}^n} \mathcal{X}_{Q-\mathbf{p}_j}(\mathbf{G}_1^{-1}\circ\cdots\circ\mathbf{G}_{n-2}^{-1}\circ\mathbf{G}_{n-1}^{-1}\circ\mathbf{G}_n^{-1}\circ\mathbf{F}^{-1}(\mathbf{y}))\,(\psi_j f)(\mathbf{h}(\mathbf{p}_i)+\mathbf{y})\,dV_n. \quad (3.33)$$

Denoting by $\mathbf{G}^{-1}$ the expression, $\mathbf{G}_1^{-1} \circ \cdots \circ \mathbf{G}_{n-2}^{-1} \circ \mathbf{G}_{n-1}^{-1} \circ \mathbf{G}_n^{-1}$,

$$\mathcal{X}_{Q-\mathbf{p}_j}\left(\mathbf{G}_1^{-1} \circ \cdots \circ \mathbf{G}_{n-2}^{-1} \circ \mathbf{G}_{n-1}^{-1} \circ \mathbf{G}_n^{-1} \circ \mathbf{F}^{-1}(\mathbf{y})\right) = 1$$

exactly when $\mathbf{G}^{-1} \circ \mathbf{F}^{-1}(\mathbf{y}) \in Q - \mathbf{p}_j$. Now recall that $\mathbf{h}(\mathbf{p}_j + \mathbf{x}) - \mathbf{h}(\mathbf{p}_j) = \mathbf{F} \circ \mathbf{G}(\mathbf{x})$ and so the above holds exactly when

$$\begin{aligned} \mathbf{y} &= \mathbf{h}\left(\mathbf{p}_j + \mathbf{G}^{-1} \circ \mathbf{F}^{-1}(\mathbf{y})\right) - \mathbf{h}(\mathbf{p}_j) \in \mathbf{h}(\mathbf{p}_j + Q - \mathbf{p}_j) - \mathbf{h}(\mathbf{p}_j) \\ &= \mathbf{h}(Q) - \mathbf{h}(\mathbf{p}_j). \end{aligned}$$

Thus 3.33 reduces to

$$\int_{\mathbb{R}^n} \mathcal{X}_{\mathbf{h}(Q)-\mathbf{h}(\mathbf{p}_j)}(\mathbf{y})\left(\psi_j f\right)(\mathbf{h}(\mathbf{p}_i) + \mathbf{y})\, dV_n = \int_{\mathbb{R}^n} \mathcal{X}_{\mathbf{h}(Q)}(\mathbf{z})\left(\psi_j f\right)(\mathbf{z})\, dV_n.$$

It follows from 3.28

$$\begin{aligned} \mathcal{U}_{\mathcal{G}}(g) - \varepsilon &\leq \mathcal{L}_{\mathcal{G}}(g) \leq \sum_{Q \in \mathcal{G}'} \int \mathcal{X}_Q(\mathbf{x})\, f(\mathbf{h}(\mathbf{x}))\, |\det D\mathbf{h}(\mathbf{x})|\, dx \\ &= \sum_{Q \in \mathcal{G}'} \sum_{j=1}^{q} \int \mathcal{X}_Q(\mathbf{x})\left(\psi_j f\right)(\mathbf{h}(\mathbf{x}))\, |\det D\mathbf{h}(\mathbf{x})|\, dx \\ &= \sum_{Q \in \mathcal{G}'} \sum_{j=1}^{q} \int_{\mathbb{R}^n} \mathcal{X}_{\mathbf{h}(Q)}(\mathbf{z})\left(\psi_j f\right)(\mathbf{z})\, dV_n \\ &= \sum_{Q \in \mathcal{G}'} \int_{\mathbb{R}^n} \mathcal{X}_{\mathbf{h}(Q)}(\mathbf{z})\, f(\mathbf{z})\, dV_n \leq \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z})\, f(\mathbf{z})\, dV_n \end{aligned}$$

which implies the inequality,

$$\int \mathcal{X}_U(\mathbf{x})\, f(\mathbf{h}(\mathbf{x}))\, |\det D\mathbf{h}(\mathbf{x})|\, dV_n \leq \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z})\, f(\mathbf{z})\, dV_n$$

But now you can use the same information just derived to obtain equality. $\mathbf{x} = \mathbf{h}^{-1}(\mathbf{z})$ and so from what was just done,

$$\int \mathcal{X}_U(\mathbf{x})\, f(\mathbf{h}(\mathbf{x}))\, |\det D\mathbf{h}(\mathbf{x})|\, dV_n$$

$$\begin{aligned} &= \int \mathcal{X}_{\mathbf{h}^{-1}(\mathbf{h}(U))}(\mathbf{x})\, f(\mathbf{h}(\mathbf{x}))\, |\det D\mathbf{h}(\mathbf{x})|\, dV_n \\ &\geq \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z})\, f(\mathbf{z})\, \left|\det D\mathbf{h}\left(\mathbf{h}^{-1}(\mathbf{z})\right)\right| \left|\det D\mathbf{h}^{-1}(\mathbf{z})\right| dV_n \\ &= \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z})\, f(\mathbf{z})\, dV_n \end{aligned}$$

from the chain rule. In fact,

$$I = D\mathbf{h}\left(\mathbf{h}^{-1}(\mathbf{z})\right) D\mathbf{h}^{-1}(\mathbf{z})$$

and so

$$1 = \left|\det D\mathbf{h}\left(\mathbf{h}^{-1}(\mathbf{z})\right)\right| \left|\det D\mathbf{h}^{-1}(\mathbf{z})\right|.$$

This proves the lemma.

The change of variables theorem follows.

Theorem C.4.15 *Let U be a bounded open set with ∂U having content 0. Also let $\mathbf{h} \in C^1\left(\overline{U}; \mathbb{R}^n\right)$ be one to one on U with $D\mathbf{h}(\mathbf{x})^{-1}$ exists for all $\mathbf{x} \in U$. Let $f \in C\left(\overline{U}\right)$. Then*

$$\int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f(\mathbf{z}) dz = \int \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx$$

Proof: You note that the formula holds for $f^+ \equiv \frac{|f|+f}{2}$ and $f^- \equiv \frac{|f|-f}{2}$. Now $f = f^+ - f^-$ and so

$$\int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f(\mathbf{z}) dz$$

$$\begin{aligned} &= \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f^+(\mathbf{z}) dz - \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f^-(\mathbf{z}) dz \\ &= \int \mathcal{X}_U(\mathbf{x}) f^+(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx - \int \mathcal{X}_U(\mathbf{x}) f^-(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx \\ &= \int \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx. \end{aligned}$$

C.5 Some Observations

Some of the above material is very technical. This is because it gives complete answers to the fundamental questions on existence of the integral and related theoretical considerations. However, most of the difficulties are artifacts. They shouldn't even be considered! It was realized early in the twentieth century that these difficulties occur because, from the point of view of mathematics, this is not the right way to define an integral! Better results are obtained much more easily using the Lebesgue integral. Many of the technicalities related to Jordan content disappear almost magically when the right integral is used. However, the Lebesgue integral is more abstract than the Riemann integral and it is not traditional to consider it in a beginning calculus course. If you are interested in the fundamental properties of the integral and the theory behind it, you should abandon the Riemann integral which is an antiquated relic and begin to study the integral of the last century. An introduction to it is in [23]. Another very good source is [12]. This advanced calculus text does everything in terms of the Lebesgue integral and never bothers to struggle with the inferior Riemann integral. A more general treatment is found in [18], [19], [24], and [20]. There is also a still more general integral called the generalized Riemann integral. A recent book on this subject is [5]. It is far easier to define than the Lebesgue integral but the convergence theorems are much harder to prove. An introduction is also in [19].

Bibliography

[1] **Apostol, T. M.,** *Calculus second edition,* Wiley, 1967.

[2] **Apostol T.** *Calculus Volume II Second edition,* Wiley *1969.*

[3] **Apostol, T. M.,** *Mathematical Analysis,* Addison Wesley Publishing Co., 1974.

[4] **Baker, Roger**, *Linear Algebra*, Rinton Press 2001.

[5] **Bartle R.G.,** *A Modern Theory of Integration,* Grad. Studies in Math., Amer. Math. Society, Providence, RI, 2000.

[6] **Chahal J. S.** , *Historical Perspective of Mathematics* 2000 B.C. - 2000 A.D.

[7] **Davis H. and Snider A.,** *Vector Analysis* Wm. C. Brown 1995.

[8] **D'Angelo, J. and West D**. *Mathematical Thinking Problem Solving and Proofs*, Prentice Hall 1997.

[9] **Edwards C.H.** *Advanced Calculus of several Variables,* Dover 1994.

[10] **Euclid**, *The Thirteen Books of the Elements*, Dover, 1956.

[11] **Fitzpatrick P. M.,** *Advanced Calculus a course in Mathematical Analysis,* PWS Publishing Company 1996.

[12] **Fleming W**., *Functions of Several Variables,* Springer Verlag 1976.

[13] **Greenberg, M**. Advanced Engineering Mathematics, Second edition, Prentice Hall, 1998

[14] **Gurtin M.** *An introduction to continuum mechanics,* Academic press 1981.

[15] **Hardy G.,** *A Course Of Pure Mathematics, Tenth edition,* Cambridge University Press 1992.

[16] **Horn R. and Johnson C.** *matrix Analysis,* Cambridge University Press, 1985.

[17] **Karlin S. and Taylor H.** *A First Course in Stochastic Processes,* Academic Press, 1975.

[18] **Kuttler K. L.**, *Basic Analysis*, Rinton

[19] **Kuttler K.L.,** *Modern Analysis* CRC Press 1998.

[20] **Lang S.** *Real and Functional analysis* third edition Springer Verlag 1993. Press, 2001.

[21] **Nobel B. and Daniel J.** *Applied Linear Algebra,* Prentice Hall, 1977.

[22] **Rose, David, A.,** The College Math Journal, vol. 22, No.2 March 1991.

[23] **Rudin, W.,** *Principles of mathematical analysis,* McGraw Hill third edition 1976

[24] **Rudin W.,** *Real and Complex Analysis,* third edition, McGraw-Hill, 1987.

[25] **Salas S. and Hille E.,** *Calculus One and Several Variables,* Wiley 1990.

[26] **Sears and Zemansky,** *University Physics, Third edition, Addison Wesley 1963.*

[27] **Tierney John,** *Calculus and Analytic Geometry,* fourth edition, Allyn and Bacon, Boston, 1969.

[28] **Yosida K.,** *Functional Analysis, Springer Verlag, 1978.*

Index