

# Course Notes for Calculus IV

Remkes Kooistra  
The King's University

February 21, 2019

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Integration</b>	<b>2</b>
2.1	The Indefinite Integral . . . . .	2
2.2	The Definite Integral . . . . .	3
2.3	Iterated Integrals . . . . .	6
2.4	Integration over Arbitrary Sets . . . . .	9
2.5	Change of Variables . . . . .	17
2.6	Polar Coordinates . . . . .	24
2.7	Cylindrical and Spherical Coordinates . . . . .	27
2.8	Applications of Multiple Integration . . . . .	31
<b>3</b>	<b>Vector Calculus</b>	<b>47</b>
3.1	Vector Fields . . . . .	47
3.2	Integral Curves . . . . .	49
3.3	Vector Operations on Vector Fields . . . . .	52
3.4	Conservative Vector Fields . . . . .	56
3.5	Line Integrals of Vector Fields . . . . .	57
3.6	Parametric Surfaces . . . . .	63
3.7	Surface Integrals . . . . .	64
3.8	Flux Integrals . . . . .	66
3.9	The Major Results . . . . .	69
3.10	Vector Calculus and Maxwell's Equations . . . . .	76

**4 Differential Forms and Differential Geometry** **81**

4.1 Manifolds and Coordinate Functions . . . . . 81

4.2 Functions on Manifolds . . . . . 83

4.3 Differential Forms . . . . . 83

4.4 Pullbacks . . . . . 88

4.5 Integration of Differential Forms . . . . . 89

4.6 Stokes' Theorem . . . . . 91

## **License**

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License.

# Chapter 1

## Introduction

Calculus III was primarily devoted to extending the derivative to multivariable scalar functions. This course starts with the parallel program of extending the integral to multivariable scalar functions. Unlike the derivative, the definition of the multivariable (definite) integral is relatively similar to the single variable case. The challenge of integration is in the geometry of the subsets of  $\mathbb{R}^n$  involved in the integration. The course proceeds to a number of applications of multivariable integration.

The second half of the course covers the material known as vector calculus. This is the calculus of multivariable vector functions: functions with multivariable outputs as well as inputs. We investigate several derivative and integral extensions for these functions.

Since the extension of the derivative to multivariable functions is multi-faceted, we lose the simple idea of an antiderivative. However, in the context of vector calculus, we can produce several new theorems which resemble the fundamental theorem of calculus, even without a simple anti-derivative. The fruit of this labour are the famous Gauss, Green and Stokes theorems.

The final section of the course is a brief introduction to differential geometry. We introduce the ideas of manifolds and differential forms (and we finally get a reasonable definition of the differential 'dx' used in integrals). We develop this material far enough to show the deep connections underlying the fundamental theorem of calculus and the Gauss, Green and Stokes theorems.

# Chapter 2

## Integration

### 2.1 The Indefinite Integral

In single variable calculus, the integral is defined as a limit of approximating Riemann sums.

**Definition 2.1.1.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function. If we partition the interval  $[a, b]$  into  $n$  equal pieces of size  $\frac{b-a}{n}$  and  $x_k^*$  is any element in the  $k$ th partition, the integral of  $f$  on  $[a, b]$  is defined to be

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k^*) \frac{b-a}{n}.$$

Recall the fundamental theorem of calculus.

**Theorem 2.1.2.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be a continuous function. Then if  $F(x)$  is any antiderivative of  $f$  (i.e.  $F'(x) = f(x)$ ), we can evaluate the definite integral using the antiderivative.

$$\int_a^b f(x)dx = F(b) - F(a)$$

*There are several version of the fundamental theorem, but all say the same thing: (single-variable) derivatives and integrals are inverse processes.*

$$\frac{d}{dx} \int_a^x f(t)dt = f(x)$$

We used the indefinite integral  $\int f dx$  to stand for *all* anti-derivatives of the single-variable function. Is there a multivariable version of this anti-derivative? Calculus III tried to extend the derivative to multi-variable functions. The result was a collection of construction: partial derivatives, tangent (hyper-)planes, gradients and the Jacobian matrix. However, none of these notions of the derivative gave a new multi-variable function. Therefore, there isn't a nice notion of anti-derivative: going back doesn't make sense, since going forward led to more complicated constructions.

We can, however, take indefinite integrals in one variable, much like partial derivatives. If  $f(x, y, z)$  is a multivariable continuous function, then the following three integrals are the counterparts of partial differentiation. In each of the following we pretend the other two variables are constant; only the variable indicated by the infinitesimal term is non-constant.

$$\int f(x, y, z) dx$$

$$\int f(x, y, z) dy$$

$$\int f(x, y, z) dz$$

Our best extension of the fundamental theorem is just a version involving the partial derivatives.

$$\int \frac{\partial}{\partial x} f(x, y, z) dx = f(x, y, z) + g(y, z)$$

$$\int \frac{\partial}{\partial y} f(x, y, z) dy = f(x, y, z) + g(x, z)$$

$$\int \frac{\partial}{\partial z} f(x, y, z) dz = f(x, y, z) + g(x, y)$$

Note that the 'constants of integration' now may involve the other variables. When we do single variable integration or partial differentiation, we pretend all the other variables are constant; therefore, the constant term can include those variables.

We will discover much deeper and more interesting extensions of the fundamental theorem of calculus later in this course. However, they will come from entirely new directions. For now, this is as far as we can extend the fundamental theorem. Nothing more can be said about the indefinite integral; it is not used for multivariable functions.

## 2.2 The Definite Integral

The single-variable definite integral measures the area under the graph of a function. We also have graphs for multivariable functions. For a function of two variables, the graph is in  $\mathbb{R}^3$  and there is volume under the graph instead of area. For functions of three or more variables, the graphs are in higher dimensional spaces and enclose a hyper-volume. The definite integral asks for the value of this volume or hyper-volume.

Single variable integration took place over intervals. We do the same for multi-variable integration, so we need intervals in  $\mathbb{R}^n$ .

**Definition 2.2.1.** An closed interval  $I$  in  $\mathbb{R}^n$  is product of single-variable closed intervals.

$$I = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$$

This is the subset of  $\mathbb{R}^n$  consisting of all  $(x_1, x_2, \dots, x_n)$  such that  $a_1 \leq x_1 \leq b_1$ ,  $a_2 \leq x_2 \leq b_2$  and so on until  $a_n \leq x_n \leq b_n$ . In  $\mathbb{R}^2$  it is a rectangle. In  $\mathbb{R}^3$  it is a rectangular prism. Similarly, by using open interval,  $(a, b)$ , in the definition, we can define an open interval in  $\mathbb{R}^n$ .

For the single-variable integral, we partitioned the interval into  $n$  pieces. We can also partition these multi-variable intervals. In general, there are several ways to do so; one of the easiest is simply to divide each of the single variable intervals into  $l$  pieces. In  $\mathbb{R}^2$ , this splits the rectangle into a checkerboard pattern, with  $l^2$  small rectangles. In  $\mathbb{R}^3$ , it splits into  $l^3$  small rectangular prisms. In  $\mathbb{R}^n$  there are  $l^n$  subdivisions.

When we have a partition of  $I$  into  $l^n$  pieces, we can use the same limit process as the single variable integral. Let  $x^*$  be any point in each piece of the partition and estimate the (hyper-)volume over that piece of the interval by a rectangular (hyper-)volume. The height is  $f(x^*)$ . The (hyper-) volume of one pieces of our approximation is this height multiplied by the size of the small partition pieces, in the appropriate dimension. We can write  $\Delta V$  for this size, thinking of  $V$  for volume or hypervolume. Then we

**Definition 2.2.2.** The integral of a function  $f : I \rightarrow \mathbb{R}$  on an interval  $I \subset \mathbb{R}^n$  is defined as the following limit, if it exists. If the limit exists, we say that  $f$  is *integrable* on the interval  $I$ .

$$\int_I f(x_1, x_2, \dots, x_n) dV = \lim_{l \rightarrow \infty} \sum_{l^n} f(x^*) \Delta V$$

In the multivariable integral, the  $\Delta V$  becomes an infinitesimal piece of area, volume or hypervolume  $dV$ .

Let's specialize this to  $\mathbb{R}^2$ . In  $\mathbb{R}^2$ , we often write  $\Delta A$  and  $dA$  for area instead of volume. Consider an interval in  $\mathbb{R}^2$ :  $I = [a_1, b_1] \times [a_2, b_2]$ .

$$\int_I f(x, y) dA = \lim_{l \rightarrow \infty} \sum_{i=1}^l \sum_{j=1}^l f(x_i^*, y_j^*) \Delta x \Delta y$$

We've clarified the nature of the sum by separating the indicies, counting the divisions of each single-variable interval. The points are also more clear:  $x_i^*$  is in the  $i$ th piece of  $[a_1, b_1]$  and  $y_j^*$  is in the  $j$ th piece of  $[a_2, b_2]$ . Finally,  $\Delta x$  is  $\frac{b_1 - a_1}{l}$ ,  $\Delta y$  is  $\frac{b_2 - a_2}{l}$ , and the product  $\Delta x \Delta y$  is area of each piece of the partion.

The same works in  $\mathbb{R}^3$ , with three sums.

$$\int_I f(x, y, z) dV = \lim_{l \rightarrow \infty} \sum_{i=1}^l \sum_{j=1}^l \sum_{k=1}^l f(x_i^*, y_j^*, z_k^*) \Delta x \Delta y \Delta z$$

Likewise in  $\mathbb{R}^n$ , with  $n$  sums.

$$\int_I f(x_1, x_2, \dots, x_n) dV = \lim_{l \rightarrow \infty} \sum_{i_1=1}^l \sum_{i_2=1}^l \dots \sum_{i_n=1}^l f(x_{1,i_1}^*, x_{2,i_2}^*, \dots, x_{n,i_n}^*) \Delta x_1 \Delta x_2 \dots \Delta x_n$$

For single variable functions, the definite integral measured the area under the graph of the function over the interval  $I$ . Here the interpretation is the same, but the notion of size increases with dimension. For functions of two variables, the definite integral is the volume under the graph over the rectangle  $I$  in question. For functions of more than two variables, we get the hyper-volume under the graph and above the interval.

There are several important similarities between the multi-variable definition and the familiar single-variable definition. Both are limits of approximation methods. All the core definitions of calculus involve some algebraic approximation and the limit of such a process. Multi-variable integrals are no different; the algebra is a bit more complicated to handle the higher dimensions, but the process is the same. Also, both definitions are very, very difficult to work with. We needed the fundamental theorem and many integration techniques to calculate single variable integrals. We will need similar theorems and techniques to calculate multivariable integrals.

For single variable functions, if the limit defining the integral existed, we said the function was integrable. The same is true for multivariable functions.

**Definition 2.2.3.** . Let  $f : I \rightarrow \mathbb{R}$  be a function of  $n$  variables. If we can break the interval  $I$  up into finitely many pieces and if  $f$  is continuous on each of the pieces, we say the function is *piecewise continuous*.

**Proposition 2.2.4.** *All piecewise continuous functions of several variables are integrable.*

There are, in fact, other functions which are also integrable. However, they can be difficult to construct. As an aside, there are also different notions of integrability and different definition of the integral. What we have discussed here is the Riemann integral, and this notion of integrability is Riemann-integrability. In higher mathematics, the most commonly used definition of the integral is a different definition, called the Lebesgue integral. It is also an approximation process, but it involve approximating the function with a special functions that only take finitely many values. The Lebesgue integral agrees with the Riemann integral on all Riemann-integrable functions. However, there are functions where the Lebesgue integral is defined but the limit in the Riemann integral fails to converge. It is interesting to observe that something as fundamental as the integral has multiple definitions and the choice of definitions has a real, measurable effect on our mathematics.

The multi-variable definite integral has some familiar properties.

**Proposition 2.2.5.** *Let  $I$  be an interval in  $\mathbb{R}^n$ ,  $f : I \rightarrow \mathbb{R}$  and  $g : I \rightarrow \mathbb{R}$  integrable functions on  $I$ , and  $c \in \mathbb{R}$ . First, the integral is linear.*

$$\begin{aligned} \int_I cf(x_1, \dots, x_n) dV &= c \int_I f(x_1, \dots, x_n) dV \\ \int_I (f \pm g) dV &= \int_I f dV \pm \int_I g dV \end{aligned}$$

Assume, in addition, that  $f \leq g$  on all of  $I$ .

$$\int_I f dV \leq \int_I g dV$$

Let  $J$  be another interval where  $f$  is integrable and assume  $J$  is adjacent to  $I$  (such that they have no open intersection and their union  $I \cup J$  is also also an interval).

$$\int_{I \cup J} f dV = \int_I f dV + \int_J f dV$$

## 2.3 Iterated Integrals

The previous section gave us a good definition, but how do we actually calculate these integrals? Let  $I = [a, b] \times [c, d]$  be an interval in  $\mathbb{R}^2$ . The infinitesimal piece of area  $dA$  in the integral can be thought of as the product  $dx dy$  of infinitesimal pieces of length. Separating  $x$  and  $y$ , we can iterate the integral, first working in  $x$  and then in  $y$ .

$$\int_I f(x, y) dA = \int_c^d \left[ \int_a^b f(x, y) dx \right] dy$$

We work from the inside out: the infinitesimal piece  $dx$  or  $dy$  that is closest to the function acts first. Its bounds are written on the right, closest to the function. That integral is a single-variable integration, pretending that the other variables are constant, as we did with partial derivatives. Then, once the first integration is finished, we proceed outward to the next integral.

The same works in higher dimensions: we just iterate through the variables. In  $\mathbb{R}^3$ , over the interval  $I = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$  with three iterations.

$$\int_I f(x, y, z) dV = \int_{a_3}^{b_3} \left[ \int_{a_2}^{b_2} \left[ \int_{a_1}^{b_1} f(x, y, z) dx \right] dy \right] dz$$

Notice that we always have to match the bounds of integration with the correct infinitesimal piece, working from the inside out.

You might ask, at this point, what happens when we change the order? We have to choose an order for these iterated integrals. Fubini's Theorem tells us that we are able to choose the order.

**Theorem 2.3.1.** *If  $f : I \rightarrow \mathbb{R}$  is an integrable function on an interval, then the iterated integral is independent of the choice or order. Any ordering of the variables will result in the same value.*

In working with iterated integrals, we must be careful with the evaluation of each integral. It's easy to confuse the variable of integration and substitute for the wrong variables.

**Example 2.3.2.** Consider the interval  $I = [0, 1] \times [3, 6]$  and the function  $f(x, y) = e^x + e^y$ . In this example, I've carefully labelled the bound to remind myself which variable to replace.

$$\begin{aligned}
 \int_I f dA &= \int_0^1 \int_3^6 f(x, y) dy dx \\
 &= \int_0^1 \int_3^6 (e^x + e^y) dy dx \\
 &= \int_0^1 \int_3^6 e^x dy dx + \int_0^1 \int_3^6 e^y dy dx \\
 &= \int_0^1 ye^x \Big|_{y=3}^{y=6} dx + \int_0^1 e^y \Big|_{y=3}^{y=6} dx \\
 &= \int_0^1 3e^x dx + \int_0^1 (e^6 - e^3) dx \\
 &= 3e^x \Big|_{x=0}^{x=1} + x(e^6 - e^3) \Big|_{x=0}^{x=1} \\
 &= 3e - 3 + e^6 - e^3
 \end{aligned}$$

**Example 2.3.3.** Consider the interval  $I = [-1, 1] \times [-2, 2]$  and the function  $f(x, y) = x^2y^3$ .

$$\begin{aligned}
 \int_I f(x, y) dA &= \int_{-2}^2 \int_{-1}^1 x^2y^3 dx dy \\
 &= \int_{-2}^2 \frac{x^3y^3}{3} \Big|_{-1}^1 dy \\
 &= \int_{-2}^2 \left( \frac{y^3}{3} + \frac{y^3}{3} \right) dy = \int_{-2}^2 \frac{2y^3}{3} dy \\
 &= \frac{y^4}{6} \Big|_{-2}^2 = \frac{16}{6} - \frac{16}{6} = 0
 \end{aligned}$$

Just to check, we could change the order.

$$\begin{aligned}
 \int_I f(x, y) dA &= \int_{-1}^1 \int_{-2}^2 x^2y^3 dy dx \\
 &= \int_{-1}^1 \frac{x^2y^4}{4} \Big|_{-2}^2 dx \\
 &= \int_{-1}^1 (2x^2 - 2x^2) dx = \int_{-1}^1 0 dx = 0
 \end{aligned}$$

We notice that the second calculation is actually a bit simpler, since we find the zero term earlier in the process. Often, one order of variables leads to an easier calculation. Fubini's theorem lets us chose any order, so we try to chose the order that has the easiest calculations. Of course, the right choice isn't always obvious before we start doing the calculations.

**Example 2.3.4.** Consider the function  $f(x, y) = \sin(x + y)$  on the interval  $I = [0, \pi] \times [0, \pi/2]$ .

$$\begin{aligned} \int_I f(x, y) dx dy &= \int_0^{\pi/2} \int_0^{\pi} \sin(x + y) dx dy \\ &= \int_0^{\pi/2} (-\cos(x + y)) \Big|_0^{\pi} dy \\ &= \int_0^{\pi/2} (\cos y - \cos(y + \pi)) dy = \int_0^{\pi/2} 2 \cos y dy \\ &= 2 \sin y \Big|_0^{\pi/2} = 2 - 0 = 2 \end{aligned}$$

**Example 2.3.5.** If we have a integrable function on  $I = [a, b] \times [c, d]$  with the special form  $f(x, y) = g(x)h(y)$  for some single variable functions  $g$  and  $h$ , then we call  $f(x, y)$  seperable. The integral seperates entirely into two single variable integrals.

$$\begin{aligned} \int_I f(x, y) dx &= \int_a^b \int_c^d g(x)h(y) dy dx \\ &= \int_a^b g(x) \left[ \int_c^d h(y) dy \right] dx \\ &= \left[ \int_c^d h(y) dy \right] \left[ \int_a^b g(x) dx \right] \end{aligned}$$

The operations above are justified because the pieces which we move in and out of a integral do not involve that variable of integration. In this way, we can seperate the integral into two single variable integrals. This kind of seperation of variables is a very important technique in differential equations.

**Example 2.3.6.** Consider the function  $f(x, y) = \frac{1}{\sqrt{(x^2+y^2)^3}}$  on the interval  $I = [0, 1] \times [0, 1]$ . We would like to try to integrate this function over this interval, but there is a problem. At one corner of the interval,  $(0, 0)$ , the function is undefined. As we approach that corner, the function diverges to infinity. Is the integral still defined? We can use a limit in the bounds, in the last step, to investigate.

$$\begin{aligned} \int_I f(x, y) dx &= \int_0^1 \int_0^1 \frac{1}{\sqrt{(x^2 + y^2)^3}} dx dy \\ &= \int_0^1 \frac{x}{y^2 \sqrt{x^2 + y^2}} \Big|_0^1 dy \\ &= \int_0^1 \frac{1}{y^2 \sqrt{1 + y^2}} dy \\ &= \lim_{a \rightarrow 0^+} \int_a^1 \frac{1}{y^2 \sqrt{1 + y^2}} dy \\ &= \lim_{a \rightarrow 0^+} \frac{-\sqrt{1 + y^2}}{y} \Big|_a^1 = \frac{-\sqrt{2}}{1} + \lim_{a \rightarrow 0^+} \frac{\sqrt{1 + a^2}}{a} = \infty \end{aligned}$$

This integral diverges; the volume under the curve grows to infinity as we include more and more of the interval approaching the point  $(0, 0)$ . This is an improper integral, where the function has an asymptote at the edge of the domain of integration. In general, we treat improper integrals as we did before: we use limits approaching the edges of the domain and calculate the limit of the integral.

## 2.4 Integration over Arbitrary Sets

**Example 2.4.1.** Consider a geometrically interesting example. The function  $f(x, y) = a - \frac{a(|x-y|+|x+y|)}{2b}$  over the interval  $[-b, b] \times [-b, b]$  describes a square pyramid with height  $a$  and side length  $2b$ . What is the volume of such a pyramid? Doing the integral directly is difficult with the various evaluations of the absolute value terms. However, we can integrate over one quarter of pyramid with the interval  $[0, b] \times [0, b]$ , then multiply by 4 to get the entire area. However, this doesn't solve the problem with the absolute value terms. We need to split the domain into regions where  $x > y$  and  $x < y$ . In the interval  $[0, b] \times [0, b]$ , those two regions are triangles, and we don't know how to deal with integration over a triangle.

Integration over intervals isn't sufficient; we need to integrate over any set  $S \subset \mathbb{R}^n$  (at least, any reasonable set). How do we do this? As with many aspects of calculus, there is a formal, proper definition which turns out to be difficult to work with. After the definition, we seek set of techniques which allow us to do calculations.

**Definition 2.4.2.** Let  $S$  be a subset of  $\mathbb{R}^n$ . The *characteristic function* of  $S$  is defined as follows, where  $v \in \mathbb{R}^n$ .

$$\chi_S(v) := \begin{cases} 1 & v \in S \\ 0 & v \notin S \end{cases}$$

The characteristic function is very simple: it has value 1 on the set  $S$  and 0 everywhere else.

**Definition 2.4.3.** Let  $S$  be a subset of  $\mathbb{R}^n$  which is contained in an interval  $I$ .  $S$  is called a *integrable set* if the following integral exists.

$$\int_I \chi_S dV$$

Almost all sets we will work with are integrable. Sets which are not integrable are very strange sets (at least, for doing anything geometric). The archetypical example is  $\mathbb{Q} \subset \mathbb{R}$ :  $\mathbb{Q}$  is not an integrable subset of  $\mathbb{R}$ . As we proceed, we won't worry too much about integrability since all reasonable geometric sets are integrable. In particular, every open subset of  $\mathbb{R}^n$  is integrable.

**Definition 2.4.4.** Let  $S$  be a integrable subset of  $\mathbb{R}^n$  with  $S \subset I$  for some interval  $I$ . Let  $f : S \rightarrow \mathbb{R}$  be an function which can be extended to an integrable function  $\tilde{f}$  on  $I$ . Then the integral of  $f$  over  $S$  is defined as:

$$\int_S f dV = \int_I \tilde{f} \chi_S dV$$

The characteristic function removes all values of  $f$  outside the set  $S$ , so all that the integral measure are the values of  $f$  over  $S$ .

As we pointed out before, this is almost entirely useless for calculation. We need a better technique; a version of iterated integrals for arbitrary subsets.

**Example 2.4.5.** Let's return to the square pyramid situation. We wanted to integrate over two triangles in the interval  $[0, b] \times [0, b]$ , separated by the line  $x = y$ . We're going to try for a similar strategy as for intervals: we want to interate single variable integrals. However, now we allow the bounds of integration to include the other variables.

Specifically, we can describe the first triangle (where  $y < x$ ) by saying that  $x \in [0, b]$ , then, once we've set an  $x$  value,  $y \in [0, x]$ . In this order, the  $y$  variable depends on the  $x$  for its bounds. The integral over this triangle is now an iterated integral.

$$\int_I f(x, y) dA = \int_0^b \int_0^x f(x, y) dy dx$$

We could have reversed the order. If we have  $y \in [0, b]$  then the condition  $y < x$  means that  $x \in [y, b]$ .

$$\int_I f(x, y) dA = \int_0^b \int_y^b f(x, y) dx dy$$

Let's now evaluate the square pyramid integral. The function was  $f(x, y) = a - \frac{a(|x-y|+|x+y|)}{2b}$ . On the triangle in  $[0, b] \times [0, b]$  where  $y < x$ , the absolute values are  $|x - y| = x - y$  and  $|x + y| = x + y$ , so the function is  $a - \frac{a(x-y+x+y)}{2b} = a - \frac{2ax}{2b} = a - \frac{ax}{b}$ .

$$\begin{aligned} \int_T f(x, y) dA &= \int_0^b \int_0^x a - \frac{ax}{b} dy dx \\ &= \int_0^b ay - \frac{axy}{b} \Big|_0^x dx \\ &= \int_0^b ax - \frac{ax^2}{b} dx \\ &= \frac{ax^2}{2} - \frac{ax^3}{3b} \Big|_0^b \\ &= \frac{ab^2}{2} - \frac{ab^3}{3b} = \frac{ab^2}{2} - \frac{ab^2}{3} = \frac{ab^2}{6} \end{aligned}$$

This measures one eighth of the total pyramid, so the total volume is  $\frac{4ab^2}{3}$ .

**Example 2.4.6.** This technique also allows us to derive the volume of a sphere. We can think of the top half of a sphere as the volume under the graph of  $f(x, y) = \sqrt{r^2 - x^2 - y^2}$ . However, it is only the volume over the circle  $x^2 + y^2 = r^2$  in the  $xy$  plane. If we take the quarter of that circle in the positive quadrant, let  $D$  be the region where  $x \in [0, r]$  and  $y \in [0, \sqrt{r^2 - x^2}]$ . One eighth of the sphere is measured by the following iterated integral.

$$\begin{aligned} \int_D f(x, y) dA &= \int_0^r \int_0^{\sqrt{r^2 - x^2}} \sqrt{r^2 - x^2 - y^2} dy dx \\ &= \int_0^r \left( \frac{y}{2} \sqrt{r^2 - x^2 - y^2} + \frac{(r^2 - x^2)}{2} \arcsin \left( \frac{y}{\sqrt{r^2 - x^2}} \right) \right) \Big|_0^{\sqrt{r^2 - x^2}} dx \\ &= \int_0^r \frac{\sqrt{r^2 - x^2}}{2} \sqrt{r^2 - x^2 - (r^2 - x^2)} + \frac{(r^2 - x^2)}{2} \arcsin \left( \frac{\sqrt{r^2 - x^2}}{\sqrt{r^2 - x^2}} \right) dx \\ &= \int_0^r \frac{r^2 - x^2}{2} \frac{\pi}{2} dx = \int_0^r \frac{\pi(r^2 - x^2)}{4} dx \\ &= \frac{\pi r^2 x}{4} - \frac{\pi x^3}{12} \Big|_0^r \\ &= \frac{\pi r^3}{4} - \frac{\pi r^3}{12} = \frac{\pi r^3}{6} \end{aligned}$$

Multiplying by 8 gives  $\frac{8\pi r^3}{6} = \frac{4\pi r^3}{3}$ , which is the familiar expression for the volume of a sphere. Of course, we could have reversed the order and taken  $y \in [0, r]$  and  $x \in [0, \sqrt{r^2 - y^2}]$  and repeated very similar steps to also get the same answer.

**Example 2.4.7.** We can also calculate the volume of a cone of height  $h$  and radius  $r$ . We integrate over the same circle of radius  $r$ , but the function is  $f(x, y) = h - \frac{h}{r} \sqrt{x^2 + y^2}$ . (This function can be derived by looking at similar triangles in vertical sections of the cone.)

$$\int_D f(x, y) dA = 4 \int_0^r \int_0^{\sqrt{r^2 - x^2}} \left( h - \frac{h}{r} \sqrt{x^2 + y^2} \right) dy dx$$

The cone integral is quite difficult to do directly; for now, we'll leave it. We will return to this integral when we do change of variables.

The general form of an iterated integral starts with constant bounds on the outside variable. Then inside bounds can include functions of variables from the outside integrals. If  $g$  and  $h$  are functions, we have two general forms in  $\mathbb{R}^3$ .

$$\begin{aligned} \int_a^b \int_{g(x)}^{h(x)} f(x, y) dy dx \\ \int_a^b \int_{g(y)}^{h(y)} f(x, y) dx dy \end{aligned}$$

Likewise, in  $\mathbb{R}^3$  we have three general forms (for functions which satisfy  $g_1 \leq h_1$  and  $g_2 \leq h_2$ ).

$$\begin{aligned} & \int_a^b \int_{g_1(x)}^{h_2(x)} \int_{g_2(x,y)}^{h_2(x,y)} f(x, y, z) dz dy dx \\ & \int_a^b \int_{g_1(x)}^{h_2(x)} \int_{g_2(x,z)}^{h_2(x,z)} f(x, y, z) dy dz dx \\ & \int_a^b \int_{g_1(y)}^{h_2(y)} \int_{g_2(x,y)}^{h_2(x,y)} f(x, y, z) dz dx dy \\ & \int_a^b \int_{g_1(y)}^{h_2(y)} \int_{g_2(y,z)}^{h_2(y,z)} f(x, y, z) dx dz dy \\ & \int_a^b \int_{g_1(z)}^{h_2(z)} \int_{g_2(x,z)}^{h_2(x,z)} f(x, y, z) dy dx dz \\ & \int_a^b \int_{g_1(z)}^{h_2(z)} \int_{g_2(y,z)}^{h_2(y,z)} f(x, y, z) dx dy dz \end{aligned}$$

**Example 2.4.8.** To integrate over an eighth of a solid sphere of radius  $r$  in  $\mathbb{R}^3$ , we might take  $x \in [0, r]$ ,  $y \in [0, \sqrt{r^2 - x^2}]$  and  $z \in [0, \sqrt{r^2 - x^2 - y^2}]$ .

$$V = \int_0^r \int_0^{\sqrt{r^2 - x^2}} \int_0^{\sqrt{r^2 - x^2 - y^2}} f(x, y, z) dz dy dx$$

This gives another way to calculate the volume of a sphere: just integrate 1 over this region (then multiply by 8).

$$\begin{aligned} V &= \int_D 1 dV = \int_0^r \int_0^{\sqrt{r^2 - x^2}} \int_0^{\sqrt{r^2 - x^2 - y^2}} 1 dz dy dx \\ &= \int_0^r \int_0^{\sqrt{r^2 - x^2 - y^2}} \sqrt{1 - x^2 - y^2} dy dx \end{aligned}$$

The understanding of volume in the previous example is archetypical.

**Definition 2.4.9.** Let  $S$  be an integrable set in  $\mathbb{R}^n$ . Then the size (area, volume, hypervolume) of  $S$  is *defined* to be

$$\int_S 1 dV.$$

Note that this is a *definition*: we have no other way, in general, to understand the size of sets. This definition is the start of another whole branch of mathematics called measure theory. It deals with various ways of measuring sizes of sets in topological spaces.

We mentioned earlier that all open sets are integrable. We have another simplifying theorem that ensure we only need to integrate over open sets.

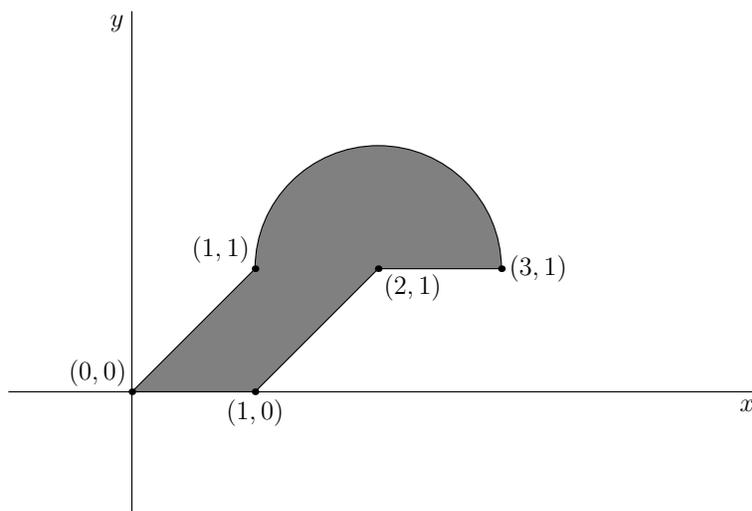


Figure 2.1: Region of Integration

**Proposition 2.4.10.** *Let  $S$  be an integrable set in  $\mathbb{R}^n$ . Let  $S^0$  be the set of interior points of  $S$ . Let  $f : S \rightarrow \mathbb{R}$  be an integrable function. Then*

$$\int_S f dV = \int_{S^0} f dV$$

This proposition means that we can ignore any inconveniences which happen on the boundaries of sets. (Note that we still do need limits for improper integrals, when the value of a function approaches  $\infty$  near the boundary of a set). The proposition also means that sets which are essentially of lower dimensions do not matter in integration. If we are integrating over subsets of  $\mathbb{R}^3$ , then integrating over points, line and planes amounts to nothing. Points, lines, and planes don't have any interior points in  $\mathbb{R}^3$ , so we might as well be integrating over the empty set, which gives zero.

Other than ignoring boundaries and smaller dimensional pieces, our approach to general sets is to break them up into pieces. It is a general result in topology, which we won't get into here, that any (reasonable) set can be broken up in to pieces (possibly infinitely many) where the previous style of integral applies. We can also be creative: for example, we could integrate over a larger circle and subtract the integral over a small circle to integrate over a ring-shaped domain. In these various ways we can set up integrals over any (reasonable) subset of  $\mathbb{R}^n$ .

**Example 2.4.11.** Say we wanted to integrate the function  $f(x, y) = xy$  over the triangle  $T$  with vertices  $(0, 0)$ ,  $(4, 5)$  and  $(6, 2)$ . There isn't any one system of iterated integrals that covers this region; therefore we have to treat it as two sections. We divide the triangle into two pieces using the line  $x = 4$ . On the left, the  $y$  coordinate is bounded by the lines  $y = x/3$  and  $y = 5x/4$ . On the right, the  $y$  coordinate is bounded by  $y = x/3$  and  $y = -3x/2 + 11$ . We calculate two integrals.

$$\begin{aligned}
 \int_0^4 \int_{\frac{x}{3}}^{\frac{5x}{4}} xy dy dx &= \int_0^4 \frac{xy^2}{2} \Big|_{\frac{x}{3}}^{\frac{5x}{4}} dx \\
 &= \int_0^4 \frac{x}{2} \left( \frac{25x^2}{16} - \frac{x^2}{9} \right) dx \\
 &= \int_0^4 \frac{25x^3}{32} - \frac{x^3}{18} = \int_0^4 \frac{209}{288} x^3 dx \\
 &= \frac{209}{288} \frac{x^4}{4} \Big|_0^4 = \frac{1196}{9} \\
 \int_4^6 \int_{\frac{x}{3}}^{-\frac{3x}{2}-11} xy dy dx &= \int_4^6 \frac{xy^2}{2} \Big|_{\frac{x}{3}}^{-\frac{3x}{2}-11} dx \\
 &= \int_4^6 \frac{x}{2} \left( \frac{-3x}{2} + 11 \right)^2 - \frac{x}{2} \frac{x^2}{9} dx \\
 &= \int_4^6 \frac{x}{2} \left( \frac{9x^2}{4} - 32x + 121 \right) - \frac{x^3}{18} dx \\
 &= \int_4^6 \frac{77x^3}{81} - \frac{33x^2}{2} + \frac{121x}{2} dx \\
 &= \frac{77x^4}{324} - \frac{11x^3}{2} + \frac{121x^2}{4} \Big|_4^6 \\
 &= \frac{77 \cdot 6^4}{324} - \frac{11 \cdot 6^3}{2} + \frac{121 \cdot 6^2}{4} - \frac{77 \cdot 4^4}{324} + \frac{11 \cdot 4^3}{2} - \frac{121 \cdot 4^2}{4} = \frac{86845}{81}
 \end{aligned}$$

The total is the sum of the two integrals.

$$\int_T f dA = \frac{1196}{9} + \frac{86845}{81} = \frac{97609}{81}$$

**Example 2.4.12.** Say we want to integrate the function  $f(x, y) = x^2 + y^2$  over the region shown in Figure 2.1.

We should divide the figure up into three pieces. The first piece is the triangle  $(0, 0)$ ,  $(1, 1)$ , and  $(1, 0)$ . The second is the half-circle above  $y = 1$ . The third is the remaining triangle  $(1, 1)$ ,  $(2, 1)$  and  $(1, 0)$ .

The first integral is a short calculation.

$$\begin{aligned}
 \int_0^1 \int_0^x x^2 + y^2 dy dx &= \int_0^1 x^2 y + \frac{y^3}{3} \Big|_0^x dx \\
 &= \int_0^1 \frac{4x^3}{3} = \frac{x^4}{3} \Big|_0^1 = \frac{1}{3}
 \end{aligned}$$

The second integral is more involved.

$$\begin{aligned}
& \int_1^3 \int_1^{1+\sqrt{1-(x-2)^2}} x^2 + y^2 dy dx \\
&= \int_1^3 x^2 y + \frac{y^3}{3} \Big|_1^{1+\sqrt{1-(x-2)^2}} \\
&= \int_1^3 x^2(1 + \sqrt{1-(x-2)^2}) + \frac{(1 + \sqrt{1-(x-2)^2})^3}{3} dx \\
&u = x - 2 \\
&= \int_{-1}^1 (u^2 + 4u + 4)(1 + \sqrt{1-u^2}) + \frac{(1 + \sqrt{1-u^2})^3}{3} du \\
&= \int_{-1}^1 u^2 + 4u + 4 + u^2\sqrt{1-u^2} + 4u\sqrt{1-u^2} + 4\sqrt{1-u^2} \\
&\quad + \frac{1}{3} + \sqrt{1-u^2} + (1-u^2) - \frac{u^2}{3}\sqrt{1-u^2} du \\
&= \int_{-1}^1 \frac{16}{3} + 4u + 5\sqrt{1-u^2} + 4u\sqrt{1-u^2} + u^2\sqrt{1-u^2} + \frac{1}{3}(1-u^2)^{\frac{3}{2}} du \\
&= \frac{16u}{3} + 2u^2 + 4(1-u^2)^{\frac{3}{2}} \frac{-2}{3} \Big|_{-1}^1 + 5 \int_{-1}^1 1\sqrt{1-u^2} du + \int_{-1}^1 u^2\sqrt{1-u^2} du + \frac{1}{3} \int_{-1}^1 (1-u^2)^{\frac{3}{2}} du \\
&= \frac{32}{5} + 0 + 0 + \frac{1}{8} 2\sqrt{1-u^2}(2u^2-1) + \arcsin u \Big|_{-1}^1 + 5(u\sqrt{1-u^2} + \arcsin u) \Big|_{-1}^1 \\
&\quad + \frac{1}{3} \frac{1}{8} (u(5-2u^2)\sqrt{1-u^2} + 3\arcsin u) \Big|_{-1}^1 \\
&= \frac{32}{3} + \frac{1}{8} \left( \frac{\pi}{2} - \frac{-\pi}{2} \right) + 5 \left( \frac{\pi}{2} - \frac{-\pi}{2} \right) + \frac{1}{8} \left( \frac{\pi}{2} - \frac{-\pi}{2} \right) \\
&= \frac{32}{3} + \frac{\pi}{8} + 5\pi + \frac{\pi}{8} = \frac{32}{3} + \frac{21\pi}{4}
\end{aligned}$$

The third integral is not quite as bad as the second.

$$\begin{aligned}
 \int_1^2 \int_{x-1}^1 x^2 + y^2 dy dx &= \int_1^2 x^2 y + \frac{y^3}{3} \Big|_{x-1}^1 dx \\
 &= \int_1^2 x^2 + \frac{1}{3} - x^2(x-1) - \frac{1}{3}(x-1)^3 dx \\
 &= \int_1^2 \frac{2}{3} - x + 3x^2 - \frac{4x^3}{3} dx \\
 &= \frac{2x}{3} - \frac{x^2}{2} + x^3 - \frac{x^4}{3} \Big|_1^2 \\
 &= \frac{4}{3} - 2 + 8 - \frac{16}{3} - \frac{2}{3} + \frac{1}{2} - 1 + \frac{1}{3} \\
 &= \frac{7}{6}
 \end{aligned}$$

The total is the sum of the three integrals.

$$\frac{1}{3} + \frac{32}{3} + \frac{21\pi}{4} + \frac{7}{6} = \frac{73}{6} + \frac{21\pi}{4}$$

**Example 2.4.13.** Here is an odd application of multiple integration. We know that  $e^{-x^2}$  has no elementary antiderivative. Therefore, the integral

$$A = \int_{-\infty}^{\infty} e^{-x^2} dx$$

cannot be evaluated directly. However, this is a very important integral:  $e^{-x^2}$  is the normal distribution and, in Statistics, we need to integrate it frequently. We'll use integrals over  $\mathbb{R}^2$ , strangely enough, to calculate this integral by squaring the single variable integral.

$$\begin{aligned}
 A &= 2 \int_0^{\infty} e^{-x^2} dx \\
 A^2 &= \left( 2 \int_0^{\infty} e^{-x^2} dx \right) \left( 2 \int_0^{\infty} e^{-y^2} dy \right)
 \end{aligned}$$

The second integral uses a new variables since variables of integration only matter inside the integral. Then we can combine the two integrals.

$$\begin{aligned}
 A^2 &= 4 \int_0^{\infty} \int_0^{\infty} e^{-x^2} e^{-y^2} dx dy \\
 &= 4 \int_0^{\infty} \int_0^{\infty} e^{-(x^2+y^2)} dx dy
 \end{aligned}$$

Now we are going to do a substitution in the  $y$  variable. Treating the  $x$  variable as a constant, we replace  $y$  with  $y = xs$  so that  $dy = xds$ . If  $y = 0$  then  $s = 0$  and as  $y \rightarrow \infty$ ,  $s \rightarrow \infty$ , so the bounds for  $s$  remain the same as the bounds for  $y$ . Remember,  $x$  is a constant through this whole substitution.

$$\begin{aligned}
 A^2 &= 4 \int_0^\infty \int_0^\infty -e^{-x^2(1+s^2)} x dx ds \\
 &= 4 \int_0^\infty \left. \frac{-1}{2(1+s^2)} e^{-x^2(1+s^2)} \right|_0^\infty ds \\
 &= 2 \int_0^\infty \frac{1}{1+s^2} ds \\
 &= \lim_{a \rightarrow \infty} 2 \arctan a - \arctan 0 = \frac{2\pi}{2} = \pi \\
 A &= \int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}
 \end{aligned}$$

We recover the area under the bell curve:  $\sqrt{\pi}$ . It's a very strange result. However, if you taken any statistics and worked on normal distributions, likely you will recall the presence of these strange  $\sqrt{\pi}$  terms. Now we know they are present to normalize the area (since we want a probability function to have area one under its graph).

## 2.5 Change of Variables

Substitution was very useful for single variable integration. In the  $e^{-x^2}$  example, we did single-variable substitution in  $y$ , treating  $x$  as a constant. One might think that since we do iterated integrals, single variable substitution is sufficient. However, treating substitution in several variables holistically turns out to be very useful and powerful.

Let's review single variable substitution, formalizing the language and showing the way forward for higher dimensions. Consider a single variable integral.

$$\int_{x=a}^{x=b} f(x) dx$$

Typically, we substitute  $u = g(x)$  using some *invertible* function  $g$ . Then we have  $du = g'(x)dx$  and we try to change all the  $x$  variables to  $u$  variables with some algebra. We also change the bounds, starting from  $u = g(a)$  to  $u = g(b)$ . In this process, the old variable is the *domain* of the transformation and the new variable is the *range*.

However, for several variables we're going other direction. We want the starting variables  $x_i$  to be the range. We do have one example from single variable calculus in this style; trigonometric substitution, such as  $x = a \sin \theta$ , have the original variable as the output of the transformation.

Therefore, in the single variable case, introduce a new variable  $u$  by  $x = h(u)$  for some *invertible* function  $h$ . Then  $dx = h'(u)du$  and the bounds become  $h^{-1}(a)$  and  $h^{-1}(b)$ . We get a new integral in the variable  $u$ .

$$\int_{u=h^{-1}(a)}^{u=h^{-1}(b)} f(h(u))h'(u)du$$

The key relationship here is  $dx = h'(u)du$ : the derivative tells us the relationships between the differential terms. In Calculus III, we did all that work on extending the notion of the derivative. We came to the conclusion that the derivative measures (in local coordinates) the best linear approximation to the function. In one variable, that was just multiplication by a number. In several variables, though, that linear approximation was a matrix, called the Jacobian matrix. Let's recall the definition.

**Definition 2.5.1.** Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a function on  $n$  variables,  $x_1, \dots, x_n$ . We can write  $F$  as its component functions  $F_1, \dots, F_n$ . Then the Jacobian Matrix of  $F$  is the matrix of all partial derivatives.

$$J(F) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \frac{\partial F_n}{\partial x_2} & \cdots & \frac{\partial F_n}{\partial x_n} \end{pmatrix}$$

The determinant  $\det J(F)$  is called the Jacobian of the function and is often written  $|J(F)|$ . Note that the notation  $|\cdot|$  is a reminder to take a determinant; it is *not* an absolute value. The Jacobian is often negative.

From the single variable case  $dx = h'(u)du$ , the Jacobian is part of the new differential  $du$ . If we start with variables  $x_1, x_2, \dots, x_n$  and we have a function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  that has  $x_i$  as its output (the original variables are the range of the transformation), we can write  $x_1 = F_1(u_1, \dots, u_n)$ ,  $x_2 = F_2(u_1, \dots, u_n)$  up to  $x_n = F_n(u_1, \dots, u_n)$ . The Jacobian determines the relationship between the differential in these variables.

$$dx_1 dx_2 \dots dx_n = |J(F)| du_1 du_2 \dots du_n$$

We can interpret the Jacobian as the change in area/volume/hyper-volume due to the change in variables. Its appearance in the integral makes sense with this interpretation: the integral is measuring area/volume/hyper-volume, so when we change variables, we need a term that keeps track of the relative change in area/volume/hypervolume.

**Example 2.5.2.** Let  $(x, y) = F(u, v) = (3u, 4v)$ . Then we have

$$\frac{\partial F_1}{\partial u} = 3$$

$$\frac{\partial F_1}{\partial v} = 0$$

$$\frac{\partial F_2}{\partial u} = 0$$

$$\frac{\partial F_2}{\partial v} = 4$$

$$J(F) = \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}$$

$$|J(F)| = (3)(4) - (0)(0) = 12$$

$$dxdy = 12dudv$$

The function is a dialation by 3 in  $u$  and by 4 in  $v$ , so the effect on area is multiplication by 12, which makes sense. The function is linear already, to the linear approximation is constant: it is exactly the function itself.

**Example 2.5.3.** Let  $(x, y) = F(u, v) = (u^2, v)$ .

$$\frac{\partial F_1}{\partial u} = 2u$$

$$\frac{\partial F_1}{\partial v} = 0$$

$$\frac{\partial F_2}{\partial u} = 0$$

$$\frac{\partial F_2}{\partial v} = 1$$

$$J(F) = \begin{pmatrix} 2u & 0 \\ 0 & 1 \end{pmatrix}$$

$$|J(F)| = 2u$$

$$dxdy = 2ududv$$

This Jacobian isn't constant. This is a stretch in  $u$ , the but effect is exagerated away from the origin due to the square term. Therefore, as  $u$  get larger, the stretch effect is greater. The Jacobian reflects that.

**Example 2.5.4.** Let  $(x, y) = F(r, \theta) = (r \cos \theta, r \sin \theta)$ . (Polar coordinates!)

$$\frac{\partial F_1}{\partial r} = \cos \theta$$

$$\frac{\partial F_1}{\partial \theta} = -r \sin \theta$$

$$\frac{\partial F_2}{\partial r} = \sin \theta$$

$$\frac{\partial F_2}{\partial \theta} = r \cos \theta$$

$$J(F) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$$

$$|J(F)| = r$$

$$dxdy = r dr d\theta$$

This Jacobian shows that the radius term gives the effect on area. This make sense: for larger circles, the differential area is part of a larger arc.

**Example 2.5.5.** Let  $(x, y) = F(u, v) = (u, uv)$ . This is the change in variables we used in the  $e^{-x^2}$  example, now treated as a two-variable changes of variables.

$$\frac{\partial F_1}{\partial u} = 1$$

$$\frac{\partial F_1}{\partial v} = 0$$

$$\frac{\partial F_2}{\partial u} = v$$

$$\frac{\partial F_2}{\partial v} = u$$

$$J(F) = \begin{pmatrix} 1 & 0 \\ v & u \end{pmatrix}$$

$$|J(F)| = u$$

$$dxdy = u du dv$$

We can use a change of variables to help us with either the integrand or the region of integration. Sometime, as in the following examples, the same transformation helps both. When we are choosing a change of variables to simplify the region of integration, we are looking for a transformation where some of the bounding curves of the region become  $u = c$  or  $v = c$ . We need constant bounds in at least one of the variable to set up the iterated integral.

**Example 2.5.6.** Let  $D$  be the region in the first quadrant between the lines  $x + y = 1$  and the line  $x + y = 2$ . In the following integral, we want to simply the integrand and integrate over a simpler region if possible.

$$\int_D \cos \left( \frac{y-x}{y+x} \right) dA$$

To simplify the integrand, we can look for transformations where  $u = x + y$  and  $v = x - y$ . Inverting these (either by linear algebra, since the transformation is linear, or just by solving with conventional algebra), we get  $x = \frac{u+v}{2}$  and  $y = \frac{u-v}{2}$ .

Then we look at the bounds. If we take  $x + y = 1$  and replace  $x$  and  $y$ , we get  $u = 1$ . Likewise,  $x + y = 2$  is  $u = 2$ . This is excellent, since we need at least one variable with constant bounds; we'll take  $u$  to be the variable of the outside integral. The other bounds are the axes, which we can express by the equations  $y = 0$  and  $x = 0$ . These turn into  $u = v$  and  $u = -v$ , respective. Therefore, the bounds of  $v$  will be these two lines:  $v \in [u, -u]$ .

Finally, we calculate the Jacobian.

$$\begin{aligned}x_u &= \frac{1}{2} \\x_v &= \frac{1}{2} \\y_u &= \frac{1}{2} \\y_v &= -\frac{1}{2} \\|J| &= -\frac{1}{2} \\dxdy &= \frac{1}{-2} dudv\end{aligned}$$

We have the new integrand, the new bounds and the Jacobian. We can proceed to the new integral.

$$\begin{aligned}\int_D \cos\left(\frac{y-x}{y+x}\right) dA &= \int_1^2 \int_{-u}^u \cos\left(\frac{-v}{u}\right) \frac{-1}{2} dv du \\&= \frac{-1}{2} \int_1^2 -\sin\left(\frac{-v}{u}\right) (-u) \Big|_{-u}^u du \\&= \frac{-1}{2} \int_1^2 u(\sin(-1) - \sin(1)) du \\&= \frac{\sin(1) - \sin(-1)}{2} \frac{u^2}{2} \Big|_1^2 \\&= \frac{\sin(1) - \sin(-1)}{2} \left(2 - \frac{1}{2}\right) = \frac{3(\sin(1) - \sin(-1))}{4}\end{aligned}$$

**Example 2.5.7.** Let  $D$  be the region in the plane bounded by the curves  $y = \frac{1}{x}$ ,  $y = \frac{3}{x}$ ,  $y = 3x$  and  $y = x$ . Consider the following integral.

$$\int_D xy dA$$

The integrand is reasonable, so we will chose a substitution to changes the bounds. We want constant bounds in at least one variables. Here is one changes of variables which gets those constant bound:  $(x, y) = F(u, v) = (uv, v)$ . The lines  $y = 3x$  and  $y = x$  become  $v = 3uv$  and  $v = uv$ . Away from  $v = 0$ , these lines are  $u = \frac{1}{3}$  and  $u = 1$ , so we have constant bounds in  $u$  and we will treat  $u$  as the outside variable. (We are save with the assumption  $v = 0$  since  $v = y$  and the region is disjoint from the  $x$ -axis, where  $y = 0$ ). The curves  $y = \frac{1}{x}$  and  $y = \frac{3}{x}$  become  $v = \frac{1}{uv}$  and  $v = \frac{3}{uv}$ , which simplify into  $uv^2 = 1$  and  $uv^2 = 3$ . If we solve for  $v$  we get  $v = \sqrt{\frac{1}{u}}$  and  $v = \sqrt{\frac{3}{u}}$ . That means we can take  $u \in [\frac{1}{3}, 3]$  and  $v \in [\sqrt{\frac{1}{u}}, \sqrt{\frac{3}{u}}]$ . The integrand  $xy$  becomes  $uv^2$ .

$$\begin{aligned}x_u &= v \\x_v &= u \\y_u &= 0 \\y_v &= 1 \\|J| &= (v)(1) - (u)(0) = v \\dxdy &= v dudv\end{aligned}$$

Then we complete the change of variables.

$$\begin{aligned}\int_D xy dA &= \int_{\frac{1}{3}}^1 \int_{\sqrt{\frac{1}{u}}}^{\sqrt{\frac{3}{u}}} uv^2 v dudv \\&= \int_{\frac{1}{3}}^1 u \frac{v^4}{4} \Big|_{\sqrt{\frac{1}{u}}}^{\sqrt{\frac{3}{u}}} du \\&= \int_{\frac{1}{3}}^1 \frac{u}{4} \left( \frac{9}{u^2} - \frac{1}{u^2} \right) du \\&= \int_{\frac{1}{3}}^1 \frac{-2}{u} du \\&= 2 \ln |u| \Big|_{\frac{1}{3}}^1 \\&= 2 \ln 1 - 2 \ln \frac{1}{3} = -2 \ln \frac{1}{3} = 2 \ln 3\end{aligned}$$

For the same integral, we could have taken  $y = v$  and  $x = \frac{u}{v}$ . Under this transformation, the bounds become  $u \in [1, 3]$  and  $v \in [\sqrt{u}, \sqrt{3u}]$ . The Jacobian is now  $\frac{1}{v}$  and the integrand is  $u$ .

$$\begin{aligned} \int_D xy dA &= \int_1^3 \int_{\sqrt{u}}^{\sqrt{3u}} \frac{1}{v} u dv du \\ &= \int_1^3 u \ln |v| \Big|_{\sqrt{u}}^{\sqrt{3u}} du \\ &= \int_1^3 u (\ln |\sqrt{3u}| - \ln |\sqrt{u}|) du \\ &= \int_1^3 \frac{1}{2} (u \ln 3u - u \ln u) du \\ &= \frac{1}{2} \left( \frac{1}{4} u^2 (2 \ln 3u - 1) \Big|_1^3 - \frac{1}{4} u^2 (2 \ln u - 1) \Big|_1^3 \right) \end{aligned}$$

If we evaluate this, it will evaluate, eventually, to  $2 \ln 3$ . However, the result is obviously more complicated. Different changes of variables can have very different effects on the integral; some will make it easier, some more difficult. Also, the answers from different change of variables can look quite different, such as the expression above, while actually being the same value.

As an aside, there is a simplification in the second version of this example that removes the complication. In the second last step, we could write  $\ln 3u$  as  $\ln 3 + \ln u$ .

$$\begin{aligned} &= \int_1^3 \frac{1}{2} (u \ln 3u - u \ln u) du \\ &= \frac{1}{2} \int_1^3 u (\ln 3 + \ln u - \ln u) du = \frac{1}{2} \int_1^3 u \ln 3 du \\ &= \frac{\ln 3}{2} \frac{u^2}{2} \Big|_1^3 = \frac{\ln 3}{4} (9 - 1) = \frac{8 \ln 3}{4} = 2 \ln 3 \end{aligned}$$

**Example 2.5.8.**

$$\begin{aligned} \int_I f(x, y) dA &= \int_0^2 \int_3^6 \frac{3x + y}{2x - y} dy dx \\ x &= \frac{u + v}{5} \\ y &= \frac{2u - 3v}{5} \\ J &= \begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{2}{5} & -\frac{3}{5} \end{pmatrix} \\ |J| &= \frac{-1}{5} \\ dy dx &= \frac{-1}{5} du dv \\ \frac{3x + y}{2x - y} &= \frac{3u + 3v + 2u - 3v}{2u + 2v - 2u + 3v} = \frac{5u}{5v} = \frac{u}{v} \end{aligned}$$

The change of variables does indeed simplify the integrand, but what about the region of integration? We know  $(x, y) \in [0, 2] \times [3, 6]$ . Looking at the boundary lines  $x = 0$  and  $x = 2$  in the changes of variables gives  $u + v = 0$  and  $u + v = 10$ . Likewise,  $y = 3$  and  $y = 6$  give  $2u - 3v = 15$  and  $2u - 3v = 30$ . The resulting shape is a parallelogram with vertices  $(9, 1)$ ,  $(12, -2)$ ,  $(6, -6)$  and  $(3, -3)$ . We would have to split up the domain of integration into three pieces to evaluate this; the result may be a much longer process than dealing with the original integral. Substitution may not always make things easier.

## 2.6 Polar Coordinates

We often use substitution in several variables to simplify the domain of integration instead of simplifying the function. Polar coordinates are ideals for a domain of integration which has any kind of circular symmetry: circles, wedges, arcs, etc.

The polar coordinate transformation is  $x = r \cos \theta$  and  $y = r \sin \theta$  with Jacobian  $|J| = r$  so that  $dx dy = r dr d\theta$ . (Notice this Jacobian also corrects units:  $dx dy$  has units of length squared, but  $\theta$  doesn't have any units, so  $r dr d\theta$  also has units of length squared.)

First, we need to understand what happens with constant bounds in polar coordinates. If  $r \in [0, R]$  and  $\theta \in [0, 2\pi]$ , we integrate over a whole circle of radius  $R$ . Integrating over a wedge with radius  $R$  from  $\theta_1$  to  $\theta_2$  is  $r \in [0, R]$  and  $\theta \in [\theta_1, \theta_2]$ . Integrating over an annulus with radii  $R_1$  and  $R_2$  is  $r \in [R_1, R_2]$  and  $\theta \in [0, 2\pi]$ . Finally, over an arc in such an annulus is  $r \in [R_1, R_2]$  and  $\theta \in [\theta_1, \theta_2]$ .

**Example 2.6.1.** Consider the function  $f(x, y) = x + y$  on the arc in the first quadrant between radii 2 and 4.

$$\begin{aligned} \int_D x + y dA &= \int_0^{\pi/2} \int_2^4 r(\cos \theta + \sin \theta) r dr d\theta \\ &= \int_0^{\pi/2} \frac{r^3}{r} (\cos \theta + \sin \theta) \Big|_2^4 d\theta \\ &= \frac{56}{3} \int_0^{\pi/2} (\cos \theta + \sin \theta) d\theta \\ &= \frac{56}{3} (\sin \theta - \cos \theta) \Big|_0^{\pi/2} \\ &= \frac{56}{3} (1 - 0 - 0 + 1) = \frac{112}{3} \end{aligned}$$

**Example 2.6.2.** Now consider the function  $e^{x^2+y^2}$  on the circle of radius  $R$  centered at the origin. Notice here that the integrand also has circular symmetry – changing to polar coordinates will help the integrand as well as the domain.

The integral is annoying in conventional Cartesian coordinates.

$$\int_D f(x, y) dA = \int_{-R}^R \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} e^{x^2+y^2} dy dx$$

This is essentially impossible. However, in polar coordinates it improves greatly.

$$\begin{aligned}\int_D f(x, y) dA &= \int_0^{2\pi} \int_0^R e^{r^2} r dr d\theta \\ &= \int_0^{2\pi} d\theta \int_0^R e^{r^2} r dr \\ &= 2\pi \left. \frac{e^{r^2}}{2} \right|_0^R = \frac{2\pi}{2} (e^{R^2} - 1) = \pi(e^{R^2} - 1)\end{aligned}$$

**Example 2.6.3.** Now recall the integral we did to calculate the volume of a sphere of radius  $R$ .

$$8 \int_0^R \int_0^{\sqrt{R^2 - x^2}} \sqrt{R^2 - x^2 - y^2} dy dx$$

This was a tricky integral. If we change to polar coordinates, an integrate over the whole circle instead of just a quarter, the integral improves.

$$\begin{aligned}A &= 2 \int_0^{2\pi} \int_0^R \sqrt{R^2 - r^2} r dr d\theta \\ &= 2 \int_0^{2\pi} d\theta \int_0^R \sqrt{R^2 - r^2} r dr d\theta \\ &= 4\pi (R^2 - r^2)^{\frac{3}{2}} \left. \frac{2 - 1}{3} \right|_0^R \\ &= \frac{4\pi}{3} (R^2)^{\frac{3}{2}} = \frac{4\pi R^3}{3}\end{aligned}$$

**Example 2.6.4.** For the cone of height  $h$  and radius  $R$ , we had the following integral.

$$\int_D f(x, y) dA = 4 \int_0^r \int_0^{\sqrt{r^2 - x^2}} \left( h - \frac{h}{r} \sqrt{x^2 + y^2} \right) dy dx$$

We didn't even evaluate the cone volume integral before, but now we can use polar coordinates to make it much more accessible.

$$\begin{aligned}A &= \int_0^{2\pi} \int_0^R \left( h - \frac{hr}{R} \right) r dr d\theta \\ &= 2\pi \left( \frac{hr^2}{2} - \frac{hr^3}{3R} \right) \Big|_0^R \\ &= 2\pi \left( \frac{hR^2}{2} - \frac{hR^2}{3} \right) = \frac{\pi R^2 h}{3}\end{aligned}$$

**Example 2.6.5.** To show how some of these techniques work together, consider integrating the function  $f(x, y) = x^2$  over the ellipse  $D = \frac{x^2}{4} + \frac{y^2}{9} = 1$ .

$$\int_D x^2 dA$$

We're going to do two changes of variables. First we can take  $x = 2u$  and  $y = 3v$ . That has Jacobian  $|J| = 6$  so that  $dx dy = 6 du dv$ . The ellipse  $D$  becomes the unit circle  $C$ .

$$\int_C (2u)^2 6 du dv = 24 \int_C u^2 du dv$$

Then we change to polar coordinates to integrate over this circle.

$$\begin{aligned} 24 \int_C u^2 du dv &= 24 \int_0^{2\pi} \int_0^1 r^2 \cos^2 \theta r dr d\theta \\ &= 24 \int_0^{2\pi} \cos^2 \theta d\theta \int_0^1 r^3 dr \\ &= 24 \left( \frac{\theta}{2} + \frac{\sin 2\theta}{4} \right) \Big|_0^{2\pi} \left( \frac{r^4}{4} \right) \Big|_0^1 \\ &= 24\pi \frac{1}{4} = 6\pi \end{aligned}$$

**Example 2.6.6.** Say we wanted to integrate the function  $x^2 + y^2$  over the circle of radius 1 centered at  $(1, 0)$ . The integrand obviously lends itself to polar coordinates. There is circular symmetry in the domain, but the offset is confusing.

The equation of such a circle is  $(x - 1)^2 + y^2 = 1$ . In polar coordinates this is  $r^2 \cos^2 \theta - 2r \cos \theta + 1 + r^2 \sin^2 \theta = 1$ , which simplifies into  $r^2 - 2r \cos \theta = 0$  or  $r = 2 \cos \theta$  for  $\theta \in [-\pi/2, \pi/2]$ . We can take a constant bound for  $\theta$  and have  $\theta$  be the outside variable. Then we can use the expression  $r = 2 \cos \theta$  as an higher bound for  $r$  in terms of  $\theta$ .

$$\begin{aligned} \int_D f(x, y) dA &= \int_{-\pi/2}^{\pi/2} \int_0^{2 \cos \theta} \frac{1}{r^2} r dr d\theta \\ &= \int_{-\pi/2}^{\pi/2} \frac{r^4}{4} \Big|_0^{2 \cos \theta} d\theta \\ &= 4 \int_{-\pi/2}^{\pi/2} \cos^4 \theta d\theta \\ &= 4 \frac{1}{32} (12\theta + 8 \sin(2\theta) + \sin(4\theta)) \Big|_{-\pi/2}^{\pi/2} \\ &= \frac{1}{8} \left( 12 \left( \frac{\pi}{2} - \frac{-\pi}{2} \right) \right) + 0 + 0 = \frac{3\pi}{2} \end{aligned}$$

**Example 2.6.7.** What if  $D$  is the circle of radius 2 excluding the circle of radius 1 centered at  $(1, 0)$  and we want to integrate  $\sqrt{x^2 + y^2}$  over this region? Here, we can think of the entire circle of radius 2 as  $D_1$  and the removed circle as  $D_2$ ; then the integral will be the integral over  $D_1$  subtracting the integral over  $D_2$ .

The first part is an integral over  $D_1$ .

$$\begin{aligned}\int_{D_1} \sqrt{x^2 + y^2} dA &= \int_0^{2\pi} \int_0^2 r r dr d\theta \\ &= 2\pi \frac{r^3}{3} \Big|_0^2 = \frac{16\pi}{3}\end{aligned}$$

The second part is an integral over  $D_2$ .

$$\begin{aligned}\int_{D_2} \sqrt{x^2 + y^2} dA &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_0^{2\cos\theta} r^2 dr d\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{r^3}{3} \Big|_0^{2\cos\theta} d\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left( \frac{8\cos^3\theta}{3} \right) = \frac{8}{3} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^3\theta d\theta \\ &= \frac{8}{3} \left( \frac{1}{3}(2 + \cos^2\theta) \sin\theta \right) \Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{32}{9}\end{aligned}$$

The result of the original integral is the difference of the two parts.

$$\int_{D_1 \setminus D_2} \sqrt{x^2 + y^2} dA = \frac{16\pi}{3} - \frac{32}{9} = \frac{48\pi - 32}{9}$$

## 2.7 Cylindrical and Spherical Coordinates

### 2.7.1 Cylindrical Coordinates

Cylindrical coordinates leave  $z$  unchanged and use polar coordinates in the  $xy$ -plane.

$$\begin{aligned}x &= r \cos \theta \\ y &= r \sin \theta \\ z &= z\end{aligned}$$

The Jacobian is  $J = r$ , using a  $3 \times 3$  determinant. We look to use cylindrical coordinates whenever a region in  $\mathbb{R}^3$  has some kind of cylindrical shape or symmetry.

**Example 2.7.1.** Say the region  $D$  is defined by  $x^2 + y^2 \leq 4$  and  $1 \leq z \leq 5$ . This is a cylinder of radius 2 of height 4. Say we want to integrate the function  $f(x, y, z) = z\sqrt{x^2 + y^2}$  over this cylinder.

$$\begin{aligned}\int_D z\sqrt{x^2 + y^2} dV &= \int_0^{2\pi} \int_0^2 \int_1^5 z r r dz dr d\theta \\ &= \int_0^{2\pi} d\theta \int_0^2 r^2 dr \int_1^5 z dz \\ &= 2\pi \frac{r^3}{3} \Big|_0^2 \frac{z^2}{2} \Big|_1^5 = 2\pi \frac{24}{2} \frac{8}{3} = 64\pi\end{aligned}$$

**Example 2.7.2.** We could have set up the volume of a cone in cylindrical coordinates. If the cone has height  $h$  and base radius  $R$ , and we set it opening upwards with its point at the origin, then we have  $z \in [0, h]$  and the radius increases linearly with  $z$  as  $r = \frac{zR}{h}$ . In  $\mathbb{R}^3$ , we integrate 1 over the cone to calculate its volume. The bounds for  $z$  are non-constant, so  $z$  must be the inside variable.

$$\begin{aligned} \int_C 1dV &= \int_0^{2\pi} \int_0^h \int_0^{\frac{Rz}{h}} r dr dz d\theta \\ &= 2\pi \int_0^h \left. \frac{r^2}{2} \right|_0^{\frac{Rz}{h}} dz \\ &= 2\pi \int_0^h \frac{R^2 z^2}{2h^2} dz \\ &= 2\pi \left. \frac{R^2}{2h^3} \frac{z^3}{3} \right|_0^h \\ &= 2\pi \frac{R^2}{2h^3} \frac{h^3}{3} = \frac{R^3 h \pi}{3} \end{aligned}$$

**Example 2.7.3.** Similarly, we can calculate the volume of a paraboloid this way. We orient it along the  $z$  axis, opening upwards. If the paraboloid has height  $h$  and base radius  $R$ , then the radius is  $r = \frac{R\sqrt{z}}{\sqrt{h}}$ . By solving for  $z$ , we have  $z = \frac{hr^2}{R^2}$ . The bounds for  $z$  are non-constant, so  $z$  must be the inside integral.

$$\begin{aligned} \int_D 1dV &= \int_0^{2\pi} \int_0^R \int_{\frac{hr^2}{R^2}}^h r dz dr d\theta \\ &= 2\pi \int_0^R r \left( h - \frac{hr^2}{R^2} \right) dr \\ &= 2\pi \left( \frac{hr^2}{2} - \frac{hr^4}{4R^2} \right) \Big|_0^R \\ &= 2\pi \left( \frac{hR^2}{2} - \frac{hR^4}{4R^2} \right) \\ &= hR^2 2\pi \left( \frac{1}{2} - \frac{1}{4} \right) \\ &= \frac{hR^2 \pi}{2} \end{aligned}$$

**Example 2.7.4.** Assume we have three cylinders of radius  $R$  along each axis in  $\mathbb{R}^3$ . What is the volume of the intersection of all three cylinders? (This is called the tricylinder Steinmetz solid). Setting up the geometry is difficult. We will integrate and height over the  $xy$  plane, and by symmetry, we can work with one 16th of the shape: the portion that lies above one 8th of the circle above the axis. In the first 8th of the circle ( $\theta \in [0, \pi/4]$ ) in polar coordinates). The  $z$ -axis cylinder restricts us to the circle in the  $xy$ -plane and the  $x$ -axis cylinder is lower than the  $y$ -axis cylinder over this particular 8th or a circle, so the  $x$ -axis cylinder is the limit on height. What is that height function? The equation of the cylinder is  $y^2 + z^2 = R^2$ , so we have  $z^2 = R^2 - y^2 = R^2 - r^2 \sin^2 \theta$ , writing  $y$  in cylindrical coordinates. This determines a region of integration and we proceed using cylindrical coordinates.

$$\begin{aligned}
V &= 16 \int_0^{\frac{\pi}{4}} \int_0^R \int_0^{\sqrt{R^2 - r^2 \cos^2 \theta}} r dz dr d\theta \\
&= 16 \int_0^{\frac{\pi}{4}} \int_0^R \sqrt{R^2 - r^2 \cos^2 \theta} r dr d\theta \\
&= 16 \int_0^{\frac{\pi}{4}} \left( (R^2 - r^2 \cos^2 \theta)^{\frac{3}{2}} \frac{2}{3} \frac{-1}{2 \cos^2 \theta} \right) \Big|_0^R d\theta \\
&= \frac{16}{3} \int_0^{\frac{\pi}{4}} \left( \frac{R^3}{\cos^2 \theta} - \frac{(R^2 - R^2 \cos^2 \theta)^{\frac{3}{2}}}{\cos^2 \theta} \right) d\theta \\
&= \frac{16}{3} \int_0^{\frac{\pi}{4}} \left( R^3 \sec^2 \theta - \frac{R^3 \sin^3 \theta}{\cos^2 \theta} \right) d\theta \\
&= \frac{16R^3}{3} \int_0^{\frac{\pi}{4}} \left( \sec^2 \theta - \frac{\sin \theta}{\cos^2 \theta} + \sin \theta \right) d\theta \\
&= \frac{16R^3}{3} \left( \tan \theta + \frac{-1}{\cos \theta} - \cos \theta \right) \Big|_0^{\frac{\pi}{4}} \\
&= \frac{16R^3}{3} \left( 1 - \frac{2}{\sqrt{2}} - \frac{\sqrt{2}}{2} - 0 + 1 + 1 \right) \\
&= \frac{16R^3}{3} \left( 3 - \frac{2\sqrt{2} + \sqrt{2}}{2} = \frac{16R^3}{3} \frac{6 - 3\sqrt{2}}{2} \right) = 8R^3(2 - \sqrt{2})
\end{aligned}$$

## 2.7.2 Spherical Coordinates

Spherical coordinates use a spherical radius and two angles:  $\phi$  as co-latitude and  $\theta$  as longitude. The ranges for the angles are  $\phi \in [0, \pi]$  and  $\theta \in [0, 2\pi]$ .

$$\begin{aligned}
x &= r \sin \phi \cos \theta \\
y &= r \sin \phi \sin \theta \\
z &= r \cos \phi
\end{aligned}$$

The Jacobian is  $J = r^2 \sin \phi$ , after a complicated  $3 \times 3$  determinant. We use spherical coordinates when the domain of integration has some kind of spherical shape or symmetry.

**Example 2.7.5.** The easiest example is the volume of a sphere of radius  $R$ .

$$\begin{aligned}\int_S 1dV &= \int_0^{2\pi} \int_0^\pi \int_0^R r^2 \sin \phi dr d\phi d\theta \\ &= \int_0^{2\pi} d\theta \int_0^\pi \sin \phi d\phi \int_0^R r^2 dr \\ &= 2\pi (-\cos \phi) \Big|_0^\pi \frac{r^3}{3} \Big|_0^R \\ &= 2\pi(1+1) \frac{R^3}{3} = \frac{4\pi R^3}{3}\end{aligned}$$

**Example 2.7.6.** We can also calculate the volume of an ellipsoid.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

We use a linear change of variables first:  $x = au$ ,  $y = vb$  and  $z = cw$ . The Jacobian is  $J = abc$  and the ellipsoid  $E$  because a unit sphere  $S$ .

$$\int_E 1dxdydz = \int_S abcduvdw$$

Then, since  $a$ ,  $b$  and  $c$  are constant, this just the volume of a sphere of radius 1.

$$\int_E 1dxdydz = \int_S abcduvdw = abc \int_S 1dudvdw = \frac{4\pi abc}{3}$$

**Example 2.7.7.** Consider a sphere of radius  $R$ . Inside this sphere is a cone with flare angle  $\pi/6$ , where the tip of the cone is located at very bottom of the sphere and the cone opens up inside the sphere. The cone, plus the portion of the sphere above the cone, gives a shape that is something like an ice-cream cone; it's a cone plus a spherically curved cap at the top of the cone. What is the volume of this object?

If we put the centre of the sphere at  $(0, 0, R)$ , then the vertex of the cone can be put at the origin. The equation of such a sphere is

$$x^2 + y^2 + (z - R)^2 = R^2.$$

Since  $r^2 = x^2 + y^2 + z^2$ , this equation simplifies into  $r^2 = 2Rz$  which is  $r^2 = 2Rr \cos \phi$  or  $r = 2R \cos \phi$ . This can be taken at the outside bound of the radius term, as  $\phi \in [0, \pi/2]$  and  $\theta \in [0, 2\pi]$ . The flare of the cone is  $\pi/6$ , which restricts  $\phi \in [0, \pi/6]$ . We integrate with these bounds for the volume, making sure that  $r$  is an inside integral since its bounds depend on  $\theta$ .

$$\begin{aligned}
\int_D 1dV &= \int_0^{2\pi} \int_0^{\frac{\pi}{6}} \int_0^{2R \cos \phi} r^2 \sin \phi dr d\phi d\theta \\
&= 2\pi \int_0^{\frac{\pi}{6}} \left. \frac{r^3}{3} \right|_0^{2R \cos \phi} \sin \phi d\phi \\
&= 2\pi \int_0^{\frac{\pi}{6}} \frac{8R^3 \cos^3 \phi \sin \phi}{3} d\phi \\
&= \frac{16\pi R^3}{3} \left( \frac{-\cos^4 \phi}{4} \right) \Big|_0^{\frac{\pi}{6}} \\
&= \frac{16R^3 \pi}{3} \left( \frac{1}{4} - \frac{9}{64} \right) = \frac{7\pi R^3}{12}
\end{aligned}$$

## 2.8 Applications of Multiple Integration

We've already spoken about the general definition of the size of sets, but let's remind ourselves. If  $S$  is any integrable set in  $\mathbb{R}^n$ , then the size of  $S$  is *defined* to be the integral of the constant function.

$$V(S) = \int_S 1dV$$

There are many physical problems involving 3D objects with variable density where we desire to know mass instead of volume.

**Definition 2.8.1.** If  $\rho(x, y, z)$  is a density function, integrable on a region  $S \in \mathbb{R}^3$ , the mass of the object  $S$  is found from by integrating the density function.

$$M(S) = \int_S \rho dV$$

This can also be done in  $\mathbb{R}^2$  for flat objects with variable density depending on only  $x$  and  $y$ . Such objects are normally referred at laminae.

A similar interesting physical problem is the problem of centre of mass. The centre of mass of an rigid object (or system) is the point in space where linear forces on the object can be accurately modeled as forces on a point-mass. In particular, forces on the centre of mass do not cause rotational acceleration. The calculation of centre of mass involves the calculation of so-called first moments. These definition vary by dimension. First we work with a lamina in  $\mathbb{R}^2$ .

**Definition 2.8.2.** Let  $S$  be a region in  $\mathbb{R}^2$  with density  $\rho(x, y)$ . Its *mass* and *first moments* are defined by the following integrals.

$$m = \int_S \rho dA$$

$$M_x = \int_S y \rho dA$$

$$M_y = \int_S x \rho dA$$

Then the *coordinates of the centre of mass* are written  $(\bar{x}, \bar{y})$  and calculated from the moments.

$$\bar{x} = \frac{M_y}{m}$$

$$\bar{y} = \frac{M_x}{m}$$

If  $S$  is a solid region in  $\mathbb{R}^3$  with density  $\rho(x, y, z)$ , then its *mass* and *first moments* are defined by the following integrals.

$$m = \int_S \rho dV$$

$$M_{yz} = \int_S x \rho dV$$

$$M_{xz} = \int_S y \rho dV$$

$$M_{xy} = \int_S z \rho dV$$

The *coordinates of the centre of mass* are written  $(\bar{x}, \bar{y}, \bar{z})$  and calculated from the moment.

$$\bar{x} = \frac{M_{yz}}{m}$$

$$\bar{y} = \frac{M_{xz}}{m}$$

$$\bar{z} = \frac{M_{xy}}{m}$$

Centre of mass and first moments are important physical properties that deal with linear acceleration; for the purpose of linear acceleration, the object acts like a point mass at the centre of mass. However, there are also moments involved in rotational movement and acceleration. These are called second moments or moments of inertia.

Let  $S \subset \mathbb{R}^2$  be a laminae with density function  $\rho$ . Its *second moments* are calculated by the following integrals.

$$I_x = \int_S y^2 \rho dA$$

$$I_y = \int_S x^2 \rho dA$$

$$I_0 = \int_S (x^2 + y^2) \rho dA$$

These moment measure the resistance to rotation:  $I_x$  is the resistance to rotation about the  $x$  axis;  $I_y$  is the resistance to rotation about the  $y$  axis; and  $I_0$  is resistance to rotation about the origin. The three *radii of gyration* are calculated from the second moments.

$$\bar{R}^2 = \frac{I_0}{m}$$

$$\bar{x}^2 = \frac{I_x}{m}$$

$$\bar{y}^2 = \frac{I_y}{m}$$

The radii of gyration are similar to centre of mass for movement. The object acts as a point mass at radius  $\bar{R}$  for rotation about the origin. Similar, it acts like a point mass at radius  $\bar{x}$  for rotation about the  $x$  axis and radius  $\bar{y}$  for rotation about the  $y$  axis.

Let  $S$  be a solid object in  $\mathbb{R}^3$  with density function  $\rho$ . Its *second moments* are calculated by the following integrals.

$$I_x = \int_S (y^2 + z^2) dV$$

$$I_y = \int_S (x^2 + z^2) dV$$

$$I_z = \int_S (x^2 + y^2) dV$$

Its *radii of gyration* are calculated from the second moments.

$$\bar{x}^2 = \frac{I_x}{m}$$

$$\bar{y}^2 = \frac{I_y}{m}$$

$$\bar{z}^2 = \frac{I_z}{m}$$

The object acts like a point mass at  $(\bar{x}, \bar{y}, \bar{z})$  for purposes of rotational physics. More specifically, for about the  $z$  axis, the particular acts as a point mass at radius  $\sqrt{\bar{x}^2 + \bar{y}^2}$ , with parallel constructions for the other two axes.

**Example 2.8.3.** Consider a quarter circle of radius  $a$  in the first quadrant with density function  $\rho = k\sqrt{x^2 + y^2}$ . What is its centre of mass?

$$\begin{aligned} m &= \int_0^{\frac{\pi}{2}} \int_0^a k r r dr d\theta \\ &= \frac{\pi}{2} \left. \frac{kr^3}{3} \right|_0^a = \frac{\pi ka^3}{6} \\ \bar{x} &= \frac{1}{m} \int_0^{\frac{\pi}{2}} \int_0^a y \rho dA = \frac{1}{m} \int_0^{\frac{\pi}{2}} \int_0^a kr^3 \sin \theta dr d\theta \\ &= \frac{k}{m} \int_0^{\frac{\pi}{2}} \sin \theta d\theta \int_0^a r^3 dr \\ &= \frac{k}{m} (-\cos \theta) \Big|_0^{\frac{\pi}{2}} \left. \frac{r^4}{4} \right|_0^a = \frac{6}{\pi a^3} \frac{1}{4} a^4 = \frac{3a}{2\pi} \\ \bar{y} &= \frac{3a}{2\pi} \end{aligned}$$

We don't have to calculate the second moment;  $\bar{x} = \bar{y}$  due to the symmetry of the situation. The centre of mass is found at  $(\frac{3a}{2\pi}, \frac{3a}{2\pi})$ .

**Example 2.8.4.** Now consider a lamina with  $y \in [-1, 1]$  and  $x$  bounded between  $\pm y^4$  with  $\rho = 1$ . What is the area and moment of inertia about the  $x$ -axis?

$$\begin{aligned} A &= \int_{-1}^1 \int_{-y^4}^{y^4} 1 dx dy \\ &= \int_{-1}^1 2y^4 dy = \left. \frac{2y^5}{5} \right|_{-1}^1 \\ &= \frac{4}{5} \\ I_x &= \int_{-1}^1 \int_{-y^4}^{y^4} y^2 dx dy \\ &= \int_{-1}^1 xy^2 \Big|_{-y^4}^{y^4} = \int_{-1}^1 2y^6 dy \\ &= \left. \frac{2y^7}{7} \right|_{-1}^1 = \frac{4}{7} \end{aligned}$$

Compare this lamina to a rectangle of height 2 and width  $2/5$ , which has the same area. We calculate the moment of inertia for the rectangle.

$$I_x = \int_{-1}^1 \int_{-\frac{1}{5}}^{\frac{1}{5}} y^2 dx dy = \frac{4}{15}$$

Our shape has twice the resistance to rotation, even though it has the same cross-section area. This physical fact partially explains the use of  $I$  beams in construction: they have more resistance to twisting and shear forces than a rectangular cross-section beam of the same size or weight.

**Example 2.8.5.** Now consider a hemisphere above the  $xy$  plane with radius  $a$  and density  $\rho = kz$ . Let's calculate its mass, centre of mass, and moments of inertia.

$$\begin{aligned}
 m &= \int_0^{2\pi} \int_0^{\pi/2} \int_0^a k(r \cos \phi) r^2 \sin^2 \phi dr d\phi d\theta \\
 &= 2\pi k \int_0^{\pi/2} \left. \frac{r^4}{4} \right|_0^a \cos \phi \sin \phi d\phi \\
 &= \frac{2\pi k a^4}{4} \int_0^{\pi/2} \frac{\sin 2\phi}{2} d\phi \\
 &= \frac{\pi k a^4}{2} \left( -\frac{\cos 2\phi}{4} \right) \Big|_0^{\pi/2} = \frac{\pi k a^4}{8} (\cos 0 - \cos \pi) = \frac{\pi k a^4}{4}
 \end{aligned}$$

$M_{yz} = 0$  due to symmetry.

$M_{xz} = 0$  due to symmetry.

$$\begin{aligned}
 M_{xy} &= \int_D z \rho dV \\
 &= \int_0^{2\pi} \int_0^{\pi/2} \int_0^a k r^2 \cos^2 \phi r^2 \sin \phi dr d\phi d\theta \\
 &= 2\pi k \int_0^a r^4 dr \int_0^{\pi/2} \cos^2 \phi \sin \phi d\phi \\
 &= \frac{2\pi k a^5}{5} \left( -\frac{\cos^3 \phi}{3} \right) \Big|_0^{\pi/2} = \frac{2\pi k a^5}{15}
 \end{aligned}$$

$$\bar{x} = 0$$

$$\bar{y} = 0$$

$$\bar{z} = \frac{\frac{2\pi k a^5}{15}}{\frac{\pi k a^4}{4}} = \frac{8a}{15}$$

The centre of mass is at  $(0, 0, \frac{8a}{15})$  Let's also calculate the moments of inertia.

$$\begin{aligned}
 I_x &= \int_D (y^2 + z^2) \rho dV = \int_D (y^2 + z^2) z k dV \\
 &= k \int_0^{2\pi} \int_0^{\pi/2} \int_0^a r^2 (\sin^2 \phi \sin^2 \theta + \cos^2 \phi) r \cos \phi r^2 \sin \phi dr d\phi d\theta \\
 &= k \int_0^a r^5 dr \int_0^{2\pi} \int_0^{\pi/2} (\sin^3 \phi \cos \phi \sin^2 \theta + \cos^3 \phi \sin \phi) d\phi d\theta \\
 &= \frac{ka^6}{6} \left[ \left( \frac{\sin^4 \phi}{4} \right) \Big|_0^{\pi/2} \left( \frac{\theta}{2} - \frac{\sin 2\theta}{4} \right) \Big|_0^{2\pi} + \left( \frac{-\cos^4 \phi}{4} \right) \Big|_0^{\pi/2} 2\pi \right] \\
 &= \frac{ka^6}{6} \left[ \left( \frac{1}{4} - 0 \right) \pi 2\pi \left( \frac{1}{4} - 0 \right) \right] = \frac{\pi ka^6}{8}
 \end{aligned}$$

$$I_y = \frac{\pi ka^6}{8} \text{ by symmetry}$$

$$\begin{aligned}
 I_z &= \int_D (x^2 + y^2) \rho dV \\
 &= k \int_0^{2\pi} \int_0^{\pi/2} \int_0^a r^2 \sin^2 \phi r \cos \phi r^2 \sin \phi dr d\phi d\theta \\
 &= \frac{2\pi ka^6}{6} \int_0^{\pi/2} \sin^3 \phi \cos \phi d\phi \\
 &= \frac{\pi ka^6}{3} \frac{\sin^4 \phi}{4} \Big|_0^{\pi/2} = \frac{\pi ka^6}{12}
 \end{aligned}$$

$$\bar{x}^2 = \frac{\frac{\pi ka^6}{8}}{\frac{\pi ka^4}{4}} = \frac{a^2}{2}$$

$$\bar{x} = \frac{a}{\sqrt{2}}$$

$$\bar{y}^2 = \frac{a^2}{2}$$

$$\bar{y} = \frac{a}{\sqrt{2}}$$

$$\bar{z}^2 = \frac{\frac{\pi ka^6}{12}}{\frac{\pi ka^4}{4}} = \frac{a^3}{3}$$

$$\bar{z} = \frac{a}{\sqrt{3}}$$

The centre of rotation is at  $\left( \frac{a}{\sqrt{2}}, \frac{a}{\sqrt{2}}, \frac{a}{\sqrt{3}} \right)$ .

**Example 2.8.6.** Now consider a paraboloid bounded by  $z = b(x^2 + y^2)$  with height  $h$  and density  $\rho = 1$ . What are its moments of inertia? We will use cylindrical coordinates, but we need a bound on the radius term at the height  $h$ . The equation of the paraboloid in cylindrical coordinates is  $z = br^2$  and when  $z = h$ , we see that  $r = \sqrt{\frac{h}{b}}$ . This is the outer bound on radius. With constant bounds on angle and radius, we can let  $z$  range from the paraboloid graph  $z = br^2$  to the constant height  $h$ .

$$\begin{aligned}
 m &= \int_R 1dV = \frac{\pi\sqrt{\frac{h}{b}}^2 h}{2} = \frac{\pi h^2}{2b} \\
 I_z &= \int_0^{2\pi} \int_0^{\sqrt{\frac{h}{b}}} \int_{br^2}^h (x^2 + y^2) r dz dr d\theta \\
 &= \int_0^{2\pi} \int_0^{\sqrt{\frac{h}{b}}} \int_{br^2}^h r^3 dz dr d\theta \\
 &= 2\pi \int_0^{\sqrt{\frac{h}{b}}} (r^3 h - r^5 b) dr \\
 &= 2\pi \left( \frac{hr^4}{4} - \frac{br^6}{6} \right) \Big|_0^{\sqrt{\frac{h}{b}}} \\
 &= 2\pi \left( \frac{h^3}{4b^2} - \frac{bh^3}{6b^3} \right) = \frac{2\pi h^3}{b^2} \left( \frac{1}{4} - \frac{1}{6} \right) = \frac{\pi h^3}{6b^2} \\
 \bar{z}^2 &= \frac{I_z}{m} = \frac{\frac{\pi h^3}{6b^2}}{\frac{\pi h^2}{2b}} = \frac{h}{3b} \\
 \bar{z} &= \sqrt{\frac{h}{3b}}
 \end{aligned}$$

$$\begin{aligned}
I_x &= \int_0^{2\pi} \int_0^{\sqrt{\frac{h}{b}}} \int_{br^2}^h (y^2 + z^2) r dz dr d\theta \\
&= \int_0^{2\pi} \int_0^{\sqrt{\frac{h}{b}}} \int_{br^2}^h (r^2 \sin^2 \theta + z^2) r dz dr d\theta \\
&= \int_0^{2\pi} \int_0^{\sqrt{\frac{h}{b}}} r^3 \sin^2 \theta (h - br^2) + r \left( \frac{h^3}{3} - \frac{b^3 r^6}{3} \right) dr d\theta \\
&= \int_0^{2\pi} \left. \frac{r^4}{4} h \sin^2 \theta - \frac{r^6}{6} b \sin^2 \theta + \frac{r^2}{2} \frac{h^3}{3} - \frac{r^8}{8} \frac{b^3}{3} \right|_0^{\sqrt{\frac{h}{b}}} d\theta \\
&= \int_0^{2\pi} \frac{h^2}{4b^2} h \sin^2 \theta - \frac{h^3}{6b^3} b \sin^2 \theta + \frac{h^4}{6b} - \frac{h^4 b^3}{24b^4} d\theta \\
&= \frac{h^3}{b^2} \int_0^{2\pi} \sin^2 \theta d\theta \left( \frac{1}{4} - \frac{1}{6} \right) + \frac{2\pi h^4}{b} \left( \frac{1}{6} - \frac{1}{24} \right) \\
&= \frac{h^3}{12b^2} \left( \frac{\theta}{2} - \frac{\sin 2\theta}{4} \right) \Big|_0^{2\pi} + \frac{3\pi h^4}{12b} \\
&= \frac{h^3 \pi}{12b^2} + \frac{3\pi h^4}{12b} = \frac{h^3 \pi}{12b} \left( \frac{1}{b} + 3h \right) \\
\bar{x}^2 &= \frac{\frac{h^3 \pi}{12b} \left( \frac{1}{b} + 3h \right)}{\frac{\pi h^2}{2b}} \\
\bar{x} &= \sqrt{\frac{a^2}{6} + \frac{h^2}{3}} = \frac{h}{6} \left( \frac{1}{b} + 3h \right) \\
\bar{x} &= \sqrt{\frac{h}{6} \left( \frac{1}{b} + 3h \right)} \\
\bar{y} &= \bar{x} \text{ by symmetry}
\end{aligned}$$

As a final example for this section, we can prove a nice theorem from physics.

**Theorem 2.8.7.** *A body of uniform density under the action of an external conservative force (such as gravity) acts like a point mass at its centre of mass.*

*Proof.* We'll just prove the theorem for the special case of a sphere. The force of gravity between two masses  $m_1$  and  $m_2$  is

$$F = \frac{Gm_1 m_2}{r^2}$$

Let's assume the source of the gravitational attraction sits at  $(0, 0, c)$  where  $c > a$  is larger than  $a$  the radius of the sphere. Let's also assume that  $\rho$  is the constant density of the sphere. Then  $\rho dV$  is an

infinitesimal piece of mass in the sphere, and we can write  $\omega$  for the distance from the infinitesimal mass to the gravitational source.

The force on the infinitesimal mass is

$$FdV = \frac{Gm\rho dV}{\omega^2}.$$

In polar coordinates, this is

$$FdV = \frac{Gm\rho r^2 \sin \phi dr d\phi d\theta}{\omega^2}.$$

By symmetry of the sphere, all lateral forces will cancel. We only care about the  $z$  component of the force. Let  $\alpha$  be the angle from the  $z$  axis of the line from  $(0, 0, c)$  to the infinitesimal mass.

$$F_z dV = \frac{Gm\rho r^2 \sin \phi \cos \alpha dr d\phi d\theta}{\omega^2}$$

Let's do some trigonometry. Consider the triangle with vertices  $(0, 0, c)$ ,  $(0, 0, 0)$  and the location of our infinitesimal mass. The angle at  $(0, 0, c)$  is  $\alpha$  by definition, and likewise the angle at  $(0, 0, 0)$  is  $\phi$ . The side lengths are  $c$ ,  $\omega$  and  $r$ , also by definition. We use the cosine law.

$$\begin{aligned}\omega^2 &= r^2 + c^2 - 2rc \cos \phi \\ \omega &= \sqrt{r^2 + c^2 - 2rc \cos \phi}\end{aligned}$$

The length  $c$  along the  $x$  axis can be divided into two pieces, so that  $c = \omega \cos \alpha + r \cos \phi$ . We can solve for  $\cos \alpha$ .

$$\cos \alpha = \frac{c - r \cos \phi}{\omega}$$

We can replace  $\omega$  with the square root expression.

$$\cos \alpha = \frac{c - r \cos \phi}{\sqrt{r^2 + c^2 - 2rc \cos \phi}}$$

We can replace both  $\cos \alpha$  and  $\omega$  in the force expression.

$$F_z dV = \frac{Gm\rho r^2 \sin \phi}{r^2 + c^2 - 2rc \cos \phi} \left( \frac{c - r \cos \phi}{\sqrt{r^2 + c^2 - 2rc \cos \phi}} \right) dr d\phi d\theta$$

The total force is this integral of this infinitesimal force over the sphere.

$$F_z = \int_0^{2\pi} \int_0^\pi \int_0^a \frac{Gm\rho r^2 \sin \phi}{r^2 + c^2 - 2rc \cos \phi} \left( \frac{c - r \cos \phi}{\sqrt{r^2 + c^2 - 2rc \cos \phi}} \right) dr d\phi d\theta$$

We'll do some substitution here, more or less reversing the trigonometry. We leave  $r$  and  $\theta$  along, but write  $\omega^2 = r^2 + c^2 - 2rc \cos \phi$  with the cosine law as before.

$$2r d\omega = 2rc \sin \phi d\phi \implies r \sin \phi d\phi = \frac{\omega}{c} d\omega$$

We have  $\omega(0) = c - r$  and  $\omega(\pi) = c + r$  for the bounds. Finally, we can solve to get  $r \cos \phi = \frac{r^2 + c^2 - \omega^2}{2c}$ . We use this as a substitution in the  $\phi$  integral.

$$\begin{aligned}
F_z &= \int_0^{2\pi} \int_0^\pi \int_0^a \frac{Gm\rho r}{(r^2 + c^2 - 2rc \cos \phi)} \frac{(c - r \cos \phi)}{\sqrt{r^2 + c^2 - 2rc \cos \phi}} (r \sin \phi d\phi) dr d\theta \\
&= \int_0^{2\pi} \int_0^a \int_{c-r}^{c+r} \frac{Gm\rho r}{\omega^2} \frac{c - \frac{r^2 + c^2 - \omega^2}{2c}}{\omega} \frac{\omega}{c} d\omega dr d\theta \\
&= \int_0^{2\pi} \int_0^a \int_{c-r}^{c+r} \frac{Gm\rho r}{\omega^2} \frac{2c^2 - r^2 - c^2 + \omega^2}{2c^2} \omega dr d\theta \\
&= \frac{2\pi Gm\rho}{2c^2} \int_0^a \int_{c-r}^{c+r} r \left( \frac{c^2 - r^2 + \omega^2}{\omega^2} \right) d\omega dr \\
&= \frac{\pi Gm\rho}{c^2} \int_0^a \int_{c-r}^{c+r} r \left( (c^2 - r^2) \frac{1}{\omega^2} + 1 \right) d\omega dr \\
&= \frac{\pi Gm\rho}{c^2} \int_0^a r \left( (c^2 - r^2) \frac{-1}{\omega} + \omega \right) \Big|_{c-r}^{c+r} dr \\
&= \frac{\pi Gm\rho}{c^2} \int_0^a (rc^2 - r^3) \left( \frac{1}{c-r} - \frac{1}{c+r} \right) + r(c+r - (c-r)) dr \\
&= \frac{\pi Gm\rho}{c^2} \int_0^a \frac{-r(c^2 - r^2)(-2r)}{c^2 - r^2} + 2r^2 dr \\
&= \frac{\pi Gm\rho}{c^2} 2r^2 + 2r^2 dr \\
&= \frac{\pi Gm\rho}{c^2} 4r^2 dr = \frac{4\pi Gm\rho a^3}{3c^2} = \frac{Gm \left( \frac{\rho 4\pi a^3}{3} \right)}{c^2}
\end{aligned}$$

The expression in brackets is the mass of the sphere and  $c$  is the distance from the centre of mass to the gravitational sources. This is exactly the expression we wanted: it is the force due to a point mass at the origin with mass equal to the total mass of the sphere.  $\square$

In the previous example, the two masses were separated from each other. We could instead consider hollow sphere, with outside radius  $a$  and inside radius  $b$ , and a point mass at  $(0, 0, c)$  with  $c < b$ , so that the point mass is inside the sphere. What is the force of gravity on that point mass? The set-up

is almost the same; the only difference is that the bounds on  $\omega$  are reversed in sign.

$$\begin{aligned}
F_z &= \int_0^{2\pi} \int_a^b \int_{r-c}^{r+c} \frac{Gm\rho r(c^2 - r^2 + \omega^2)}{2\omega^2 c^2} d\omega dr d\theta \\
&= \frac{\pi Gm\rho}{c^2} \int_a^b (r^3 - rc^2) \left. \frac{1}{\omega} \right|_{r-c}^{r+c} + r\omega \left. \right|_{r-c}^{r+c} dr \\
&= \frac{\pi Gm\rho}{c^2} \int_a^b r(r^2 - c^2) \left( \frac{1}{r+c} - \frac{1}{r-c} \right) + 2rcdr \\
&= \frac{\pi Gm\rho}{c^2} \int_a^b r(r^2 - c^2) \frac{-2c}{r^2 - c^2} + 2rcdr \\
&= \frac{\pi Gm\rho}{c^2} \int_a^b -2rc + 2rcdr = \frac{\pi Gm\rho}{c^2} \int_a^b 0dr = 0
\end{aligned}$$

Everything cancels out. We reach a fairly strange conclusion: no matter the location of a point mass is inside a hollow sphere (of uniform density), it experiences no force of gravity.

### 2.8.1 Moments and Probability Distributions

Let's recall some definitions about continuous probability.

**Definition 2.8.8.** A *probability distribution* on an integrable set  $D \subset \mathbb{R}^n$  is a function  $\rho : D \rightarrow [0, \infty)$  such that

$$\int_D \rho dV = 1.$$

The set  $D$  is called the *set of states*; each point represents the state of the system. For any integrable subset  $A \subset D$ , the probability that the state of the system is in the subset  $A$  is the integral over that subset.

$$P(A) = \int_A \rho dV$$

**Definition 2.8.9.** An integrable function  $f : D \rightarrow \mathbb{R}$  is called an *observable* or *measurable*; for each state, an observable measure some property of that state. The *expectation value* of the observable  $f$  is the continuous version of the average value of  $f$  over the state. It is written  $\langle f \rangle$  and calculated by integration.

$$\langle f \rangle = \int_D f \rho dV$$

If  $D \subset \mathbb{R}^3$ , then  $D$  may be a domain of positions. These states are simply the possible places where a particle may be found.

$$\begin{aligned}\langle x \rangle &= \int_D x \rho dV \\ \langle y \rangle &= \int_D y \rho dV \\ \langle z \rangle &= \int_D z \rho dV \\ \langle r \rangle &= \langle \sqrt{x^2 + y^2 + z^2} \rangle = \int_D \sqrt{x^2 + y^2 + z^2} \rho dV\end{aligned}$$

The first three of these are the expectation values of each coordinate of position and the last is the expectation value of the distance to the origin.

Alternatively, with  $D \subset \mathbb{R}^3$ ,  $\rho$  could be an ordinary mass-density. We can think of this as the probability of finding mass at each point in the set. In that context,  $\langle x \rangle$ ,  $\langle y \rangle$  and  $\langle z \rangle$  are just the first moment and the coordinates of the centre of mass. That also makes sense: the centre of mass is the ‘average’ location of the mass of the object.

The second moments are also expectation values. For example,  $\langle x^2 + y^2 \rangle$  is the average (square) distance from the  $z$  axis. Objects further away from the  $z$  axis are more difficult to rotate around that axis, so this naturally measure the resistance to rotation around the  $z$  axis. Likewise for  $\langle x^2 + z^2 \rangle$  about the  $y$  axis and  $\langle y^2 + z^2 \rangle$  about the  $x$  axis.

The use of the term ‘moment’ is historical. The entire subject of continuous probability can be developed using the terminology of moments. For these notes, we’ll use the term ‘expectation value’ instead of ‘moment’.

**Definition 2.8.10.** The standard deviation of an observable is a measure of the width of the distribution. This can be expressed as the expectation value of the distance from the average. In practice, we use the square of distance and calculate the square of standard deviation (much like the pythagorean theorem). Let  $f$  be an observable. Its standard deviation  $\sigma_f$  is calculate by the following integral.

$$\sigma_f^2 = \int_D (f - \langle f \rangle)^2 \rho dV$$

With some algebra, we can uncover some interesting properties of standard deviation.

$$\begin{aligned}\sigma_f^2 &= \int_D (f - \langle f \rangle)^2 \rho dV \\ &= \int_D (f^2 - 2f\langle f \rangle + \langle f \rangle^2) \rho dV \\ &= \int_D f^2 \rho dV - 2 \int_D f \langle f \rangle \rho dV + \int_D \langle f \rangle^2 \rho dV \\ &= \langle f^2 \rangle - 2\langle f \rangle \int_D f \rho dV + \langle f \rangle^2 \int_D \rho dV \\ &= \langle f^2 \rangle - 2\langle f \rangle \langle f \rangle + \langle f \rangle^2 = \langle f^2 \rangle - \langle f \rangle^2 \\ \sigma_f^2 &= \langle f^2 \rangle - \langle f \rangle^2\end{aligned}$$

The last line above is an important identity for standard deviation and expectation values.

## 2.8.2 Quantum Mechanics

One of the most well-known applications of continuous probability is quantum mechanics. The whole field is built on the assumption that states of physical systems are probabilistically determined and that we should study the physics of the system by studying the probability distribution.

The state of a physical system in quantum mechanics is measured by a function  $\Psi$ , called a wave function.  $\Psi$  itself is not exactly the probability density; quantum mechanics works with a  $\mathbb{C}$ -valued function, which adds another layer of confusion. With  $\mathbb{C}$ -valued function, there is an operation called complex conjugation, which changes the sign of the imaginary piece of the function. It is written with a bar, so  $\bar{\Psi}$  is the conjugate. Then, expectation values in quantum mechanics are given by the following integral, where  $D$  is the domain of states and  $f$  is a observable on that domain of states.

$$\langle f \rangle = \int_D \bar{\Psi} f \Psi dV$$

All observables in quantum mechanics are calculated from the wave function. Determining the behaviour of the wave function over time is the goal of the discipline; that behaviour is given by the famous Schrodinger equation. In this equation  $V$  is a potential energy function on the space of states,  $\hbar$  is a constant,  $m$  is the mass and  $i$  is the imaginary number, with  $i^2 = -1$ .

$$i\hbar \frac{\partial \Psi}{\partial t} = \frac{-\hbar^2}{2m} \nabla^2 \Psi + V \Psi$$

We'll assume that  $\Psi(x, t)$  has only one variable of position. (The following derivation works in three variables of position with roughly the same steps, but the notation becomes much more challenging). The expectation value for position is  $\langle x \rangle$ . What is momentum? It should be (up to a mass term) the rate of change of position. But the only available sense of position is the expectation value. Therefore, we should try to calculate  $\frac{\partial \langle x \rangle}{\partial t}$ . We use the compatibility of integration and differentiation to exchange the operations. (There are theorems, which we've omitted in this course, which allow this exchange).

$$\frac{\partial \langle x \rangle}{\partial t} = \frac{\partial}{\partial t} \int_{\mathbb{R}} \bar{\Psi} x \Psi dx = \int_{\mathbb{R}} \frac{\partial}{\partial t} x \bar{\Psi} \Psi dx = \int_{\mathbb{R}} x \frac{\partial}{\partial t} (\bar{\Psi} \Psi) dx$$

We need to calculate this derivative term. We make use of the Schrodinger equation to change the time derivatives into space derivatives.

$$\begin{aligned}
\frac{\partial}{\partial t} (\bar{\Psi}\Psi) &= \bar{\Psi} \frac{\partial}{\partial t} \Psi + \Psi \frac{\partial}{\partial t} \bar{\Psi} \\
&= \bar{\Psi} \frac{1}{i\hbar} \left[ \frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi + V\Psi \right] + \Psi \frac{-1}{i\hbar} \left[ \frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \bar{\Psi} + V\bar{\Psi} \right] \\
&= \frac{i\hbar}{2m} \bar{\Psi} \frac{\partial^2 \Psi}{\partial x^2} - \frac{iV\Phi\bar{\Phi}}{h} - \frac{i\hbar}{2m} \Psi \frac{\partial^2 \bar{\Psi}}{\partial x^2} + \frac{iV\Psi\bar{\Psi}}{h} \\
&= \frac{\partial}{\partial x} \left[ \frac{i\hbar}{2m} \left( \bar{\Psi} \frac{\partial \Psi}{\partial x} - \frac{\partial \bar{\Psi}}{\partial x} \Psi \right) \right] \\
\frac{d\langle x \rangle}{dt} &= \frac{i\hbar}{2m} \int x \frac{\partial}{\partial x} \left( \bar{\Psi} \frac{\partial \Psi}{\partial x} - \frac{\partial \bar{\Psi}}{\partial x} \Psi \right) dx
\end{aligned}$$

Integrate by parts.

$$= \frac{i\hbar}{2m} x \left( \bar{\Psi} \frac{\partial \Psi}{\partial x} - \frac{\partial \bar{\Psi}}{\partial x} \Psi \right) \Big|_{-\infty}^{\infty} - \frac{i\hbar}{2m} \int \left( \bar{\Psi} \frac{\partial \Psi}{\partial x} - \frac{\partial \bar{\Psi}}{\partial x} \Psi \right) dx$$

The evaluation terms decays to 0 due to normalization limits.

Integrate by parts again. Half the terms cancel.

$$\begin{aligned}
&= \frac{-i\hbar}{m} \int \bar{\Psi} \frac{\partial \Psi}{\partial x} dx = \frac{\hbar}{im} \int \bar{\Psi} \frac{\partial \Psi}{\partial x} dx \\
\langle p \rangle &= \frac{d\langle x \rangle}{dt} = \int \bar{\Psi} \frac{\hbar}{i} \frac{\partial}{\partial x} \Psi dx = \left\langle \frac{\hbar}{i} \frac{\partial}{\partial x} \right\rangle
\end{aligned}$$

This is something new: the expectation value of an *operator* instead of a function. This is well defined because the operator acts on the wave function. We are led to a general correspondence.

Expectation Values	→	Operators on Wave Functions
Position	→	Multiplication by $x$
Momentum	→	Operator $\frac{\hbar}{i} \frac{\partial}{\partial x}$

Any observable in quantum mechanics can be reduced to an operator  $F$  on the space of wave functions. Its expectation value is its integral.

$$\langle F \rangle = \int_D \bar{\Psi} (F\Psi) dV$$

Moreover, all operators in quantum mechanics can be derived from combination of the position and momentum operators.

### 2.8.3 The Uncertainty Principle

The study of quantum mechanics thus becomes the study of operators on wave functions. This is a mathematically intense study, leading to whole new branches of mathematics focused on operator algebra. One of the first and most important question concerning operator is this: given two operators, do they commute? That is, if  $F$  and  $G$  are operators, is  $F(G\Psi) = G(F\Psi)$ ?

**Definition 2.8.11.** For any objects with multiplication  $a$  and  $b$ , the commutator bracket is defined as

$$[a, b] = ab - ba.$$

The two items commute if and only if  $[a, b] = 0$ .

We use the commutator bracket to study the commutativity of operators. In studying operator algebra and continuous probability, we can derive the following inequality. (We don't have the time or machinery for the proof, unfortunately).

**Theorem 2.8.12.** Let  $F$  and  $G$  be operators on wave functions.

$$\sigma_F \sigma_G \geq \frac{1}{2i} \langle [F, G] \rangle$$

The commutator bracket is still an operator, so it has an expectation value. Let's look at position and momentum, to see if they commute. Let  $f$  be a test function (something to act upon, for an operator).

$$\begin{aligned} \left[ x, \frac{\hbar}{i} \frac{\partial}{\partial x} \right] f &= x \frac{\hbar}{i} \frac{\partial f}{\partial x} - \frac{\hbar}{i} \frac{\partial}{\partial x} (xf) \\ &= x \frac{\hbar}{i} \frac{\partial f}{\partial x} - x \frac{\hbar}{i} \frac{\partial f}{\partial x} - \frac{\hbar}{i} f \\ &= -\frac{\hbar}{i} f \\ \left[ x, \frac{\hbar}{i} \frac{\partial}{\partial x} \right] &= -\frac{\hbar}{i} = i\hbar \end{aligned}$$

So we know the commutator of position and momentum. What is its expectation value?

$$\langle i\hbar \rangle = \int \bar{\Psi} i\hbar \Psi dx = i\hbar \int \bar{\Psi} \Psi dx = i\hbar$$

This makes some sense: this operator is simply multiplication by a constant; it doesn't change the wave functions at all. Its expectation is simply itself – it is a constant.

Then we apply the theorem to get the next result, where  $x$  stands for position and  $p$  for momentum:

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}$$

What does this mean? The  $\sigma$  is the standard deviation: it measures how wide the probability is for each of the measurements. This can be thought of as error. If a  $\sigma$  is very small, we have a very precise observable. If  $\sigma$  is large, the possible values in a reasonable probability are much larger. The product of these two  $\sigma$  is the product of the uncertain in our measurement of position and moment. This is the famous uncertainty principle. Because the operators do not commute, we can't measure them both precisely at the same time. That fact comes directly the the operator mathematics; this makes it intrinsic to this quantum mechanical model. The uncertainty is not a problem of measurement, but a mathematical fact of the systems.

This can be thought of in terms of particle/wave duality. A particle, as a probability distribution, is just a single peak. A wave, as a probability distribution, has a sinusoidal graph. Momentum is essentially wavelength in this interpretation. A particle has a definite position: the single peak is located somewhere. However, with no repetition, it has no wavelength. A wave has a wavelength, but since the graph extends outward, it has no fixed position. The uncertainty principle reflects how elementary objects have both wave and particle like behaviour, one to the exclusion of the other.

This commutator analysis works for all operators. Two operators are compatible observables if they commute. If they do not, a version of the uncertainty principles holds for them.

We are led to some of the standard philosophical problems of measurement in QM. What does a measurement do? Why does it collapse a wave function? Is it human observation? Machine observation? Consciousness?

Historically, there were three main camps. The Realist camp calimed that things existed in reality and that probability is an illusion and weakness of the model. The Orthodox camp said that nothing exists before a measurement. The system *is* a probability—nothing more or less. The Agnostic camp said that before measurement, any such question is meaningless, since measurement is all we have to interact with the universe. Many feel that none of the three answers are entirely satisfactory and the mystery of quantum mechanics remains with us.

# Chapter 3

## Vector Calculus

### 3.1 Vector Fields

So far, we studied scalar fields  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . When defining derivatives of various types, we only considered single-valued functions. In multiple integration, the integrands were functions with a single real output.

We have mentioned functions  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  a few times, but only as transformations of spaces. Those who have taken Linear Algebra are familiar with linear functions  $\mathbb{R}^n \rightarrow \mathbb{R}^m$ , which are encoded in  $m \times n$  matrices. We used functions  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  as change of variables transformations for integrals. Their Jacobian matrices  $J(F)$  were square  $n \times n$  matrices, so we could take the determinant to define the Jacobian  $|J(F)|$ , which acted as a measure of the change in local size/area/volume for the purposes of integration.

Interpreting functions  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  as transformations of space is very useful and valuable. It connects well with linear algebra, where matrices are thought of as transformations. It serves multiple integration by allowing coordinate transformations. However, it is not the only conceptual way to understand such functions. We are now going to re-interpret these functions  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  as vector fields.

The use of the word ‘field’ here refers to a function on subsets of  $\mathbb{R}^n$ . So far, we have worked with *scalar fields*, functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . On  $\mathbb{R}^n$  (or a subset), these functions output scalars. Conceptually, any scalar quantity that might differ throughout a three-dimensional region is a scalar field. Familiar examples are temperature, pressure, concentration of some material in a solution, density, and elevation. Now we define the other type of field.

**Definition 3.1.1.** A function on a region of  $S \subset \mathbb{R}^n$  which outputs vectors is a *vector field*. If the output are  $m$ -vectors, we write  $f : S \rightarrow \mathbb{R}^m$ .

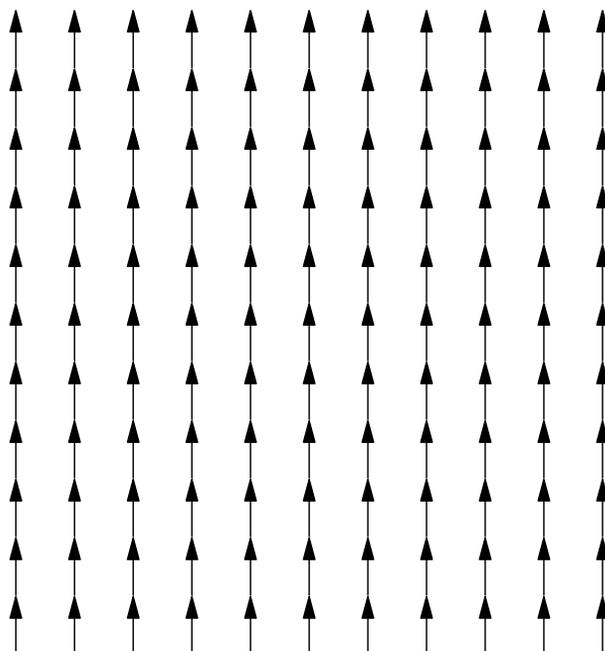


Figure 3.1: The Vector Field  $F(x, y) = (0, 1)$

Familiar examples are force, movement, acceleration, wind speed, ocean current speed, fluid flow, or the gradient of any differentiable scalar field. All these can vary over a region of space but need to be represented by a vector: by a magnitude and a direction.

Vector fields rely on the idea of local coordinates: the vector output is always a direction as if the current location is a local origin.

**Example 3.1.2.** Consider the constant vector field  $F(x, y) = (0, 1)$  on  $\mathbb{R}^3$ , as seen in figure 3.1. The vector  $(0, 1)$  is a unit vector in the vertical  $y$ -axis direction in  $\mathbb{R}^2$ . This vector field associates that unit vertical vector to each point in  $\mathbb{R}^2$  (in local coordinates). At each point, the vector points vertically upward from that point. If this field represented a fluid flow, the fluid would be flowing upwards (in  $y$ ) at a uniform pace everywhere, following these local direction vectors.

**Definition 3.1.3.** If  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a vector field, then we can write  $F = (F_1, F_2, \dots, F_n)$  in terms of its *components*. Each component  $F_i$  is itself a scalar field, has partial derivatives and gradients, and is subject to all the tools we already have for scalar fields.

We can also use all the tools of vector analysis to understand vector fields. If  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a vector field, then  $|F|$  is the length of each vector.  $\frac{F}{|F|}$  is the unit vector in each direction of  $F$  as long as  $|F|$  is not zero. If  $G$  is another vector field, then we can calculate the scalar field  $F \cdot G$  at each point. If  $n = 3$ , we can also calculate  $F \times G$ . If  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a scalar field, then the multiplication  $fF$  is a scalar multiplication of the vector  $F$  on each point in the domain of both functions.  $F + G$  and  $F - G$

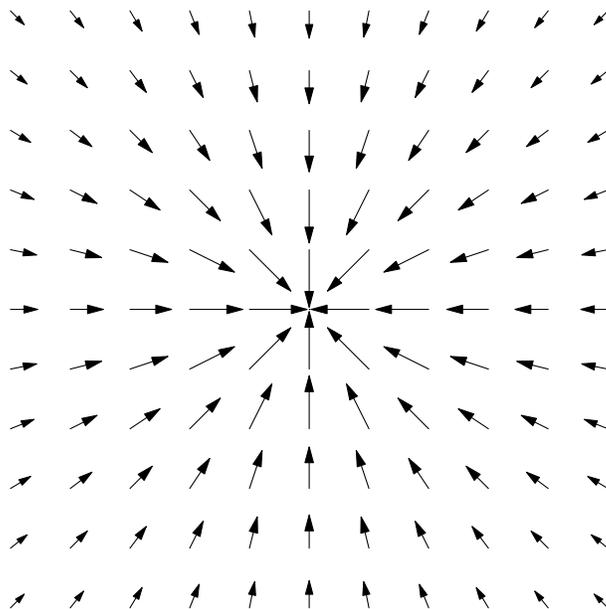


Figure 3.2: The Vector Field Describing Gravitational Force

are also vector fields, formed by adding or subtracting vectors. The expression  $FG$  doesn't make any sense, since we can't multiply vectors.

**Example 3.1.4.** An excellent example of a vectors field is the force of gravity, shown in Figure 3.2. The magnitude of the force of gravity per unit mass due to a mass  $M$  at the origin is a scalar field.  $(x, y, z)$  is:

$$f = \frac{MG}{x^2 + y^2 + z^2}$$

However, the force itself also includes direction, so it is a vector field.

$$F = \frac{MG}{(x^2 + y^2 + z^2)^{\frac{3}{2}}}(-x, -y, -z)$$

The direction  $(-x, -y, -z)$  is back towards the origin. We can recover the magnitude by taking the length of the vector.

$$|F| = \frac{MG}{(x^2 + y^2 + z^2)^{\frac{3}{2}}} \sqrt{x^2 + y^2 + z^2} = \frac{MG}{x^2 + y^2 + z^2} = \frac{MG}{r^2}$$

## 3.2 Integral Curves

If we interpret a vector field as a fluid flow, imagine a bouyant object floating in the flow. Moving with the flow, what path would the boyant object take? More mathematically, given the arrows of direction of a vector field, can we find parametric curves  $\gamma(t)$  which follow these arrows? What does follow mean?

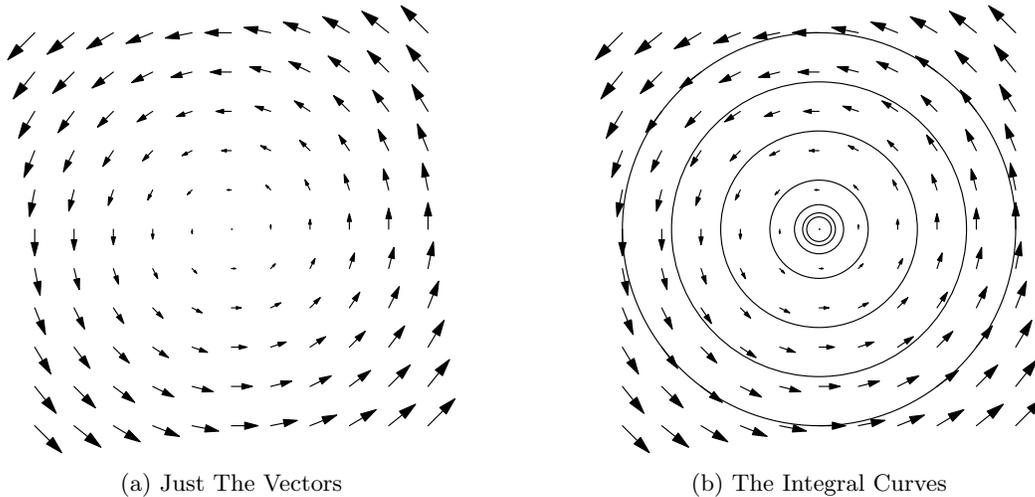


Figure 3.3: The Vector Field  $F(x, y) = (-y, x)$

**Definition 3.2.1.** Let  $F : S \rightarrow \mathbb{R}^m$  be a vector field on a region  $S \subset \mathbb{R}^n$ . If there exists a family (depending on some variable  $a \in \mathbb{R}^k$ ) of parametric curves  $\gamma_a$  such that, over the whole family,  $\gamma'(t) = F$ , then these parametric curves are called the *integral curves* of the vector field. That is, the *tangents* of the parametric curves are the same as the vector field everywhere. Equivalently, a parametric curve  $\gamma(t)$  is an integral curve for a vector field  $F$  if  $\gamma'(t) = F(\gamma(t))$ .

**Example 3.2.2.** Consider the vector field  $F(x, y) = (-y, x)$ , as shown in Figure 3.3, with its integral curves. These integral curves of  $F(x, y) = (-y, x)$  aren't surprising. Looking just at the vector fields, we can clearly tell that the fluid is travelling in a circular paths. Mathematically, the circles are curves  $\gamma(t) = (a \cos t, a \sin t)$  for a parameter  $a > 0$ . We can calculate  $\gamma'(t) = (-a \sin t, a \cos t) = (-y, x)$ . In terms of the curve coordinates, these tangents are exactly the vector field.

**Example 3.2.3.** Let  $F(x, y) = (x, y)$ , as in Figure 3.4. The curves  $\gamma(t) = (ae^t, be^t)$  where  $a, b \in \mathbb{R}$  are integral curves. They have  $x = ae^t$  and  $y = be^t$ , and we can calculate tangents:  $\gamma'(t) = (ae^t, be^t) = (x, y)$ .

Integral curves are a very powerful conceptual tool. However, they are usually very hard to calculate and most examples are beyond the scope of this course. However, there will be a few we can attempt. I'll describe the general process in  $\mathbb{R}^3$ .

If  $F = (F_1, F_2, F_3)$  is a vector field and  $\gamma(t) = (\gamma_1, \gamma_2, \gamma_3)$  is a parametric curve, then  $\gamma$  is an integral curve if it satisfies the following system.

$$\begin{aligned} \gamma_1'(t) &= F_1(\gamma_1, \gamma_2, \gamma_3) \\ \gamma_2'(t) &= F_2(\gamma_1, \gamma_2, \gamma_3) \\ \gamma_3'(t) &= F_3(\gamma_1, \gamma_2, \gamma_3) \end{aligned}$$

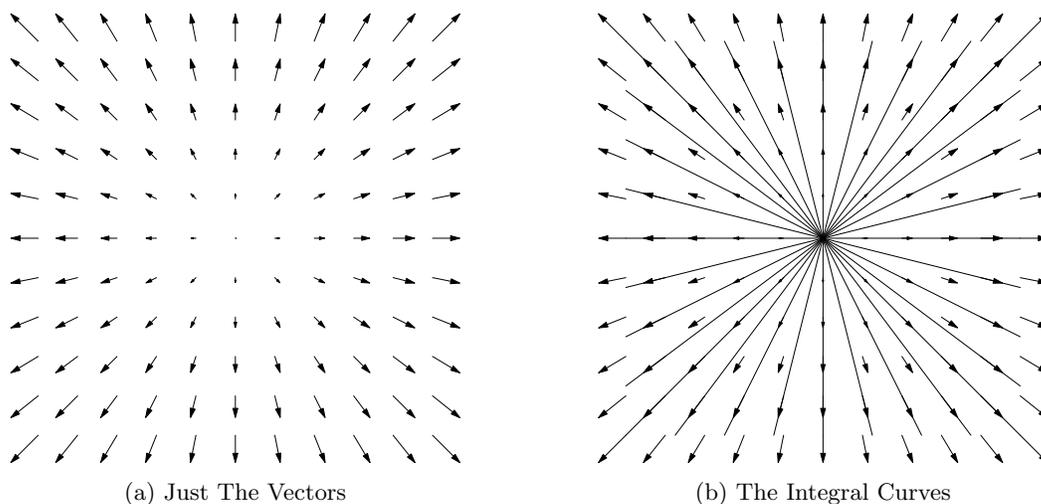


Figure 3.4: The Vector Field  $F(x, y) = (x, y)$

This is a system (generally non-linear) of three differential equations in three functions, the  $\gamma_i$ . Even if it is linear, this is still a difficult system to solve, if not impossible. Even a simple field such as  $F(x, y, z) = (y, z, x)$  leads to a difficult system.

$$\begin{aligned}\gamma_1' &= \gamma_2 \\ \gamma_2' &= \gamma_3 \\ \gamma_3' &= \gamma_1\end{aligned}$$

Currently, we have no method for approaching this system. The solution would be a function which is its own third derivative, and we are not currently aware of any such functions. However, there are some approach examples with reasonable systems of equations.

**Example 3.2.4.** Let  $F = (x, y, z)$ .

$$\begin{aligned}\gamma_1' = \gamma_1 &\implies \gamma_1(t) = ae^t \\ \gamma_2' = \gamma_2 &\implies \gamma_2(t) = be^t \\ \gamma_3' = \gamma_3 &\implies \gamma_3(t) = ce^t\end{aligned}$$

**Example 3.2.5.** Let  $F = (1, 2x, 3y)$ .

$$\begin{aligned}\gamma_1' &= 1 \\ \gamma_2' &= 2\gamma_1 \\ \gamma_3' &= 3\gamma_2\end{aligned}$$

This can be solved iteratively.

$$\gamma_1 = \int 1 dt = t + a$$

$$\gamma_2 = 2 \int \gamma_1 dt = \int 2t + 2adt = t^2 + 2at + b$$

$$\gamma_3 = 3 \int \gamma_2 dt = \int 3t^2 + 6at + 3bdt = t^3 + 3at^2 + 3bt + c$$

$$\gamma(t) = (t + a, t^2 + 2ta + b, t^3 + 3at^2 + 3bt + c)$$

### 3.3 Vector Operations on Vector Fields

Recall the gradient  $\nabla f$  of a scalar field  $f$ . It's useful to recall the definition of the operator  $\nabla$ . (stated in  $\mathbb{R}^3$  for convenience).

$$\nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$$

If  $f$  is a scalar field, then the gradient  $\nabla f$  is a *vector field* describing the direction of greatest change. We didn't investigate it's properties as a vector field in Calculus III, but it was the first vector field we used.

There are new operators we can define using  $\nabla$  on vector fields  $F$ .  $\nabla$  operates a vector, since it has components which are differential operators. Therefore, we use vector notation involving  $\nabla$ .

**Definition 3.3.1.** Let  $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be a vector field. The *curl* of  $F$  is defined as the the cross product of  $\nabla$  and  $F$ . Note that this outputs a new vector field, not a scalar field.

$$\nabla \times F = \left( \frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right)$$

Curl measures the tendency of the vector field to cause *local* rotation. If we think of the vector field as a fluid flow and if we drop an object in the fluid, it will flow along the integral curves of the vector field. However, as it flows along, it may also start spinning about its own axis. Curl measure the tendency of the vector field to cause such a spin. (This is very different from global rotation. The paths of rotation themselves may be circular without actually causing the object itself to spin. Likewise, the paths can be totally straight but still cause rotation.)

**Definition 3.3.2.** A vector field with zero curl is called *irrotational*.

**Example 3.3.3.** Consider  $F(x, y, z) = (y, 0, 0)$ . This is a field which moves objects in the  $x$  direction, but the speed of movement varies with the  $y$  coordinate. The curl is  $\nabla \times F = (0, 0, -1)$ . This field causes a clockwise rotation about the  $z$  axis; as particles in the fluid move in the  $x$  direction, they start spinning around a vertical axis.

**Definition 3.3.4.** Let  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a vector field in any dimension. The *divergence* of  $F$  is the dot product  $\nabla \cdot F$ . Note that this outputs a new scalar field, not a vector field.

$$\nabla \cdot F = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \dots + \frac{\partial F_n}{\partial x_n}$$

The divergence measure the tendency of a vector field to diffuse. Thinking in terms of gaseous fluids, a positive divergence at a point means that the density of the gas is *decreasing*. Some directions of flow may be inward and some outward, but there is a net diffusion of the gas. If the divergence is negative, the density is *increasing* and there is a net gathering of the gas.

**Definition 3.3.5.** A vector where where the divergence is zero is called *incompressible*.

Many liquids are incompressible, at least locally and under reasonable energy circumstances. Water is usually treated as an incompressible fluid. The major difference in fluid dynamics between gases and liquid is compressibility.

**Definition 3.3.6.** Let  $f$  be is a scalar field  $\mathbb{R}^n \rightarrow \mathbb{R}$ . It's *Laplacian* is the divergence of its gradient:  $\nabla \cdot (\nabla f) = \nabla^2 f$ . Note that this outputs a scalar field.

The Laplacian was mentioned in the previous term, since the input and output are both scalar fields. However, the intermediate state (the gradient) is a vector field, and the second part of the operation (the divergence) is a vector operation. This persepctive of the Laplacian is quite useful.

The Laplacian is very important for many differential equations in physics. As the divergence of the gradient field, it measures where the scalar field leads to a gathering or spreading. In a potential energy field, it measures the sources (attractors) and repellers which generate the field.

There were two important differential equations introduced in Calculus III: the heat equation and the wave equation. Extending these DES to several dimensions uses the Laplacian. Recall the 1-variable and multi-variable versions of those two equations.

$$\begin{aligned} \frac{\partial f}{\partial t} &= \alpha \frac{\partial^2 f}{\partial x^2} \\ \frac{\partial^2 f}{\partial t^2} &= \alpha \frac{\partial^2 f}{\partial x^2} \end{aligned}$$

Replacing the 1-variable position derivative with a multivariable Laplacian extends these DEs to several variables.

$$\begin{aligned} \frac{\partial f}{\partial t} &= \alpha \nabla^2 f \\ \frac{\partial^2 f}{\partial t^2} &= \alpha \nabla^2 f \end{aligned}$$

We've also seen the Laplacian in the Schrödinger equation and the Navier-Stokes equation as the space derivative term.

In the wave equation, if  $\nabla^2 f = 0$  then  $\frac{\partial^2 f}{\partial t^2} = 0$ . Therefore,  $f$  has, at most, a linear dependence in  $t$ . Since the dependence in  $x$  is often sinusoidal, a common solution here is a standing wave or a wave with constant velocity. This (among other, more confusing reasons) leads to the terminology of harmonic functions.

**Definition 3.3.7.** A scalar field  $f$  is *harmonic* if  $\nabla^2 f = 0$

The harmonic condition is a very restrictive condition. In particular, harmonic functions satisfy a maximum modulus principle: if  $f$  has an strict extrema in an open set, then it has a point with non-zero Laplacian and cannot be harmonic. Therefore, the extrema of  $f$  must always come on the boundary of any open set. In complex analysis, all differentiable (holomorphic) functions are harmonic; this fact leads to very different behaviour between real and complex differentiable functions.

### 3.3.1 Interaction between Vector Operations

All uses of  $\nabla$  are linear.

**Proposition 3.3.8.** Let  $f$  and  $g$  be scalar fields on  $\mathbb{R}^3$ , and let  $F$  and  $G$  be vector fields on  $\mathbb{R}^3$ . Let  $a$  and  $b$  be constants. Then:

$$\begin{aligned}\nabla(a f \pm b g) &= a \nabla f \pm b \nabla g \\ \nabla \times (a F \pm b G) &= a(\nabla \times F) \pm b(\nabla \times G) \\ \nabla \cdot (a F \pm b G) &= a(\nabla \cdot F) \pm b(\nabla \cdot G) \\ \nabla^2(a f \pm b g) &= a \nabla^2 f \pm b \nabla^2 g\end{aligned}$$

In the previous identity (and similar identities), pay careful attention to which outputs are scalar fields and which are vector fields.

Since  $\nabla$  is a differential operator, it doesn't distribute over multiplication. However, there are several generalizations of the Leibniz rule.

**Proposition 3.3.9.** Let  $f$  be a scalar field on  $\mathbb{R}^3$  and let  $F$  and  $G$  be vector fields on  $\mathbb{R}^3$ .

$$\begin{aligned}\nabla \cdot (F \times G) &= (\nabla \times F) \cdot G - F \cdot (\nabla \times G) \\ \nabla \cdot (f F) &= f(\nabla \cdot F) + (\nabla f) \cdot F \\ \nabla \times (f F) &= f(\nabla \times F) + (\nabla f) \times F\end{aligned}$$

In the first Leibniz rule we have a difference instead of the expected sum. In general versions of the Leibniz rule, this negative sign is quite common. Many versions of the rule have  $(-1)^k$  for some  $k$ , showing that it is equally easy to have a sum or difference.

There are two important results about composition of vector operations.

**Proposition 3.3.10.** *Let  $f$  be a scalar field on  $\mathbb{R}^3$  and  $F$  a vector field on  $\mathbb{R}^3$ . Then  $\nabla \times (\nabla f) = 0$ : the curl of a gradient is zero. Also,  $\nabla \cdot (\nabla \times F) = 0$ : the divergence of a curl is zero.*

This proposition is easy to prove by just calculating all the terms and using the compatibility of order in mixed second partial derivatives. The proofs are left as exercises.

This proposition is also a glimpse of a very general result in analysis. If we work with the right definitions, applying the same (or similar) differential operator twice in a row should give the zero operator. Obviously, this isn't always true: partial derivatives certainly don't satisfy this, and even for  $\nabla$ ,  $\nabla \cdot \nabla f \neq 0$  for a general scalar field. Near the end of these notes, we will give a general structure for understanding why differential operators should compose to zero.

**Example 3.3.11.** Consider the scalar field  $f(x, y) = \ln(x^2 + y^2)$ , which is defined everywhere except the origin.

$$\begin{aligned}\nabla f &= \left( \frac{2x}{x^2 + y^2}, \frac{2y}{x^2 + y^2} \right) \\ \nabla^2 f &= \frac{2}{x^2 + y^2} - \frac{4x^2}{(x^2 + y^2)^2} + \frac{2}{x^2 + y^2} - \frac{4y^2}{(x^2 + y^2)^2} = 0\end{aligned}$$

This is a harmonic scalar field. In particular, it can have no extrema on an open set. Values approach  $-\infty$  near the origin and  $\infty$  as we get very far from the origin.

**Example 3.3.12.** Consider the vector field  $F(x, y, z) = (-z, x, -y)$ .  $\nabla \times F = (-1, 1, 1)$ , so the vector field causes the same local rotation at all points, about the axis  $(-1, 1, 1)$  (as a local axis direction).  $\nabla \cdot F = (0, 0, 0)$ , so the flow is incompressible. This is a good field to potentially model a liquid flow that induces a particular local rotation everywhere.

**Example 3.3.13.** Consider the vector field  $F(x, y, z) = (x^2, y^2, z^2)$ .  $\nabla \times F = (0, 0, 0)$ , so the field is irrotational. The divergence is  $\nabla \cdot F = (2x + 2y + 2z)$ . In the positive octant, the flow accelerates away from the origin, so the rate of diffusion increases away from the origin. In other octants, we may have negative divergence, reflecting the fact that the local vector field directions are always positive.

**Example 3.3.14.** Consider the vector field  $F(x, y, z) = (-y, x, 0)$  (which should be fairly familiar by this time).  $\nabla \times F = (0, 0, 2)$ , which shows the creation of local rotation about the  $z$  axis. Note that this is local rotation, which is separate from the global rotation of the integral curves about the origin. Also,  $\nabla \cdot F = 0$ , so the flow is incompressible. As the flow spins around, it neither collects or diffuses anywhere.

**Example 3.3.15.** Consider the vector field  $F(x, y, z) = (\sin x, \cos y, 0)$ .  $\nabla \times F = (0, 0, 0)$ , so the flow is irrotational. This is interesting, given the trigonometric term; trigonometry might cause us to expect spin.  $\nabla \cdot F = \cos x - \sin y$ , so there are various areas of diffusion and collection. The trigonometric term here causes diffusion/collection, not rotation.

**Example 3.3.16.** Consider the vector field  $F(x, y, z) = \left(\frac{1}{x+y}, \frac{1}{x+y}, 0\right)$ .  $\nabla \times F = \left(0, 0, \frac{-1}{(x-y)^2} - \frac{-1}{(x+y)^2}\right)$ : this curl is always negative and always about the  $z$  axis.  $\nabla \cdot F = \frac{-1}{(x+y)^2} + \frac{1}{(x-y)^2}$ : therefore,  $F$  collects when  $x$  and  $y$  are dissimilar and diffuses when  $x$  and  $y$  are quite close.

## 3.4 Conservative Vector Fields

An important class of vector fields are gradients of scalar fields. The most common case is a potential energy scalar field, where  $f$  measures potential energy. A principle of motion in physics states that particles will seek to lower their potential energy in the most efficient way possible. That is, they will move in the direction of greatest descent of potential energy. That movement is precisely along the integral curves of the vector field  $\nabla f$ . So, if we try to explain a physical situation by a potential energy field, then the movement of particles is equivalent to finding integral curves of  $\nabla f$ . Moreover, movement along these paths creates kinetic energy. The speed of movement can be determined by conservation of energy: the movement along the paths has speed and kinetic energy gained at exactly the same rate as potential energy is lost. Conservation of energy explains the terminology of this section.

**Definition 3.4.1.** Let  $f$  be a scalar field. Then field  $\nabla f$  created from a potential energy field is called a *conservative vector field*.

If we are given a general vector field  $F$ , we would like to check whether or not it is conservative. We have the identity  $\nabla \times \nabla f$ , which shows that all conservative fields are irrotational. Is the converse true? We need a technical definition before we can answer this question.

**Definition 3.4.2.** Let  $U$  be an open subset of  $\mathbb{R}^n$ .  $U$  is called *path connected* if there is a parametric curve connecting any two points  $a, b \in U$ . (Recall in our definition in Calculus III that parametric curves are always continuous.) If  $U$  is path connected, it is also called *simply connected* if any closed path (a path which starts and ends at the same spot) can be contracted down to a point. (Think of the path as a loop in a rope, and contraction as pulling the rope so that the loop disappears. This needs to happen in such a way that all the intermediate steps are still paths in the set  $U$ ).

Naïvely, a simply connected set has no holes in it. A solid cylinder is simply connected. However, the hollow cylinder is not, since a loop around the cylinder can never be contracted.

**Proposition 3.4.3.** Let  $U$  be a simply connected open set in  $\mathbb{R}^3$  and let  $F : U \rightarrow \mathbb{R}^3$  be a differentiable vector field. Then  $F$  is conservative on  $U$  if and only if  $\nabla \times F = 0$ .

The condition  $\nabla \times F$  works both ways under a reasonable condition for the domain of the vector field. We can check if a vector is conservative by taking the curl; if the curl is zero, it is conservative.

If  $F$  is conservative, then there exists at least one (in fact, infinitely many!) scalar fields  $f$  with  $F = \nabla f$ . This  $f$ , inspired by the physics situation, is called a (scalar) potential for  $F$ . If  $f$  is a potential energy

field,  $F$  is the associated field of force. Usually, both the potential energy and force fields are defined per unit mass or charge, so that they can act on any mass or charge present in the field.

How do we calculate  $f$ ? The equation  $F = \nabla f$  can be expanded.

$$\begin{aligned}F_1 &= \frac{\partial f}{\partial x} \\F_2 &= \frac{\partial f}{\partial y} \\F_3 &= \frac{\partial f}{\partial z}\end{aligned}$$

We can integrate each of these equations. We have to be careful with the ‘constants’: in each integral, the constants can be functions of both of the remaining variables.

$$\begin{aligned}f &= \int F_1 dx + g_1(y, z) \\&= \int F_2 dx + g_2(x, z) \\&= \int F_3 dx + g_3(x, y)\end{aligned}$$

Finding  $f$  amounts to finding a scalar field that fits this system. For reasonable  $F$ , this isn’t too terrible. We expect a family of solutions: since we are integrating and there will be constants of integration introduced. Initial values can determine the constant to give a unique potential. The constants here make sense: potential energy is always a relative measure, so we have to set a base level of potential energy. Please note: energy is always relative! Unlike mass and charge, energy is not an intrinsic measure. Energy is a fiction, but a very useful fiction for explaining motion.

### 3.5 Line Integrals of Vector Fields

As the start of this course, We integrated scalar fields over various regions in  $\mathbb{R}^3$ . Now that we have defined vector fields, we can ask if there is a reasonable definition for integration of a vector field. This seems like an odd question: integrating scalar fields, such a density, made sense. What does it mean to integrate a vector field?

We will start with integrating along a parametric curve. If we think of the vector field as a force field, then the integral of the vector field along the path will be interpreted as the *work* to move through the field. Alternatively, we can interpret the line integral as the work that the field accomplished to move the object. (The sign of the vector field can be changed to accomodate this variable perspective: working with or working against the field.)

If  $F$  represents resistance to movement, like walking into the wind, then we look at the vectors as directions pushing with or against the movement. If  $\gamma$  is movement through a vector field, then its

local direction at any moment is its tangent  $\gamma'(t)$ . To determine whether or not we are moving into the wind, with the wind, or sideways through it, we have to compare the vector field and the tangent. We use dot product:  $F \cdot \gamma'(t)$  is large if the vectors share a direction and zero when the vectors are perpendicular.

However, we want this integral to only depend on the path, not the parametrization. Therefore, we should use attributes of the curve which are not dependant on the parametrization. The unit tangent  $T(s)$  is one such attribute (in terms of the arclength parameter  $s$ .  $F \cdot T$  measures the interaction of the movement along the curve with the vector field. The arclength infinitesimal  $ds$  is independent of parametrization. We use these tools to give the definition.

**Definition 3.5.1.** The *line integral* of a vector field  $F$  along a path  $\gamma(s) : [0, L] \rightarrow \mathbb{R}^n$  is the integral of the scalar  $F(\gamma(s)) \cdot T(s)$  along the length of the curve.

$$\int_{\gamma} F \cdot T ds = \int_{\gamma} F \cdot ds := \int_0^L F(\gamma(s)) \cdot T(s) ds$$

The first notation is more complete, but the shorter second notation is conventional.

Using the arclength parameter for the definition if appropriate but inconvenient for calculation. Let  $\gamma(t)$  be an arbitrary parametrization. Then we can think of  $s(t)$  (the arclength of the curve in terms of  $t$ ) as a substitution of single variable integration. This substitution has  $ds = |\gamma'(t)|dt$  and  $T = \gamma'(t)/|\gamma'(t)|$ , and the new bounds are 0 and  $L$ , the length of the curve. We can then transform the integral in the definition.

$$\int_{\gamma} F \cdot T ds = \int_0^L F(\gamma(s)) \cdot T(s) ds = \int_a^b F(\gamma(t)) \cdot \frac{\gamma'(t)}{|\gamma'(t)|} |\gamma'(t)| dt = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt$$

The line integral measures the work or effort it takes to move along the path through the vector field. It integral doesn't depend at all on the speed of moving through the vector field, since it is independent of the parametrization. The line integral is positive if it generally goes with the direction of the field and negative if it generally goes against. Since moving with  $F$  is positive and moving against  $F$  is negative, we sometime will want to replace  $F$  with  $-F$  when we want to think of working against the field as positive work.

**Example 3.5.2.** Let's do a classic example: the force of gravity. We've already discussed the field (per unit mass) generated by a mass  $M$  at the origin.

$$F = \frac{GM}{\sqrt{(x^2 + y^2 + z^2)^3}}(-x, -y, -z)$$

We're going to reverse the sign here, since we want to do positive work when we go against gravity.

$$F = \frac{GM}{\sqrt{(x^2 + y^2 + z^2)^3}}(x, y, z)$$

Let's take an outward path  $\gamma(t) = (t, t, t)$  for  $t \in [a, b]$  and  $b > a > 0$ . Then

$$\begin{aligned}\gamma' &= (1, 1, 1) \\ F(\gamma(t)) &= \frac{GM}{\sqrt{(3t^2)^3}}(t, t, t) \\ F(\gamma(t)) \cdot \gamma'(t) &= \frac{3tGM}{3\sqrt{3}t^3} = \frac{GM}{t^2\sqrt{3}} \\ \int_{\gamma} F \cdot ds &= \int_b^a \frac{GM}{t^2\sqrt{3}} = \frac{GM}{\sqrt{3}} \int_a^b \frac{1}{t^2} dx \\ &= \frac{GM}{\sqrt{3}} \left. \frac{-1}{t} \right|_a^b = \frac{GM}{\sqrt{3}} \left( \frac{1}{a} - \frac{1}{b} \right) \\ &= \frac{GmM(b-a)}{ab\sqrt{3}}\end{aligned}$$

If the distances  $a$  and  $b$  are large, but  $a$  and  $b$  are close to each other, then the change  $a - b$  is much more significant than the change in the term  $ab$  in the denominator. If we call  $g = \frac{GM}{\sqrt{3}ab}$  and pretend this is locally constant, the line integral (approximately) evaluates to  $g(b - a)$ . Recall we were working with force per unit mass, so if we act on a mass  $m$  we get a change in potential of  $mg(b - a)$ . This is the familiar high-school physics result of  $mgh$ , the increase in potential energy, here  $h$  is the change in height and  $g$  the local acceleration due to gravity. (Proving that our expression for  $g$  is the local acceleration due to gravity is a little tricky, but we can see, using the fact that  $G$  has units  $m^3/(kg \cdot s^2)$ , that the units of  $g$  are  $m/s^2$ , so it is an acceleration term).

**Example 3.5.3.** For another example, consider the rotational flow  $F(x, y) = (-y, x)$  and  $\gamma(t)$  the counterclockwise circle  $(r \cos t, r \sin t)$  for  $t \in [0, 2\pi]$ .

$$\begin{aligned}\gamma'(t) &= (-r \sin t, r \cos t) \\ F(\gamma(t)) &= (-r \sin t, r \cos t) \\ F(\gamma(t)) \cdot \gamma'(t) &= r^2 \sin^2 t + r^2 \cos^2 t = r^2 \\ \int_{\gamma} F \cdot ds &= \int_0^{2\pi} r^2 dr = 2\pi r^2\end{aligned}$$

The work to move around the circle (with the vector field) of radius  $r$  is  $2\pi r^2$ . Since it is positive and follows the integral curve, we can think of it as the work the field accomplished to move the object along its integral curve.

### 3.5.1 Fundamental Theorems

Let's return to a version of the Fundamental Theorem of Calculus. Let  $f(x)$  be a differentiable function.

$$\int_a^b \frac{df}{dx} dx = f(b) - f(a)$$

This integration takes place on the interval  $[a, b]$  and is evaluated at the end points  $a$  and  $b$ . We can think of the endpoints as the *boundary* of the interval. With this perspective, the left side is the integral of some kind of derivative and the right side is the evaluation of the original function on a boundary. If we let  $df$  be any kind of derivative, and  $\int_{\partial S}$  be an integral which relates to the boundary of a set, then when can write the archtypical form of this theorem.

$$\int_S df = \int_{\partial S} f$$

For the fundamental theorem, the ‘integral’ on the right-hand side is just evaluation, since the boundary is only the two points. Integration over a zero-dimensional regions, such as a set of points, is just evaluation. Using this integration notation, the fundamental theorem is only the first of a whole family of theorems of this type. They all relate a derivative operator on the left to a boundary operator on the right.

### 3.5.2 The Fundamental Theorem of Line Integrals

The first of these new fundamental theorems comes from thinking about line integrals of paths through conservative vector fields.

**Theorem 3.5.4.** *Let  $F = \nabla f$  be a conservative vector field with potential  $f$  and let  $\gamma(t) : [a, b] \rightarrow \mathbb{R}^n$  be a parametric curve. The fundamental theorem of line integrals shows we can solve this integral by evaluation on the endpoints of the curve.*

$$\int_{\gamma} F \cdot ds = f(\gamma(b)) - f(\gamma(a))$$

*Proof.* We calculate the integral in steps.

$$\begin{aligned} \int_{\gamma} F \cdot ds &= \int_{\gamma} \nabla f \cdot T ds \\ &= \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t) dt \\ &= \int_a^b \left( \frac{\partial f}{\partial x}(\gamma(t)), \frac{\partial f}{\partial y}(\gamma(t)), \frac{\partial f}{\partial z}(\gamma(t)) \right) \cdot (\gamma'_1(t), \gamma'_2(t), \gamma'_3(t)) dt \\ &= \int_a^b \frac{\partial f}{\partial x}(\gamma(t))\gamma'_1(t) + \frac{\partial f}{\partial y}(\gamma(t))\gamma'_2(t) + \frac{\partial f}{\partial z}(\gamma(t))\gamma'_3(t) dx \\ &= \int_a^b \frac{d}{dt} (f(\gamma(t))) dt \\ &= f(\gamma(b)) - f(\gamma(a)) \end{aligned}$$

We used multi-variable chain rule in the fourth line of the calculation. □

Side by side, we can see the similarity of the theorem with the original fundamental theorem: derivatives on one side and (evaluation on) boundaries on the other side.

$$\int_a^b \frac{df}{dx} dx = f(b) - f(a)$$

$$\int_{\gamma} \nabla f \cdot ds = f(\gamma(b)) - f(\gamma(a))$$

### 3.5.3 Implications for Conservative Vector Fields

The fundamental theorem for line integral applies to conservative vector fields: it says that the integral of a conservative vector field can be calculated simply by evaluation on the end points. That means all the points on the path inbetween are irrelevant!

**Proposition 3.5.5.** *Let  $F = \nabla f$  be a conservative vector field. Then line integrals of  $F$  are path-independent. Their value only depends on the endpoints.*

No matter how strange the path is, no matter how many times it loops around and goes elsewhere, only the endpoints matter. If  $f$  is potential energy, then  $f(\gamma(b)) - f(\gamma(a))$  is the potential energy at the end less the potential energy at the start. This difference is the change in potential energy. The work to move through the force field is equal to the change in potential energy. The work can be defined in terms of kinetic energy lost or gained, so this says that the kinetic energy lost or gained is equal to the change in potential energy. Therefore, energy is conserved! This explains the term ‘conservative vector field’: these are vector fields where conservation of energy makes sense. (As before, there are sign issues in the fiction of energy and work: we may need to take the negative of the usual or expected potential energy to make conservation of energy make sense.)

We can calculate very explicitly in terms of the gain in kinetic energy as well. Let  $F = \nabla f$  be a conservative force. Then  $F = ma$ , where  $a$  is acceleration. Also, if  $\gamma$  is a curve representing movement, then  $\gamma''$  is the acceleration. Then we look carefully at the change in potential energy using our new theorem.

$$\begin{aligned} -\Delta PE &= f(\gamma(b)) - f(\gamma(a)) \\ &= \int_{\gamma} F \cdot T ds \\ &= \int_a^b m\gamma''(t) \cdot \gamma'(t) dt \\ &= \frac{m}{2} \int_a^b \frac{d}{dt} (\gamma'(t) \cdot \gamma'(t)) dt \\ &= \frac{m}{2} \int_a^b \frac{d}{dt} |\gamma'(t)|^2 dt \\ &= \frac{m}{2} |\gamma'(t)|^2 \Big|_a^b \\ &= \frac{mv_b^2}{2} - \frac{mv_a^2}{2} = \Delta KE \end{aligned}$$

Previously, we saw that conservative vector fields were irrotational and, conversely, an irrotational field defined on a simply connected set was guaranteed to be conservative. We now have a new characterization of conservative fields: their line integrals are path independent. We can similarly ask if the converse holds and again, we find a topological condition. First, we start with a slightly more narrow proposition.

**Proposition 3.5.6.** *Let  $F$  be a vector field on  $U$  an open path-connected set in  $\mathbb{R}^n$ . If  $\int_{\gamma} F = 0$  for all closed paths  $\gamma$  in  $U$ , then  $F$  is conservative.*

A full characterization of conservative fields is given by this more-complete proposition.

**Proposition 3.5.7.** *Let  $F$  be a vector field on  $U$  an open simply-connected set in  $\mathbb{R}^n$  and let  $\gamma : [a, b] \rightarrow U$  be a parametric curve. The following four conditions are equivalent.*

- (a)  $F$  is conservative on  $U$ . (By definition, this means  $F = \nabla f$  for a scalar field  $f$ ).
- (b) The line integrals of  $F$  are path independent in  $U$ .
- (c) The line integral of  $F$  is zero for any closed path in  $U$ .
- (d)  $\nabla \times F = 0$ .

### 3.5.4 Examples of Conservative Vector Fields

**Example 3.5.8.** If  $F = (y \cos(xy), x \cos(xy), 2z)$  then  $F$  is defined on  $\mathbb{R}^3$ , which is simply connected.

$$\begin{aligned} \nabla \times F &= \left( \frac{\partial}{\partial y} 2z - \frac{\partial}{\partial z} x \cos(xy), \frac{\partial}{\partial z} y \cos(xy) - \frac{\partial}{\partial x} 2z, \frac{\partial}{\partial x} x \cos(xy) - \frac{\partial}{\partial y} y \cos(xy) \right) \\ &= (0, 0, \cos xy - xy \sin xy - \cos xy + xy \sin xy) = (0, 0, 0) \end{aligned}$$

Therefore,  $F$  is a conservative field. We can try to calculate its potential.

$$f = \int y \cos(xy) dx + g_1(y, z) = \sin(xy) + g_1(y, z)$$

$$f = \int x \cos(xy) dy + g_2(x, z) = \sin(xy) + g_2(x, z)$$

$$f = \int 2z dz + g_3(x, y) = z^2 + g_3(x, y)$$

$$f = \sin xy + z^2 + c$$

We can check  $\nabla f$  to see that we recover the original field  $F$ .

**Example 3.5.9.** If  $F = (3yz - 8xz^2 + 3x^2y, 6xyz + x^3 + 2yz^2, 3xy^2 - 8x^2z + 2y^2z)$ , then  $F$  is defined on all  $\mathbb{R}^3$ .

$$\nabla \times F = (6xy + 4yz - 6xy - 4yz, 3y^2 - 16z - 3y^2 + 16z, 6yz + 3x^2 - 3yz - 3x^2) = (0, 0, 0)$$

Therefore,  $F$  is conservative. We can try to calculate its potential.

$$f = \int F_1 dx = 3xy^2z - 4x^2 + z^2 + x^3y + g_1(y, z)$$

$$f = \int F_2 dx = 3xy^2z + x^3y + y^2z^2 + g_2(x, z)$$

$$f = \int F_3 dx = 3xy^2z - 4x^2z^2 + y^2z^2 + g_3(x, y)$$

$$f = 3xy^2z - 4x^2z^2 + x^3y + y^2z^2 + c$$

### 3.6 Parametric Surfaces

We've defined integrals of vector fields along parametric curves. However, parametric curves are only one dimensional objects. What does it mean to integrate a vector field over higher dimensional objects?

Vector fields are not like scalar fields, where we can just integrate over regions of  $\mathbb{R}^3$ . We take a hint from line integrals using parametric curves: we need parametric objects. We can define parametric objects in any dimension, though the theory gets more and more difficult with added dimensions. In this section we define parametric surfaces, which are the 2-dimensional analogue to parametric curves.

If  $[a, b]$  is an interval on  $\mathbb{R}$ , a curve was a function on that interval into a vector space.

$$\gamma : [a, b] \rightarrow \mathbb{R}^3$$

The curve is a *continuous* (usually differentiable) function on the interval. For surfaces, we need two directions of freedom, so we start in  $\mathbb{R}^2$  instead of  $\mathbb{R}$ . We might start with intervals  $I$  in  $\mathbb{R}^2$ , but we can be a bit more general.

**Definition 3.6.1.** Let  $D \subset \mathbb{R}^2$  be a *simply connected* set. A parametric surface in  $\mathbb{R}^3$  is a *continuous* (usually differentiable) function on  $D$  into a vector space.

$$\sigma : D \rightarrow \mathbb{R}^3$$

Like parametric curves, we visualize surfaces only by their outputs. The input set  $D$  isn't visualized; rather, it is thought of as the parameter space. The difference, compared to curves, is that we now have a 2-dimensional parameter space. This allows two independent directions of movement on the surface.

We need to refine the previous definition a little bit: we want to avoid places where the surface has strange behaviour by insisting that the surface is 'smooth'. Differentiability is required here, but the condition is actually a bit stronger. We need to think about tangents.

For a curve  $\gamma(t)$ , the only derivative was the time derivative and  $\gamma'(t)$  was the tangent. For a surface, let  $u$  and  $v$  be the coordinates in the region  $D$ . Then we have two partial derivatives:  $\sigma_u = \frac{\partial}{\partial u} \sigma(u, v)$  and  $\sigma_v = \frac{\partial}{\partial v} \sigma(u, v)$ . These are both tangents.

**Definition 3.6.2.** A parametric surface is called *smooth* if it has two (linearly) independent tangent directions at any point. Working in  $\mathbb{R}^3$ , we can calculate  $\sigma_u \times \sigma_v$ . This is perpendicular to both tangents, so this direction should be normal to the surface. We will use that name, and call  $\sigma_u \times \sigma_v$  the *normal* to a smooth parametric surface.

The smoothness condition can be rephrased: the normal must never be zero. We will be using the normal extensively and zero normals will cause many problems. Therefore, from this point, we assume all surfaces are piece-wise smooth, with non-zero normals. Piece-wise is enough, since the cross-over sections will be lower-dimensions and thus can be ignored in integration. (For curves, we made implicit use of a similar convention that the tangent should never be the zero vector, though we didn't say so at the time.) Since we are using the cross-product, this definition applies to  $\mathbb{R}^3$ . For the rest of this entire chapter, we will exclusively be working with parametric surfaces in  $\mathbb{R}^3$ .

**Example 3.6.3.** If  $f(x, y)$  is a 2-variable function on a simply-connected domain  $D$ , its graph is a surface. We can use  $D$  as the domain and  $\sigma(u, v) = (u, v, f(u, v))$  is the graph expressed as a parametric surface.

**Example 3.6.4.** Consider a surface of rotation. If we rotate  $y = f(x)$  about the  $x$  axis for  $x \in [a, b]$ , then we can take  $\theta$  to be other parameter and  $D = [a, b] \times [0, 2\pi]$  to be the parameter domain. Then  $\sigma(x, \theta) = (x, f(x) \cos \theta, f(x) \sin \theta)$  is the surface of revolution.

**Example 3.6.5.** If  $(\theta, \phi) \in D = [0, 2\pi] \times [0, \pi]$  then  $\sigma(\theta, \phi) = (r \sin \phi \cos \theta, r \sin \phi \sin \theta, r \cos \phi)$  is a parametric description of the sphere with radius  $r$ .

**Example 3.6.6.** If  $(\theta, z) \in D = [0, 2\pi] \times [0, h]$  then  $\sigma(\theta, z) = (r \cos \theta, r \sin \theta, z)$  is a parametric description of a cylinder with height  $h$  and radius  $r$  oriented along the  $z$ -axis.

**Example 3.6.7.** If  $(u, v) \in D = [0, h] \times [0, 2\pi]$  then  $\sigma(u, v) = (u \cos v, u \sin v, u)$  is a parametric description of a cone opening upwards in the  $z$  direction with height  $h$  and base radius  $h$  as well.

## 3.7 Surface Integrals

Recall the definition of the line integral.

$$\int_{\gamma} F \cdot T ds = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt$$

On the left, we started with intrinsic information, using the arclength parametrization and infinitesimal  $ds$  to build a definition which didn't depend on parametrization. Then we showed, using  $ds = |\gamma'(t)| dt$  how to calculate the line integral for an arbitrary description of length in terms of the parametrization. We have something similar for surfaces. Instead of length, we have an infinitesimal area  $dA$  which is independent of the parametrization. If  $\sigma(u, v)$  is a function of the parameters  $u$  and  $v$  on a simply connected domain  $D$ , through a change of variables, we can express the intrinsic  $dA$  in terms of the specific parameters  $u$  and  $v$ . For curves, the mediating factor was the length of the tangent; for surfaces, it is the length of the normal.

$$dA = |\sigma_u \times \sigma_v| du dv$$

This gives us a local area differential for integration.

**Definition 3.7.1.** Let  $\sigma(u, v)$  be a parametric surface defined on a simply connected open set  $D$ . Let  $f$  be a scalar field defined in a neighbourhood of the surface  $\sigma$ . The integral of  $f$  over  $\sigma$  is defined as follows.

$$\int_{\sigma} f dA = \int_D f(\sigma(u, v)) |\sigma_u \times \sigma_v| dudv$$

This definition lets us calculate surface area of parametric surfaces. To find area, we just integrate the constant scalar field  $f = 1$ .

$$A = \int_{\sigma} 1 dA = \int_D |\sigma_u \times \sigma_v| dudv$$

If we expand this out, we can get the full expression in all its complexity.

$$A = \int_D \sqrt{\left(\frac{\partial \sigma_2}{\partial u} \frac{\partial \sigma_3}{\partial v} - \frac{\partial \sigma_2}{\partial v} \frac{\partial \sigma_3}{\partial u}\right)^2 + \left(\frac{\partial \sigma_3}{\partial u} \frac{\partial \sigma_1}{\partial v} - \frac{\partial \sigma_3}{\partial v} \frac{\partial \sigma_1}{\partial u}\right)^2 + \left(\frac{\partial \sigma_1}{\partial u} \frac{\partial \sigma_2}{\partial v} - \frac{\partial \sigma_1}{\partial v} \frac{\partial \sigma_2}{\partial u}\right)^2} dudv$$

**Example 3.7.2.** Let  $\sigma$  be the graph of a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  as defined in Example 3.6.3. Then we can use the area formula to write the general equation of the surface area of the graph of  $f$  over a simply-connected region  $D$  in its domain.

$$A = \int_D \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1} dx dy$$

**Example 3.7.3.** Let  $f : [a, b] \rightarrow \mathbb{R}$  be a differentiable function and let  $\sigma$  is a surface of revolution as defined in Example 3.6.4. Let  $D = [a, b] \times [0, 2\pi]$ . Then we can define the surface of such a general surface of revolution.

$$A = \int_D \sqrt{1 - \left(\frac{\partial f}{\partial x}\right)^2} f(x) dx$$

**Example 3.7.4.** The paraboloid  $z = k(x^2 + y^2)$  can be described by  $\sigma(u, v) = (u, v, k(u^2 + v^2))$  over  $D$  a circle of radius  $r$ . We can calculate it surface area.

$$\begin{aligned} A &= \int_D \sqrt{4k^2 u^2 + 4k^2 v^2 + 1} dudv \\ &= \int_0^{2\pi} \int_0^r \sqrt{4k^2 r^2 + 1} r dr d\theta \\ u &= 4k^2 r^2 + 1 \implies du = 8k^2 r dr \\ &= 2\pi \int_1^{4hk} \sqrt{u} \frac{1}{8k^2} du \\ &= \frac{\pi}{4k^2} \left. \frac{2u^{\frac{3}{2}}}{3} \right|_1^{4hk} \\ &= \frac{\pi}{6k^2} \left( (4kh)^{\frac{3}{2}} - 1 \right) = \frac{2\sqrt{h^3 k^3} - 1}{6k^2} \end{aligned}$$

**Example 3.7.5.** The surface of revolution under  $f(x) = \frac{1}{x}$  for  $x \in [1, \infty)$  is called the Horn of Gabriel. It can be described parametrically for  $(x, \theta) \in [1, \infty) \times [0, 2\pi)$  as the following surface.

$$\sigma(x, \theta) = \left( x, \frac{\cos \theta}{x}, \frac{\sin \theta}{x} \right)$$

The surface area of the Horn of Gabriel is quite interesting. (The comparison results for single-variable integrals are useful in this calculation.)

$$\begin{aligned} \sigma_x &= \left( 1, \frac{-\cos \theta}{x^2}, \frac{-\sin \theta}{x^2} \right) \\ \sigma_\theta &= \left( 0, \frac{-\sin \theta}{x}, \frac{\cos \theta}{x} \right) \\ \sigma_x \times \sigma_\theta &= \left( \frac{-1}{x^3}, \frac{\cos \theta}{x}, \frac{-\sin \theta}{x} \right) \\ |\sigma_x \times \sigma_\theta| &= \sqrt{\frac{1}{x^6} + \frac{1}{x^2}} = \frac{1}{x} \sqrt{\frac{1}{x^4} + 1} \\ A &= \int_\sigma 1 = \int_1^\infty \int_0^{2\pi} \frac{1}{x} \sqrt{\frac{1}{x^4} + 1} d\theta dx \\ &= 2\pi \int_1^\infty \frac{1}{x} \sqrt{\frac{1}{x^4} + 1} dx \geq 2\pi \int_1^\infty \frac{1}{x} dx = \infty \end{aligned}$$

Compare this surface area with the volume of the Horn of Gabriel.

$$\begin{aligned} V &= \int_1^\infty \int_0^{\frac{1}{x}} \int_0^{2\pi} r d\theta dr dx \\ &= 2\pi \int_1^\infty \left. \frac{r^2}{2} \right|_0^{\frac{1}{x}} = 2\pi \int_1^\infty \frac{1}{2x^2} dx \\ &= \left. \frac{2\pi}{2} \frac{-1}{x} \right|_1^\infty = \pi \end{aligned}$$

This is a very strange situation: an object with finite volume and infinite surface area.

## 3.8 Flux Integrals

In addition to integrating scalars along surfaces, we also want to integrate vector fields. For the definition, we again look at line integrals. We took a dot product with the tangent to the curve to measure how much the field acts with or against the motion of the curve. For surfaces, we also take a dot product, but now with the normal of the surface. By doing so, we are measuring how much the vector field passes through the surface: if the dot product is large, the field direction is similar to the normal and almost perpendicular to the surface. If the dot product is small, the field moves almost in parallel with the surface, moving across it instead of through it.

We used the dot product with the unit tangent to calculate a line integral.

$$\int_{\gamma} F \cdot ds = \int_{\gamma} F \cdot T ds = \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt$$

We need something similar to the unit tangent to calculate integrals of fields over surfaces.

**Definition 3.8.1.** Let  $\sigma(u, v)$  be a parametric surface in  $\mathbb{R}^3$  defined over a simply-connected domain  $D$ . The *unit normal* of the surface is the vector  $\frac{\sigma_u \times \sigma_v}{|\sigma_u \times \sigma_v|}$ . (Our convention that the normal is never zero helps to avoid division by zero in this definition.)

**Definition 3.8.2.** Let  $\sigma(u, v)$  be a parametric surface in  $\mathbb{R}^3$  defined over a simply-connected domain  $D$ . Let  $F$  be a vector field defined on a neighbourhood of the surface. The *flux integral* is defined to be the following integral.

$$\int_{\sigma} F \cdot NdA = \int_{\sigma} F \cdot dA$$

The first notation is more complete, but the second notation is conventional.

As with curve, we don't want to calculate with the intrinsic unit normal; we'd rather calculate with a specific parametrization. In an arbitrary parametrization,  $da = |\sigma_u \times \sigma_v| dudv$ . Then we perform a change of variables on the integral in the definition.

$$\int_{\sigma} F \cdot NdA = \int_D F(\sigma(u, v)) \cdot \frac{\sigma_u \times \sigma_v}{|\sigma_u \times \sigma_v|} |\sigma_u \times \sigma_v| dudv = \int_D F(\sigma(u, v)) \cdot (\sigma_u \times \sigma_v) dudv$$

Like the line integral, this is independent of the parametrization since it is equal (by change of variables) to the intrinsic definition.

Flux measures how much of the vector field passes through the surface. This is mostly easily seen for fluids if we think of the surface as a net. In this case the flux is exactly the amount of fluid passing through the net per unit time. For fields of force, the interpretation is slightly less obvious. However, flux will still matter a great deal for calculations involving forces.

**Example 3.8.3.**  $F(x, y, z) = (0, 0, k)$  is a constant vertical flow with the same flow rate. We can think of this flowing through a vertical pipe with radius  $a$  around the  $z$  axis by restricting the field to the domain where  $x^2 + y^2 \leq a^2$ . The surface is a cross-section of the pipe, which we can parametrize as  $\sigma(r, \theta) = (r \cos \theta, r \sin \theta)$  for  $(r, \theta) \in [0, a] \times [0, 2\pi]$ .

$$\begin{aligned} \sigma_r &= (\cos \theta, \sin \theta, 0) \\ \sigma_{\theta} &= (-r \sin \theta, r \cos \theta, 0) \\ \sigma_r \times \sigma_{\theta} &= (0, 0, r) \\ \int_{\sigma} F \cdot dA &= \int_0^a \int_0^{2\pi} k r d\theta dr \\ &= \frac{2\pi a^2}{2} = \pi a^2 k = (\pi a^2) k \end{aligned}$$

$\pi a^2$  is the cross-sectional area and  $k$  is the rate of flow, so we have a uniform rate of volume of water flowing through the pipe per unit time.

**Example 3.8.4.** We could ask the same question with a different flow, one which is not uniform. Consider something like this, still moving vertically, but with variable speed.

$$\begin{aligned}
 F(x, y, z) &= \left( 0, 0, \frac{k(a^2 - x^2 - y^2)}{a^2} \right) \\
 \int_{\sigma} F \cdot dA &= \int_0^a \int_0^{2\pi} \frac{k(a^2 - r^2)}{a^2} r d\theta dr \\
 &= \frac{2\pi k}{a^2} \int_0^a a^2 r - r^3 dr \\
 &= \frac{2\pi k}{a^2} \left( \frac{a^2 r^2}{2} - \frac{r^4}{4} \right) \Big|_0^a \\
 &= \frac{2\pi k}{a^2} \left( \frac{a^4}{2} - \frac{a^4}{4} \right) = \frac{\pi k a^2}{2}
 \end{aligned}$$

This is half the original flux. The lower rate of flow is justified since the fluid speed goes to zero near the edge of the pipe.

**Example 3.8.5.** We can also consider a turbulent flow.

$$\begin{aligned}
 F(x, y, z) &= \left( \sin \left( \left( \frac{a-r}{a} \right) \pi \right), \sin \left( \left( \frac{a-r}{a} \right) \pi \right), \cos \left( \left( \frac{a-r}{a} \right) \pi \right) \right) \\
 F \cdot \sigma_r \times \sigma_z &= k \cos \left( \left( \frac{a-r}{a} \right) \pi \right) r \\
 \int_{\sigma} F \cdot dA &= \int_0^a \int_0^{2\pi} k \cos \left( \left( \frac{a-r}{a} \right) \pi \right) r dr d\theta \\
 &= 2\pi k \int_0^a \cos \left( \pi - \frac{\pi r}{a} \right) r dr \\
 &= 2\pi k \int_0^a -r \cos \left( \frac{\pi r}{a} \right) dr \\
 &= -2\pi k \frac{a}{\pi} \int_0^a \frac{\pi r}{a} \cos \left( \frac{\pi r}{a} \right) dr \\
 u &= \frac{\pi r}{a} \\
 &= -2\pi k \frac{a}{\pi} \int_0^{\pi} u \cos u \frac{adu}{\pi} \\
 &= \frac{-2\pi a^2}{\pi} (\cos u + u \sin u) \Big|_0^{\pi} = \frac{4ka^2}{\pi}
 \end{aligned}$$

**Example 3.8.6.** Consider a paddle with a roughly rectangular cross section 15cm wide and 30cm tall. Say it moves through the water in the  $z$  direction with speed  $v$ . We can reinterpret the situation by letting the water move with field  $F = (0, 0, v)$  in which the paddle is stationary. We can think of the force the paddle causes in terms of the flux of this field through the paddle.

Now, we can perform a paddle stroke directly perpendicular to the direction of movement or at an angle  $\theta$ . Our question is: how does our forward force (hence flux) vary due to  $\theta$ ?

If the paddle is angled at angle  $\theta$ , then it is represented as a surface with normal  $(\sin \theta, 0, \cos \theta)$ . Therefore  $F \cdot N = v \cos \theta$ .

$$\int_{\sigma} F \cdot N dA = \int_D V \cos \theta dA = v \cos \theta \int \sigma_1 dv = 450v \cos \theta$$

Compared to the perpendicular force or  $450v$ , we just multiply by  $\cos \theta$ .

## 3.9 The Major Results

We we discussed the fundamental theorem of calculus earlier in these notes, we identified a particular archetypical style.

$$\int_{\partial D} f = \int_D df$$

Here,  $\partial$  is a boundary operator for sets and  $df$  is some kind of derivative of  $f$ . There are several very important theorems involving flux integrals which have this style. Before we state them, we have to talk a little about boundaries of sets and orientation.

### 3.9.1 Boundaries and Orientation

Topologically, we have already defined the boundary of a set. The boundaries we want to use are slightly different from the topological definition. Our sets here are parametric objects, so the boundaries are the edges in terms of the parameter. For curves, these were the endpoints, and the boundary integral was evaluation on the endpoints. However, evaluation on one endpoint was positive and the other was negative. These boundaries have orientation, which affects the integrals. For curves, orientation is direction and curves move from negative to positive. This gives the positive or negative signs for evaluation on the endpoints.

For surfaces, orientation is a choice of an outward or inward direction. It is indicated by the direction of the (outward facing) normal vector. If the surface has an edge, then we can think of the edge as a closed parametric curve; the curve traces the path around the edge. We need the direction of the curve and the orientation of the surface to compatible. This is accomplished by a right-hand-rule. If we trace the path around the surface, then the outward pointing normal of the surface must satisfy a right-hand-rule for this rotation.

Lastly, we have solid regions in  $\mathbb{R}^3$ . By themselves, solid regions have no orientation. However, solid regions have boundaries which are closed surfaces (surfaces without edges). By convention, we insist that the surface which bounds a solid region has an outward pointing normal: the normal points away from the region, not into it.

Note that on the objects discussed in this section, the boundary operator  $\partial$  satisfies  $\partial^2 = 0$ . If we have a solid region, its boundary is a closed surface, i.e., a surface with no boundary. If we have a surface, its boundary is a closed curve, i.e., a curve with no boundary. And the boundary of a curve is two points, which don't have boundaries for dimension reasons.

### 3.9.2 Gauss-Green-Stokes

Now we can state the two main theorems for vector calculus. The theorems are stated for piece-wise smooth objects, which means we allow for some number of sharp corners. The first is called Stokes' Theorem.

**Theorem 3.9.1.** *Let  $\sigma$  be a piecewise-smooth oriented parametric surface in  $\mathbb{R}^3$ , with boundary  $\partial\sigma = \gamma$  a piecewise smooth oriented closed curve. Let  $F$  be a  $C^1$  vector field defined on a neighbourhood of  $\sigma$ .*

$$\int_{\partial\sigma} F \cdot ds = \int_{\sigma} (\nabla \times F) \cdot dA$$

This theorem has the desired form. We have a boundary on the left, and a differential operator (curl, in this case) on the right. The second result is called Gauss' Theorem.

**Theorem 3.9.2.** *Let  $D$  be a simply-connected region in  $\mathbb{R}^3$  whose boundary  $\partial D = \sigma$  is a piecewise smooth oriented closed parametric surface. Let  $F$  be a  $C^1$  vector field defined on a neighbourhood of  $D$ .*

$$\int_{\partial D} F \cdot dA = \int_D (\nabla \cdot F) dV$$

This theorem also has the desired form. We have a boundary on the left and a differential operator (divergence, in this case) on the right.

Lastly, there is a specialization of Stokes' theorem for  $\mathbb{R}^2$ , called Green's theorem.

**Theorem 3.9.3.** *Let  $D$  be a simply-connected region in  $\mathbb{R}^2$  and let  $F = (F_1, F_2)$  be a  $C^1$  vector field defined on a neighbourhood of  $D$ . We can interpret  $\mathbb{R}^2$  as the  $xy$  plane in  $\mathbb{R}^3$  and extend  $F$  as  $F = (F_1, F_2, 0)$  as a vector field in  $\mathbb{R}^3$ . We can interpret  $D$  as a parametric surface in  $\mathbb{R}^3$  which is restricted to the  $xy$  plane, using  $x$  and  $y$  as parameters and the identity function  $D \rightarrow D$  as  $\sigma$ . In this setting, we can apply Stokes' to get Green's theorem.*

$$\int_{\partial D} F \cdot ds = \int_D (\nabla \times F) \cdot dA = \int_D \left( \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) dx dy$$

Since these three named theorems are of similar form, they are often referred to as Gauss-Green-Stokes.

### 3.9.3 Examples of Gauss-Green-Stokes

**Example 3.9.4.** Let's start in  $\mathbb{R}^2$  with Green's theorem. Let  $F(x, y) = (\sin x, x^2 y^3)$  and let  $\sigma$  be the triangle with vertices  $(0, 0)$ ,  $(2, 0)$  and  $(2, 2)$ . Then  $\gamma$  is the path in three straight lines from  $(0, 0)$  to  $(2, 0)$  to  $(2, 2)$  and back to  $(0, 0)$ . (Now we see why we need piece-wise smooth conditions: allowing these sharp corners on the boundary path is very reasonable.) The line integral of  $F$  over  $\gamma$  could

be calculated in three steps, by parametrizing each line segments and calculating those line integrals. Green's theorem gives us an alternate way to calculate the line integral.

$$\begin{aligned}\int_{\gamma} F \cdot ds &= \int_{\sigma} (\nabla \times F) \cdot dA = \int_{\sigma} (2xy^3 - -0) dx dy \\ &= \int_0^2 \int_0^x 2xy^3 dy dx \\ &= 2 \int_0^2 \frac{xy^4}{4} \Big|_0^x dx \\ &= 2 \int_0^2 \frac{x^5}{4} dx = \frac{x^6}{12} \Big|_0^2 = \frac{2^6}{12} = \frac{16}{3}\end{aligned}$$

**Example 3.9.5.** Let  $F = (\sin x^2, e^{y^2} + x^2, z^2 + 2x^2)$  and let  $\gamma$  be the counterclockwise path around the triangle in  $\mathbb{R}^3$  with vertices  $(3, 0, 0)$ ,  $(0, 2, 0)$  and  $(0, 0, 1)$ . We could parametrize the three pieces of the triangle, but lets use Stokes' theorem with the triangle as the surface. The triangle is flat, so it may be parametrized with a constant normal. The equation of the plane containing the triangle is  $\frac{x}{3} + \frac{y}{2} + z = 1$ , with normal  $(\frac{1}{3}, \frac{1}{2}, 1)$ . The curl of  $F$ , after some algebra is  $(0, -4x, 2x)$ . (Here the curl is much easier to work with than the original  $F$ !) Then  $\nabla \times F \cdot N = (0, -4x, 2x) \cdot (\frac{1}{3}, \frac{1}{2}, 1) = 0$ . This lets us treat the original (very complicated) line integral as a very straightforward surface integral.

$$\int_{\gamma} F \cdot ds = \int_{\sigma} F \cdot dA = \int_{\sigma} F \cdot N dA = \int_{\sigma} 0 dA = 0$$

**Example 3.9.6.** Let  $F$  be any  $C^1$  vector field and let  $\sigma = S^2$ , the unit sphere in  $\mathbb{R}^3$ .

$$\int_{\sigma} \nabla \times F = \int_{\partial\sigma} F \cdot ds = 0$$

The last integral is zero because the sphere doesn't have any boundary. This is true for any closed surface: Stokes' theorem tells us that the integral of a curl over any closed surface must be zero.

We can ask: what does it mean to consider the flux of curl? If there is a positive flux of  $\nabla \times F$  through a surface, this indicates the (local) tendency of the flow to move in a counterclockwise rotation across the surface.

**Example 3.9.7.** Let  $\sigma(u, v) = (u, v, 4 - u^2 - v^2)$  for  $(u, v)$  in the disc about the origin of radius 2,  $(u^2 + v^2 \leq 4)$ . This  $\sigma$  is a paraboloid above the  $xy$  plane. Its bounding curve is the circle of radius 2 about the origin in the  $xy$  plane, going counterclockwise (to fit the upward normal of the surface with a right-hand-rule). The boundary curve is parametrized as  $\gamma(t) = (2 \cos t, 2 \sin t, 0)$ . Let  $F(x, y, z) = (z^2, x^2, y^2)$ . Then we can calculate the line integral of  $F$  over the boundary curve.

$$\begin{aligned}\int_{\gamma} F \cdot ds &= \int_0^{2\pi} F(\gamma(t)) \cdot \gamma'(t) dt \\ &= \int_0^{2\pi} (0, 4 \cos^2 t, 4 \sin^2 t) \cdot (-2 \sin t, 2 \cos t, 0) dt \\ &= \int_0^{2\pi} 8 \cos^3 t dt = \frac{8}{3} (2 + \cos^2 2t) \sin t \Big|_0^{2\pi} = 0\end{aligned}$$

This implies, by Stokes' theorem, a zero result about the surface integral of  $\nabla \times F$  over the paraboloid.

$$\int_{\sigma} (\nabla \times F) \cdot dA = \int_{\sigma} (2y, 2z, 2x) \cdot dA = 0$$

If we calculated the flux directly, it would be as follows.

$$\begin{aligned}\sigma_u &= (1, 0, -2u) \\ \sigma_v &= (0, 1, -2v) \\ \sigma_u \times \sigma_v &= (2u, 2v, 1) \\ \nabla \times F &= (8v, 8 - 2u^2 - 2v^2, 2u) \\ \int_{\sigma} \nabla \times F \cdot dA &= \int_D (4uv - 16v - 4vu^2 - 4v^3 + 2u) dudv\end{aligned}$$

Now we already know this is zero, without doing the long polar coordinate integration. Curiously, this would be zero for *any* surface with this bounding curve  $\gamma$ . A cylinder closed at the top, a cone, a very complicated ballon; it doesn't matter. Only the boundary matters.

**Example 3.9.8.** Similarly, lets say that we want to integrate  $F = (z^2, y, xz)$  over a closed curve  $\gamma$ . Stokes' theorem says we can calculate the integral of  $\nabla \times F$  over any surface with boundary  $\gamma$ .  $\nabla \times F = (0, z, 0)$ . But  $\nabla \times (\frac{z^2}{2}, 0, 0)$  is also  $(0, z, 0)$ . Look at Stokes' theorem for this case.

$$\int_{\gamma} F \cdot ds = \int_{\sigma} \nabla \times F \cdot dA = \int_{\gamma} (\frac{z^2}{2}, 0, 0) \cdot dA$$

We can replace  $F$  with a much less complicated field and get the same line integral. This is true because  $F$  and this new field both give the same curl. We have some terminology to generalize this strange situation.

**Definition 3.9.9.** Let  $G$  and  $F$  be  $C^1$  vectors fields with  $G = \nabla \times F$ . We call  $F$  a *vector potential* for  $G$ . If we need to differentiate between the two notion of potential, we will refer to the previous use of potential as *scalar potential*.

The above argument shows that if  $G_1$  and  $G_2$  are vector potentials for the same  $F$ , then they have equivalent line integrals over any closed curve.

How do we know if a field has a vector potential? For scalar potentials (on a simply connected set), we checked  $\nabla \times F = 0$ . If  $F = \nabla \times G$  then  $\nabla \cdot F = \nabla \cdot (\nabla \times G) = 0$ . As with scalar potentials, under reasonable conditions, the inverse is true as well. A field has a vector potential (on a simply connected set) if it have zero divergence.

How do we find a vector potential? If we know  $G$  and we assume  $\nabla \times F = G$ , then we have to solve this system of three differential equations:

$$\begin{aligned}\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z} &= G_1 \\ \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x} &= G_2 \\ \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} &= G_3\end{aligned}$$

This is usually prohibitively difficult. Moreover, this is a very underdetermined system. If  $F$  is a vector potential, then for any scalar field  $f$ ,  $F + \nabla f$  is also a vector potential, because

$$\nabla \times (F + \nabla f) = \nabla \times F + \nabla \times \nabla f = \nabla \times F + (0, 0, 0) = \nabla \times F = G$$

The inverse is true as well: if two have two vector potentials, they must differ by some  $\nabla f$ .

**Example 3.9.10.** To do a tractable example, let's return to the previous field. Let  $G = (0, z, 0)$ .  $F_a = (z^2, y, xz)$ ,  $F_b = \left(\frac{z^2}{2}, 0, 0\right)$  are both vector potentials for  $G$ . Their difference must be a conservative field.

$$F_a - F_b = \left(\frac{z^2}{x}, y, xz\right) = \nabla \left(\frac{xz^2}{2} + \frac{y^2}{2}\right)$$

If  $\gamma$  is any closed curve entirely in the  $xy$  plane, then since  $F_b = 0$  on the  $xy$  plane, we have a general vanish result about line integrals along  $\gamma$ .

$$\int_{\gamma} F_a \cdot ds = \int_{\gamma} F_b \cdot ds = \int_{\gamma} 0 \cdot ds = 0$$

**Example 3.9.11.** If  $\sigma$  is the sphere of radius 4 in  $\mathbb{R}^3$  and  $F = (2x^3, 2y^3, 2z^3)$ , then let  $D$  be the solid ball of radius 4. We start with a flux integral and use Gauss' theorem to solve it.

$$\begin{aligned} \int_{\sigma} F \cdot dA &= \int_D \nabla \cdot F dV = \int_D 6(x^2 + y^2 + z^2) dV \\ &= \int_0^4 \int_0^{\pi} \int_0^{2\pi} 6r^2 r^2 \sin \phi d\theta d\phi dr \\ &= 12\pi (-\cos \phi) \Big|_0^{\pi} \frac{r^5}{2} \Big|_0^4 = \frac{12288\pi}{5} \end{aligned}$$

**Example 3.9.12.** Let  $D$  is the region inside the cylinder  $x^2 + y^2 = 4$  bounded below by the  $xy$  plane and above by the plane  $x + z = 6$ . Let  $F = (x^2 + \sin z, xy + \cos z, e^y)$ . Again, we start with a flux integral and use Gauss' theorem to solve it.

$$\begin{aligned} \int_{\sigma} F \cdot dA &= \int_D \nabla \cdot F dV = \int_D 3x dV \\ &= \int_0^{2\pi} \int_0^2 \int_0^{6-r \cos \theta} 3r \cos \theta r dz dr d\theta \\ &= \int_0^{2\pi} \int_0^2 3 \cos \theta r^2 (6 - r \cos \theta) dr d\theta \\ &= \int_0^{2\pi} \left( 16 \frac{r^3}{3} \cos \theta - \frac{3r^4}{4} \cos^2 \theta \right) \Big|_0^2 d\theta \\ &= \frac{128}{3} (-\sin \theta) \Big|_0^{2\pi} - 12 \left( \frac{\theta}{2} + \frac{\sin 2\theta}{4} \right) \Big|_0^{2\pi} = -12\pi \end{aligned}$$

**Example 3.9.13.** If  $F = (z^2 + xy^2, \cos(x + z), e^{-y} + zy^2)$ , then  $\nabla \cdot F = 0$ , so  $F$  is incompressible. Then, if  $D$  is any solid region in  $\mathbb{R}^3$  with boundary  $\sigma$ , Gauss' theorem gives us a vanishing result.

$$\int_{\sigma} F \cdot dA = \int_D \nabla \cdot F dV = 0$$

That calculation means that the flux of an incompressible field over any closed surface is zero. The property of being incompressible means that equal amounts of the field flow in and flow out of any closed surface, no matter how complicated the geometry.

Working with Stokes' theorem previously, we realized we could change line integrals over a closed curve by a conservative field and not alter the line integral. We can do the same over closed surfaces. Let  $\sigma$  be a closed surface and  $F$  any field. Recall that  $\nabla \cdot (\nabla \times G) = 0$ . The following calculation shows how we can alter the field  $F$  without changing the value of the flux integral.

$$\int_{\sigma} F \cdot dA = \int_D \nabla \cdot F dV = \int_D \nabla \cdot (F + \nabla \times G) dV = \int_{\sigma} (F + \nabla \times G) \cdot dA$$

**Example 3.9.14.** Consider a windsock which is the surface of revolution of the function  $y = \sqrt{1 - \frac{x}{3}}$  for  $x \in [0, 3]$ , with outward oriented normal. Let  $F = (7, -z, y)$  be a field describing the wind near the windsock. What is the flux through the windsock?

First, we can check  $\nabla \cdot F = 0$ , so  $F$  is incompressible and has a vector potential. Assume  $F = \nabla \times G$  for some vector potential  $G$ . If we call  $D$  the solid interior of the windsock, then the boundary of  $D$  is the windsock and the circle at its mouth. We label the circle  $\mu$  and consider it a parametric surface with normal in the *negative*  $x$  direction. Then we apply Gauss twice, once in each direction.

$$\int_{\sigma} F \cdot dA = \int_D \nabla \cdot F dV - \int_{\mu} F \cdot dA = - \int_{\mu} F \cdot dA$$

Alternatively, if  $\gamma$  is the boundary of the circle, we have the following calculation using Stokes' theorem (with the negative sign accounting for the change in direction of the normal).

$$\int_{\sigma} F \cdot dA = \int_{\gamma} G \cdot ds = - \int_{\mu} F \cdot dA$$

Very curiously, both Stokes' and Gauss' give the same assistance here. In any case, we can replace our surface with a much easier surface: a disc of radius 1 in the  $yz$  plane, parametrized as  $\mu(r, \theta) = (0, r \cos \theta, r \sin \theta)$  for  $r \in [0, 1]$  and  $\theta \in [0, 2\pi]$ . We calculate the flux, reversing the conventional order of the cross product to account for the normal in the negative  $x$  direction.

$$\begin{aligned} \sigma_r &= (0, \cos \theta, \sin \theta) \\ \sigma_{\theta} &= (0, -r \sin \theta, r \cos \theta) \\ \sigma_r \times \sigma_{\theta} &= (-r \sin^2 \theta - r \cos^2 \theta, 0, 0) = (-r, 0, 0) \\ \int_{\mu} F \cdot dA &= - \int_0^{2\pi} \int_0^1 (7, -r \sin \theta, r \cos \theta) \cdot (-r, 0, 0) dr d\theta \\ &= - \int_0^{2\pi} \int_0^1 -7r dr d\theta = 2\pi 7 \left. \frac{r^2}{2} \right|_0^1 = 7\pi \end{aligned}$$

This is substantially easier than working with the windsock directly.

**Example 3.9.15.** Let's calculate the flux of  $F = \frac{1}{\sqrt{(x^2+y^2+z^2)^3}}(x, y, z)$  through the ellipsoid  $4x^2 + 9y^2 + 6z^2 = 36$ .

$$\begin{aligned} \frac{\partial}{\partial x} F_1 &= \frac{\partial}{\partial x} \frac{x}{\sqrt{(x^2+y^2+z^2)^3}} \\ &= \frac{(x^2+y^2+z^2)^{\frac{3}{2}} 0 - x \frac{3}{2}(x^2+y^2+z^2)^{\frac{1}{2}} 2x}{(x^2+y^2+z^2)^3} \\ &= \frac{(x^2+y^2+z^2)^{\frac{1}{2}}}{(x^2+y^2+z^2)^3} (x^2+y^2+z^2 - 3x^2) \\ &= \frac{-2x^2+y^2+z^2}{(x^2+y^2+z^2)^{\frac{5}{2}}} \\ \nabla \cdot F &= \frac{1}{(x^2+y^2+z^2)^{\frac{3}{2}}} (-2x^2+y^2+z^2 - 2y^2+x^2+z^2 - 2z^2+x^2+y^2) = \\ \int_{\sigma} F \cdot dA &= \int_D \nabla \cdot F dV = 0 \end{aligned}$$

This is an incompressible field over a closed surface, so it has zero flux.

**Example 3.9.16.** Let  $F = (y, ze^{y^2}, x + z^2e^{z^2})$  and let  $\gamma$  be the semicircle from  $(-3, 0, 0)$  to  $(3, 0, 0)$  which lies in the positive  $y$  part of the  $xy$  plane. What is the line integral?

First, we can check that  $\nabla \times F = (-1, 1, 0)$ , so the field is not conservative. However, let's also consider the path  $\delta$  from  $(3, 0, 0)$  to  $(-3, 0, 0)$  along the  $x$  axis. Then  $\gamma$  followed by  $\delta$  is a closed path which is the boundary of the half-circle of radius 3 in the  $xy$  plane. Call the half circle  $\sigma$ . We apply Stokes' theorem.

$$\int_{\gamma+\delta} F \cdot ds = \int_{\sigma} (\nabla \times F) \cdot dA$$

We can solve for the original line integral.

$$\int_{\gamma} F \cdot ds = \int_{\sigma} (\nabla \times F) \cdot dA - \int_{\delta} F \cdot ds$$

These two integrals on the right are notably easier than the first integral.  $\delta$  is entirely on the  $x$  axis, so  $F = (0, 0, x)$  when  $y = z = 0$ . The tangent to  $\delta$  is  $(1, 0, 0)$  so  $F \cdot N = 0$ , which means the second integral vanishes.

The normal to  $\sigma$  is  $(0, 0, -1)$ , pointing in the negative  $z$  direction to match orientation.  $(\nabla \times F) \cdot N = (-1, 1, 0) \cdot (0, 0, -1) = 0$ , so the first integral vanishes. Therefore, our original line integral also vanishes.

Note that the field is not conservative, so a different path from  $(-3, 0, 0)$  to  $(3, 0, 0)$  might have a different value. If we worked in the  $xz$  plane instead of the  $xy$  plane with the same setup, the second integral would still vanish but the normal to  $\sigma$  would be  $(0, -1, 0)$ , so  $(\nabla \times F) \cdot N = (-1)$  and the value of the first integral would be non-zero.

$$\int_{\sigma} -1 dA = -\frac{9\pi}{2} \implies \int_{\gamma} F \cdot ds = \frac{-9\pi}{2}$$

## 3.10 Vector Calculus and Maxwell's Equations

### 3.10.1 19th Century Electromagnetic Observations

A very general setup in physics is to observe the interactions of a system, write down a differential equation which describes those interactions, and try to solve the DE to find a function which will predict future action. In vector calculus, those differential equations are almost always stated in terms of the  $\nabla$  operator. The origin of  $\nabla$  and the vector calculus was the attempt to understand electricity and magnetism in the later part of the 19th century. In the investigation of electricity and magnetism in the 19th century, several important experimental observations were made by a number of different scientists and engineers.

- Electric charge exists and creates an electric field.
- Flow of electric charge is possible (current).
- Magnetic charge only comes in dipoles – isolated magnetic charge is impossible. Magnetic dipoles create magnetic fields.
- Magnetic and electric fields interact.
- A current (moving electric charge) through a wire creates a magnetic field.
- If the wire is wrapped into a solenoid, it creates a field inside the coil which obeys a RHR.
- A changing magnetic field induces an electromotive force.
- Transformers can be built out of two solenoids: the first generates a magnetic field, which induces electric flow in the second.
- Electric fields act on any charged particles by attracting or repelling them.
- Magnetic fields act only on moving charged particles and induce torque to change the direction of movement.

### 3.10.2 Mathematics Formalism for Electromagnetism

We label the electric field  $E$  and the magnetic field  $B$ . In terms of the fields, we can restate the observations. The first set of observations involve statics.

- Isolated electric charges  $q$  exist.
- Isolated electric charge creates a field  $E$  of force per unit charge.
- Charge can be a point charge or a charge density  $\rho$  over a region. Charge density has units of  $C/m^3$  or  $A \cdot s/m^3$ .
- The field  $E$  has unit of volts per meter, which in SI units is  $kg \cdot m/s^3 \cdot A$ .
- There is a constant permittivity of the vacuum, which is  $\epsilon_0 = 8.85 \times 10^{-12}$  with units  $s^4 \cdot A^2/m^3 \cdot kg$ .
- If a charge density  $\rho$  is creating the electric field  $E$ , we can observe that  $\nabla \cdot E = \frac{\rho}{\epsilon_0}$ .
- Isolated magnetic charge does not exist.

- A magnetic dipole creates a force  $B$  (per charge-velocity) which acts on a moving charged particle by rotation it away from its direction of movement.
- The units of  $B$  are teslas, which are  $kg/A \cdot s^2$ .
- There is a constant permeability of the vacuum, which is  $\mu_0 = 4\pi \times 10^{-7}$  with units  $m \cdot kg/s^2 \cdot A^2$ .
- Since there are no monopoles, all charged is balanced and  $\nabla \times B = 0$  in any magnetic situation.

Both static situations (electric and magnetic) produce fields and forces on charged particle. The action on charged particle by both  $E$  and  $B$  is summarized in the Lorentz force law. If  $q$  is a charged particle travelling with velocity  $v$ , then the force satisfies this law.

$$F = Eq + q(v \times B)$$

The second set of observations involve dynamics. What happens when  $E$  and  $B$  are changing?

- A changing  $E$  fields induces a  $B$  field and vice-versa.
- In addition, a current induces a  $B$  field.
- Magnetic current is impossible, so it cannot induce a  $E$  field.
- We can observe that the induced  $E$  field due to  $B$  satisfies a differential equation.

$$\nabla \times E = \frac{-\partial B}{\partial t}$$

- We can observe that the induced  $B$  field due to an electric field  $E$  and a current (where  $J$  is current density) also satisfies a differential equation.

$$\nabla \times B = \mu_0 J + \mu_0 \epsilon_0 \frac{\partial E}{\partial t}$$

These observations give us the first presentation of Maxwell's equations: the presentation via differential operators. There are four equations: two for statics, which calculate divergence of the fields, and two for dynamics, which calculate curl of the fields.

$$\begin{aligned} \nabla \cdot B &= 0 \\ \nabla \cdot E &= \frac{\rho}{\epsilon_0} \\ \nabla \times B &= \mu_0 J + \mu_0 \epsilon_0 \frac{\partial E}{\partial t} \\ \nabla \times E &= \frac{-\partial B}{\partial t} \end{aligned}$$

### 3.10.3 Gauss-Green-Stokes and Maxwell's Equations

We can apply Gauss' and Stokes' theorem to Maxwell's equation to get a second presentation. Let  $D$  be a solid region of space and  $\sigma$  a parametric surface (not necessarily the boundary of  $D$ ). We define four new symbols.

- $Q_D$  is the total charge in the solid region  $D$ .
- $I_\sigma$  is the total current flowing through a surface.
- $\Phi_{E,\sigma}$  is the flux of  $E$  through  $\sigma$ .
- $\Phi_{B,\sigma}$  is the flux of  $B$  through  $\sigma$ .

If  $\sigma$  is the bounding surface for the region  $D$ , then we can apply Gauss' theorem to the field  $E$ .

$$\Phi_{E,\sigma} = \int_{\sigma} E \cdot dA = \int_D \nabla \cdot E dV = \int_D \frac{\rho}{\epsilon_0} dV = \frac{Q_D}{\epsilon_0}$$

We can do the same for  $B$ , remembering that  $\nabla \cdot B = 0$ .

$$\Phi_{B,\sigma} = \int_{\sigma} B \cdot dA = \int_D \nabla \cdot B dV = \int_D 0 dV = 0$$

Now let  $\sigma$  be any surface (not the boundary of  $D$ ) and let  $\gamma$  be its boundary. We can think of  $\gamma$  as a loop of wire, and  $\sigma$  the surface inside the loop. The line integral of  $E$  along the wire is the work done on by the field moving electrons along the wire. We can use Stokes' to calculate the line integral.

$$\int_{\gamma} E \cdot ds = \int_{\sigma} (\nabla \times E) \cdot dA = - \int_{\sigma} \frac{\partial B}{\partial t} = - \frac{\partial}{\partial t} \Phi_{B,\sigma}$$

Thinking the other way around, this tells us the flux through the wire of the magnetic field induced by the current. If the loop is now a series of loops in a solenoid, the flux adds up, so more loops gives more surface area, hence more flux, hence a greater induced magnetic field. In an electric transformer we adjust the ingoing and outgoing current by having different numbers of coils on each solenoid.

For magnetic field, we have also have a Stokes' theorem calculation.

$$\int_{\gamma} B \cdot ds = \int_{\sigma} (\nabla \times B) \cdot dA = \int_{\sigma} \mu_0 J \cdot dA + \int_{\sigma} \mu_0 \epsilon_0 \frac{\partial E}{\partial t} = \mu_0 I_\sigma + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \Phi_{E,\sigma}$$

We can put these all together into the second form of Maxwell's equations. In the first part,  $\sigma$  is the closed boundary of  $D$ , in the second part,  $\sigma$  is an open surface with boundary  $\gamma$ .

$$\begin{aligned} \int_{\sigma} E \cdot dA &= \frac{Q_D}{\epsilon_0} \\ \int_{\sigma} B \cdot dA &= 0 \\ \int_{\gamma} E \cdot ds &= - \frac{\partial}{\partial t} \Phi_{B,\sigma} \\ \int_{\gamma} B \cdot ds &= \mu_0 I_\sigma + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \Phi_{E,\sigma} \end{aligned}$$

### 3.10.4 Maxwell's equations in a Vacuum

A strange question: what happens we we consider Maxwell's equation in a vacuum with no external influence? Without charge or current, the differential form of the equations is as follows.

$$\begin{aligned}\nabla \cdot B &= 0 \\ \nabla \cdot E &= 0 \\ \nabla \times B &= \mu_0 \epsilon_0 \frac{\partial E}{\partial t} \\ \nabla \times E &= \frac{-\partial B}{\partial t}\end{aligned}$$

This is a difficult system of differential equations, but it does have a solution. Therefore, at least mathematically, we can have electric and magnetic field existing in a vacuum, even without any charge, current, or magnetic dipoles. We're going to do some work to try to solve this system of DEs. In what looks like random calculations, let's dive in.

$$\begin{aligned}\nabla \times (\nabla \times E) &= \nabla \times -\frac{\partial B}{\partial t} = \frac{-\partial}{\partial t}(\nabla \times B) \\ &= -\frac{\partial}{\partial t} \mu_0 \epsilon_0 \frac{\partial E}{\partial t} \\ &= -\mu_0 \epsilon_0 \frac{\partial^2 E}{\partial t^2}\end{aligned}$$

We have the identity  $\nabla \times (\nabla \times E) = \nabla(\nabla \cdot E) - \nabla^2 E$ , where the Laplacian here means the Laplacian of each of the three components of  $E$ . The first term is 0, since  $\nabla \cdot E = 0$ , so we get  $\nabla \times (\nabla \times E) = -\nabla^2 E$ . Let's put this new piece together with the previous calculation.

$$\nabla^2 E = \mu_0 \epsilon_0 \frac{\partial^2 E}{\partial t^2}$$

A similar calculation exists for  $B$ .

$$\nabla^2 B = \mu_0 \epsilon_0 \frac{\partial^2 B}{\partial t^2}$$

These are wave equations, so we expect that we will have wave solutions for  $E$  and  $B$ . The full solutions are very general, so let's make some assumptions which are justified by observation. First, we assume  $E \perp B$ . Second, let  $z$  be the direction of propagation of the waves, so the waves themselves are only in the  $xy$  plane. By readjusting coordinates, we can assume  $E = (E_1, 0, 0)$  and  $B = (0, B_2, 0)$ . Then we have two differential equations in the scalar fields  $E_1$  and  $B_2$ .

$$\begin{aligned}\frac{\partial}{\partial z^2} E_1 &= \mu_0 \epsilon_0 \frac{\partial^2 E_1}{\partial t^2} \\ \frac{\partial}{\partial z^2} B_2 &= \mu_0 \epsilon_0 \frac{\partial^2 B_2}{\partial t^2}\end{aligned}$$

We'll note solve these DEs, but they have the following reasonable solutions.

$$E_1 = a_1 \cos\left(z + \frac{1}{\sqrt{\mu_0 \epsilon_0}} t\right) + b_1 \sin\left(z + \frac{1}{\sqrt{\mu_0 \epsilon_0}} t\right)$$

$$B_2 = a_2 \cos\left(z + \frac{1}{\sqrt{\mu_0 \epsilon_0}} t\right) + b_2 \sin\left(z + \frac{1}{\sqrt{\mu_0 \epsilon_0}} t\right)$$

These are waves propogating in the  $z$  direction. The coefficient of  $t$  is the wave speed. What is this coefficient?

$$\frac{1}{\sqrt{\epsilon_0 \mu_0}} = \frac{1}{8.86 \times 10^{-12} \cdot 4\pi \times 10^{-9}} = 2.997 \times 10^8 \text{ m/s}$$

The units also work out to recover a velocity.

$$\frac{1}{\sqrt{\frac{\text{m} \cdot \text{kg}}{\text{s}^2 \text{A}^2} \frac{\text{s}^4 \text{A}^2}{\text{m}^3 \text{kg}}}} = \frac{1}{\sqrt{\frac{\text{s}^2}{\text{m}^2}}} = \frac{\text{m}}{\text{s}}$$

This is a familiar number:  $c$ , the speed of light. And light is exactly what this wave is: propagation of electromagnetic fields through a vacuum is light. At the time of Maxwell, light was not understood as electromagnetic radiation. Maxwell argued that his system predicted the electromagnetic nature of light, which was eventually proved correct. The relationship in the basic constants (the permittivity, permeability and the speed of light) is an important fact about the universe.

## Chapter 4

# Differential Forms and Differential Geometry

### 4.1 Manifolds and Coordinate Functions

The last section of these notes is a brief introduction to differential geometry. The main purpose of this introduction is to explain the connection between the major theorems of the previous chapter. By defining objects called *differential forms*, we can reduce all the fundamental theorems and Gauss-Green-Stokes to one very simply statement.

First, however, we need to a new geometric object: manifolds. Differential geometry is essentially calculus on manifolds. So far, we've worked with parametric curves and surfaces (we could imagine parametric objects in higher dimensions as well). Manifolds are a generalization of parametric objects. The major novelty of manifolds is that we allow different parts of the manifolds to have different parametrizations, as long the parametrizations cooperate where they overlap.

Consider the sphere, which we parametrized as  $\sigma(\theta, \phi) = (r \sin \phi \cos \theta, r \sin \phi \sin \theta, r \cos \phi)$ . This was a good parametrization, but it had problems at the poles, where  $\theta$  was no longer defined. The poles present a problem for doing calculus on the sphere. We might try to rework the parametrization, but we would find that *any* parametrization of the sphere will have at least one point with this polar problem. To fully perform calculus on the sphere, we have to allong different regions to have different parametrization. The sphere is one of the first and most important manifolds in differential geometry.

We can understand the problems of parametrization in terms of coordinate lines. At most points on the sphere, the coordinate lines (looking very locally) resemble the coordinates lines of  $\mathbb{R}^2$ : a grid. However, at the poles, the coordinate lines look like rays and concentric circles. For manifolds, we can parametrizations where the coordinate lines always locally look like a grid.

**Definition 4.1.1.** Let  $U \subset \mathbb{R}^n$  be any set and let  $D$  be a simply connected open subset of  $\mathbb{R}^k$ . A *coordinate function* on  $U$  is an injective  $C^1$  function  $\phi : D \rightarrow U$ . Let  $M \subset \mathbb{R}^n$ .  $M$  is called a *manifold* if it can be covered by open sets  $U$  which have coordinate functions. The number of variables  $k$  in the *dimension* of the manifold.

The complete definition of a manifold also includes information about how the coordinate functions cooperate where they overlap. Though overlaps and compatibility are important and central parts of the definition, we are going to ignore these issues for the purposes of these notes. We will work only in the local situation, considering one open set  $U$  and its coordinate function  $\phi$ .

$\phi$  is called a coordinate function because it allows us to use the variables  $u_i \in D(0, r)$  as coordinates on the manifold. Much like parametrized object, the input coordinates to the coordinate function tell us how to identify points on the manifold. Everything on a manifold is determined by how it works in local coordinates, accessed by use of a coordinate function.

**Example 4.1.2.** Any open set  $U$  in  $\mathbb{R}^n$ , is itself a manifold with  $\phi = \text{Id}$ . If  $k = 1$ , then  $\phi$  defines a curve where the local coordinate is the parameter  $t$ . If  $k = 2$ , then  $\phi$  defines a surface and the parameters  $u$  and  $v$  are the local coordinates.

**Example 4.1.3.** Any parametric curve  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  is also a manifold. The function  $\gamma$  is the coordinate function.

**Example 4.1.4.** If  $D$  is a simply connected open set in  $\mathbb{R}^2$ , then any parametric surface  $\sigma : D \rightarrow \mathbb{R}^n$  is also a manifold. The function  $\sigma$  is the coordinate function.

For parametric curves,  $\gamma$ , we calculated  $\gamma'$ , which was the tangent vector. All multiples of  $\gamma'$  gave a tangent line to the curve. For parametric surfaces,  $\sigma$ , we calculated  $\sigma_u$  and  $\sigma_v$ . All linear combinations of  $\sigma_u$  and  $\sigma_v$  give a plane which is tangent to the surface.

**Definition 4.1.5.** Let  $M$  be a manifold of dimension  $k$ . Let  $D$  be a simply connected open set in  $\mathbb{R}^k$  with  $\phi : D \rightarrow U$  a coordinate function on  $M$ . Let  $u_i$  be the variables of  $\mathbb{R}^k$  on the set  $D$ . For each such  $u_i$ , the partial derivative  $\frac{\partial \phi}{\partial u_i}$  is a *tangent vector* to the manifold. The *tangent space*  $T_p$  at a point  $p \in U$  is the span all of the tangent vectors  $\frac{\partial \phi}{\partial u_i}$  where  $i = 1, \dots, k$ .

There is some strange but conventional notation for tangent vectors to manifolds. Using the conditions of the previous definition, instead of  $\frac{\partial \phi}{\partial u_i}$ , we often simply write  $\partial_i$  for the tangent vector. If we want to specify the point  $p$  on the manifold, we write  $\partial_i(p)$ . Notice, in this definition, that the subscript  $i$  refers to the coordinates defined by the coordinate function  $\phi$ : the coordinate function is implicit in every manifold definition. We access all the information about a manifold via its coordinate functions.

For parametric curves, we used the condition  $\gamma'(t) \neq 0$  as an important condition in line integral. Similarly, for parametric surfaces, we insisted on non-zero normals. We called such curves and surfaces smooth. For manifolds, the situation is similar.

**Definition 4.1.6.** A manifold  $M$  of dimension  $k$  is called *smooth* if all of its tangent spaces  $T_p$  are vector spaces of dimension  $k$  (i.e., copies of  $\mathbb{R}^k$ ).

Many definition of manifolds assume the smoothness condition. We will do the same: for the remainder of these notes, all manifolds are assumed to be smooth.

## 4.2 Functions on Manifolds

**Definition 4.2.1.** Let  $M$  be a  $k$ -dimensional manifold in  $\mathbb{R}^n$ . A *scalar field* on  $M$  is a function  $f : M \rightarrow \mathbb{R}$ .

A scalar field is simply defined: it assigns a scalar to each point on the manifold. We might think that vector fields on the manifold are similarly easy, but we are going to take a very different approach.

**Definition 4.2.2.** A *vector-valued function* or *vector field* on a manifold is defined to be a function

$$F : M \rightarrow \cup_{p \in M} T_p$$

such that  $f(p) \in T_p$  for all  $p \in M$ .

What does this strange notation mean? Previously, our vector fields had outputs in  $\mathbb{R}^n$  for some  $n$ . This definition is different in two ways. First, the output dimension can only be  $k$ , the dimension of the manifold. Second, instead of assigning vectors in  $\mathbb{R}^k$  to each point, we assign vectors in the tangent space to each point. Each  $T_p$  is a copy of  $\mathbb{R}^k$  (since we assumed all manifolds are smooth!), so a vector field is assigning vectors in  $\mathbb{R}^k$ . Though it seems strange, there are excellent reasons for restricting outputs to tangent spaces.

Any vector field  $F$  can be decomposed into components  $F = (f_1, f_2, \dots, f_k)$  in terms of the coordinates of each of the  $T_p$ . Since the  $\partial_i$  are a basis for each  $T_p$ , we can write vector fields as:

$$F = \sum_{i=1}^k f_i \partial_i = (f_1, f_2, \dots, f_k) \cdot (\partial_1, \dots, \partial_k)$$

Each  $f_i$  is a scalar function. Vector fields are combinations of the  $\partial_i$  where the coefficients are scalar functions.

**Example 4.2.3.** Let's make this a little bit more explicit. If  $u, v$  are the local coordinates on a 2-dimensional manifold, then the following are all vector fields on the manifold.

- $(u^2 + v^2)\partial_u + 3\partial_v$
- $\sin(v)\partial_u + \cos(u)\partial_v$
- $(4u - 3v)\partial_u + (2u - 9v)\partial_v$

## 4.3 Differential Forms

Now we can start to define the main object of this chapter: the differential form. The definition is very odd; be warned. We start with a little bit of linear algebra.

**Definition 4.3.1.** Let  $L$  be a linear space over  $\mathbb{R}$ . The *dual space* of  $L$  is the space of all linear function  $f : L \rightarrow \mathbb{R}$ . It is a linear space of the same dimension as  $L$  and is written  $L^\vee$ . If  $\{v_1, \dots, v_k\}$  is a basis for  $L$ , then the *dual basis* for  $L^\vee$  is the set of function  $f_i$  such that  $f_i(v_j) = \delta_{ij}$ .

Differential forms are classified by their degree, which is a non-negative integer. Instead of attempting a full definition, we will start with the lowest degree forms and work our way up.

**Definition 4.3.2.** A degree 0 differential form (0-form) on a manifold  $M$  is just a scalar function on  $M$ .

**Definition 4.3.3.** Let  $M$  be a manifold of dimension  $k$  and  $\phi : D \rightarrow U$  a coordinate function on  $M$ . A degree 1 differential form (1-form) on a manifold  $M$  is a function  $\omega$  defined on  $U$  as follows.

$$\omega : U \rightarrow \cup_{p \in U} T_p^\vee$$

This is very strange:  $T_p^\vee$  is the set of linear functions on each tangent space  $T_p$ . Therefore, a 1-form is a function that, at each point of the manifold, outputs a new linear function on its own tangent space  $T_p$ . Functions that output other functions – it’s a valid definition, but difficult to parse.

Since the  $\partial_i$  are a basis for each  $T_p$  and differential forms create linear function on the  $T_p$ , it is sufficient to know their effect on the  $\partial_i$ . We can make use of the dual basis defined above.

**Definition 4.3.4.** Let  $du_i$  to be a differential form that acts on each  $\partial_i$  by

$$du_i(\partial_i) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The  $du_i$  for the dual basis to  $\partial_i$ . A general 1-form  $\omega$  can always be written in terms of this basis. If the  $f_i$  are scalar functions, then all 1-forms are written as a linear combination of the  $du_i$ .

$$\omega = \sum_{i=1}^k f_i du_i$$

This defines 1-forms. Since they output linear function on the tangent space  $T_p$ , we think of them acting on tangent vectors. Informally, higher degree forms are defined similarly; the only difference is that they act on  $m$ -tuples of vectors. A 1-form acts on one vector, but an  $m$ -forms acts on a  $m$ -tuples of vectors. Instead of giving a precise definition for this action, we’ll define higher order forms out of 1-form by the operations in the next section.

### 4.3.1 Products and Derivatives

There are four important operations that we can perform on differential forms. First, differential forms are linear, so we can add, subtract and multiply by constants.

Second, there is a product of forms which is called the exterior product. We’ll define it on 1-forms, but it can be extended to any degree. In general, the exterior product acts the degrees of the forms; in particular, the product of two 1-forms will be a 2-form.

**Definition 4.3.5.** Let  $\omega$  and  $\eta$  be 1-forms. They each act on vectors in tangent spaces. Their *exterior product* is a 2-form, so it acts on two vectors from each tangent space. It is defined by the following formula.

$$\omega \wedge \eta(v_1, v_2) = \omega(v_1)\eta(v_2) - \omega(v_2)\eta(v_1)$$

**Proposition 4.3.6.** *The exterior product is associative. if  $\omega$ ,  $\eta$  and  $\sigma$  are all forms, then*

$$\omega \wedge \eta \wedge \sigma = \omega \wedge (\eta \wedge \sigma) = (\omega \wedge \eta) \wedge \sigma$$

**Proposition 4.3.7.** *The exterior product is not commutative. Instead, it is anti-commutative.*

$$\omega \wedge \eta = -\eta \wedge \omega$$

The anti-commutativity is easy to see from the definition. It implies a strange but very important property of the exterior product.

$$\omega \wedge \omega = 0$$

With the basis  $du_i$ , we had a description of any 1-form as a linear combination of the  $du_i$  (where the scalars were scalar functions). The exterior product gives a basis for higher degree forms. However, anti-commutativity implies that  $du_i \wedge du_i = 0$ , so the basis for the 2-forms are all  $du_i \wedge du_j$  such that  $i \neq j$ . Likewise, the 3-forms has a basis  $du_i \wedge du_j \wedge du_k$  where all three indexed are distinct. In general,  $l$ -forms are spanned by wedges of  $du_i$  of length  $l$  where no index is repeated. When any index is repeated, the form becomes 0. After we have these basis, any form is a linear combination of the basis elements, where the coefficients are scalar functions.

In addition to a product, there is a derivative for differential forms.

**Definition 4.3.8.** Let  $\omega$  be a 1-form, which we write with the basis  $du_i$ .

$$\omega = \sum_{i=1}^k f_i du_i$$

The *exterior derivative* of  $\omega$  is written  $d\omega$  and is expressed as follows.

$$d\omega = \sum_{i=1}^k \left( \sum_{j=1, j \neq i}^k \frac{\partial f_i}{\partial x_j} du_j \wedge du_i \right)$$

For higher degree forms, a similar equation holds: we express the form in a basis of wedges of  $du_i$  and differentiate the coefficient scalar functions. Each time we differentiate  $\frac{\partial}{\partial u_i}$  we add a  $du_i \wedge$  to the start of the wedge product.

In the previous chapter, we talked about how certain differential operators composed to zero: the curl of a gradient is zero and the divergence of a curl is zero. The exterior derivative lives up to this legacy perfectly.

**Proposition 4.3.9.** *If  $\omega$  is any differential form, then  $d(d\omega) = 0$ . Succinctly, we write  $d^2 = 0$  for the exterior derivative.*

*Proof.* There is nothing radical or insightful about the proof. We just apply the expression for the exterior derivative twice and do a lot of tedious algebra with the indicies. All the terms cancel out and we get zero.  $\square$

There is one last operator on forms.

**Definition 4.3.10.** Let  $\omega$  be a  $k$ -form and  $v$  is a vector field on the manifold. The *interior product* of  $\omega$  and  $v$  is written  $i_v\omega$ . It is a  $k-1$ -form and is defined as follows.

$$i_v\omega(v_1, \dots, v_{k-1}) = \omega(v, v_1, \dots, v_{k-1})$$

As a differential  $k$ -form,  $\omega$  acts on  $k$  vectors. The interior product simply insist that the first vector acted on is always the vector  $v$ , leaving the remainin  $k - 1$  vectors undetermined.

### 4.3.2 Differential Forms in $\mathbb{R}^3$

The definitions so far have all been very abstract, so lets try to specialize back to  $\mathbb{R}^3$  to see what happens. In this section, our manifold will simple be an open set  $U$  in  $\mathbb{R}^3$ . Its coordinate function is simply the identity  $\text{Id} : U \rightarrow U$ . This lets us use the familiar coordinates  $x, y, z$  as local coordinates on  $U$  and define fields and forms interms of these local coordinates. (Here, the local and ambient coordinate are the same.)

- Vector Fields  $\partial_x, \partial_y, \partial_z$
- 1-Forms  $dx, dy, dz$
- 2-Forms  $dy \wedge dz, dz \wedge dx, dx \wedge dy$
- 3-Forms  $dx \wedge dy \wedge dz$

It is impossible to have a differential form of degree 4 or higher, since a 4-form must have at least one repeated  $dx, dy$  or  $dz$  term and forms with repeated basis terms are zero. The degree of forms is bounded by the dimension of the object. For the 2-forms, the ordering of the basis is related to orientation. We will stick with the order as presented above.

If  $\omega_0, \omega_1, \omega_2$  and  $\omega_3$  are 0, 1, 2, and 3 forms respectively, they are written in terms of the basis by using scalar fields  $f_i$  as follows.

$$\begin{aligned} \omega^0 &= f \\ \omega^1 &= f_1 dx + f_2 dy + f_3 dz \\ \omega^2 &= f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy \\ \omega^3 &= f dx \wedge dy \wedge dz \end{aligned}$$

Now we can consider the exterior derivatives of these forms. This will start to show us the connection to the vector calculus.

First, let's differentiate  $\omega_0$ , which is just a scalar field. Its derivative, however, will be a 1-form.

$$d\omega^0 = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz = df = \nabla f \cdot (dx, dy, dz)$$

The exterior derivative of a 0-form (scalar field) recovers the gradient operator. The gradient is the *coefficients* of the 1-form.

Now let's take  $\omega^1$  a 1-form and calculate its exterior derivative. This derivative will be a 2-form.

$$\begin{aligned} \omega^1 &= f_1 dx + f_2 dy + f_3 dz \\ d\omega^1 &= \frac{\partial f_1}{\partial y} dy \wedge dx + \frac{\partial f_1}{\partial z} dz \wedge dx + \frac{\partial f_2}{\partial x} dx \wedge dy + \frac{\partial f_2}{\partial z} dz \wedge dy + \frac{\partial f_3}{\partial x} dx \wedge dz + \frac{\partial f_3}{\partial y} dy \wedge dz \\ &= -\frac{\partial f_1}{\partial y} dx \wedge dy + \frac{\partial f_1}{\partial z} dz \wedge dx + \frac{\partial f_2}{\partial x} dx \wedge dy - \frac{\partial f_2}{\partial z} dy \wedge dz - \frac{\partial f_3}{\partial x} dz \wedge dx + \frac{\partial f_3}{\partial y} dy \wedge dz \\ &= \left( \frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z} \right) dy \wedge dz + \left( \frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x} \right) dz \wedge dx + \left( \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) dx \wedge dy \\ &= (\nabla \times (f_1, f_2, f_3)) \cdot (dy \wedge dz, dz \wedge dx, dx \wedge dy) \end{aligned}$$

If we do the orientation correctly and choose the basis order as we did, we recover the curl operation. Again, the curl gives the *coefficients* of this 2-form.

The exterior derivative of a 0-form involved the gradient. For a 1-form, it involved the curl. If we take a 0-form  $\omega^0$  and take the exterior derivative twice, then the fact  $d(d\omega^0) = 0$  is equivalent to the identity  $\nabla \times \nabla f = 0$ .

Now let's take  $\omega^2$  a 2-form and calculate its exterior derivative.

$$\begin{aligned} \omega^2 &= f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy \\ d\omega^2 &= \frac{\partial f_1}{\partial x} dx \wedge dy \wedge dz + \frac{\partial f_2}{\partial y} dy \wedge dz \wedge dx + \frac{\partial f_3}{\partial z} dz \wedge dx \wedge dy \\ &= \left( \frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z} \right) dx \wedge dy \wedge dz \\ &= (\nabla \cdot (f_1, f_2, f_3)) dx \wedge dy \wedge dz \end{aligned}$$

The exterior derivative of a 2-form recovers the divergence operator. Again,  $d^2 = 0$  on a 1-form recovers the identity that  $\nabla \cdot (\nabla \times F) = 0$  from before.

All three operators: gradient, divergence and curl, are found in the exterior derivative in the right degree. Our vector fields defined previously can now be interpreted as the component functions of either 1-forms or 2-forms in  $\mathbb{R}^3$ .

The idea of potentials also generalizes. On a simply connected open set, we can prove that  $d\omega^1 = 0 \implies \omega = df$ , i.e., we can find scalar potentials. Similarly,  $d\omega^2 = 0 \implies \omega^2 = d\omega^1$ , i.e., we can find vector potentials. This is part of a very general theorem that any differential form on a simply connected open set which satisfies  $d\omega = 0$  must have another form as its potential, i.e.,  $\omega = d\eta$ . There is some language for this situation.

**Definition 4.3.11.** A differential form  $\omega$  is called *closed* if  $d\omega = 0$ . It is called *exact* if there exists another form  $\eta$  with  $\omega = d\eta$ . On a simply connected open set, all closed forms are exact. On an arbitrary set, there are often closed forms which are not exact.

These results about finding potentials are local on manifolds, where we can also work with these simply connected open sets. Whether we can find a potential that works everywhere is a major question in differential geometry. The problems that prevent us from finding such global potentials are called *obstruction* and obstructions are the building blocks of structures called homology theories. A great deal of effort is spent studying homology theories on manifolds.

One of the great advantages of working with differential forms is that they tell us how to generalize to any dimensions. Curl is particular to  $\mathbb{R}^3$ , and it is not at all obvious what the generalization of curl is in higher dimensions. Now that curl is just the exterior derivative of differential 2-form, the extension of curl is the same: the exterior derivative of a 2-form in any higher dimension. The exterior derivative will give the formula in terms of components. The question of potentials is also now extendable, since potentials are just the question of when a  $k$ -form  $\omega^k$  which satisfies  $d\omega^k = 0$  also has  $\omega^k = d\omega^{k-1}$  for  $\omega^{k-1}$  some  $k-1$ -form.

## 4.4 Pullbacks

When we worked with parametric curves and parametric surfaces, there were two kinds of coordinates: coordinates in the parameter space ( $t$  for curves,  $u$  and  $v$  for surfaces) and coordinates in the ambient space ( $x$ ,  $y$  and  $z$  in  $\mathbb{R}^3$ ). We needed to work with both sets of coordinates to properly understand these parametric objects. In line and flux integrals, we had to evaluate the vector fields, given in  $x$ ,  $y$  and  $z$  ambient coordinates, in terms of the parameters, so that we could integrate in those parameters. We developed an interplay between the ambient coordinates and the parameters.

If  $U$  is an open set in a dimension  $k$  manifold in  $\mathbb{R}^n$ , then we like have two sets of coordinates: local coordinates (parameters) on the manifold  $u_i$  for  $1 \leq i \leq k$  coming from the domain  $D(0, r)$  of a coordinate function; and the ambient coordinates  $x_i$  for  $1 \leq i \leq n$  in  $\mathbb{R}^n$ . In this section, we are going to build a general understanding of the interaction between the two sets of coordinate.

Since the manifold under consideration has dimensions  $k$ , the higher degree differential form on  $U$  is a  $k$ -form. In the local coordinates  $u_1, \dots, u_k$ , this form is simply  $f(u_1, \dots, u_k)du_1 \wedge du_2 \wedge \dots \wedge du_k$ .

However, the form may also be expressed in terms of the ambient coordinates  $x_1, \dots, x_n$ . (Note that  $n \geq k$ : there may be more of these coordinates than the parameters!) The form can be expressed in terms of wedges of  $k$  different  $dx_i$  in the ambient coordinates. For example, if  $k = 2$  and  $n = 3$ , then the two form could be  $\omega = (x^2yz)dy \wedge dz$  or  $\omega = (x + y + z)dx \wedge dz$ .

In general, a  $k$ -form in the ambient coordinates can be a wedge of any subset consisting of exactly  $k$  of the  $x_i$ .

$$\omega^k = f(x_1, \dots, x_n)dx_{j_1} \wedge dx_{j_2} \wedge \dots \wedge dx_{j_k}$$

For parametric curves and surfaces, we replace the ambient coordinates with the parameters using the definition of the curves and surface. When we went to integrate, we used a Jacobian term to relate the integral of the ambient coordinates to the integral of the parameters. For the curve, that Jacobian was  $|\gamma'(t)|$ , and for a surface, that Jacobian was  $|\sigma_u \times \sigma_v|$ .

For a manifold, we will do the same. We will replace the ambient coordinates with the local coordinates using the coordinate function. This is a change of coordinates: ambient going to local. Since the local coordinates are the *domain* of the coordinate functions, this is sort-of going backwards with the coordinate functions. For this reason, we call the operation a *pullback* of the differential form. Like the curve and surface case, we will need a Jacobian for this pullback.

**Definition 4.4.1.** Let  $U$  be an open set on a manifold of dimension  $k$  inside  $\mathbb{R}^n$ . Let  $u_1, \dots, u_k$  be its local coordinates and  $\phi : D \rightarrow U$  is coordinate function. Let  $\omega^k$  be a differential  $k$ -form expressed in the ambient coordinates (where the  $j_i$  are any subset of exactly  $k$  of the ambient coordinates).

$$\omega^k = f(x_1, \dots, x_n) dx_{j_1} \wedge dx_{j_2} \wedge \dots \wedge dx_{j_k}$$

The *pullback* of  $\omega^k$  is the same form expressed in the local coordinates.

$$\phi^* \omega^k = (f \circ \phi(u_1, \dots, u_k)) \frac{\partial(x_{j_1}, x_{j_2}, \dots, x_{j_k})}{\partial(u_1, u_2, \dots, u_k)} du_1 \wedge \dots \wedge du_k$$

The intermediate terms is the *Jacobian* of the pullback. It is defined as the determinant of a particular matrix of partial derivatives of the coordinate function.

$$\frac{\partial(x_{j_1}, x_{j_2}, \dots, x_{j_k})}{\partial(u_1, u_2, \dots, u_k)} = \left| \begin{pmatrix} \frac{\partial \phi_{j_1}}{\partial u_1} & \frac{\partial \phi_{j_1}}{\partial u_2} & \dots & \frac{\partial \phi_{j_1}}{\partial u_k} \\ \frac{\partial \phi_{j_2}}{\partial u_1} & \frac{\partial \phi_{j_2}}{\partial u_2} & \dots & \frac{\partial \phi_{j_2}}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_{j_k}}{\partial u_1} & \frac{\partial \phi_{j_k}}{\partial u_2} & \dots & \frac{\partial \phi_{j_k}}{\partial u_k} \end{pmatrix} \right|$$

## 4.5 Integration of Differential Forms

Differential forms are built to serve as integrands. A major contention of differential geometry is that an differential form is the *only* object that should be integrated and that any previously defined integrations can be efficiently and clearly realized as the integrals of differential forms. The notation reflects this: the fact that the bases for forms uses the symbols  $dx_i$  is explicitly because these have always been part of the notation for integration. The differential terms  $dx$  in an integral have always been problematic: historically, they were infinitesimal, and they have lingered as a vestige in modern calculus. Differential forms finally give those pieces of notation proper, clear definitions in the modern treatment of integration.

This definition of integration assumes we already know single-variable integral and multiple integral of scalar functions by way of iterated integrals (i.e., the first chapter of these notes). This definition does not assume any integration of vector fields or integration over parametric surfaces.

**Definition 4.5.1.** Let  $\omega^k$  is a  $k$ -form on  $U$ , an open set in a  $k$ -dimensional manifold. Let  $u_1, \dots, u_k$  be local coordinates and let  $x_1, \dots, x_n$  be ambient coordinates. We can write  $\omega = f dx_{j_1} \wedge \dots \wedge dx_{j_k}$  in ambient coordinates. Let  $\phi : D \rightarrow U$  is a coordinate function. The integral of  $\omega$  over  $U$  is defined as the integral of the pullback of  $\omega$  over the domain of the coordinate function.

$$\int_U \omega = \int_U f dx_{j_1} \wedge \dots \wedge dx_{j_k} = \int_D \phi^* \omega du_1 \dots du_l$$

So we pullback to a simply connected set in  $\mathbb{R}^k$  and do the integral there. The coefficient  $\phi^* f(u_1, \dots, u_k)$  is just a scalar function, so the integral is just a multiple-integral of a scalar function. Notice that the wedge  $dx_{j_1} \wedge \dots \wedge dx_{j_k}$  turns into  $du_1 \dots du_k$  in the integral over  $D$ ! We finally treat the differential terms  $du$  is integration properly: they are just differential forms written without the wedge product.

Let's work this out in dimensions 1 and 2. In dimension 1, a manifold is locally a parametric curve and the coordinate function is  $\gamma : [a, b] \rightarrow \mathbb{R}^3$ . We want to integrate a 1-form on  $\gamma$ , since the degree has to match the dimension. Such a form in ambient coordinates is  $\omega = f_1 dx + f_2 dy + f_3 dz$ . Let's calculate the integral.

$$\begin{aligned} \int_\gamma \omega &= \int_{[a,b]} \gamma^* \omega \\ &= \int_{[a,b]} \left( f_1 \circ \gamma \frac{dx}{dt} dt + f_2 \circ \gamma \frac{dy}{dt} dt + f_3 \circ \gamma \frac{dz}{dt} dt \right) \\ &= \int_a^b f(\gamma(t)) \cdot \gamma'(t) dt \end{aligned}$$

If we treat the coefficient of the form as a vector field, we recover the integration formula for line integrals of a vector field over parametric curves.

In dimension 2, a manifold is locally a parametric surface and the coordinate function is  $\sigma : D \rightarrow \mathbb{R}^3$  in variables  $(u, v)$ . We want to integrate a 2-form on  $\sigma$ , since the degree has to match the dimensions. Such a form can be written in ambient coordinates.

$$\omega = f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy$$

Its integral over a surface  $\sigma$  is the integral of the pullback over the parameter domain.

$$\begin{aligned} \int_\sigma \omega &= \int_D \sigma^* \omega \\ &= \int_D \left( f_1 \circ \sigma \frac{\partial(y, z)}{\partial(u, v)} + f_2 \circ \sigma \frac{\partial(z, x)}{\partial(u, v)} + f_3 \circ \sigma \frac{\partial(x, y)}{\partial(u, v)} \right) du \wedge dv \\ &= \int_D (f \circ \sigma) \cdot \left( \frac{dy}{dv} \frac{dz}{du} - \frac{dy}{du} \frac{dz}{dv}, \frac{dz}{dv} \frac{dx}{du} - \frac{dz}{du} \frac{dx}{dv}, \frac{dx}{dv} \frac{dy}{du} - \frac{dx}{du} \frac{dy}{dv} \right) du \wedge dv \\ &= \int_D f(\sigma(u, v)) \cdot (\sigma_u \times \sigma_v) dudv \end{aligned}$$

If we treated the coefficients of the differential form as a vector field, we recover the integration formula for the flux integral of a field over a parametric surfaces.

In this way, the integration definition for differential forms recovers all the vector field integrals we defined (and any similar vector field integrals one might define in any higher dimensions). Differential geometry argues, based on the fact that we can recover all these integration definition, that differential forms are the proper elements for integration. This formalism simply isn't possible if we stick with vector fields: we need the strangeness of differentials forms (with their dual-space definition, action on tangent vectors, exterior products and derivatives) in order to make this unification work. We need pullbacks to change variables and have well-behaved Jacobians. In differential geometry, pullbacks always exist for differential forms, but usually do not exist for vector fields. Differential forms are simply the right things to integrate.

## 4.6 Stokes' Theorem

Having done all this work, we can state the final result. This is called Stokes' Theorem, but it generalizes Stokes' Green's, Gauss', and all the fundamental theorems we have so far. It is one theorem that clearly, succinctly and elegantly describe the relationships between integration, differentiation and boundary operators, all in the language of differential forms.

**Theorem 4.6.1.** *Let  $S$  be a oriented smooth  $k$ -dimensional manifold in  $\mathbb{R}^n$  with boundary  $\partial S$  a piecewise-smooth oriented  $k-1$ -dimensional manifold with compatible orientation. Let  $\omega$  be a degree  $k-1$  differential form on  $\partial S$  with  $C^2$  coefficients which can be extended to a neighbourhood of  $S$  (so that  $d\omega$  is a degree  $k$  form on  $S$  with  $C^1$  coefficients.) Stoke's theorem is this succinctly stated result.*

$$\int_S d\omega = \int_{\partial S} \omega$$

That's the whole theorem. If  $k = 1$  then we have  $\omega = f$  a scalar field,  $\gamma$  a curve and  $d\omega$  had coefficients which were  $\nabla f$ , so the theorem turns into the fundamental theorem for line integrals.

$$\int_{\gamma} \nabla f = \int_{\gamma} d\omega = \int_{\partial\gamma} \omega = \int_{\partial\gamma} f = f(\gamma(b)) - f(\gamma(a))$$

Integration over a zero dimensional set, such as the endpoint of a curve, is just evaluation. The sign (positive or negative) is given by the orientation.

If  $k = 2$ , then we have  $\omega = f_1 dx + f_2 dy + f_3 dz$  a 1-form,  $\sigma$  is a surface and  $\partial\sigma$  is a closed curve. Stokes' theorem turns into the vector-field version of Stokes' theorem. (We already argued that Green's theorem was just a special case of Stokes' theorem).

$$\int_{\sigma} d\omega = \int_{\partial\sigma} \omega \implies \int_S \nabla \times (f_1, f_2, f_3) = \int_{\partial\sigma} (f_1, f_2, f_3)$$

Lastly, if  $k = 3$ , then we have  $\omega = f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy$ ,  $S$  is a solid region and  $\partial S$  is a closed surface. Stokes' theorem turns into Gauss' theorem.

$$\int_S dw = \int_{\partial S} \omega \implies \int_S \nabla \cdot (f_1, f_2, f_3) = \int_{\partial S} (f_1, f_2, f_3)$$

In this way, using the general theorem for differential forms, we recover Gauss-Green-Stokes and the fundamental theorem of line integrals.