# Multivariable Calculus
# My Way

Michael P. Windham

October 8, 2008

# Contents

# Introduction

This book is about multivariable calculus. It is not a textbook - no exercises, no practice quizzes, not many examples, just calculus. Not all the topics you might expect are covered and not all those covered are covered completely. You see, this is my book. I am going to talk about the topics I want to talk about the way I like to talk about them.

I am fast and loose with hypotheses and, for that matter, grammar. I am more interested in getting a feel for what in some cases and why in others without worrying about technicalities.

Sometimes I emphasize not so important things, and don't emphasize important things. Like the title says "My Way".

You may find it useful, I hope you do.

# Chapter 1

# Algebra and Geometry

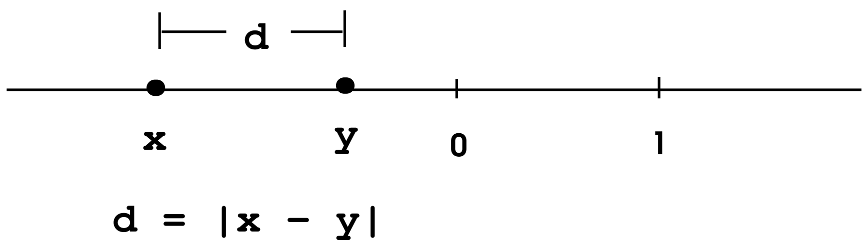
One of the most important things that happened to contribute to the development of modern society was the linking of algebra and geometry. This statement may sound a little pompous, if not ridiculous, but I believe it. We think and feel geometrically. We interact with our environment geometrically, every time we walk through a door without hitting the sides. But, to precisely deal with with the world, accurately and efficiently, we describe it algebraically. We have the ability to link the two, associating geometric objects, usually sets of points, to algebraic objects, usually sets of equations. The equations allow us to reason with precision and the geometry gives us insight into the direction of the calculations and the meaning of the results.

## 1.1 Euclidean spaces

The linking begins with the association of the set of real numbers $\mathbb{R}$ with the points on a straight line.

Simply pick a point and assign it to zero. Pick another point and assign it to one, usually chosen to the right of zero, since most western cultures move from left to right, I suppose. The distance between the two points becomes the unit of distance and determines the associations of the rest of the real numbers to the rest of the points by insisting that

- the distance between the point associated to the number $x$ and the point associated to the number $y$ is given by $|x - y|$
- if $y > x$, then the point associated to $y$ is to the right of the point associated to $x$.



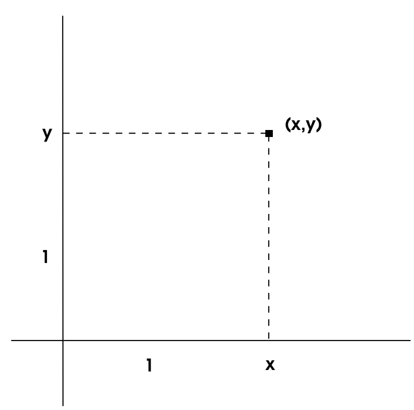

This association is so strong and natural that we almost immediately say the point 3 rather that the point associated to the number 3. They have become for all practical purposes the same. We call the line the real (number) line.

The power of algebra is based on its ability to handle as many quantitative variables as are necessary to deal with a situation. The link between more than one variable and geometry is the following.

### 1.1.1 Rectangular coordinate systems

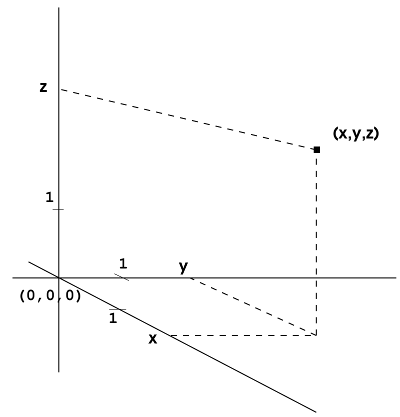
Two variables are described by ordered pairs of real numbers, the set of which is denoted by $\mathbb{R}^2$. These ordered pairs are assigned to points in a euclidean plane. Each pair is assigned to each point in a one-to-one way by first of all building a *coordinate system* by putting in two copies of the real line, intersecting at their zeroes. One of the lines is horizontal and the other is usually perpendicular to it, as illustrated below. There is also a right-handed orientation built into the system by the placement of the ones on the lines, which by the way are called the *axes*. A pair $(x, y)$ is associated to a point by finding $x$ on the horizontal axis and $y$ on the vertical axis and using lines parallel to the axes to pin down the point. But, you already know all this and have been doing it for years or you would not be interested in reading this in the first place.

$$\mathbb{R}^2 = \{(x, y) : x, y \text{ in } \mathbb{R}\}$$

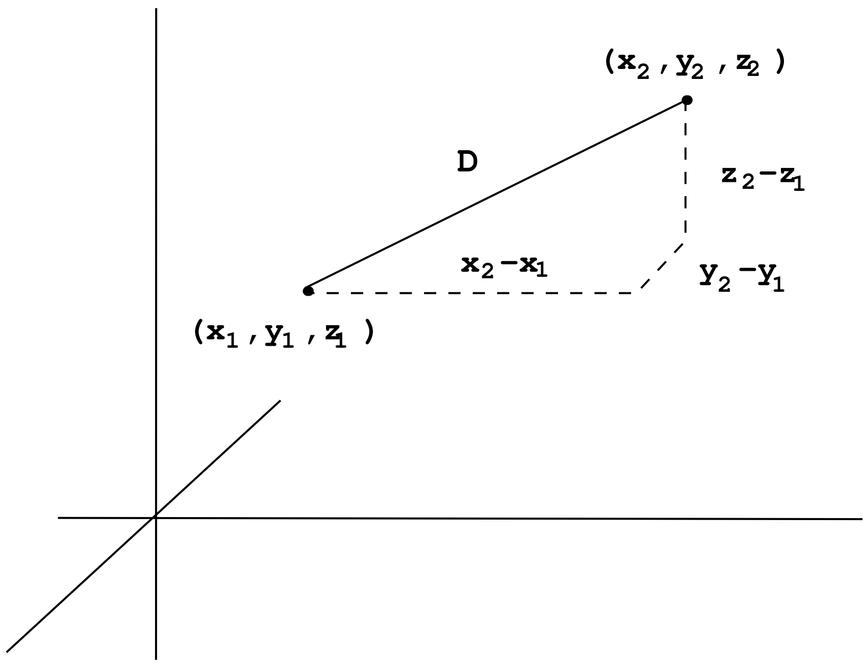



The next step may not be so familiar, but is certainly necessary since the world is not so simple as two things happening at once. So we have for three variables, $\mathbb{R}^3$ and euclidean space, but the link is done in exactly the same way.

$$\mathbb{R}^3 = \{(x, y, z) : x, y, z \text{ in } \mathbb{R}\}$$




We can now begin to link the basic geometrical concepts to algebraic objects and vice versa. Perhaps the most basic is *distance* between two points, which with a coordinate system in place can be calculated in $\mathbb{R}^3$ using the Pythagorean theorem by




$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_1 - z_2)^2}$$

We simply identify the algebraic objects $\mathbb{R}, \mathbb{R}^2$, and $\mathbb{R}^3$ with the line, plane and space, and call them euclidean spaces. That is, $\mathbb{R}^3$ refers to both the algebraic set of ordered triples and geometric space with a coordinate system.

Of course, the world is not always so simple as to be described by three variables, so that, in general we have

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_1, \ldots, x_n \text{ in } \mathbb{R}\}$$

for describing $n$ variables algebraically. But, we have run out of geometry, at least a tangible geometry. We do not give it up, however. We will call $\mathbb{R}^n$ *$n$-dimensional* *euclidean space*, and refer to $(x_1, \ldots, x_n)$ as a point in space and somehow use our geometrical intuition to reason in $\mathbb{R}^n$. This process will be carried even further by talking about calculations in $\mathbb{R}^n$ that make sense geometrically in $\mathbb{R}^3$, as if they made the same sense geometrically in any number of dimensions. Pretty soon you think they do. For example, we call

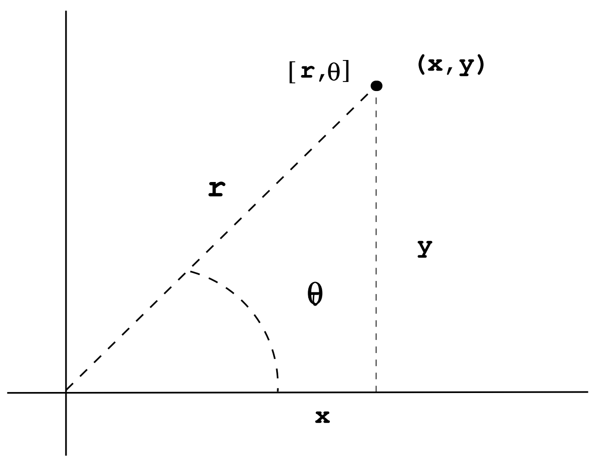
$$\sqrt{(y_1 - x_1)^2 + \ldots (y_n - x_n)^2}$$

the *distance* between two points $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ in $n$-dimensional space.

The coordinate systems I have described are called *rectangular* because of the rectangular relationships between the axes and the rectangles that describe the association to points. There are other approaches that are useful, so let me describe the most common.

### 1.1.2 Polar coordinates in $\mathbb{R}^2$

The *polar* coordinates of a point in the plane are two numbers $r$ and $\theta$ that tell you how to get to the point. The number $r$ tells you the distance from the origin to the point and $\theta$ tells you how far to swing up or down from the $x$-axis to head in the direction of the point.




I have written the polar description of the point with square brackets to distinguish polar coordinates from rectangular. In other words, $[2, \frac{\pi}{3}]$ refers to the polar coordinates of the point whose rectangular coordinates are $(\sqrt{3}, 1)$.

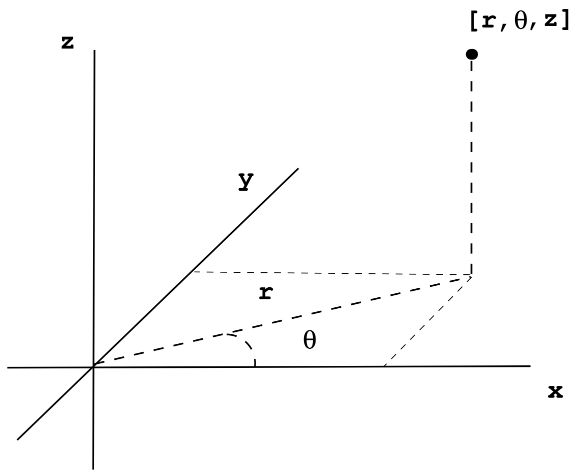
One must be able to convert back and forth between the two descriptions, which can be done with

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

One need only use $r \geq 0$ and $0 \leq \theta < 2\pi$ to identify all the points.

### 1.1.3 Cylindrical coordinates in $\mathbb{R}^3$

*Cylindrical* coordinates are obtained by simply replacing the $x$ and $y$ rectangular coordinates with their polar counterparts.



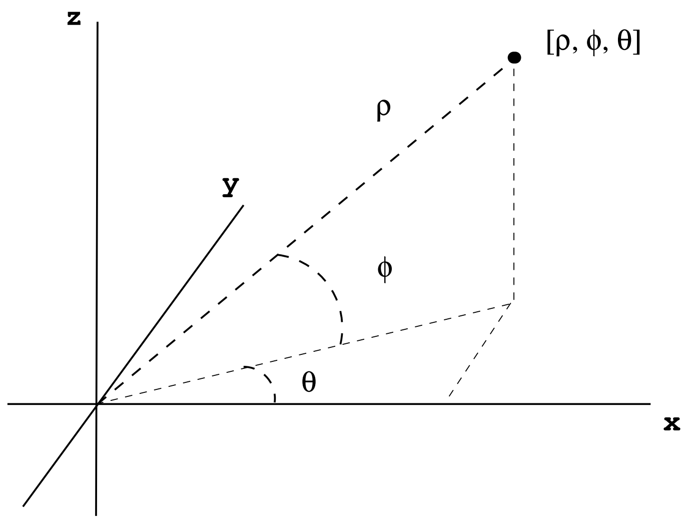

The transformation from one to the other is

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

for $r \geq 0$ and $0 \leq \theta < 2\pi$ and $-\infty < z < \infty$.

### 1.1.4 Spherical coordinates in $\mathbb{R}^3$

*Spherical* coordinates locate a point in terms of $\rho$, its distance from the origin, and two angles, $\phi$ the angle to rotate up (or down) from the $x$-$y$ plane and $\theta$, the angle to swing around the $z$-axis to head in the direction of the point.




The transformation between spherical and rectangular coordinates is

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

for $\rho \geq 0$, $-\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2}$, and $-\pi \leq \theta \leq \pi$.

I must warn you that most calculus text books use the angle between the $z$-axis and the line from the origin to the point for $\phi$, which would cause some differences in the transformation as well as other calculations that occur later on. So, when comparing what I say to others, watch for the difference. My choice allows interpretation of $\phi$ as latitude and $\theta$ as longitude. Seems like a good idea to me.

## 1.2 Algebra and geometry

The serious linking of algebra and geometry involves associating algebraic objects in the form of systems of equations to geometric objects that are the points whose coordinates satisfy the equations.

You have been doing this for a while yourself. In fact, if I asked you what $y = 2x - 1$ was you would likely say it is a straight line with slope 2 and $y$-intercept $-1$. I would agree with you, but technically we are both wrong. $y = 2x - 1$ is an equation and the line refers to the points whose coordinates satisfy the equation. But it is useful and, in fact, quite comfortable to ignore the technical distinction, so I will.

Describing a line with an equation such as $y = 2x - 1$ is one of the first kinds of associations ones learns. One quickly, realizes that the equation $2x - y = 1$ also describes the line. Seen, perhaps, not so quickly, is that

$$x = 2 + 2t$$

$$y = 3 + 4t$$

also describes the line by putting various values for $t$ on the right to produce points $(x, y)$ on the line. I will clarify this later. These three are the basic kinds of descriptions and I want to formalize them for posterity.

We will use the following terminology to classify algebraic and the corresponding geometrical situations.

1. **Explicit:** A system of equations where some of the variables are given explicitly in terms of the others is called an *explicit* system and the corresponding set of points is explicitly described by the system. For example, these two equations in four unknowns

$$\begin{aligned} z &= 3x^2 - xy + 3 \\ w &= xe^{-y} \end{aligned}$$

   explicitly describe $z$ and $w$ in terms of $x$ and $y$, in that substituting values for $x$ and $y$ produces directly values for $z$ and $w$. The geometry sits in $\mathbb{R}^4$, so I don't think I'll worry about it.

2. **Implicit:** A system of equations relating variables in any way is called an *implicit* description. For example,

$$\begin{aligned} x^2 + y^2 + z^2 &= 4 \\ 3x + 2y - z &= 1 \end{aligned}$$

   relates the variables $x$, $y$ and $z$ implicitly in that values of one or more of the variables produce values of the other that satisfy the equations, but not without a little work. Note that geometry here is in $\mathbb{R}^3$, but again we will delay the details.

3. **Parametric:** A *parametric* description uses external variables called *parameters* to produce values for the variables of interest. For example,

$$\begin{aligned} x &= 2\cos t \\ y &= 3\sin t \end{aligned}$$

   describes the variables $x$ and $y$ using the parameter $t$. The points we want are the $(x, y)$'s produced by the values of the parameter.

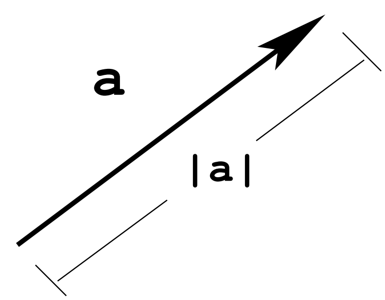
# Chapter 2

# Vectors

A *vector* is an object that characterizes a magnitude and a direction. There are both geometric and algebraic versions. Vectors are used to described such things as forces which act with a certain strength (magnitude) in a particular direction, and velocity which describes your speed (magnitude) and direction you are going.

I will begin with the geometric version of a vector which simply specifies an arrow in the plane or space with the length of the arrow given by the magnitude and the direction it is pointing, well, that's the direction.

I will use boldface letters to denote vectors, **a**, **b**, etc. You will also see $\vec{a}$. The front end of the vector is called the *head* and the back end the *tail*. The magnitude of a vector **a** is denoted |**a**| and is also called its length.




Vectors have no location assigned to them. You can put them wherever you want, only the direction and magnitude matter. If you are heading southwest at thirty miles per hour it does not matter where you are, the description of your velocity is complete.

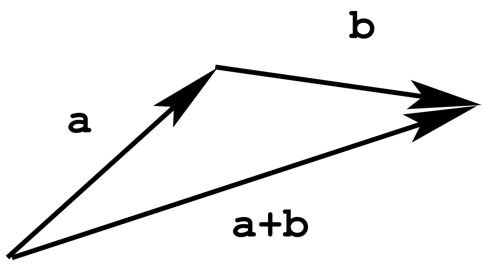
Vectors are useful for describing change. For example, a vector drawn from a point $P$ to a point $Q$ is called a *displacement vector* and denoted $\vec{PQ}$. It can be thought of as describing the change in location in moving from $P$ to $Q$.

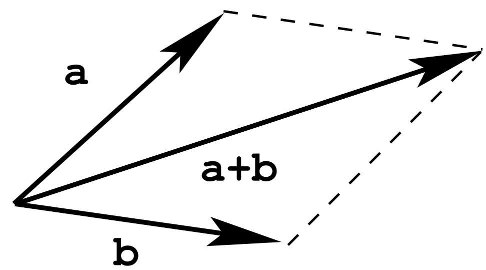
## 2.1 Vector arithmetic

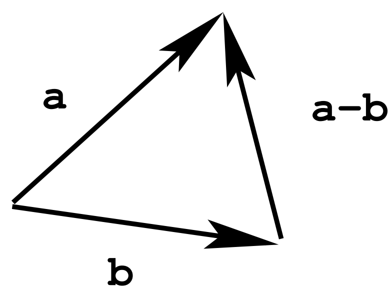
There is, in fact, a variety of ways of manipulating vectors, using operations with arithmetic names and geometric definitions.

### 2.1.1 Addition and subtraction

Adding two vectors **a** and **b** is done by putting the tail of **b** to the head of **a** and connecting the tail of **a** to the head of **b**. Got that? Try the illustration on the left below. This formulation is useful for combining displacements to get a total displacement. The illustration to the right of that gives an another way to look at addition. This formulation is useful for working with velocities or forces. Vectors **a** and **b** "act" at the same location and $\mathbf{a} + \mathbf{b}$ is the resultant effect of their actions.




Subtraction just produces the vector $\mathbf{a} - \mathbf{b}$ that one must add to **b** to get **a**.

I am not going to spend too much time motivating this arithmetic. We will see it everywhere in time and motivation will occur naturally. For the moment I just want to accumulate the information necessary to get going. In particular, how do these operations work? What are the rules?

The basic rules for addition of vectors are the following.

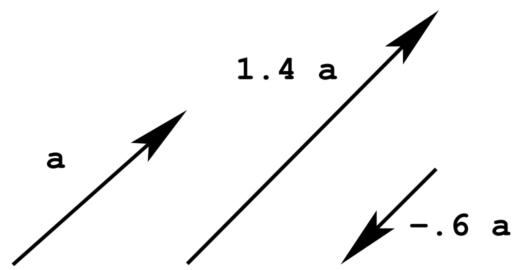
1. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$
2. $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$
3. The is a vector $\mathbf{0}$, called the *zero* vector, so that $\mathbf{a} + \mathbf{0} = \mathbf{a}$ for any vector **a**. The zero vector is just the one with length zero, it points in any direction since it does not go anywhere.
4. For each **a** there is a vector $-\mathbf{a}$ satisfying $\mathbf{a} + -\mathbf{a} = \mathbf{0}$. The vector $-\mathbf{a}$ is just a copy of **a** with the head and tail reversed. Moreover, $\mathbf{a} - \mathbf{b} = \mathbf{a} + -\mathbf{b}$.

You can verify these rules easily by just drawing a couple of pictures, so I will leave you to do so.

These rules do not appear to say too much, and, of course, there are many more. But, what they tell you is that vector addition and subtraction act exactly like real number addition and subtraction, so just go for it.

### 2.1.2 Scalar multiplication

Scalar multiplication combines a real number $r$ with a vector $\mathbf{a}$ to produce a vector denoted $r\mathbf{a}$, with length $|r||\mathbf{a}|$ and same direction as $\mathbf{a}$ for $r \geq 0$ and the opposite direction for $r < 0$. In other words, scalar multiplication stretches or shrinks a vector with possibly a reversal of direction.



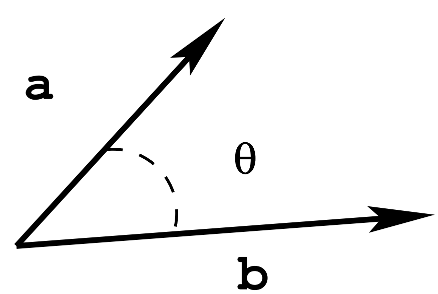

The rules are

1. $r(\mathbf{a} + \mathbf{b}) = r\mathbf{a} + r\mathbf{b}$
2. $(r + s)\mathbf{a} = r\mathbf{a} + s\mathbf{a}$
3. $(rs)\mathbf{a} = r(s\mathbf{a})$
4. $1\mathbf{a} = \mathbf{a}$
5. $0\mathbf{a} = \mathbf{0}$
6. $(-1)\mathbf{a} = -\mathbf{a}$

Not very exciting, and there are more rules. Even though you are multiplying different kinds of objects the arithmetic acts like multiplication in the reals.

### 2.1.3 Dot product

The *dot product* of vectors $\mathbf{a}$ and $\mathbf{b}$ is the real number $|\mathbf{a}||\mathbf{b}| \cos \theta$, where $\theta$ is the angle between the two vectors when their tails are touching. It is denoted $\mathbf{a} \cdot \mathbf{b}$




The dot product is probably somewhat mysterious at this point, but you will find that it is incredibly useful. It, too, has its rules. Some are a bit difficult to believe, but they will become clearer.

1. $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$

2. $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$

3. $(r\mathbf{a}) \cdot \mathbf{b} = \mathbf{a} \cdot (r\mathbf{b}) = r(\mathbf{a} \cdot \mathbf{b})$

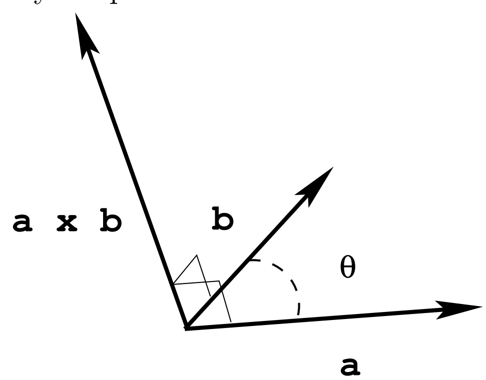
4. $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$

When the dot product of two vectors is zero the vectors are said to be *orthogonal, normal* or *perpendicular* to each other since the angle between them is $\frac{\pi}{2}$.

### 2.1.4 Cross product

The next operation is truly mysterious at first, but you will learn to love it. It is the *cross product*. Everything defined to this point works equally well in the plane or in space, but the cross product is defined only in space.

It is denoted $\mathbf{a}\times\mathbf{b}$ and is the vector with length $|\mathbf{a}||\mathbf{b}|\sin\theta$, where $\theta$ is the angle between the vectors and direction perpendicular to both vectors satisfying the so-called right-hand rule, which is best explained by carefully looking at the illustration.




If you put your pointing finger along $\mathbf{a}$ and your middle finger along $\mathbf{b}$, then your thumb points in the direction of $\mathbf{a} \times \mathbf{b}$, provided you are doing these gyrations with your *right* hand, hence the term "right-hand rule". Note also that the length of the cross product is the area of the parallelogram determined by using $\mathbf{a}$ and $\mathbf{b}$ as adjacent sides.

The rules are not as straightforward as for the other operations, be careful.

1. $\mathbf{a} \times \mathbf{a} = \mathbf{0}$

2. $\mathbf{b} \times \mathbf{a} = -\mathbf{a} \times \mathbf{b}$ (Your fingers have switched places, so your thumb switches directions.)

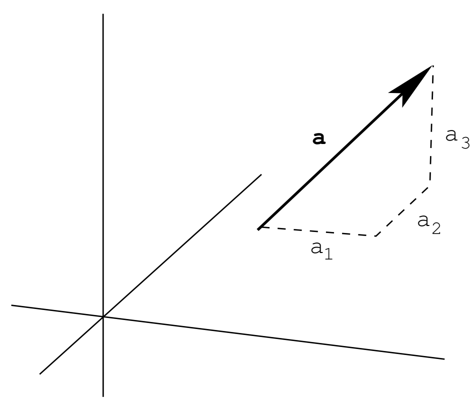
3. $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$

4. $\mathbf{a} \times (r\mathbf{b}) = (r\mathbf{a}) \times \mathbf{b} = r(\mathbf{a} \times \mathbf{b})$

The first two are believable, but the last two, although things you would do instinctively, are not so obviously true.

### 2.1.5 The way

It may appear that I have said too little about what is going on and I cannot disagree. In fact, the situation is much worse than it appears. You have probably been thinking very geometrically about this arithmetic and that is what I intended. Much of the value of these operations is in the geometrical insight they will provide in a wide variety of situations. But, have you actually, made any precise "calculations". I doubt it, because all I have told you is how to draw pictures. Pictures are our friends, but they do not get the work done. In order to effectively work with vectors we need more structure. In particular, we need to put a coordinate system in space and use it to precisely describe vectors in a useful way.




The description provides a recipe for drawing the vector using three numbers that delineate a path from the tail to the head. The first number, $a_1$, gives the length and direction of the path parallel to the $x$-axis, the second, $a_2$, parallel to the $y$-axis and the third, $a_3$, parallel to the $z$-axis. There is no pre-ordained starting point - your choice.

As is the custom we will identify the geometrical vector with these three numbers, called its *components* and write

$$\mathbf{a} = \langle a_1, a_2, a_3 \rangle$$

where the brackets emphasize that this is a description of a vector, not a point.

Now for the really good news. We have the following for $\mathbf{a} = \langle a_1, a_2, a_3 \rangle$, $\mathbf{b} = \langle b_1, b_2, b_3 \rangle$, and $r$ in $\mathbb{R}$,

1. $|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}$
2. $-\mathbf{a} = \langle -a_1, -a_2, -a_3 \rangle$
3. $\mathbf{0} = \langle 0, 0, 0 \rangle$
4. $\mathbf{a} + \mathbf{b} = \langle a_1 + b_1, a_2 + b_2, a_3 + b_3 \rangle$
5. $\mathbf{a} - \mathbf{b} = \langle a_1 - b_1, a_2 - b_2, a_3 - b_3 \rangle$
6. $r\mathbf{a} = \langle ra_1, ra_2, ra_3 \rangle$
7. $\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3$

8. $\mathbf{a} \times \mathbf{b} = \langle a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1 \rangle$

The first six are easy enough to believe with at most a picture or two. The dot product and cross product are a bit more interesting.

For the dot product, look at the triangle using $\mathbf{a}$ and $\mathbf{b}$ as adjacent sides, then $\mathbf{a} - \mathbf{b}$ is the side opposite the angle $\theta$ between $\mathbf{a}$ and $\mathbf{b}$. The law of cosines gives

$$\begin{aligned} |\mathbf{a} - \mathbf{b}|^2 &= |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}| \cos\theta \\ &= |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2\, \mathbf{a} \cdot \mathbf{b} \end{aligned}$$

Multiply everything out using 1 and 5, and the required miracle happens when everything cancels but 7.

The cross product is a mess no matter how you look at it. The easiest way out is to continue to wonder how to derive the result and merely check that it works. First check that the vector $\langle a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1 \rangle$ in 8, call it $\mathbf{c}$ for the moment is orthogonal to $\mathbf{a}$ and $\mathbf{b}$ by calculating $\mathbf{a} \cdot \mathbf{c}$ and $\mathbf{b} \cdot \mathbf{c}$ to see that they are both zero. Next check the length where the fact that $|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2|\mathbf{b}|^2(1 - \cos^2\theta) = |\mathbf{a}|^2|\mathbf{b}^2| - (\mathbf{a} \cdot \mathbf{b})^2$ will be useful. Finally, check that the right-hand rule is satisfied.

These formulas are the basis for precise calculation in vector arithmetic and that together with the insight provided by the geometrical descriptions are what we need to get going.

### 2.1.6 The standard basis

The *standard basis* in $\mathbb{R}^3$ consists of the three vectors $\mathbf{i} = \langle 1, 0, 0 \rangle$, $\mathbf{j} = \langle 0, 1, 0 \rangle$ and $\mathbf{k} = \langle 0, 0, 1 \rangle$. Their role is to provide an alternative way of writing vectors , namely

$$\langle a_1, a_2, a_3 \rangle = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$$

## 2.2 Miscellaneous

### 2.2.1 Area and volume

Recall that the length of the cross product of two vectors is the area of the parallelogram formed using the vectors as adjacent sides. If the vectors are $\langle a_1, a_2, 0 \rangle$ and $\langle b_1, b_2, 0 \rangle$ the the area is given by $|a_1b_2 - a_2b_1|$. In fact what we have shown here is how to calculate the area of a parallelogram in $\mathbb{R}^2$ determined by vectors $\langle a_1, a_2 \rangle$ and $\langle b_1, b_2 \rangle$.

You may recognize the calculation $a_1b_2 - a_2b_1$ as the *determinant* of a $2 \times 2$ array of numbers, which I will denote by

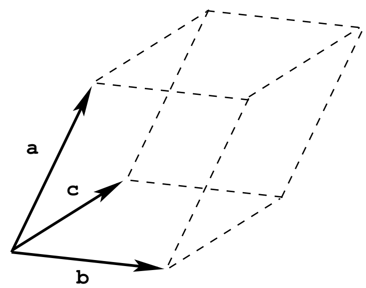
$$\det \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} = a_1b_2 - a_2b_1$$

You may have seen it in solving systems of equations. You may not have known that it computed area. If not, you do now.

The determinant for a $3 \times 3$ array also has geometrical implications for us. It is calculated by

$$\det \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1)$$

Note that if the entries of the array are the components of vectors **a**, **b** and **c**, then the determinant computes $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$, which can be seen to be the volume of the parallelepiped determined by using the vectors as adjacent sides.



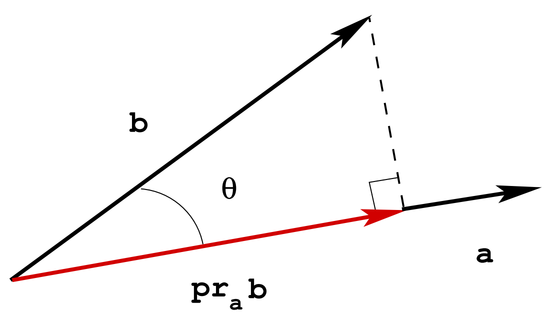

And a final note, if the components of **a** are replaced by the standard basis vectors the determinant calculates $\mathbf{b} \times \mathbf{c}$.

### 2.2.2 Projections

The *projection* of a vector **b** onto a vector **a** is given by

$$\mathrm{pr}_{\mathbf{a}}\mathbf{b} = |\mathbf{b}| \cos\theta \frac{\mathbf{a}}{|\mathbf{a}|} = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|^2}\mathbf{a}$$

.




If **b** represented a force, its projection onto **a** would describe the effect of the force in the **a** direction.

# Chapter 3

# Algebra and Geometry - seriously

So, how do you link algebra and geometry? Any way you can.

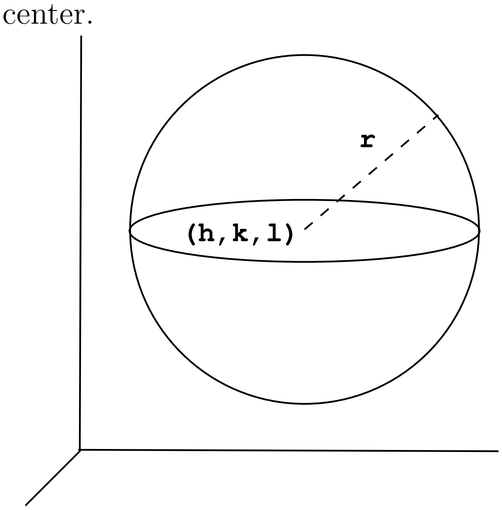
## 3.1 Geometry to algebra

Let us begin by looking at how you describe some familiar geometrical objects algebraically.

### 3.1.1 Sphere

To specify a sphere you need to know its center and radius, call it $r$. The sphere is just the set of points a distance $r$ from the center.

To describe anything algebraically you need a coordinate system. With the coordinate system, the center can be identified by its coordinates, $(h, k, l)$.




Now, just use the distance formula, and the sphere "becomes" the set of points $(x, y, z)$ satisfying

$$(x-h)^2 + (y-k)^2 + (z-l)^2 = r^2$$

which describes the sphere *implicitly*.

From here, you can get an *explicit* description by simply solving for one of the variables.

$$z = \begin{cases} l + \sqrt{r^2 - (x-h)^2 - (y-k)^2} \\ l - \sqrt{r^2 - (x-h)^2 - (y-k)^2} \end{cases}$$

In this case, there are two possibilities, a bit messy. The first describes the top of the sphere and the other the bottom.

How about a *parametric* description. Sometimes, you just have to get lucky. In this case, luck is having already been told about spherical coordinates, from which we obtain

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

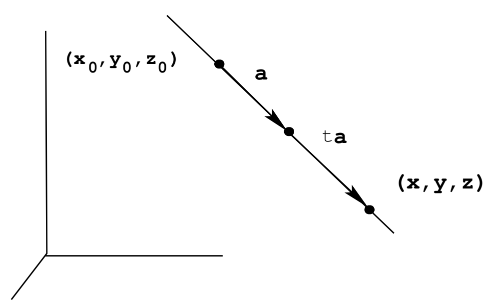
where $r$ is the (constant) radius and the parameters are $\theta$ and $\phi$.

### 3.1.2 Line

To specify a line in space you need one of the following

- two different points on the line, say $(x_0, y_0, z_0)$ and $(x_1, y_1, z_1)$.
- one point on the line to locate it, say $(x_0, y_0, z_0)$ and a description of the direction the line goes. The direction could easily be give by a vector parallel to the line.
- the intersection of two planes.

As it happens, the second is easy to use.




Using a given point $(x_0, y_0, z_0)$ and a vector $\mathbf{a} = \langle a_1, a_2, a_3\rangle$ we need to describe the coordinates of any point $(x, y, z)$ on the line in terms of them. Now, the displacement vector from $(x_0, y_0, z_0)$ to $(x, y, z)$ is parallel to $\mathbf{a}$, so it is some scalar multiple of $\mathbf{a}$, and we are almost done.

From

$$\langle x - x_0, y - y_0, z - z_0\rangle = t\mathbf{a}$$

with a little arithmetic, we extract

$$\begin{aligned} x &= x_0 + a_1 t \\ y &= y_0 + a_2 t \\ z &= z_0 + a_3 t \end{aligned}$$

a *parametric* description of a line in space. You can think that letting the parameter $t$ range over the reals draws the line.

If you have two points you can use the displacement vector from one to the other, $\langle x_1 - x_0, y_1 - y_0, z_1 - z_0 \rangle$ for **a**.

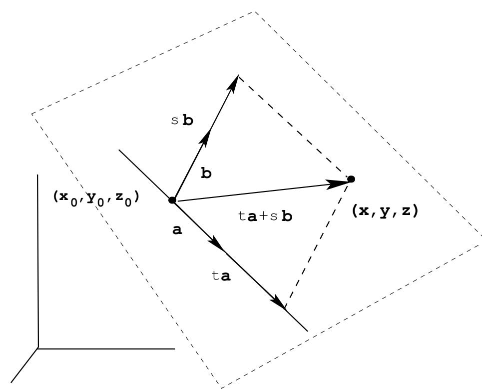
I will come back to the problems of finding *explicit* and *implicit* descriptions. Also, the third way of specifying a line requires that we know how to specify a plane. So, let's do that.

### 3.1.3 Plane

The most famous ways of specifying a plane in space are

- two intersecting lines
- three points in the plane that do not lie on the same line
- a point in the plane and a line perpendicular to the plane.

The first one looks attractive since we are now familiar with lines. We need the point of intersection and two vectors one parallel to each line. So suppose they are $(x_0, y_0, z_0)$, **a** and **b**.




The same trick that worked for the line works for the plane, namely express the displacement vector to any point in the plane $(x, y, z)$ from $(x_0, y_0, z_0)$ in terms of **a** and **b**. The figure shows how to build the displacement vector as a sum of scalar multiples of **a** and **b**. That's it.

We have, then

$$\langle x - x_0, y - y_0, z - z_0 \rangle = t\mathbf{a} + s\mathbf{b}$$

with a little arithmetic, we extract

$$\begin{aligned} x &= x_0 + a_1 t + b_1 s \\ y &= y_0 + a_2 t + b_2 s \\ z &= z_0 + a_3 t + b_3 s \end{aligned}$$

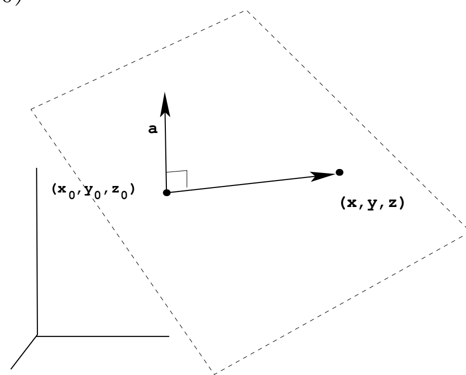
a *parametric* description of a plane in space using parameters $t$ and $s$.

If we are given three points, then we pick one to use for $(x_0, y_0, z_0)$, use it as the tail for displacement vectors to the other two, use them for **a** and **b** and be done with it.

Let me delay finding an *explicit* description for the moment and look instead at the third option.

As usual, we want a vector **a** to define the direction of the line perpendicular to the plane and a point on the plane $(x_0, y_0, z_0)$.

And, as usual we relate the displacement vector from $(x_0, y_0, z_0)$ to any point $(x, y, z)$ in the plane to **a**. As the figure indicates, they are perpendicular to each other, so that their dot product is zero.




That is,

$$\begin{aligned} \mathbf{a} \cdot \langle x - x_0, y - y_0, z - z_0 \rangle &= 0 \\ a_1(x - x_0) + a_2(y - y_0) + a_3(z - z_0) &= 0 \end{aligned}$$

which is an *implicit* description of the plane.

## 3.2 Algebra to geometry

Drawing pictures is what it is all about.
**Explicit and implicit:**

You have a great deal of experience in these cases in $\mathbb{R}^2$. You have the graphs of many functions, and a few other special cases that will be very useful in the near future, such as

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

an ellipse and

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

an hyperbola.

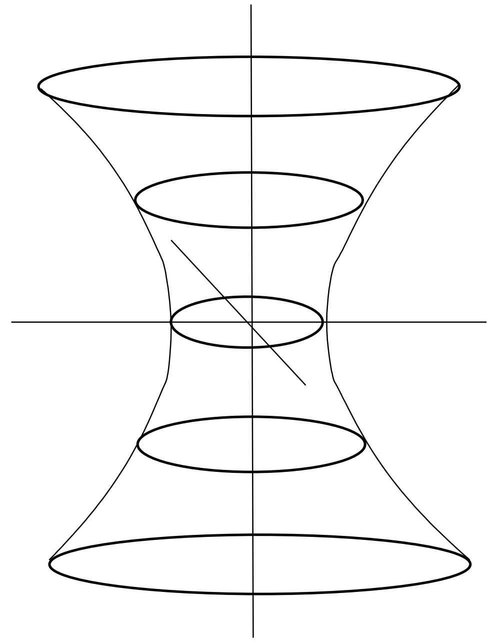
You may have no experience in $\mathbb{R}^3$, so let me give you a trick. The idea is to reduce the problem to $\mathbb{R}^2$ where you can use your experience. You slice the object with planes perpendicular to the axes to get two dimensional cross-sections.

For example, what geometrical object does

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

correspond to?

Let $z = 0$, then you have $x^2+y^2 = 1$ which is just the unit circle in the plane. But, we are in space, so the points that satisfy the equation also have $z = 0$. So, the circle is a horizontal slice made by the $x$-$y$ plane of the object we are looking for. Slice at various heights along the $z$-axis to get cross-sections of the object. In particular for $z = b$ you have $x^2+y^2 = b^2+1$, another circle. The horizontal cross-sections are circles getting large as $|b|$ gets larger.



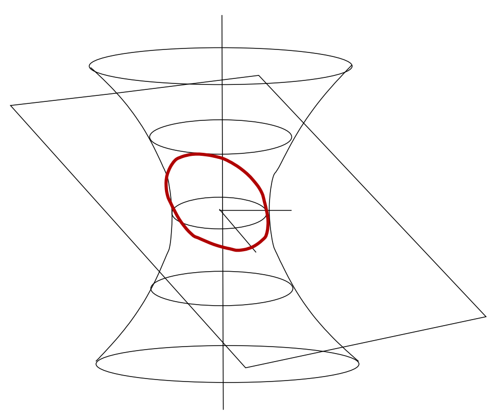
You can take vertical slices as well, at least one or two to clarify how the horizontal slices stack up. For example, letting $x = 0$ in $x^2 + y^2 - z^2 = 1$ shows that slicing the object with the $y$-$z$ plane is the hyperbola, $y^2 - z^2 = 1$ which clarifies the shape of the surface.

If you have more than one equation, the geometrical object is the intersection of the objects described by each equation

For example,

$$\begin{aligned} x^2 + y^2 - z^2 &= 1 \\ x + y + z = 0 \end{aligned}$$

The first equation describes the hyperboloid as before and the second a plane through the origin. The two together describe the intersection, which happens to be an ellipse



This example also suggests how to describe a line as the intersection of two planes. Each plane is represented by an equation of the form $ax + by + cz = d$, and the line is represented by the system of the two equations.

**Parametric:** Quite frankly, the best way to handle parametric descriptions for $\mathbb{R} \to \mathbb{R}^2$ is to get a graphing calculator and look up parametric graphing in the manual. For

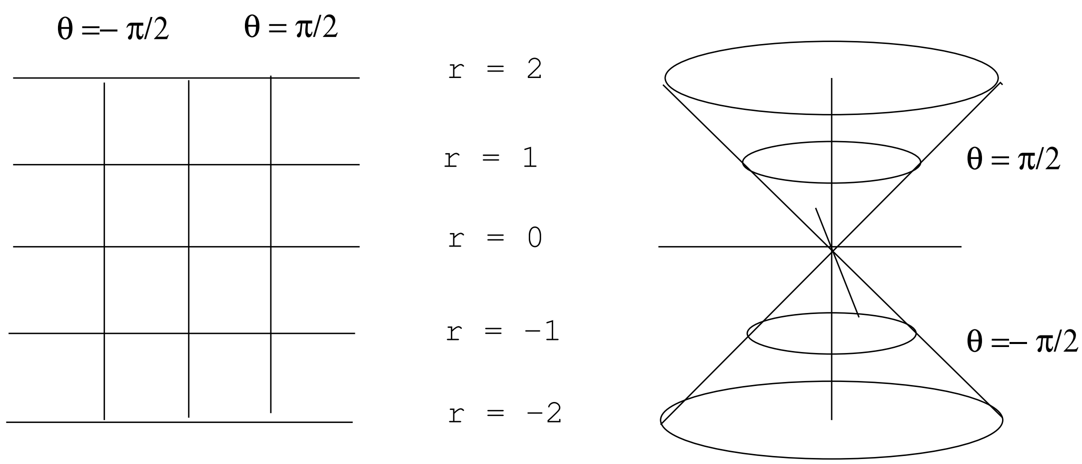
$\mathbb{R} \to \mathbb{R}^3$ you can sometimes get insight by using the graphing calculator with two of the equations at a time. You are looking down the axis of the variable you ignored. Then, see if you can put it all together.

For $\mathbb{R}^2 \to \mathbb{R}^3$ you will generally be drawing surfaces, By letting one of the parameters be constant, you draw a curve in the surface using the other parameter. Several of these curves should help you visualize the surface.

For example, suppose you have

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

For a fixed $r$ you draw a circle with radius $r$ at height $r$ on the $z$-axis. For a fixed $\theta$ you draw a straight line through the origin parallel to the vector $\langle \cos\theta, \sin\theta, 1\rangle$,




## 3.3 Going from one to another

You may have one kind of algebraic description, but need another. Let me show you some ways to do that.

**Explicit to implicit:** An explicit description is simply a special case of an implicit description. That was easy.

**Explicit to parametric:** This one is easy too. You have some variables explained by others. Use the explainers for the parameters. For example, you have $z = x^2 + y^2$, so let

$$\begin{aligned} x &= x \\ y &= y \\ z &= x^2 + y^2 \end{aligned}$$

and there you have it. The parameters are $x$ and $y$. They are also the variables being parameterized, which may be confusing and perhaps even embarrassing. If you want to feel more comfortable do this.

$$\begin{aligned} x &= s \\ y &= t \\ z &= s^2 + t^2 \end{aligned}$$

to distinguish things properly.

**Parametric to implicit:** Eliminate the parameters. For example,

$$\begin{aligned} x = \tan t \to t &= \tan^{-1} x \\ &\qquad\downarrow \\ y = \sec t \to y &= \sec(\tan^{-1} x) \end{aligned}$$

which is, in fact, an explicit relationship between $y$ and $x$.

You may also be able to take advantage of structure specific to the situation. For the example, you could observe that

$$y^2 = \sec^2 t = \tan^2 + 1 = x^2 + 1$$

and obtain a nice implicit description

$$y^2 - x^2 = 1$$

which tells you you have an hyperbola.

The first approach may seem more straightforward, but you may lose information. For example, for $x = \sqrt{3}$, the first gives $y = 2$, but the second gives $y = \pm 2$.

Another example will finish off planes. suppose you have the plane thru $(1, 2, 0)$ parallel to $\langle -1, 1, 1\rangle$ and $\langle 2, -1, 1\rangle$, then the parametric equations for the plane are

$$\begin{aligned} x &= 1 - s + 2t \\ y &= 2 + s - t \\ z &= s + t \end{aligned}$$

Solve the first two for $s$ and $t$ in terms of $x$ and $y$ to get

$$\begin{aligned} s &= x + 2y - 5 \\ t &= x + y - 3 \end{aligned}$$

and substitute in to the third equation to get an implicit (in fact, explicit) description of the plane

$$z = x + 2y - 5$$

**Implicit to explicit:** Try to solve for some of the variables in terms of the others, if you are lucky.

This is a problem that can be attacked with calculus, amazingly enough. More on that later.

**Parametric to explicit:** Go to implicit and convert that to explicit, if you are lucky.

**Implicit to parametric:** This can be really tough. You want somehow to come up with parameters to describe something some way. There is no systematic way to do that. You call on experience and magic.

For example,

$$x^2 + y^2 = 1$$

describes the unit circle. Trigonometry tells you that the coordinates of a point on the circle can be described in terms of distance along the circle, defining the sine and cosine functions. This leads to the parametrization of the circle

$$\begin{aligned} x &= \cos t \\ y &= \sin t \end{aligned}$$

Once you have a parameterization of something you can see ways to expand its use.

For example,

$$\begin{aligned} x &= h + r\cos t \\ y &= k + r\sin t \end{aligned}$$

draws a circle with radius $r$ centered at $(h, k)$.

$$\begin{aligned} x &= h + a\cos t \\ y &= k + b\sin t \end{aligned}$$

draws an ellipse centered at $(h, k)$. It is, in fact, the ellipse described implicitly by

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$$

# Chapter 4

# Functions

We have spent a lot of time looking at algebra and geometry, but this is supposed to be a book about calculus. Calculus you do to functions, so now it is time to look at them.

You, by now, are experts on real-valued functions of a real variable,

$$\begin{aligned} f: \quad & \mathbb{R} \to \mathbb{R} \\ & x \to \; y = f(x) \end{aligned}$$

You input $x$ and the function outputs $y$. You are now ready to increase the numbers of both inputs and outputs. We will be looking at all of the following at some point or another.

$$\begin{array}{lll} \mathbb{R} \to \mathbb{R} & \mathbb{R}^2 \to \mathbb{R} & \mathbb{R}^3 \to \mathbb{R} \\ x \to \; y & (x,y) \to \; z & (x,y,z) \to \; w \\ & & \\ \mathbb{R} \to \mathbb{R}^2 & \mathbb{R}^2 \to \mathbb{R}^2 & \mathbb{R}^3 \to \mathbb{R}^2 \\ t \to \; (x,y) & & \\ & & \\ \mathbb{R} \to \mathbb{R}^3 & \mathbb{R}^2 \to \mathbb{R}^3 & \mathbb{R}^3 \to \mathbb{R}^3 \\ t \to \; (x,y,z) & & \end{array}$$

We will focus on these cases because we will have geometry to help with the study, but, in fact, we could do it all with

$$\begin{aligned} f: \quad & \mathbb{R}^n \to \mathbb{R}^m \\ & x = (x_1, \ldots \; x_n) \to \; y = (y_1, \ldots, y_m) = f(x) \end{aligned}$$

## 4.1 Sets for functions

To help organize the study of functions and more importantly to bring geometrical insight into the picture, there are sets associated to a function and they are the following.

**Domain:** $\{x \text{ in } \mathbb{R}^n : f(x) \text{ is defined } \}$
The domain is the set of possible inputs, which may be specified in the definition of the

function or assumed to be any $x$ for which nothing goes wrong when it is put into the function.

**Range:** $\{y \text{ in } \mathbb{R}^m : y = f(x) \text{ for some } x\}$
The range is just the set of possible outputs. It is also called the *image*.

**Level sets:** Level sets come in two varieties.
For $b$ in $\mathbb{R}^m$, the *level set at* $b = \{x \text{ in } \mathbb{R}^n : f(x) = b\}$
For $a$ in $\mathbb{R}^n$, the *level set through* $a = \{x \text{ in } \mathbb{R}^n : f(x) = f(a)\}$
The first is the set of all the points in the domain that produce the output $b$, and the second is set of inputs that produce the same output as $a$. You have actually worked with level sets all your algebraic life. They are just solution sets to equations, now appearing in a different context.

**Graph:** $\{(x, y) \text{ in } \mathbb{R}^{n+m} : y = f(x)\}$
The graph is your old friend, in one variable calculus, at least. I need to say that $(x, y)$ means $(x_1, \ldots\ x_n, y_1, \ldots, y_m)$, which says that, for example, the graph of a function from $\mathbb{R}^3$ to $\mathbb{R}^2$ is in $\mathbb{R}^5$, which may be difficult to draw.

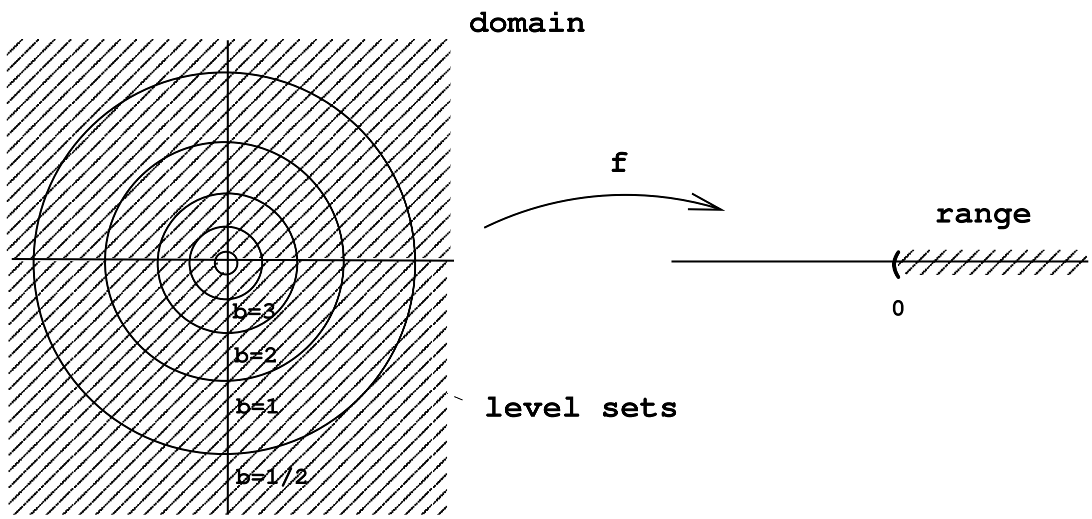
Before going on let me point out again exactly where these sets are. The domain and level sets are in $\mathbb{R}^n$. The range is in $\mathbb{R}^m$. The graph is in $\mathbb{R}^{n+m}$.

## 4.2 Examples

Let's take a look at $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $z = f(x, y) = \dfrac{1}{x^2 + y^2}$.
**Domain:** $\{(x, y) \text{ in } \mathbb{R}^2 : f(x, y) = \frac{1}{x^2+y^2} \text{ is defined }\}$

Since nothing about the domain is specified in the definition of the function, all we need do is ask what can go wrong. Certainly if $x$ and $y$ are both zero, then the function is undefined. Otherwise, there is no problem, so that the domain of $f$ is $\{(x, y) \neq (0, 0)\}$.

**Range:** $\{z \text{ in } \mathbb{R} : z = f(x, y) = \frac{1}{x^2+y^2} \text{ for some } (x, y)\}$
We are being asked what are the possible values of $\frac{1}{x^2+y^2}$. Because of the squares the value will always be positive. In fact, any positive value is possible. For $b > 0$, just let $(x, y) = (1/\sqrt{b}, 0)$, then $f((1/\sqrt{b}, 0)) = 1/(1/\sqrt{b})^2 = b$. So, the range of $f$ is $\{z : z > 0\} = (0, \infty)$.
All of these are illustrated in the diagram below. This picture is a new one, since it separates the domain space and range space. In one variable calculus these are just real lines, and have little structure to provide insight. With more variables, they become interesting in their own right.



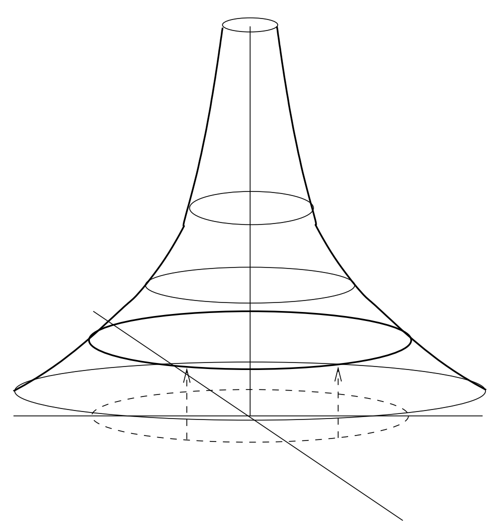

**Level sets:** For $b$ in $\mathbb{R}$, the level set at $b = \{x \text{ in } \mathbb{R}^n : \frac{1}{x^2+y^2} = b\}$

One need only consider $b$'s in the range, in this case, $b > 0$. We want the $(x, y)$'s so that

$$\begin{aligned} \frac{1}{x^2+y^2} &= b \\ x^2+y^2 &= \frac{1}{b} \end{aligned}$$

which shows that geometrically, the level sets are circles centered at the origin with radius $1/\sqrt{b}$. The circles become smaller as $b$ increases. You will find level sets ocurring more than you might expect, but for the moment their claim to fame will be their ability to help with the graph.

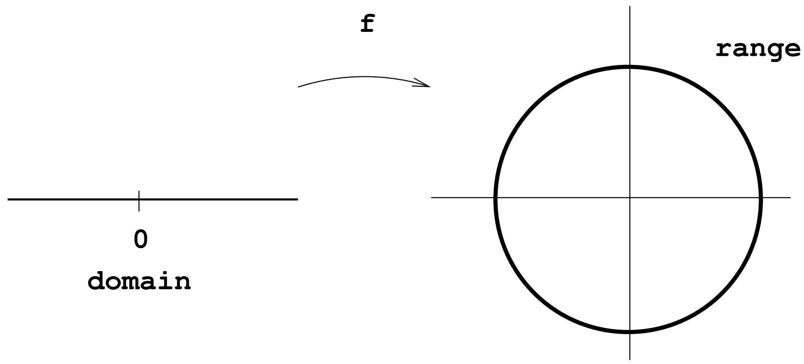
**Graph:** $\{(x, y, z) \text{ in } \mathbb{R}^3 : z = f(x, y) = \frac{1}{x^2+y^2}\}$
The secret is that for $(x, y)$ in the level set at $b$, the point $(x, y, b)$ is on the graph. So, the level set is a picture of a horizontal slice of the graph at height $b$. To construct the graph, just lift the level sets to their levels.



This approach is just the horizontal slicing technique in the previous chapter, using different terminology. Taking a vertical slice or two can ensure that you have a good picture.

Let us look at another example, $f : \mathbb{R} \to \mathbb{R}^2$ defined by $f(t) = (2\cos t, 2\sin t)$
**Domain:** The domain is $\mathbb{R}$.

**Range:** $\{(x, y) \text{ in } \mathbb{R}^2 : x = 2\cos t \text{ and } y = 2\sin t\}$, in other words the range is defined by familiar parametric equations for the circle centered at the origin with radius 2.



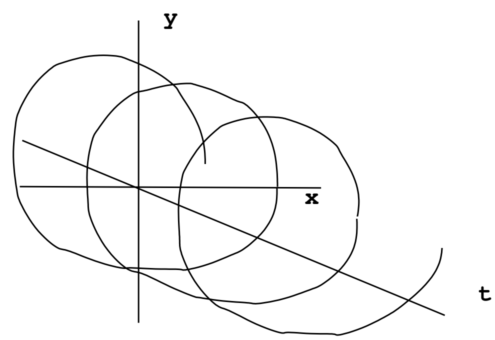

**Level sets:** For example, the level set at $(\sqrt{3}, 1)$ is the set of solutions to the equations

$$\begin{array}{lcl} 2\cos t = \sqrt{3} & \quad\text{or}\quad & \cos t = \frac{\sqrt{3}}{2} \\ 2\sin t = 1 & & \sin t = \frac{1}{2} \end{array}$$

So, the level set is $\{t : t = \frac{\pi}{6} + 2\pi n \text{ for any } n\}$

**Graph:** $\{(t, x, y) : x = 2\cos t \text{ and } y = 2\sin t\}$

If you look down the $t$-axis you see the range in the $x$-$y$ plane. So, as $t$ moves out along the axis, $x$ and $y$ go around the circle.




## 4.3 Functions for sets

You may have noticed familiar geometric objects in the examples. In fact, we can incorporate functions into the algebra-geometry classification scheme.

For a set $S$

| Algebra-geometry | | Functions |
|---|---|---|
| $S$ is described *explicitly* | $\longleftrightarrow$ | $S$ is the *graph* of a function |
| $z = x^2 + xy - 3$ | | $f(x, y) = x^2 + xy - 3$<br>$\mathbb{R}^2 \to \mathbb{R}$ |
| $y = 3x - 2$<br>$z = x + 1$ | | $f(x) = (3x - 2, x + 1)$<br>$\mathbb{R} \to \mathbb{R}^2$ |
| $S$ is described *parametrically* | $\longleftrightarrow$ | $S$ is the *range* of a function |
| $x = 1 - 2\cos t$<br>$y = 2 + 3\sin t$ | | $f(t) = (1 - 2\cos t, 2 + 3\sin t)$<br>$\mathbb{R} \to \mathbb{R}^2$ |
| $S$ is described *implicitly* | $\longleftrightarrow$ | $S$ is a *level set* of a function |

I will need a little room to illustrate the implicit relationship. Suppose you have a set $S$ defined implicitly by the system of equations

$$\begin{aligned} x^2 &= \sin(yz) + 2 \\ xy + z &= e^{2z} \end{aligned}$$

Rearrange the equations so that only constants appear on one side, which constants does not matter,

$$\begin{aligned} x^2 - \sin(yz) &= 2 \\ xy + z - e^{2z} &= 0 \end{aligned}$$

The function you need is on the left and the $b$ is on the right. In other words, for $F : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $F(x, y, z) = (x^2 - \sin(yz), xy + z - e^{2z})$, the set $S$ is the level set of $F$ at $b = (2, 0)$.

Being able to describe geometric and algebraic situations in terms of functions is a very important capability. Calculus is built to work with functions, so now calculus can be used to attack algebra. This may sound a bit strange, but try to solve the equation $x + y + e^y = 0$ for $y$ in terms of $x$. Good luck. Using a result from calculus called the Implicit Function Theorem, even though you cannot solve the equations you may, at least, be able to determine if a solution exists.

## 4.4 Vector-valued functions

We have been discussing functions whose values are in $\mathbb{R}^n$, but we will also have need of functions whose values are vectors. At the beginning of the next chapter we will need vector-valued functions of one variable and toward the end vector-valued functions of two and three variables.

So, let me say we have a function that assigns points in $\mathbb{R}^n$ to vectors in space (or the plane).

$$t \to \mathbf{v}(t)$$

One useful aspect of vector-valued functions is that they have the same arithmetic structure that vectors do. The functions inherit the structure from their values. In particular, for $t \to \mathbf{v}(t)$, $t \to \mathbf{w}(t)$ and real valued $t \to c(t)$ we define the following.

$$\begin{aligned}
\mathbf{v} + \mathbf{w}: &\quad t \to \mathbf{v}(t) + \mathbf{w}(t) \\
\mathbf{v} - \mathbf{w}: &\quad t \to \mathbf{v}(t) - \mathbf{w}(t) \\
c\mathbf{v}: &\quad t \to c(t)\mathbf{v}(t) \\
\mathbf{v} \cdot \mathbf{w}: &\quad t \to \mathbf{v}(t) \cdot \mathbf{w}(t) \\
\mathbf{v} \times \mathbf{w}: &\quad t \to \mathbf{v}(t) \times \mathbf{w}(t)
\end{aligned}$$

There are many wonderful uses for these kinds of functions, but there is one that I do not like. People use them to parameterize curves and surfaces. Curves and surfaces have a location, they are made up of points, so I prefer to use point-valued functions to draw them. The other point of view is to use vector-valued functions and an artificial rule to draw things. The rule is

> The set of points described by a set of vectors are the points at the heads of the vectors when there tails are at the origin.

I admit it is fun to imagine a vector with its tail stuck at the origin frantically swinging around drawing a curve in space with its head. It is not necessary and often confusing. I think it is done to keep from having two different kinds of functions. I will use point-valued functions to draw things and vector-valued functions for things that are really vectors.

# Chapter 5

# Derivatives

Now to calculus. As usual, we begin with differential calculus.

## 5.1 Curves

A nice place to begin is to look at curves in the plane and space. To do anything quantitative you need a coordinate system and an algebraic description of the curve. We will be assuming in this section that curves are defined parametrically, so that the curve we are looking at is the range of a function.

Consider $r : \mathbb{R} \to \mathbb{R}^3$ with $r(t) = (x(t), y(t), z(t))$.

You can think of the parameter as drawing the curve in space. An even more comfortable interpretation is to think of $t$ as time and $r(t)$ as the location of a particle traveling through space at time $t$.

Calculus is the mathematics of change. For a particle moving through space, the interest could be in change of location, and the rate of change of location, which would be velocity. So, how do we describe these concepts?

Change in location is easy enough to describe with a displacement vector. In moving from $r(t)$ to $r(t + \Delta t)$, the net change in location is

$$\Delta\mathbf{r}(t) = \langle x(t + \Delta t) - x(t), y(t + \Delta t) - y(t), z(t + \Delta t) - z(t)\rangle$$

and the rate of change per unit change $t$,

$$\begin{aligned} \frac{\Delta\mathbf{r}(t)}{\Delta t} &= \frac{1}{\Delta t}\langle x(t + \Delta t) - x(t), y(t + \Delta t) - y(t), z(t + \Delta t) - z(t)\rangle \\ &= \left\langle \frac{x(t + \Delta t) - x(t)}{\Delta t}, \frac{y(t + \Delta t) - y(t)}{\Delta t}, \frac{z(t + \Delta t) - z(t)}{\Delta t} \right\rangle \end{aligned}$$

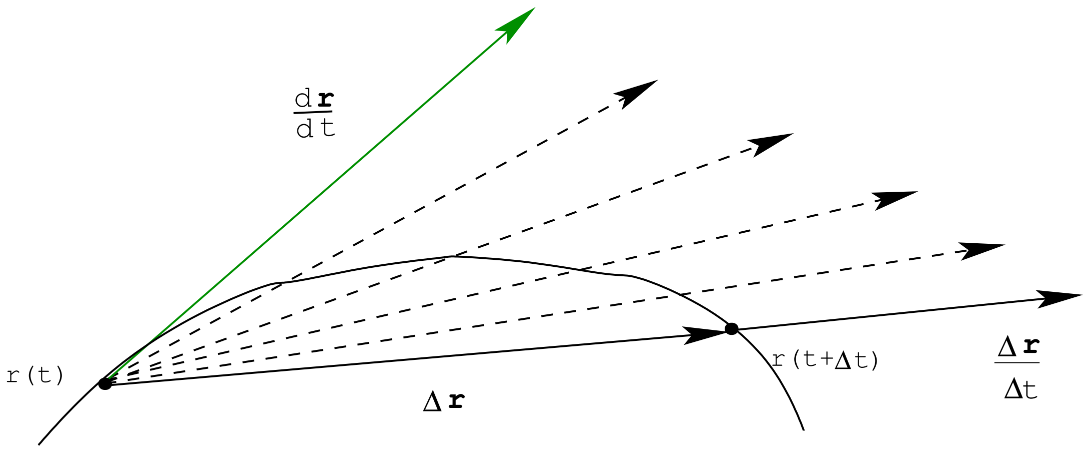
But, this is calculus. We want instant, or rather instantaneous, gratification, so we let $\Delta t \to 0$ and we have derivatives.

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}'(t) = \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right\rangle$$

One variable calculus derivatives at that. So, the obvious question is: So what?

OK, I'll take that to mean what good is the vector $\dfrac{d\mathbf{r}}{dt}$?

What direction does it point and what is its length?
For direction, a picture suffices.




The vector is *tangent* to the curve at $r(t)$, that is put the tail of $\mathbf{r}'(t)$ at $r(t)$ and it is tangent to the curve, pointing in the direction of increasing $t$, the direction of travel along the curve.

So, what is its length? It is easy enough to write down

$$\left|\frac{d\mathbf{r}}{dt}\right| = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}$$

Let's back up a minute. $|\Delta r(t)|$ is the length of the displacement vector, which approximates the change in distance along the curve $\Delta s$, $s$ denoting the distance or arclength as it is sometimes called. So, dividing by $\Delta t$ and letting it go to zero, gives the instantaneous rate of change in distance along the curve per unit increase in $t$. For our moving particle, that is just its speed!

Fantastic! The vector $\mathbf{r}'(t)$ points in the direction the curve is being drawn, that is, in the direction of increasing $t$. Its length measures how fast the curve is being drawn. For a moving particle it points in the direction the particle is going and its length is the speed of the particle, in other words, it is the particle's velocity.

## 5.2 Vector-valued functions

I introduced vector-valued functions in the previous chapter, and now you see their value. You can do arithmetic to the functions, but can do calculus as well. We just did.

For $t \to \mathbf{v}(t) = \langle x(t), y(t), z(t)\rangle$ we define

$$\mathbf{v}'(t) = \frac{d\mathbf{v}}{dt} = \langle x'(t), y'(t), z'(t)\rangle$$

In particular, for $t \to \mathbf{v}(t)$, $t \to \mathbf{w}(t)$, and real-valued $t \to c(t)$,

$$\begin{aligned}
(\mathbf{v}+\mathbf{w})' &= \mathbf{v}'+\mathbf{w}' \\
(\mathbf{v}-\mathbf{w})' &= \mathbf{v}'-\mathbf{w}' \\
(c\mathbf{v})' &= c'\mathbf{v}+c\mathbf{v}' \\
(\mathbf{v}\cdot\mathbf{w})' &= \mathbf{v}'\cdot\mathbf{w}+\mathbf{v}\cdot\mathbf{w}' \\
(\mathbf{v}\times\mathbf{w})' &= \mathbf{v}'\times\mathbf{w}+\mathbf{v}\times\mathbf{w}'
\end{aligned}$$

## 5.3 Real-valued functions

First, a quick review.

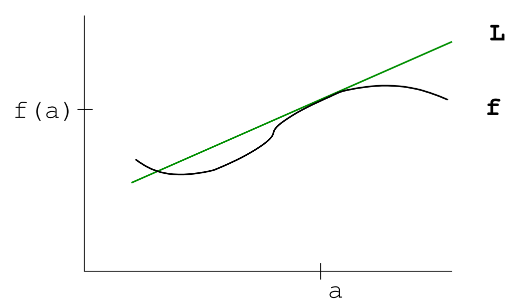
For a function $f:\mathbb{R}\to\mathbb{R}$ and $a$ in its domain, the derivative of $f$ at $a$ is

$$f'(a) = \lim_{x\to a}\frac{f(x)-f(a)}{x-a} = \left.\frac{dy}{dx}\right|_{x=a}$$

which does the following collection of wonderful things.

1. $f'(a)$ is the instantaneous rate of change $f(x) = y$ per unit increase in $x$ at $a$. This fact alone makes the derivative worthwhile. Calculus is the mathematics of change and the derivative is the tool that does the work.

2. $y = f(a) + f'(a)(x-a)$ is the equation of the tangent line to the graph of $f$ at the point $(a, f(a))$, a nice geometrical way to look at the derivative.




3. The function $L:\mathbb{R}\to\mathbb{R}$ defined by $L(x) = f(a) + f'(a)(x-a)$ is the "best" linear approximation to $f$ near $a$. The graph of $L$ is the tangent line so the two functions fit together well. They are doing much the same thing, at least near $a$.

   The function $Q:\mathbb{R}^2\to\mathbb{R}$ defined by $Q(x) = f(a)+f'(a)(x-a)+\frac{1}{2}f''(a)(x-a)^2$ is the best quadratic approximation to $f$ near $a$.

4. Algebraic properties of $f'$ and $f''$ identify qualitative properties of $f$. In particular,

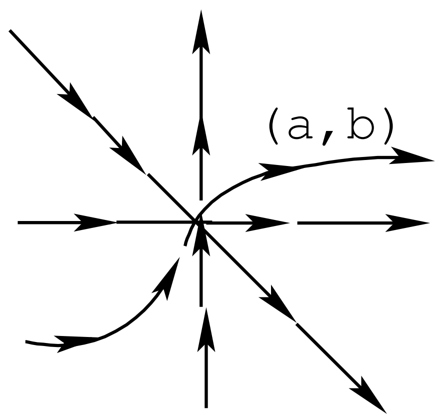
| | | |
|---|---|---|
| $f' > 0$ | $\implies$ | $f$ is increasing |
| $f' < 0$ | $\implies$ | $f$ is decreasing |
| $f'' > 0$ | $\implies$ | $f$ is concave up (convex) |
| $f'' < 0$ | $\implies$ | $f$ is concave down (concave) |
| $f'(c) = 0$ | $\impliedby$ | $f(c)$ is a local maximum or minimum |
| $f''(c) = 0$ | $\impliedby$ | $(c, f(c))$ is an inflection point on the graph |

We will begin doing these things for $f : \mathbb{R}^n \to \mathbb{R}$ for at least $n = 2$ or 3, mostly $n = 2$, since for that case two things happen. First, we will have geometry for everything, including the graph. Second, what happens in $\mathbb{R}^2$ is what happens for $n > 2$, you just have more alphabet.

### 5.3.1 Partial derivatives

Suppose you have $f : \mathbb{R}^2 \to \mathbb{R}$ with $z = f(x, y)$ and $(a, b)$ in the domain of $f$. How does $f(x, y)$ change as you pass through $(a, b)$?

That question is not so simple. What do you mean by "pass through" $(a, b)$? In $\mathbb{R}$, there is only one way to go through $a$, along the line. In $\mathbb{R}^2$ there are many different ways to pass through $(a, b)$ as the picture shows.




Maybe we better start slowly. Suppose we look at changing one variable at a time, then we define what are called the *partial derivatives* of $f$, as follows.

$$\frac{\partial f}{\partial x}(a, b) = \frac{\partial z}{\partial x}(a, b) = f_x(a, b) = \lim_{x \to a} \frac{f(x, b) - f(a, b)}{x - a}$$

is the *partial derivative of $f$ (or $z$) with respect to $x$ at* $(a, b)$ and

$$\frac{\partial f}{\partial y}(a, b) = \frac{\partial z}{\partial y}(a, b) = f_y(a, b) = \lim_{y \to b} \frac{f(a, y) - f(a, b)}{y - b}$$

is the *partial derivative of $f$ (or $z$) with respect to $y$ at* $(a, b)$.

What we have is the instantaneous rate of change $f(x, y) = z$ per unit increase in $x$ at $(a, b)$ with the first, and the instantaneous rate of change $f(x, y) = z$ per unit increase in $y$ at $(a, b)$ with the second. That is a very good start.

Calculating these derivatives is easy. Since, only one variable is changing, all of the one variable rules apply. All you need to do is to remember which variable you are using and to treat all others as constants.

For example,

$$f(x,y) = 3x^2 - x^3y^2 + 5y^3 + 2 \begin{array}{l} \Rightarrow \quad f_x(x,y) = 6x - 3x^2y^2 \\ \Rightarrow \quad f_y(x,y) = -2x^3y + 15y^2 \end{array}$$

We can also define higher order derivatives

$$\begin{aligned} f_{xx} &= \tfrac{\partial^2 z}{\partial x^2} = \tfrac{\partial^2 f}{\partial x^2} \\ f_{xy} &= \tfrac{\partial^2 z}{\partial y \partial x} = \tfrac{\partial^2 f}{\partial y \partial x} \\ f_{yx} &= \tfrac{\partial^2 z}{\partial x \partial y} = \tfrac{\partial^2 f}{\partial x \partial y} \\ f_{yy} &= \tfrac{\partial^2 z}{\partial y^2} = \tfrac{\partial^2 f}{\partial y^2} \end{aligned}$$

For the example, $f_{xx} = 6 - 6xy^2$, $f_{xy} = -6x^2y$, $f_{yx} = -6x^2y$, and $f_{yy} = -2x^3 + 30y$. Note that $f_{xy} = f_{yx}$. If the function is nice enough, that will always happen. Our functions will be nice enough.

For the record, for $f : \mathbb{R}^n \to \mathbb{R}$ with $y = f(x_1, \ldots, x_n)$ we have

$$f_{x_i} = \frac{\partial y}{\partial x_i} = \frac{\partial f}{\partial x_i}$$

for $i = 1, \ldots n$

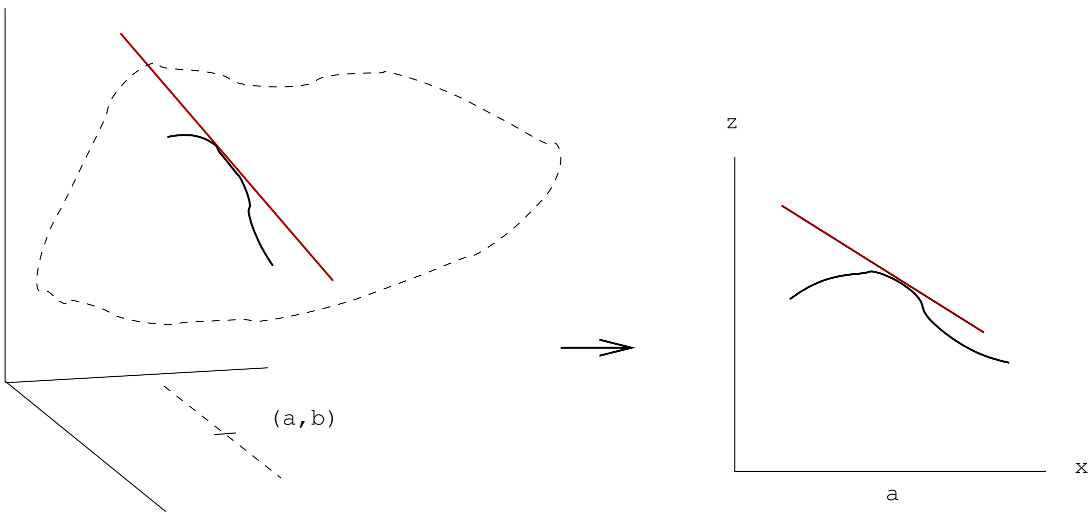
Of course, you can keep going to derivatives of any order.

### 5.3.2 Tangent plane

The equation of the tangent plane to the graph of $f$ at $(a, b, f(a, b))$ is

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$$

The way to see this is to look at one variable at a time. For example, $x$ is the variable and $y = b$, then slice $\mathbb{R}^3$ with a vertical plane through $(a, b)$ parallel to the $x$-axis.




The slice is just a one variable problem, for which we know the answer, the equation of the tangent line would be $z = f(a,b) + f_x(a,b)(x-a)$, where $b$ just goes along for the ride. But, this is what you get when you put $y = b$ into $z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$, the equivalent of slicing the thing I have called the tangent plane with the vertical plane. So, this plane contains a line that is tangent to the graph. Similar reasoning gives a line tangent to the graph determined by letting $y$ vary and $x = a$. These two lines determine the tangent plane and its algebraic description.

### 5.3.3 Approximation

If we let $L : \mathbb{R}^2 \to \mathbb{R}$ be defined by $L(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$, then the graph of $L$ is precisely the tangent plane to the graph of $f$. In other words, the two functions fit together well for $(x,y)$ near $(a,b)$. How near and how well depends on $f$, but $L$ is about as good as it gets with a linear function.

We could calculate a few approximations for the record, but that is not what is really going on here. The idea is that knowing $L$ and how it acts, can give information about $f$ and $L$ is easy to work with. Unfortunately, some of the best examples are beyond the scope of these notes. If you look at the list at the beginning of the chapter of algebraic properties of derivatives that give qualitative information about $f$, you can restate some of them in terms $L$. For example, $f' > 0 \Rightarrow f$ is increasing, could be said $L$ is increasing $\Rightarrow f$ is increasing near $a$.

You can improve the approximation by adding more terms to it. $L$ is a polynomial of degree one. It would seem that using a polynomial of degree two to approximate $f$ would be an improvement without too much extra effort. The one to use is

$$\begin{aligned} Q(x,y) &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &\quad + \frac{1}{2}(f_{xx}(a,b)(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)(y-b)^2) \end{aligned}$$

$Q$ is not as bad as it looks. The first line is just $L$, and the rest are the second degree terms whose coefficients are the second derivatives of $f$ evaluated at $(a, b)$. The real question is, why should this be a good approximation. At this point in your career, the best I can do for you is to point out $Q$ and $f$ have the same value, and the same first and second derivatives at $(a, b)$. There is more, but you will have to wait.

So, what good is $Q$? That is such an interesting question I am going to dwell on it.

First, we need to get a feel for second degree polynomials in two variables, suppose we have

$$P(x, y) = Ax^2 + 2Bxy + Cy^2$$

The lower order terms would just be in the way, and the 2 will be convenient later.

In fact, let me simplify further by assuming, for the moment, that $B = 0$, so that $z = P(x, y) = Ax^2 + Cy^2$.

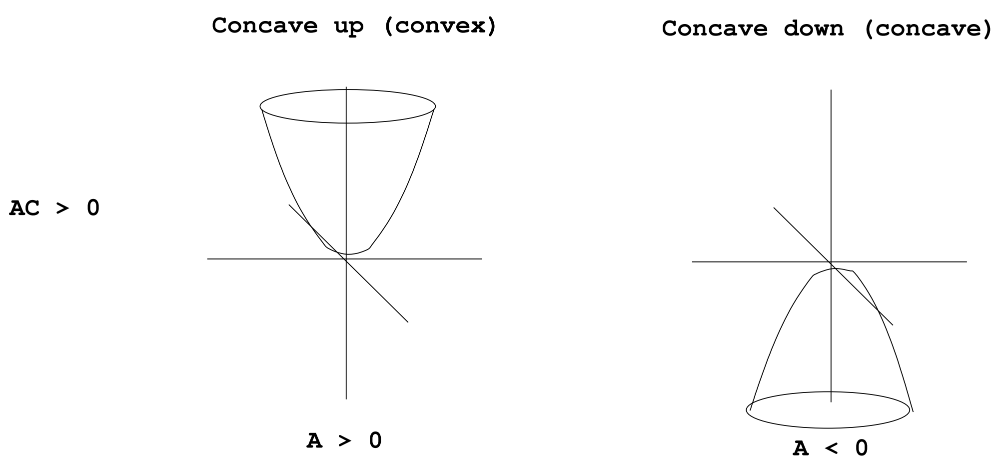
The level sets of $P$ should be familiar to you. One is enough to tell the story.

$$Ax^2 + Cy^2 = k$$

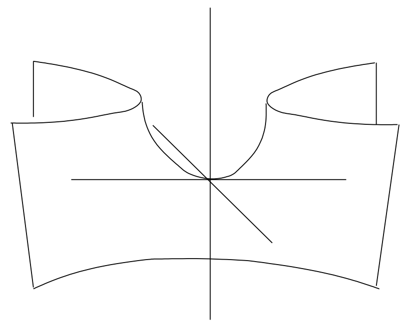
is one of three things, depending on what $A$ and $C$ are.

- If $A$ and $C$ have the same sign, the level sets are *ellipses*, the condition on $A$ and $C$ can be summarized by $AC > 0$.
- If $A$ and $C$ have opposite signs, that is $AC < 0$, then the level sets are *hyperbolas*.
- If $A$ or $C$ is zero, that is $AC = 0$, the level sets are pairs of *parallel lines*.

From this we can build the graphs,

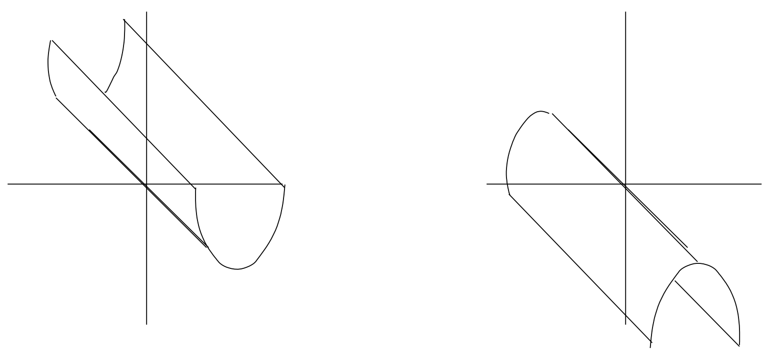


AC < 0



Saddle point

AC = 0



It is important to note that you can tell which of the three you have just by looking at $AC$.

Now to $B \neq 0$, if both $A$ and $C$ are zero, we have $z = 2Bxy$, the level sets are $y = k/x$ for some constant k, which are also hyperbolas.

Now, for the general case $Ax^2 + 2Bxy + Cy^2$. First, a little maneuvering. Let me assume that $A \neq 0$, if it is, then do what I am about to do with $C$. If both are, we just took care of that.

$$\begin{aligned} Ax^2 + 2Bxy + Cy^2 &= A(x^2 + 2\frac{B}{A}xy + \frac{B^2}{A^2}y^2 - \frac{B^2}{A^2}y^2) + Cy^2 \\ &= A(x + \frac{B}{A}y)^2 + (C - \frac{B^2}{A})y^2 \\ &= Au^2 + (C - \frac{B^2}{A})y^2 \end{aligned}$$

where $u = x + \frac{B}{A}y$. The new variable $u$ is not really necessary, but it makes a little clearer that the general case is just like the cases where $B = 0$. In other words, the graphs of quadratic polynomials look like one of the three pictured above, and you can tell which from

- $AC - B^2 > 0$, strictly concave up for $A > 0$ and strictly concave down for $A < 0$
- $AC - B^2 < 0$, a saddle point on the graph
- $AC - B^2 = 0$, a trough, concave up or down, but not strictly.

$AC - B^2$ is called the discriminant, and I will denote it by $D$.

The discriminant for the best quadratic approximation is

$$D = f_{xx}(a,b)f_{yy}(a,b) - (f_{xy}(a,b))^2$$

Since $Q$, fits the function well, the graph of $f$ will look approximately like one of these three near $(a,b,f(a,b))$ and you can tell which one by looking at $D$ and $f_{xx}(a,b)$.

There is one catch. If $D = 0$, you could have, for example, a cubic polynomial which would not be a trough looking object and would not fit into the scheme of things. So, if $D = 0$, you do not learn anything much about $f$. But, this is not new to you, $D$ plays the role that $f''$ plays in one variable calculus. $f''(c) = 0$ gives no information about what happens to $f$ at $c$. In any case, we have qualitative properties of $f$ identified from algebraic statements about derivatives, namely

- $D > 0$, $f$ is concave up if $f_{xx}(a,b) > 0$ and concave down if $f_{xx}(a,b) < 0$, or if $Q$ is strictly concave up (down), then $f$ is strictly concave up (down) near $(a,b)$.
- $D < 0$, we say $(a,b,f(a,b))$ is a *saddle point* on the graph, the analog to an inflection point in one variable, or if $Q$ has a saddle point at $(a,b,f(a,b))$, then $f$ has a saddle point at $(a,b,f(a,b))$.

### 5.3.4 Chain rule

I said that you calculate partial derivatives using one variable calculus rules, but that is not quite correct. There is one rule that is slightly more complicated in a multivariable setting than for one variable, and that is the chain rule.

The one variable chain rule deals with two functions or three variables, acting one after the other.

$$x \xrightarrow{g} y \xrightarrow{f} z$$

In terms of functions the action of $g$ is followed by the action of $f$. In terms of variables, $z$ is determined by $y$, which is determined by $x$, therefore $z$ is a function of $x$.

The chain rule describes how the derivative behaves in this situation.

The chain rule in terms of functions is

$$(f(g(x)))' = f'(g(x))g'(x)$$

in terms of variables

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$$

The general multivariable analog is

$$(x_1, \ldots, x_n) \xrightarrow{g} (y_1, \ldots, y_m) \xrightarrow{f} (z_1, \ldots, z_p)$$

I am not going to justify the chain rule, but what you do is natural, you will not be too uncomfortable using it. So, here it is

$$\frac{\partial z_k}{\partial x_i} = \frac{\partial z_k}{\partial y_1}\frac{\partial y_1}{\partial x_i} + \frac{\partial z_k}{\partial y_2}\frac{\partial y_2}{\partial x_i} + \ldots + \frac{\partial z_k}{\partial y_m}\frac{\partial y_m}{\partial x_i}$$

For example, $z = f(x, y)$ and

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

So,

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

Another example: $t \to (x, y) \to z = f(x, y)$, then we have

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

or

$$\frac{dz}{dt} = f_x\frac{dx}{dt} + f_y\frac{dy}{dt}$$

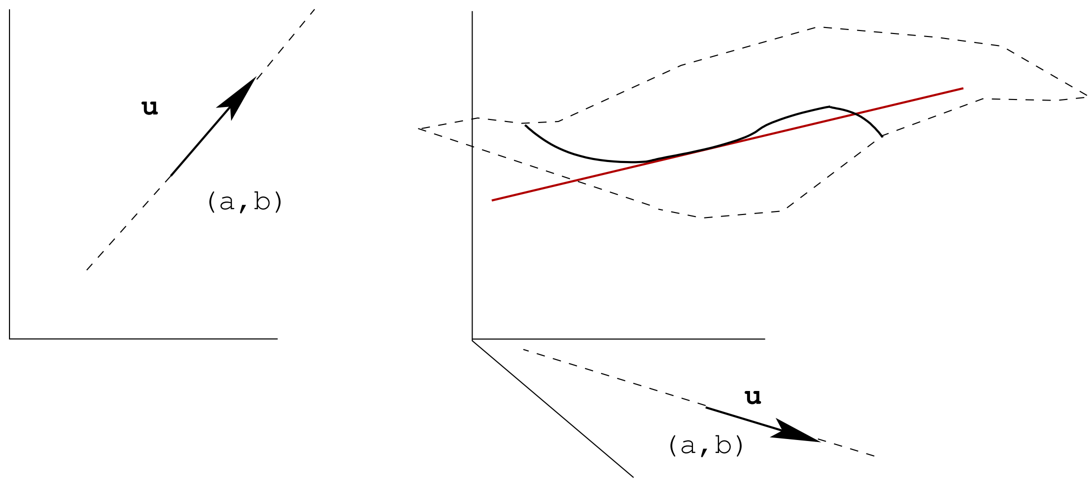
### 5.3.5 Directional derivative and gradient

The partial derivatives do measure change but only in one variable. You might want to see what happens as you pass through $(a, b)$ in a direction other than parallel to one of the axes.

All you need to do is specify the direction, which you do, as you might guess, by specifying a vector $\mathbf{u} = \langle u_1, u_2 \rangle$ and using it to parameterize a line through $(a, b)$ in the direction of $\mathbf{u}$ by

$$\begin{aligned} x &= a + u_1 t \\ y &= b + u_2 t \end{aligned}$$

We use a unit vector, that is $|\mathbf{u}| = 1$, so that the parameter measures distance along the line.




For $f : \mathbb{R}^2 \to \mathbb{R}$ and $(a, b)$ in the domain of $f$, the *directional derivative* of $f$ in the direction $\mathbf{u}$ at $(a, b)$ is

$$\frac{\partial f}{\partial \mathbf{u}}(a, b) = \lim_{t \to 0} \frac{f(a + u_1 t, b + u_2 t) - f(a, b)}{t}$$

What we have here is the instantaneous rate of change of $f(x, y)$ per unit increase in distance in the direction $\mathbf{u}$ at $(a, b)$. All we are really doing is taking the derivative of $f(a + u_1 t, b + u_2 t)$. Using the chain rule, we can calculate it easily.

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{u}} &= f_x \frac{dx}{dt} + f_y \frac{dy}{dt} \\ &= f_x u_1 + f_y u_2 \end{aligned}$$

That last equation looks like a dot product of two vectors, and, in fact, we make it so.

The *gradient* of $f$ is the vector $\langle f_x(x, y), f_y(x, y) \rangle$ denoted $\text{grad}(f)(x, y)$ or $\nabla f(x, y)$ Then,

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{u}} &= \nabla f \cdot \mathbf{u} \\ &= |\nabla f| \cos \theta \end{aligned}$$

where $\theta$ is the angle between the gradient of $f$ and the direction $\mathbf{u}$. This angle gives us some useful information about what the gradient does. In particular,

- $(\theta = 0)$: The gradient points in the direction to go to experience the greatest increase and the rate of increase is $|\nabla f|$.

- ($\theta = \pi$): $-\nabla f$ points in the direction of greatest decrease and rate of decrease is $-|\nabla f|$.
- ($\theta = \frac{\pi}{2}$) : $\nabla f(a,b)$ is perpendicular to the level set of $f$ through $(a,b)$ at $(a,b)$ as explained below.

To see that the gradient is perpendicular to the level set, build $r(t) = (x(t), y(t))$ a parameterization of the level set of $f$ through $(a,b)$. So, $f(x(t), y(t)) = f(a,b)$. Taking the derivative with respect to $t$,

$$\begin{aligned} f_x \frac{dx}{dt} + f_y \frac{dy}{dt} &= 0 \\ \nabla f \cdot \frac{d\mathbf{r}}{dt} &= 0 \end{aligned}$$

$r(t)$ draws a level set of $f$, and therefore $\frac{d\mathbf{r}}{dt}$ is the tangent vector to the level set. The last equation above shows that the gradient at a point is perpendicular to the tangent vector to the level set through the point, which is what one means by perpendicular to the level set at the point.

### 5.3.6 More tangent

For a function $f : \mathbb{R}^2 \to \mathbb{R}$ with $z = f(x,y)$, we already have the equation of the tangent plane to the graph at a point $(a, b, f(a,b))$

$$z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) = L(x,y)$$

The function $L$ is the best linear approximation and one can say that graph of $L$ is the tangent plane to the graph of $f$. But, there is more.

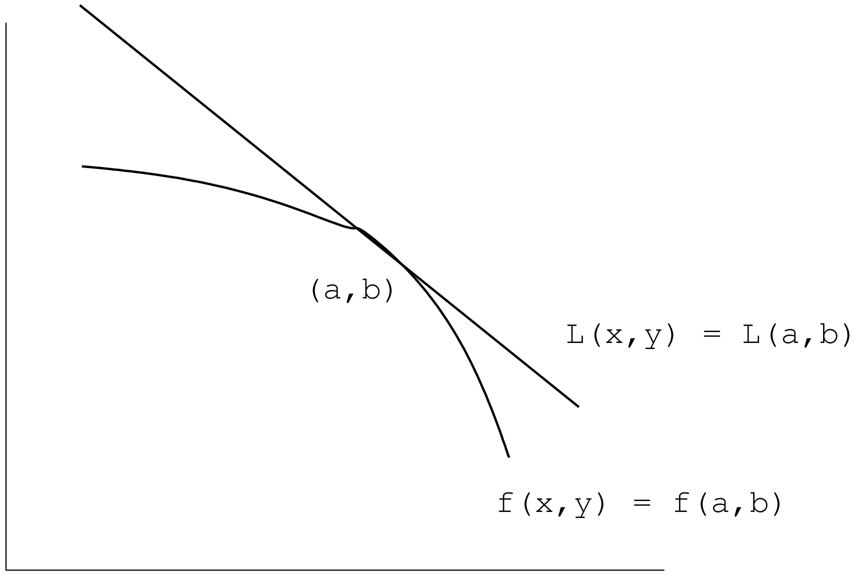
The level set of $L$ through $(a,b)$ intersects the level set of $f$ through $(a,b)$ at $(a,b)$, since $L(a,b) = f(a,b)$. For $(x,y)$ in the level set of $L$ through $(a,b)$,

$$L(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) = L(a,b) = f(a,b)$$

or

$$f_x(a,b)(x-a) + f_y(a,b)(y-b) = \nabla f(a,b) \cdot \langle x-a, y-b \rangle = 0$$

which says that the level set of $L$ through $(a,b)$ is perpendicular to the gradient, hence tangent to the the level set of $f$ at $(a,b)$.




For a function $r : \mathbb{R} \to \mathbb{R}^2$ given by $r(t) = (x(t), y(t))$ and $t_0$ in the domain of $r$, $\mathbf{r}'(t_0) = \langle x'(t_0), y'(t_0)\rangle$ is the tangent vector to the range of $r$ at $r(t_0)$. Therefore, the line through $r(t_0)$ parallel to $\mathbf{r}'(t_0)$ is tangent to the range of $r$ at $r(t_0)$. The line has parametric description

$$x = x(t_0) + x'(t_0)(t - t_0)$$
$$y = y(t_0) + y'(t_0)(t - t_0)$$

Let $L : \mathbb{R} \to \mathbb{R}^2$ be defined by $L(t) = (x(t_0) + x'(t_0)(t - t_0), y(t_0) + y'(t_0)(t - t_0))$, the the range of $L$ is the tangent line to range of $r$. $L$ is the best linear approximation to $r$ near $t_0$.

The bottom line is that any function $f : \mathbb{R}^n \to \mathbb{R}^m$, with reasonably well-behaved partial derivatives, has a best linear approximation at each point in its domain. The graph, the range and level set of the linear approximation through the point are the tangent planes to the corresponding sets for $f$.

## 5.4 Optimization

Local optimization refers to finding local maxima and minima.

In one-variable calculus, you have seen the following.

- If $f(c)$ is a local optimum, then $f'(c) = 0$ or does not exist.
- If $f'(c) = 0$, then if $f''(c) > 0$, then $f(c)$ is a local minimum,
  if $f''(c) < 0$, then $f(c)$ is a local maximum.

The situation for $f : \mathbb{R}^2 \to \mathbb{R}$ is essentially the same. Look back at the graphs of the quadratic polynomials. The discriminant $D = f_{xx}f_{yy} - (f_{xy})^2$ plays the role of the second derivative. The key observation is that at a local optimum, the tangent plane is horizontal, which happens if $f_x(a,b) = f_y(a,b) = 0$ or $\nabla f(a,b) = \mathbf{0}$.

- If $f(a,b)$ is a local optimum, then $\nabla f(a,b) = \mathbf{0}$ or does not exist.
- If $\nabla f(a,b) = \mathbf{0}$, then

$$\begin{array}{lll} \text{if } D > 0 & f_{xx}(a,b) > 0 & \rightarrow \quad f(a,b) \text{ is a local minimum.} \\ & f_{xx}(a,b) < 0 & \rightarrow \quad f(a,b) \text{ is a local maximum.} \\ \text{if } D < 0 & f(a,b) \text{ is not a local optimum.} & \end{array}$$

The discriminant is useful for two variable calculus, but for more variables linear algebra is need. I can tell you this. If the best quadratic approximation near $a$ has a strict minimum at $a$ ($Q(x) > Q(a), x \neq a$), then $f$ has a strict local minimum at $a$ and so on.

## 5.5 Infinitesimals - a way to think

If you were moving along at a constant speed, you could calculate the speed by measuring the time, $\Delta t$, and the distance, $\Delta s$, traveled in that time, then your speed would be

$$\text{speed } = \frac{\Delta s}{\Delta t}$$

In the early days of calculus people thought of the derivative $\frac{dy}{dx}$ in exactly the same way, as a ratio of elapsed distance and elapsed time. The difference was that the distance and times were *infinitesimal*, something smaller than any number, but not zero. They used $dx$ as an entity in its own right. It was infinitesimally small. It could be interpreted as the width of a point. They would say that as you travel an infinitesimal distance $dx$ in an instant of time $dt$ your speed is $\frac{dx}{dt}$, the ratio of the infinitesimals.

There were rules for calculation with these infinitesimal things that were not entirely justified, but seemed to work. Such as, the product rule,

$$d(xy) = ydx + xdy$$

They were called *differentials* because they measure an amount of change, but an infinitesimal amount as opposed to a real change which was called a *difference*.

Differences had their rules, too. The product rule is

$$\Delta(xy) = y\Delta x + x\Delta y + \Delta x\Delta y$$

The difference includes the so called "higher order terms". In the example, there is one higher order term $\Delta x\Delta y$. Order counts the number of changes appearing in a given term. For example, $x\Delta y$ is a first order term and $\Delta x\Delta y$ is a second order term.

The product for differentials would be derived as follows. $x \to x+dx$ and $y \to y+dy$, so that the change in the product $xy$ is

$$\begin{aligned} d(xy) &= (x+dx)(y+dy) - xy \\ &= xy + xdy + ydx + dxdy - xy \\ &= xdy + ydx + dxdy \end{aligned}$$

then you "throw away the higher order terms" to get the product rule for differentials.

$$d(xy) = ydx + xdy$$

It works, but why it works wasn't clear to many people who used it.

It was in the middle of the eighteenth century that the "correct" way was developed. It involved using real differences

$$\Delta(xy) = y\Delta x + x\Delta y + \Delta x\Delta y$$

then dividing through by a difference and letting it go to zero.

$$\begin{aligned} \frac{\Delta(xy)}{\Delta t} &= \frac{y\Delta x}{\Delta t} + \frac{x\Delta y}{\Delta t} + \frac{\Delta x\Delta y}{\Delta t} \\ \frac{\Delta(xy)}{\Delta t} &= y\frac{\Delta x}{\Delta t} + x\frac{\Delta y}{\Delta t} + \Delta y\frac{\Delta x}{\Delta t} \end{aligned}$$

Letting $\Delta t$ go to zero, one obtains the product rule for derivatives.

$$\frac{d(xy)}{dt} = y\frac{dx}{dt} + x\frac{dy}{dt}$$

Dividing through by a difference, gives ratios of differences and just differences. The limit process, letting a difference go to zero, finishes the job. The ratios go to derivatives. The other differences just go away and take anything they multiply with them, which is why "throwing away higher order terms" works.

Why am I mentioning this? Well, it is intuitive to think in terms of infinitesimals.

- $dt$ is an instant of time
- $dx$ is the width of a point along an axis
- $ds = \sqrt{dx^2 + dy^2 + dz^2}$ is the length of a point along a curve
- $dA = dxdy$ is the area of a point in the plane
- $dV = dxdydz$ is the volume of a point in space

We will have opportunity to think in terms of differentials quite a bit, but that will come later when we deal with integration.

## 5.6 Curves and surfaces

### 5.6.1 Geometry of curves

I have mentioned that calculus can be used to solve algebra problems, but it can also be used to study geometry. Let me give you a hint by looking at curves, parameterized curves.

Suppose, $r : \mathbb{R} \to \mathbb{R}^3$ with $r(t) = (x(t), y(t), z(t))$

I want to decompose the tangent vector $\mathbf{r}'$ into two pieces, one that that describes direction and the other change in distance, by letting

$$\mathbf{T} = \frac{1}{|\mathbf{r}'|}\mathbf{r}' \qquad \text{and} \qquad \frac{ds}{dt} = |\mathbf{r}'|$$

then

$$\mathbf{r}' = \frac{ds}{dt}\mathbf{T}$$

The vector $\mathbf{T}$ has length one, that is it contains only directional information. This fact makes it a geometrical object in that it is independent of the parameterization, but depends only on the geometry of the curve. Well, OK, if you reverse the parameterization it goes along the curve in the opposite direction but still tells you about the geometry without any other effect of the parameter.

The derivative of $\mathbf{T}$ with respect to arclength tells you how the direction of the curve is changing per unit distance along the curve, in other words, how the curve is curving. Break it into pieces as I did above

$$\mathbf{N} = \frac{1}{|\frac{d\mathbf{T}}{ds}|}\frac{d\mathbf{T}}{ds} \qquad \text{and} \qquad \kappa = \left|\frac{d\mathbf{T}}{ds}\right|$$

then

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}$$

The vector $\mathbf{N}$ is a unit vector. Differentiating $|\mathbf{T}|^2 = \mathbf{T}\cdot\mathbf{T} = 1$ to obtain

$$2\mathbf{T}\cdot\frac{d\mathbf{T}}{ds} = 2\kappa\mathbf{T}\cdot\mathbf{N} = 0$$

shows that $\mathbf{N}$ is perpendicular to $\mathbf{T}$. In fact, $\mathbf{N}$ points in the direction the curve is curving.

And, $\kappa$ measures the rate at which the curve is curving and is called the *curvature*.

An example will give you an even better feel for it.

Parameterize a circle with radius $a$, with $r(t) = (a\cos t, a\sin t)$, then

$$\mathbf{r}'(t) = \langle -a\sin t, a\cos t\rangle \Rightarrow \mathbf{T}(t) = \langle -\sin t, \cos t\rangle \text{ and } \frac{ds}{dt} = a$$

We have, then

$$\frac{d\mathbf{T}}{ds} = \frac{d\mathbf{T}}{dt}\frac{dt}{ds} = \frac{1}{a}\langle -\cos t, -\sin t\rangle$$

This equation says that the curvature of a circle $\kappa = \frac{1}{a}$ is the reciprocal of the radius, the larger the circle, the more slowly it curves as you walk around it. The direction the circle is curving $\mathbf{N} = \langle -\cos t, -\sin t\rangle$, is toward the center of the circle.

I could go on and talk about things like torsion, how the curve is twisting, and that curves are are essentially determined, if you know their curvature, torsion, starting point and initial direction. But, I won't, just an inkling is all you get.

I do think a look at physics may be interesting. If $r(t)$ describes the location of a particle at time $t$, then the velocity of the particle is

$$\mathbf{v}(t) = \mathbf{r}'(t) = v\mathbf{T} \text{ where } v = s' \text{ the speed}$$

and the acceleration is

$$\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{r}''(t) = v'\mathbf{T} + v\mathbf{T}' = a\mathbf{T} + v\mathbf{T}'$$

where $a$ is the scalar acceleration along the curve, the derivative of the speed with respect to time.

Now,

$$\frac{d\mathbf{T}}{dt} = \frac{ds}{dt}\frac{d\mathbf{T}}{ds} = v\frac{d\mathbf{T}}{ds} = \kappa v\mathbf{N}$$

Now we can write the acceleration in a rather interesting way,

$$\mathbf{a} = a\mathbf{T} + \kappa v^2\mathbf{N}$$

The first component is tangent to the curve and its length is the linear acceleration along the curve.

The second component is perpendicular to the curve pointing in the direction the particle is turning. Recall that the curvature of a circle is one over the radius and the meaning of $\kappa v^2$ become clear. It is the centripetal acceleration.

So, an application of calculus to geometry to physics.

### 5.6.2 Surfaces and tangent planes

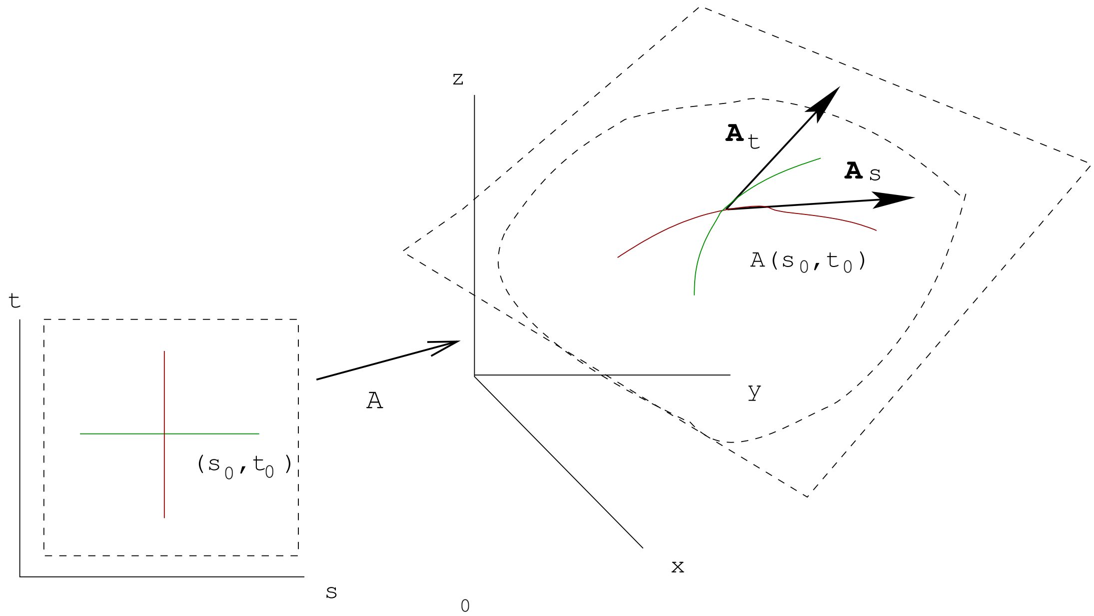
We can now add a little useful structure to surfaces.

Suppose $A : \mathbb{R}^2 \to \mathbb{R}^3$ parameterizes a surface with $A(s,t) = (x(s,t), y(s,t), z(s,t))$ and $(s_0, t_0)$ is a point in the domain of $A$.

I want to draw two curves on the surface through $A(s_0, t_0)$, namely, $s \to A(s, t_0)$ and $t \to A(s_0, t)$. These intersect at $A(s_0, t_0)$ and their tangent vectors are given by the usual partial derivatives,

$$\begin{aligned}\mathbf{A}_s &= \langle x_s, y_s, z_s\rangle \\ \mathbf{A}_t &= \langle x_t, y_t, z_t\rangle\end{aligned}$$

evaluated at $(s_0, t_0)$. Since the vectors are tangent to the curves they are tangent to the surface and together with $A(s_0, t_0)$ determine a plane, the plane through $A(s_0, t_0)$ parallel to $\mathbf{A}_s(s_0, t_0)$ and $\mathbf{A}_t(s_0, t_0)$, which more importantly, it is more than reasonable to call the *tangent plane* to the surface at $A(s_0, t_0)$.



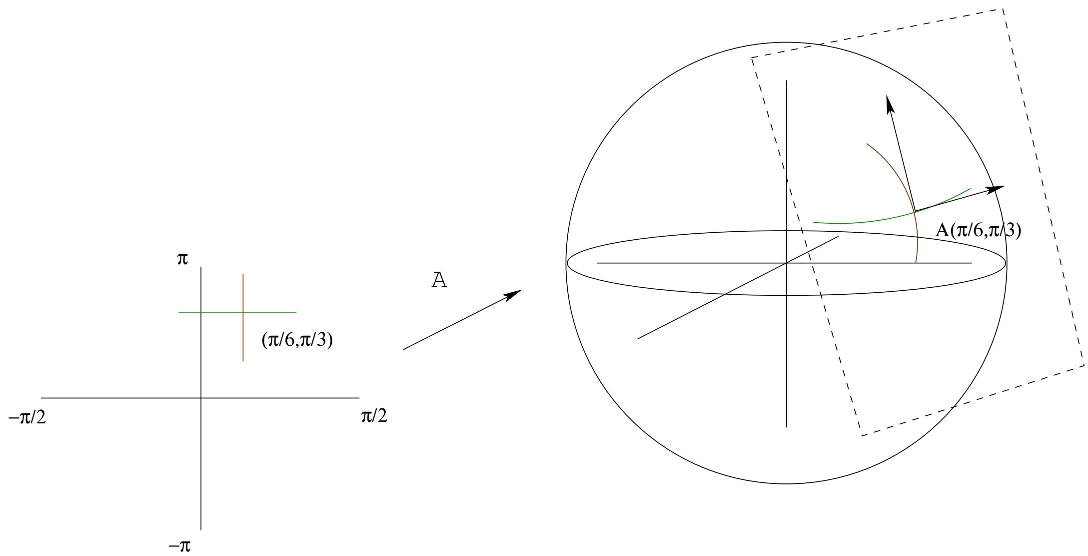

The vector $\mathbf{A}_s(s_0, t_0) \times \mathbf{A}_s(s_0, t_0)$ is normal to the surface at $A(s_0, t_0)$ and could be used to describe the tangent plane implicitly. We will see more interesting uses for it later.

For example, the unit sphere parameterized by $A(\phi, \theta) = (\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi)$ for which

$$\begin{aligned}
\mathbf{A}_\phi &= \langle -\sin\phi\cos\theta, -\sin\phi\sin\theta, \cos\phi\rangle \\
\mathbf{A}_\theta &= \langle -\cos\phi\sin\theta, \cos\phi\cos\theta, 0\rangle \\
\mathbf{A}_\phi \times \mathbf{A}_\theta &= -\cos\phi\langle \cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi\rangle
\end{aligned}$$

Look at the point $A(\pi/6, \pi/3) = (\sqrt{3}/4, 3/4, 1/2)$, then,

$$\begin{aligned}
\mathbf{A}_\phi &= \left\langle -\frac{1}{4}, -\frac{\sqrt{3}}{4}, \frac{\sqrt{3}}{2}\right\rangle \\
\mathbf{A}_\theta &= \left\langle -\frac{3}{4}, \frac{\sqrt{3}}{4}, 0\right\rangle \\
\mathbf{A}_\phi \times \mathbf{A}_\theta &= \left\langle -\frac{3}{8}, -\frac{3\sqrt{3}}{8}, -\frac{\sqrt{3}}{4}\right\rangle = -\frac{\sqrt{3}}{2}\left\langle \frac{\sqrt{3}}{4}, \frac{3}{4}, \frac{1}{2}\right\rangle
\end{aligned}$$




The tangent plane is described parametrically by

$$
\begin{aligned}
x &= \frac{\sqrt{3}}{4} - \frac{1}{4}s - \frac{3}{4}t \\
y &= \frac{3}{4} - \frac{\sqrt{3}}{4}s + \frac{\sqrt{3}}{4}t \\
z &= \frac{1}{2} + \frac{\sqrt{3}}{2}s
\end{aligned}
$$

implicitly by

$$\frac{3}{8}\left(x - \frac{\sqrt{3}}{4}\right) + \frac{3\sqrt{3}}{8}\left(y - \frac{3}{4}\right)\frac{\sqrt{3}}{4}\left(z - \frac{1}{2}\right) = 0$$

and explicitly by

$$z = \frac{1}{2} - \frac{\sqrt{3}}{2}\left(x - \frac{\sqrt{3}}{4}\right) - \frac{3}{2}\left(y - \frac{3}{4}\right)$$

The structure we have described can be used to study the geometry of surfaces in much the same way we used calculus to study the geometry of curves.

For example, the curvature of a curve was measured by noting how the unit tangent vector changed as it moved along the curve. The curvature of a surface can be measured by looking at how the unit normal vector changes as it moves about the surface, which can be calculated in terms of directional derivatives of the normal vector.

One difference between curves and surfaces is that at a single point on a surface you can go in infinitely many directions with the normal vector, so that there are infinitely many curvatures at a point. There are several that are more important than others. The principal curvatures are the maximum and minimum curvatures you would see by looking at directional derivatives of the normal vector at the point. These two curvatures probably make the most sense, but the most famous is the so-called Gaussian curvature, which is the product of the principal curvatures. Why it is the most famous I'll leave you to discover. It is worth the safari. The second most famous, and possibly the most famous for non-mathematicians, is the Riemann curvature. It is this curvature that Einstein used to describe the effect that mass and energy have on the geometry of the universe, that we perceive as gravity - the theory of general relativity.

We can look at a couple of examples. The calculations involved are usually quite complicated, so I will leave them out.

First look at a surface that is the graph of a function $z = f(x, y)$, which will be parameterized in the usual way with $x = x$, $y = y$, and $= f(x, y)$. The normal vector is $\langle -f_x, -f_y, 1\rangle$ divided by its length, $(1 + f_x^2 + f_y^2)^{\frac{1}{2}}$. After a nontrivial calculation you would find that the Gaussian curvature, $K$, is given by

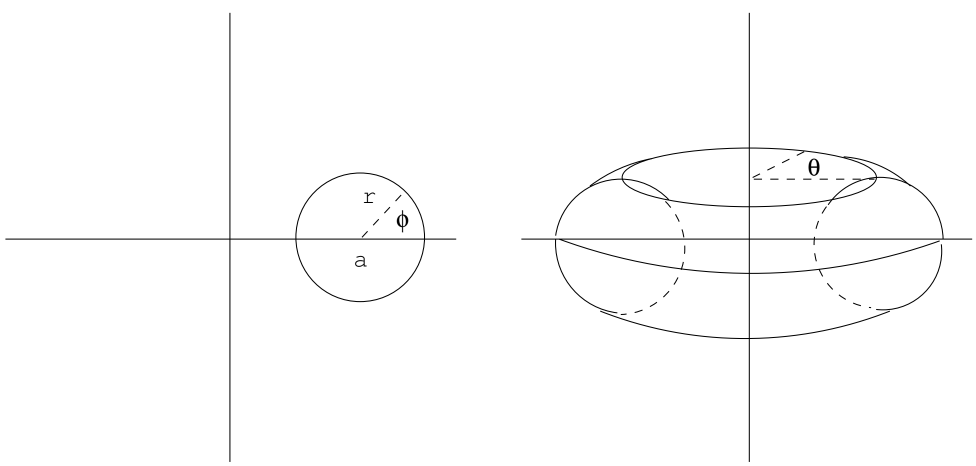
$$K = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2}$$

The numerator is nothing more than our old friend the discriminant, which as we have already seen, measures the shape of the graph near a point.

Another interesting example is the torus, which is obtained by revolving a circle around an axis outside the circle to obtain a donut shape. For example, for $r < a$, the circle in the $x$-$z$ plane given by $x = a + r\cos\phi, z = r\sin\phi$ for $-\pi \le \phi \le \pi$ is centered at $(a, 0)$ and has radius $r$. When revolved around the $z$ axis the surface generated is parameterized by

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

for $-\pi \le \phi \le \pi$ and $-\pi \le \theta \le \pi$.




The principal and Gaussian curvatures turn out to be

$$\kappa_1 = \frac{1}{r}$$

$$
\begin{aligned}
\kappa_2 &= \frac{\cos\phi}{a + r\cos\phi} \\
K &= \frac{\cos\phi}{r(a + r\cos\phi)}
\end{aligned}
$$

Notice that for $-\frac{\pi}{2} < \phi < \frac{\pi}{2}$, on the outside of the torus, $\kappa_2$ and $K$ are positive. For $\phi = \pm\frac{\pi}{2}$, the circles around the top and bottom of the torus, $\kappa_2 = K = 0$. On the inside of the torus, $\kappa_2$ and $K$ are negative.

We also have, $\kappa_1$ is constant and is, in fact, the curvature of the circle we revolved to get the torus. $\kappa_2$ is almost the curvature of a circle. If you take a single point on the generating circle, that is fix $\phi$, and revolve the point around the $z$ axis the circle you get has radius $a + r\cos\phi$, so that $\kappa_2$ is the curvature of this circle multiplied by $\cos\phi$. Where the $\cos\phi$ comes from is a long, but interesting story, involving things like geodesics, curves in the surface that do not appear to curve in the surface. We are already pretty far afield, so I will stop here.

# Chapter 6

# Integrals

For $f : \mathbb{R} \to \mathbb{R}$ and real numbers $a$ and $b$, you can talk about the *integral* of $f$ from $a$ to $b$,

$$\int_a^b f(x)\,dx$$

and what do you say?

What does the integral do?

- Everybody's favorite use of the integral is that it computes the area under the graph of $f$ from $a$ to $b$. A very comforting, easy to grasp idea.

  Unfortunately, that is not exactly what the integral does. It actually computes the area under the graph and above the $x$-axis minus the area above the graph and below the $x$-axis unless $a > b$ in which case it is minus all that. Well, there goes the comfort part. The reason for this mess is that the integral was not built to compute area, it was built to do the next thing on the list.

- If $f$ is continuous on an interval $I$ and $a$ is in $I$, then $F : \mathbb{R} \to \mathbb{R}$ defined by

  $$F(x) = \int_a^x f(u)\,du$$

  is an *antiderivative* of $f$ on $I$. That is, $F' = f$, the integral undoes differentiation.

Antiderivatives are what the integral is all about and the way most people actually use the integral, but they don't necessarily realize it. They have a quantity represented by $x$ and another quantity by $y$ that is instantaneously changing with $x$. They know the rate of change and want to know $y$.

They have even been known to talk this way. For an infinitesimal amount of $x$, $dx$, an infinitesimal amount of $y$ is given by

$$dy = \frac{dy}{dx}\,dx$$

so that, to get $y$ all you need to do is add up the infinitesimal pieces by integration, or

$$y = \int dy = \int \frac{dy}{dx}\, dx$$

For example, suppose you have a reasonably uniform piece of wire and you know the density, $\rho$ in terms of length. You want to know the mass of the wire, so you take the density, which is a rate, namely, mass per unit length, multiply by an infinitesimal length to get an infinitesimal amount of mass $dm = \rho(x)\, dx$, then add them to get the mass $m$ of the wire,

$$m = \int dm = \int \rho(x)\, dx$$

You might even say $dx$ is the length of a point and $dm$ is the mass of a point-size piece of wire - and I would not laugh at you.

People actually did reason in terms of infinitesimal pieces of things in the beginnings of calculus. It was intuitive and worked. Mathematically, it is not rigorous, that is, not legal, so we mathematicians fixed it. But, infinitesimals are still intuitively appealing, and people still think and work that way and do great things. You will find me doing it, too. Soon, I will find you doing it.

The wire example is unrealistic, wire has volume, not just length, but you have only one variable calculus, hence the uniform assumption. What you really want to do is to take the density of an object, mass per unit volume, multiply by an infinitesimal volume and add them up. That requires the Riemann integral for three variables, but we will start with two variable version. When we finish we will have the following integrals.

| | Integrating over a set of | | |
|---|---|---|---|
| | 1 dimension | 2 dimensions | 3 dimensions |
| $\mathbb{R}$ | $\int_a^b f(x)\, dx$ | | |
| $\mathbb{R}^2$ | $\int_C f(x,y)\, ds$ | $\int_D f(x,y)\, dA$ | |
| $\mathbb{R}^3$ | $\int_C f(x,y,z)\, ds$ | $\int_S f(x,y,z)\, dS$ | $\int_B f(x,y,z)\, dV$ |

In other words, you will be able to integrate functions over all possible dimensional sets in any of the three euclidean spaces where they make sense.

## 6.1 The Riemann integral

### 6.1.1 The hard way

In one variable calculus, you no doubt began by looking at the definition of the Riemann integral. You have $f : \mathbb{R} \to \mathbb{R}$ and an interval $[a, b]$.

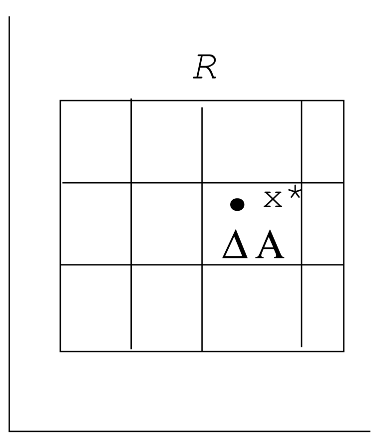
- You chop up the interval into pieces,

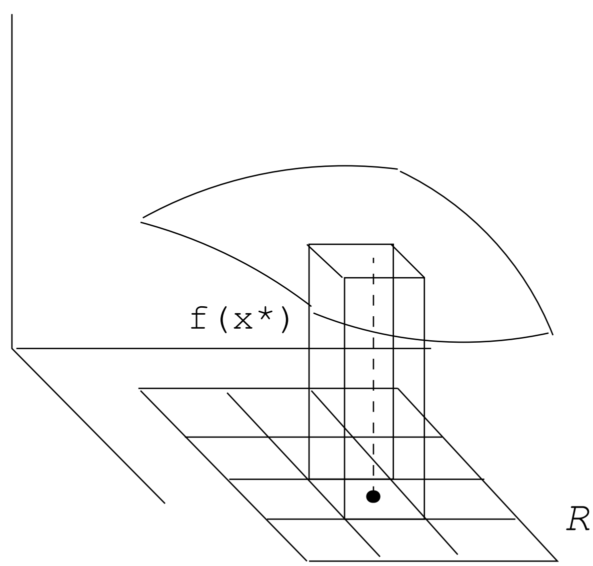
- pick a point in each piece and evaluate the function there,
- multiply the function value for the piece by the length of the piece,
- add them all up and you have a Riemann sum.
- Finally, you take the limit of the Riemann sums as you chop the interval finer and finer and that is the integral

which you computed this way once, maybe twice. Fortunately, you can calculate integrals much more easily using antiderivatives, so that is for the most part what you did.

Be that as it may, we define the Riemann integral for $f : \mathbb{R}^2 \to \mathbb{R}$ the same way. Basically, I'll give you a sketch. Suppose you want to integrate the function $f : \mathbb{R}^2 \to \mathbb{R}$ over a rectangle $R$

- You chop up the rectangle into pieces,
- pick a point in each piece and evaluate the function there,
- multiply the function value for the piece by the area of the piece,
- add them all up and you have a Riemann sum.
- Finally, you take the limit of the Riemann sums as you chop the rectangle finer and finer and that is the integral




and we have vaguely

$$\int_R f(x, y)\, dA = \lim \sum f(x^*)\, \Delta A$$

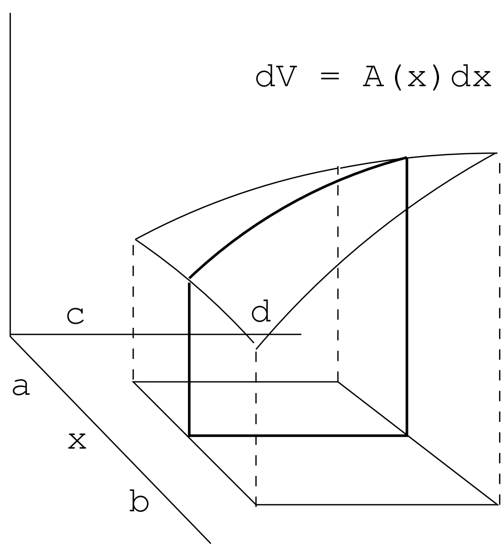
One advantage to this approach is that you believe you have calculated something you understand, the volume under the graph. You have a solid with $R$ as the base, and the graph of $f$ as the top, and vertical sides. Well, provided the graph is above the $x$-$y$ plane. If it goes below, then the volume is subtracted, but comes as no surprise.

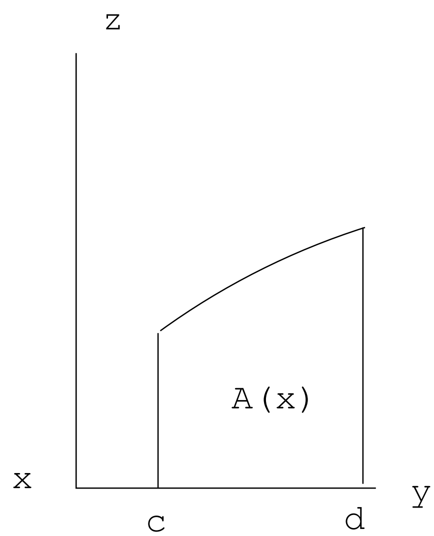
This time you will not even have to calculate it once, we proceed immediately to ...

### 6.1.2 The easy way - iterated

Let's reconsider $\int_R f(x,y)\,dA$

Suppose we pick an $x$ and slice the $x$-axis with a plane perpendicular to it at the point $x$, then we get a cross-section of the solid under the graph of $f$,




The area of this cross-section is the one variable integral

$$A(x) = \int_c^d f(x,y)\,dy$$

where $x$ is treated as a constant. In space we have an infinitesimally thin slice whose volume is $A(x)\,dx$, the area of the base, $A(x)$, times the thickness, $dx$. To get the volume just add up the volumes of the slices from one side to the other.

$$\int_a^b A(x)\,dx = \int_a^b \left( \int_c^d f(x,y)\,dy \right) dx$$

Essentially, we are simply integrating away each variable, one at a time. The integral is called an *iterated integral*, written

$$\int_a^b \int_c^d f(x,y)\,dydx$$

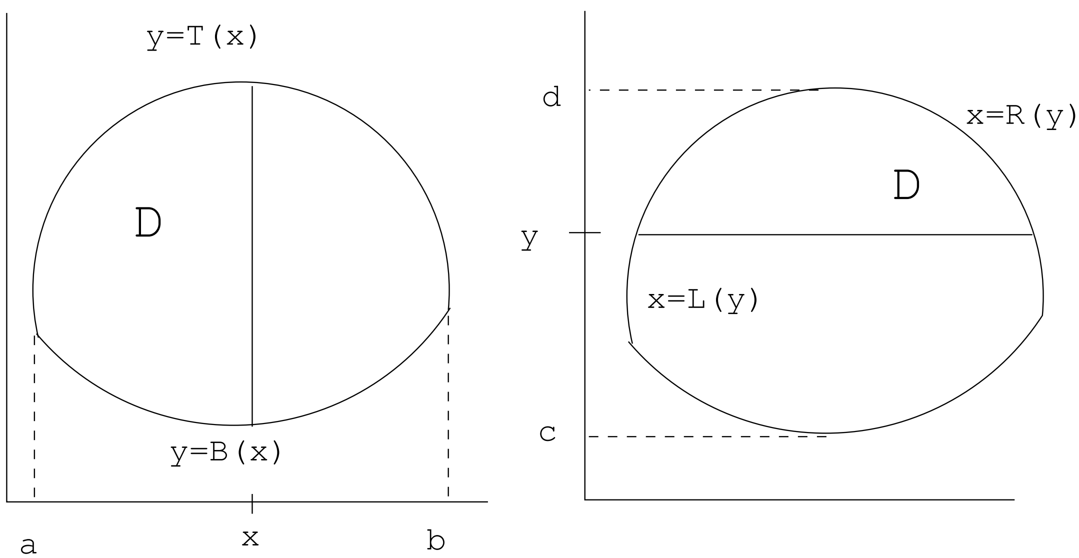
You could also slice perpendicular to the $y$-axis to get

$$\int_c^d \int_a^b f(x,y)\,dxdy$$

We have the following wonderful fact that we will assume applies at all times.

$$\int_R f(x,y)\,dA = \int_a^b \int_c^d f(x,y)\,dydx = \int_c^d \int_a^b f(x,y)\,dxdy$$

Not only do we get an easier way to calculate, we can easily expand to more general regions. A set of points is *convex* means that for any two points in the set, the line segment connecting the points is entirely in the set. Suppose $D$ is a bounded, convex subset of $\mathbb{R}^2$, the fact that it is convex means that we can realize the top and bottom as the graphs of functions of $x$, we can also realize the sides as functions of $y$. Perhaps a picture is worth a few words.




So, to integrate over $D$ we can do it either of two ways. First slice at a fixed $x$ between $a$ and $b$. At that $x$, $B(x) \le y \le T(x)$ so the area of the slice is

$$\int_{B(x)}^{T(x)} f(x,y)\,dy$$

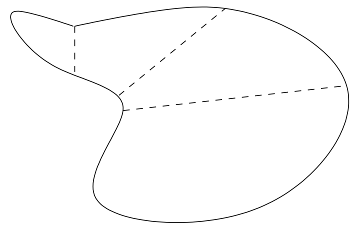
and then adding up the slices gives

$$\int_D f(x,y)dA = \int_a^b \int_{B(x)}^{T(x)} f(x,y)\,dydx$$

similarly

$$\int_D f(x,y)dA = \int_c^d \int_{L(y)}^{R(y)} f(x,y)\,dxdy$$

If the region you want to integrate over is not convex, you may be able to chop it up into convex sets and deal with them individually, then add up all the results



And that's how its done.

Now there is no reason to restrict to two variables either the Riemann integral or iterated integrals. The geometry of the situation disappears rapidly, but the arithmetic is the same. You can define Riemann sums and take their limits, and iterate integrals in any number of variables as well. We will stick to 2 or 3 variables.

In $\mathbb{R}^3$ you would want a Riemann integral over a three dimensional solid $B$, if the solid is convex, the you could imagine doing something like this.

$$\int_B f(x,y,z)\,dV = \int_a^b \int_{L(x)}^{R(x)} \int_{B(x,y)}^{T(x,y)} f(x,y,z)\,dzdydx$$

### 6.1.3 Changing the variable

Changing variables in an integral is one of the most useful techniques of integration. Recall, that for $u = g(x)$,

$$\int_{g(a)}^{g(b)} f(u)\,du = \int_a^b f(g(x))g'(x)\,dx$$

You could go from right to left and eliminate $x$ or you could go from left to right and introduce $x$. Either way you change the problem and hope it gets better.

Let me look at the pieces of the formula.

1. $f(u) = f(g(x))$ are the values of the function, given in terms of $u$ or $x$, they are the same.

2. $[a,b]$ the set of $x$'s being used, $[g(a), g(b)]$ (or possibly $[g(b), g(a)]$) the set of $u$'s that are being used, and are the range of $g$.

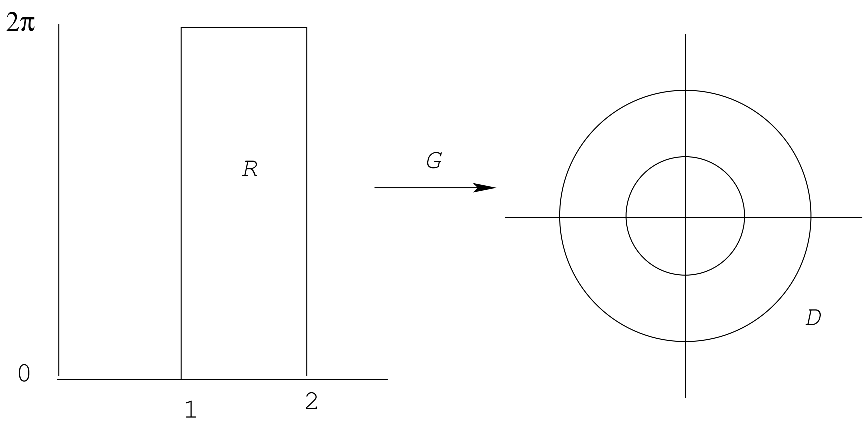
3. $du = g'(x)\,dx$, this is a critical one. You could say you are integrating $f$ in $u$, that is what you want to calculate, so in changing to $x$, you need to know what an infinitesimal piece of $u$, $du$, is in terms of $x$ and it is precisely $\frac{du}{dx}dx = g'(x)dx$. You want to change the variable, but not the geometry.

So, to $\mathbb{R}^2$. You have

$$\int_D f(x,y)dA$$

and think a change of variables is worthwhile. When would that be?

Suppose you need to integrate over the annulus, $D = \{(x, y) : 1 \leq x^2 + y^2 \leq 4\}$. This set is not at all fun to integrate over. You would need to chop it up into four pieces, with unpleasant functions describing the sides. However, look at $G(r, \theta) = (r \cos \theta, r \sin \theta)$. The rectangle $R = \{(r, \theta) : 1 \leq r \leq 2, 0 \leq \theta \leq 2\pi\}$ draws $D$.




and integrating over $R$ is easy.

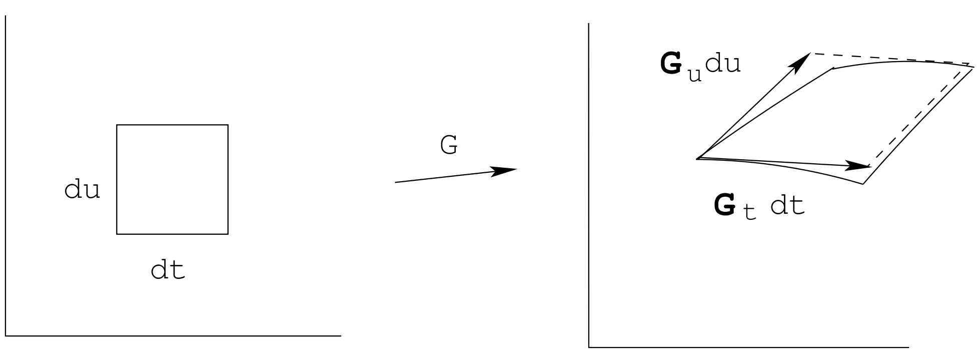
In general, suppose we want to change variables in $\int_D f(x, y) dA$ from $(x, y)$ to $(t, u)$ using $G(t, u) = (x(t, u), y(t, u))$. We need that part of the domain of $G$ that draws $D$, call it $E$. We can now say that

$$\int_D f(x, y) dA_{x,y} = \int_E f(x(t, u), y(t, u)) \boxed{\phantom{xx}} dA_{t,u}$$

where the box must contain whatever scale factor is needed so that

$$\boxed{\phantom{xx}} dA_{t,u} = dA_{x,y}$$

So, imagine, if you will, an infinitesimal box in $t$ an $u$ and its image under $G$.




We can approximate the image with a parallelogram built from the first derivatives of $G$, tangent vectors to the image. This step is also known as "throwing away higher order terms". The parallelogram is an infinitesimal piece of area in $(x, y)$ described in terms of $t$ and $u$. The sides of the parallelogram are $\mathbf{G}_t dt = \langle x_t dt, y_t dt\rangle$ and $\mathbf{G}_u du = \langle x_u du, y_u du\rangle$. The area, according to 3.2.1 earlier in the book, is

$$dA_{x,y} = |x_t dt\, y_u du - x_u du y_t dt| = |x_t y_u - x_u y_t| dt du = |x_t y_u - x_u y_t| dA_{t,u}$$

and

$$\int_D f(x,y) dA_{x,y} = \int_E f(x(t,u), y(t,u)) |x_t y_u - x_u y_t|\, dA_{t,u}$$

and that's it.

$J(t,u) = x_t y_u - x_u y_t$ is called the *Jacobian determinant*.

For the annulus example, $J(r,\theta) = x_r y_\theta - x_\theta y_r = r$, so that

$$\int_D f(x,y) dA = \int_R f(r\cos\theta, r\sin\theta)\, r\, dA$$

If you want to change variables in $\mathbb{R}^3$ from $(x,y,z)$ to $(t,u,w)$. Another look at 3.2.1 might convince you that the Jacobian determinant should be

$$J(t,u,w) = \det \begin{bmatrix} x_t & x_u & x_w \\ y_t & y_u & y_w \\ z_t & z_u & z_w \end{bmatrix}$$

## 6.2 On lower dimensional sets

You could want to add things up along a curve or over a surface. The philosophy is the same, function value times an infinitesimal amount of the space, add them up.

### 6.2.1 Curves

We want to integrate a function $f : \mathbb{R}^2 \to \mathbb{R}$ along a curve $C$ between two points on the curve, $P$ and $Q$. The integral will be denoted

$$\int_C f(x,y)\, ds$$

It is defined just as it is written, add up values of $f$ times an infinitesimal amount of arclength along the curve.

It may come as no surprise that we will find it productive to have a parameterization, $r : \mathbb{R} \to \mathbb{R}^2$ with $r(t) = (x(t), y(t))$. Now we describe everything in terms of $r$.

In particular, $P = r(a)$ and $Q = r(b)$ for some real numbers $a$ and $b$ and

$$ds = \frac{ds}{dt}\, dt = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\; dt$$

and the integral of $f$ over $C$ is defined to be

$$\int_C f(x,y)\,ds = \int_a^b f(x(t),y(t))\frac{ds}{dt}\,dt$$

It should be clear how to define the integral for curves in space, just add the $z$ stuff.

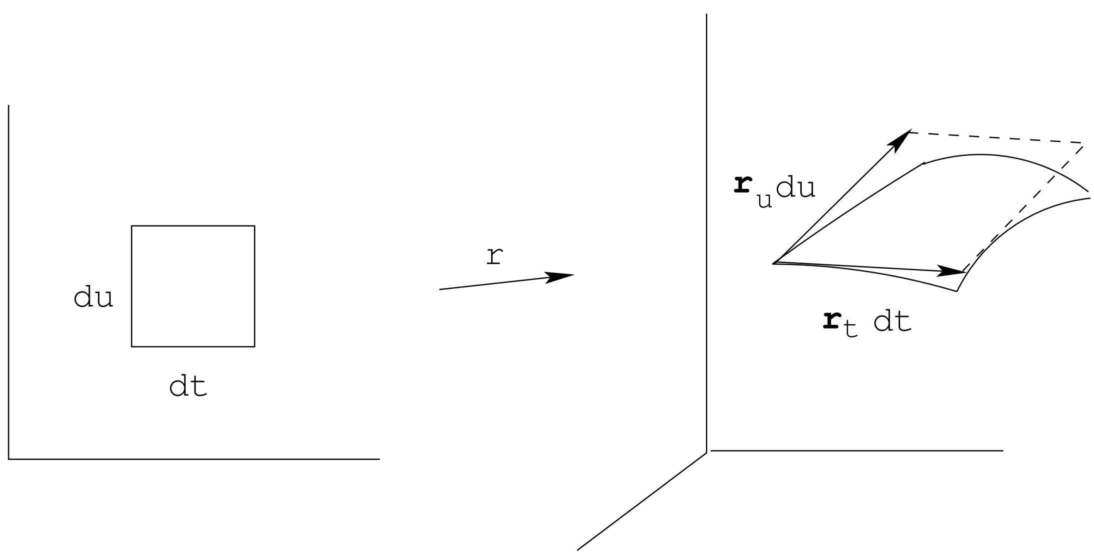
We can now integrate functions on one dimensional sets in $\mathbb{R}^2$ and $\mathbb{R}^3$ by using a parameterization to reduce the calculation to an ordinary Riemann integral in $\mathbb{R}$.

### 6.2.2 Surfaces

Defining an integral over a surface proceeds just as with the curves. You have a surface $S$ in $\mathbb{R}^3$ and a function $f : \mathbb{R}^3 \to \mathbb{R}$. You want to multiply the values of $f$ by an infinitesimal piece of area on the surface and add them up.

Naturally, find a parameterization $r : \mathbb{R}^2 \to \mathbb{R}^3$ with $r(t,u) = (x,y,z)$ and a region $D$ in $\mathbb{R}^2$, so that the values of $r$ at points in $D$ produce the surface $S$ we want to work with. We will reduce the problem to a Riemann integral over $D$ in $\mathbb{R}^2$.

The first step of the reduction is to determine how to measure area on $S$ in terms of area in $D$. Take an infinitesimal rectangle in $D$ with sides $dt$ and $du$ whose area is $dA = dtdu$. Its image under $r$ would be some infinitesimal region, which we can approximate by a parallelogram built from the tangent vectors to the surface. Using the parallelogram instead of the image amounts to "throwing away higher order terms."




The area of this parallelogram would be the infinitesimal area we need. The area is

$$dS = |(\mathbf{r}_t dt) \times (\mathbf{r}_u du)| = |\mathbf{r}_t \times \mathbf{r}_u| dtdu = |\mathbf{r}_t \times \mathbf{r}_u|\,dA$$

The integral of the function over the surface is

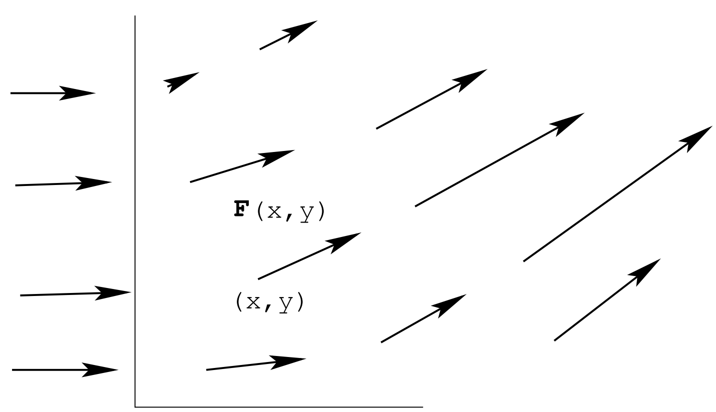
$$\int_S f(x,y,z)\,dS = \int_D f(x(t,y),y(t,u),z(t,u))|\mathbf{r}_t \times \mathbf{r}_u|\,dA$$

# Chapter 7

# Vector fields

A *vector field* is a function that assigns to a point in $\mathbb{R}^n$ an $n$-dimensional vector. In general, of course, vectors have no location, but in this case one usually puts the tail of the vector field at the point where it is defined.

For example, for $\mathbf{F} : \mathbb{R}^2 \to$ 2-dimensional vectors, with $(x, y) \to \mathbf{F}(x, y)$ we might have a picture like this.




You can no doubt see why vector fields and this visualization could be useful. The field in this picture might represent the velocity vectors of a fluid flowing the through the plane, or a force field acting at each point in the plane.

We are already familiar with a most important vector field, for $f : \mathbb{R}^2 \to \mathbb{R}$, its gradient $\nabla f$ is a vector field. We have typically drawn $\nabla f(x, y)$ with its tail at $(x, y)$ because it shows the direction of maximum increase from there.

## 7.1 The antiderivative problem

For a function $f$, $\nabla f$ is a vector field. It would be reasonable to ask if a given vector field $\mathbf{F}$ is somebody's gradient.

More precisely, given a vector field $\mathbf{F}$ is there a function $f$, so that $\mathbf{F} = \nabla f$?

This is the basic antiderivative problem. In one variable calculus, the Fundamental Theorem of Calculus says if $F$ is continuous, then the answer is yes. Look at the beginning of the chapter on integration (except that the roles of $F$ and $f$ are reversed).

Let's look at a few examples. I give you $\mathbf{F}(x,y) = \langle P(x,y), Q(x,y)\rangle$ and you find $f$ so that $f_x = P$ and $f_y = Q$.

- $\mathbf{F}(x,y) = \langle 2x, 2y\rangle$
  OK, how about $f(x,y) = x^2 + y^2$, which could be obtained by just thinking about it.

- $\mathbf{F}(x,y) = \langle y, x\rangle$
  OK, how about $f(x,y) = xy$.

- $\mathbf{F}(x,y) = \langle -y, x\rangle$
  Got anything - no - there is a reason for that. If there is an $f$ so that $f_x = -y$ and $f_y = x$, then the first equation says $f_{xy} = -1$ and the second says $f_{yx} = 1$, but they should be equal. So, $\mathbf{F}$ does not have an antiderivative.

The first two fields had antiderivatives. In fact, adding any constant gives another antiderivative. The third, a harmless, simple vector field does not.

We do have a test.

- For $F = \langle P, Q\rangle$ a vector field on $\mathbb{R}^2$, $\mathbf{F}$ has antiderivative $\Longrightarrow Q_x = P_y$.

- For $F = \langle P, Q, R\rangle$ a vector field on $\mathbb{R}^3$, $\mathbf{F}$ has antiderivative $\Longrightarrow \text{curl}(\mathbf{F}) = \mathbf{0}$. The curl will be discussed in detail a bit later, but for the moment I'll just tell you what it is and let you determine why the statement is true.

$$\text{curl}(\mathbf{F}) = \nabla \times \mathbf{F} = \langle R_y - Q_z, P_z - R_x, Q_x - P_y\rangle$$

The test is easy to apply, but gives only negative information. If $Q_x \neq P_y$, then there is no antiderivative. If $Q_x = P_y$, then you don't know what happens. Of course, you could give it a try and see what you can do.

For example $\mathbf{F}(x,y) = \langle 3x^2 - 2y, -2x + 2y\rangle$ so that, $P_y = Q_x = -2$.

If $P = f_x$ for some $f$, then

$$f(x,y) = \int f_x(x,y)\,dx = \int P(x,y)\,dx = \int (3x^2 - 2y)\,dx = x^3 - 2xy + g(y)$$

where $g$ is any function of $y$ alone, since $g_x$ would be zero, and

$$f(x,y) = \int f_y(x,y)\,dy = \int Q(x,y)\,dy = \int (-2x + 2y)\,dy = -2xy + y^2 + h(x)$$

where $h$ is any function of $x$ alone.

We have two versions of what $f$ should be. Do they say the same thing? If $g(y) = y^2$ and $h(x) = x^3$, then they do. So that, $f(x, y) = x^3 - 2xy + y^2 + c$, for any constant $c$ is an antiderivative for $\mathbf{F}$.

A little more subtle example is the following.

$$\mathbf{F}(x, y) = \left\langle \frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right\rangle$$

then

$$P_y = \frac{y^2 - x^2}{(x^2 + y^2)^2} = Q_x$$

So, we go for it.

$$\begin{aligned} \int \frac{-y}{x^2 + y^2}\, dx &= -\tan^{-1}\left(\frac{x}{y}\right) + g(y) \\ \int \frac{x}{x^2 + y^2}\, dy &= \tan^{-1}\left(\frac{y}{x}\right) + h(x) \end{aligned}$$

It would appear that these two do not match up, but they do, if you let $g(y) = \frac{\pi}{2}$. Think of a right triangle, if one of the angles has tangent $y/x$, then the other, $\pi/2-$ that angle has tangent $x/y$. Well, they agree, at least when $x/y$ and $y/x$ are tangents of angles in a right triangle, that is for $x, y > 0$. In other quadrants they do not necessarily agree. For example, if $x = 1$ and $y = -1$, then $\frac{\pi}{2} - \tan^{-1}\frac{x}{y} = \frac{\pi}{2} + \frac{\pi}{4} = \frac{3\pi}{4}$, but $\tan^{-1}\frac{y}{x} = -\frac{\pi}{4}$.

We can do the following.

$$f(x, y) = \begin{cases} \tan^{-1}\left(\frac{y}{x}\right) & x > 0 \ \boxed{\text{A}} \\ \frac{\pi}{2} - \tan^{-1}\left(\frac{x}{y}\right) & y > 0 \ \boxed{\text{B}} \\ -\frac{\pi}{2} - \tan^{-1}\left(\frac{x}{y}\right) & y < 0 \ \boxed{\text{C}} \end{cases}$$

We are using $\boxed{\text{A}}$ to cover everything to the right of the $y$-axis and $\boxed{\text{B}}$ to cover everything above the $x$-axis, and the only place we are using both at the same time is for $x > 0$ and $y > 0$ where we know they agree. To go across the bottom we just pick the constant that gives agreement with $\boxed{\text{A}}$ when $x > 0$ and $y < 0$, giving $\boxed{\text{C}}$. All we have left is to patch across the negative $x$-axis, but that will be impossible. As you approach the $x$-axis from above, $\boxed{\text{B}}$ approaches $\pi$. Approaching from below $\boxed{\text{C}}$ approaches $-\pi$, so that there is a $2\pi$ discontinuous jump that can't be patched.

The original vector field is defined everywhere but (0,0). The antiderivative is defined everywhere except the negative $x$-axis. We have found an antiderivative, on part but not all of the domain of the vector field. So, we found an antiderivative - sort of. It may not seem like much of a problem, but it is just the kind of thing that makes a mathematician mad and we will come back to it.

Much of the early work in calculus was done to solve physics problems. Physics people would say that a vector field that is somebody's gradient is *conservative* and the somebody is its *potential function*. You will see these terms used in mathematics books as well.

## 7.2 Derivatives of vector fields

### 7.2.1 Curl and divergence

Earlier I mentioned something called the curl of a vector field. It is actually one of two ways to differentiate a vector field on $\mathbb{R}^3$.

$$\text{curl}(\mathbf{F}) = \langle R_y - Q_z, P_z - R_x, Q_x - P_y \rangle$$

The curl is also written $\nabla \times \mathbf{F}$. The curl measures the tendency for the vector field to be rotating in the vicinity of a point. The magnitude measures the amount of rotation and the direction the axis of rotation. More on this shortly.

Technically the curl is defined in $\mathbb{R}^3$, but you can use it on two-dimensional fields. For $\mathbf{F}(x, y) = \langle P(x, y), Q(x, y) \rangle$ look at the vector field $\hat{\mathbf{F}} = \langle P, Q, 0 \rangle$. Note that any derivatives with respect to $z$ of $P$ an $Q$ would be zero since they are functions of $x$ and $y$ alone. So, we have $\text{curl}(\hat{\mathbf{F}}) = \langle 0, 0, Q_x - P_y \rangle$.

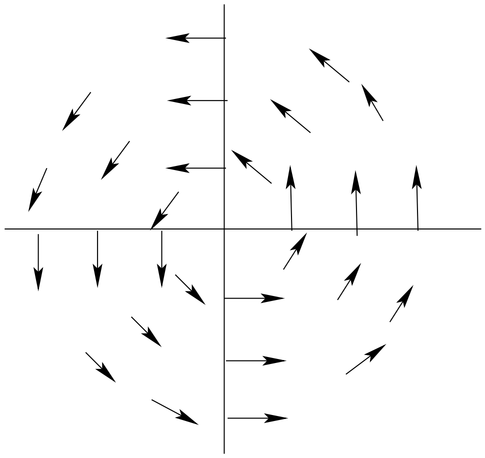
The other way of differentiating a vector field is called the *divergence* and is defined to be

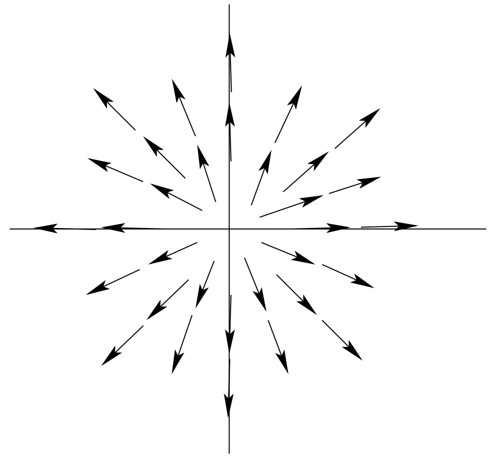
$$\text{div } \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$$

The divergence is also denoted $\nabla \cdot \mathbf{F}$. It is defined on $\mathbb{R}^2$ as well, just delete the $z$ stuff. The divergence measures the extent to which the field is flowing into or out of the point where it is calculated.

Let's look at two examples,

$$\mathbf{F}(x,y) = \left\langle \frac{-y}{\sqrt{x^2+y^2}}, \frac{x}{\sqrt{x^2+y^2}} \right\rangle \quad \mathbf{G}(x,y) = \left\langle \frac{x}{\sqrt{x^2+y^2}}, \frac{y}{\sqrt{x^2+y^2}} \right\rangle$$





$$\text{curl}(\hat{\mathbf{F}}) = \left\langle 0, 0, \frac{1}{\sqrt{x^2+y^2}} \right\rangle \qquad \text{curl}(\hat{\mathbf{G}}) = \mathbf{0}$$

$$\text{div}(\mathbf{F}) = 0 \qquad \text{div}(\mathbf{G}) = \frac{1}{\sqrt{x^2+y^2}}$$

These two examples pretty well tell what these two derivatives measure, if not why anyone cares.

These are differentiation operations, so they have their rules. Such as the following.

- $\nabla \times (\mathbf{F} + \mathbf{G}) = \nabla \times \mathbf{F} + \nabla \times \mathbf{G}$
- $\nabla \times (f\mathbf{F}) = \nabla f \times \mathbf{F} + f\nabla \times \mathbf{F}$
- $\nabla \cdot (\mathbf{F} + \mathbf{G}) = \nabla \cdot \mathbf{F} + \nabla \cdot \mathbf{G}$
- $\nabla \cdot (f\mathbf{F}) = \nabla f \cdot \mathbf{F} + f\nabla \cdot \mathbf{F}$
- $\nabla \times (\nabla f) = \text{curl}(\nabla f) = \mathbf{0}$
- $\nabla \cdot \nabla \times \mathbf{F} = \text{div}(\text{curl}(\mathbf{F})) = 0$

There are, of course, many more, you see here the basics.

The last two are interesting for antiderivative problems. They yield tests for not having an antiderivative, the first of which I have already mentioned. They are, for a given $\mathbf{F}$

- $\mathbf{F} = \nabla f$ for some $f \Rightarrow \text{curl}(\mathbf{F}) = \mathbf{0}$
- $\mathbf{F} = \text{curl}(\mathbf{G})$ for some $\mathbf{G} \Rightarrow \text{div}(\mathbf{F}) = 0$

That covers the antiderivative problem for the gradient and the curl, for the moment, what about the divergence? For a function $f$ is there a vector field $\mathbf{F}$ so that $\text{div}(\mathbf{F}) = f$? Yes,

$$\mathbf{F} = \left\langle p_1\int f(x,y,z)\,dx,\ \ p_2\int f(x,y,z)\,dy,\ \ p_3\int f(x,y,z)\,dz \right\rangle$$

where $p_1 + p_2 + p_3 = 1$

### 7.2.2 The Del operator

A function that acts on functions is often called an *operator*. You have seen many, but probably have not called them operators,

- $(\ )' : f \to f'$
- $\int(\ ) : f \to \int f(x)\,dx$
- $\frac{\partial}{\partial x} : z \to \frac{\partial z}{\partial x}$
- $\nabla : f \to \nabla f$
- $\text{curl} : \mathbf{F} \to \ \text{curl}(\mathbf{F})$
- $\text{div} : \mathbf{F} \to \ \text{div}(\mathbf{F})$

Perhaps the most important to us of late is the so called *del* operator, $\nabla$. We will begin to think of it as a vector of operators and build an arithmetic around it, that will at least help us remember how to calculate some of the other operators we are using. So, the del operator is officially

$$\nabla = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle$$

We are now going to use the usual vector arithmetic with operators thrown in. There is a rule for multiplication that must be used.

> Multiplying an operator on the left by a function is ordinary multiplication. The action produces an operator.
> Multiplying an operator on the right by a function means apply the operator to the function. The action produces a function.

For example,

$$f\nabla g = f\left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle g = \left\langle f\frac{\partial}{\partial x}, f\frac{\partial}{\partial y}, f\frac{\partial}{\partial z} \right\rangle g = \left\langle f\frac{\partial g}{\partial x}, f\frac{\partial g}{\partial y}, f\frac{\partial g}{\partial z} \right\rangle$$

Now, the notation for curl and divergence using the del operator makes more sense and helps you remember how to compute them. In particular,

$$\begin{aligned}
\nabla \times \mathbf{F} &= \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle \times \langle P, Q, R \rangle \\
&= \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{bmatrix} \\
&= (R_y - Q_z)\mathbf{i} + (P_z - R_x)\mathbf{j} + (Q_x - P_y)\mathbf{k}
\end{aligned}$$

$$\begin{aligned}
\nabla \cdot \mathbf{F} &= \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle \cdot \langle P, Q, R \rangle \\
&= \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}
\end{aligned}$$

## 7.3 Integrals of vector fields

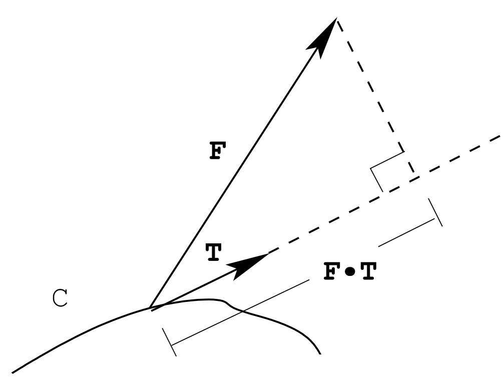
If there is differentiation, there should be integration. For vector fields integration is very interesting. For one thing you integrate vector fields only on lower dimensional sets, that is on curves in $\mathbb{R}^2$ and curves and surfaces in $\mathbb{R}^3$. Well, let's get on with it.

### 7.3.1 Line integrals

We begin by looking at integrating vector fields along curves. If you have a curve $C$ in $\mathbb{R}^3$ and a vector field $\mathbf{F} = \langle P, Q, R \rangle$, then the *(line) integral* of $\mathbf{F}$ over $C$ is

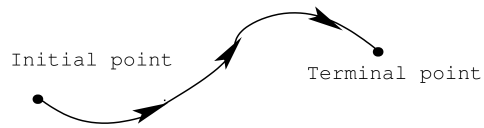
$$\int_C \mathbf{F} \cdot \mathbf{T}\, ds$$

In other words, we are not introducing a new integral. This one is just the integral of a function along a curve, where the function is the tangential component of $\mathbf{F}$ at points on the curve, that is the magnitude of the action of $\mathbf{F}$ in the direction of travel along the curve.




Wait a minute, what was that about the direction of travel? Looking at the picture the "direction of travel" appears to be from left to right, if we had gone the other way, the vector $\mathbf{T}$ would be pointing to the left and $\mathbf{F} \cdot \mathbf{T}$ would be the negative of the component in the other direction, so that the way the curve is traversed makes a difference in the value of the integral.

In order to straighten out this mess we introduce the *oriented curve*, which is a curve and a direction of travel along the curve.




The direction is indicated by arrows along the curve or by specifying a starting point, the *initial* point, and the ending or *terminal* point of the curve.

So, the real definition of a line integral is the following

For an oriented curve $C$ in $\mathbb{R}^3$ and a vector field $\mathbf{F} = \langle P, Q, R\rangle$, then the *(line) integral* of $\mathbf{F}$ over $C$ is

$$\int_C \mathbf{F}\cdot\mathbf{T}\,ds$$

with the usual analogous definition in $\mathbb{R}^2$.

If $C$ is an oriented curve, the the same set of points traveled in the opposite direction is denoted $-C$ and

$$\int_{-C} \mathbf{F}\cdot\mathbf{T}\,ds = -\int_C \mathbf{F}\cdot\mathbf{T}\,ds$$

So, what good is it. Well this should hold you for a bit. If the curve is the path of a particle and the vector field is a force field, the line integral is the work done in moving from the initial to the terminal point along the curve. In fact, you could say the $\mathbf{F}\cdot\mathbf{T}\,ds$ is the work done in moving through a point, the force acting on the particle $\mathbf{F}\cdot\mathbf{T}$ times the distance traveled $ds$.

There is another way of writing the line integral,

$$\int_C P\,dx + Q\,dy + R\,dz$$

It is a little dangerous, because one is tempted to say things like $\int_C (2x+y^3)\,dx + 2xy\,dy$ has something to do with $x^2 + xy^3 + xy^2$ but it doesn't.

By the way, how do you calculate this thing? First you parameterize the curve with $r(t) = (x(t), y(t), z(t))$. and then

- $\mathbf{T} = \frac{1}{|\mathbf{r}'|}\mathbf{r}'$
- $ds = |\mathbf{r}'|dt$, so that
- $\mathbf{T}\,ds = \frac{1}{|\mathbf{r}'|}\mathbf{r}'|\mathbf{r}'|dt = \mathbf{r}'\,dt$
- The initial point is $r(a)$ for some $a$ and the terminal point is $r(b)$ for some $b$.

and we have

$$\int_C \mathbf{F}\cdot\mathbf{T}\,ds = \int_a^b \mathbf{F}(t)\cdot\mathbf{r}'(t)\,dt = \int_a^b \left(P\frac{dx}{dt} + Q\frac{dy}{dt} + R\frac{dz}{dt}\right)\,dt$$

One should come up with a parameterization that draws the curve in the correct direction. If you have $r(t)$ with $a \le t \le b$ that is backwards there are two ways to get out of trouble. The "correct" way is to use the parameterization $r(a+b-t)$, check it out. The quick and dirty way is to use the one you have and change the sign of the result.

In $\mathbb{R}^2$ there is also a line integral of the normal component of $\mathbf{F} = \langle P, Q\rangle$, which is

$$\int_C \mathbf{F}\cdot\mathbf{n}\,ds = \int_C -Q\,dx + P\,dy$$

What is happening is that if the tangent vector to the curve is $\mathbf{T} = \langle x', y'\rangle/|\mathbf{r}'|$, then the normal vector being used is $\langle y', -x'\rangle/|\mathbf{r}'|$. The reason for this choice will be discussed later.

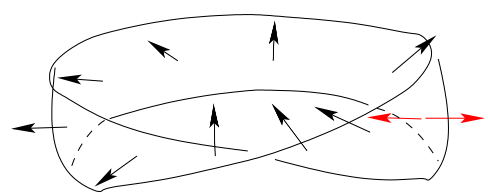
### 7.3.2 Surface integrals

Integrating vector fields over surfaces proceeds much the same as integrating over curves.

In particular, there is need for oriented surfaces. An *oriented surface* is a surface in $\mathbb{R}^3$ with a unit normal (perpendicular) vector chosen at each point so that it is continuously defined as you move about the surface. This last condition means the normal vector does not change direction abruptly. So, to orient a sphere you would choose the normal that always points outward for one orientation or the always pointing inward normal for the other.

Each surface has, then, two orientations, if $S$ has one then $-S$ has the other.

Or none! There are so-called non-orientable surfaces. One "simple" one is the Möbius strip. Take a belt and give it a half twist and buckle it. You have a closed band with a twist in it. Now walk a normal vector around the band when you get back to where you started your normal will point in the opposite direction from when it began. Not good.



Ok, so we will restrict ourselves to orientable surfaces, there are plenty.

For an oriented surface $S$ with normal vector $\mathbf{n}$, and a vector field $\mathbf{F}$, the *(surface) integral* of $\mathbf{F}$ over $S$ is

$$\int_S \mathbf{F}\cdot\mathbf{n}\,dS$$

Again the integral is not new, it is the integral of a function over a surface, but the function is the normal component of a vector field.

Why the normal component? Well, I could say it is a physics thing, which was probably where it got its start. For example, if **F** is the velocity vector of a fluid flowing through space, the $\mathbf{F} \cdot \mathbf{n}$ is the speed of the fluid through the surface $S$. I think it is useful in physics because of what it does mathematically, which we'll see in the next chapter.

You calculate with a parameterization $r(t, u) = (x(t, u), y(t, u), z(t, u))$, for which

- $\mathbf{n} = \frac{1}{|\mathbf{r}_t \times \mathbf{r}_u|}\mathbf{r}_t \times \mathbf{r}_u$
- $dS = |\mathbf{r}_t \times \mathbf{r}_u| dA$
- $\mathbf{n}\, dS = \frac{1}{|\mathbf{r}_t \times \mathbf{r}_u|}\mathbf{r}_t \times \mathbf{r}_u |\mathbf{r}_t \times \mathbf{r}_u|\, dA = \mathbf{r}_t \times \mathbf{r}_u\, dA$
- A region $D$ in the parameter space that draws with $r$ the surface $S$,

then

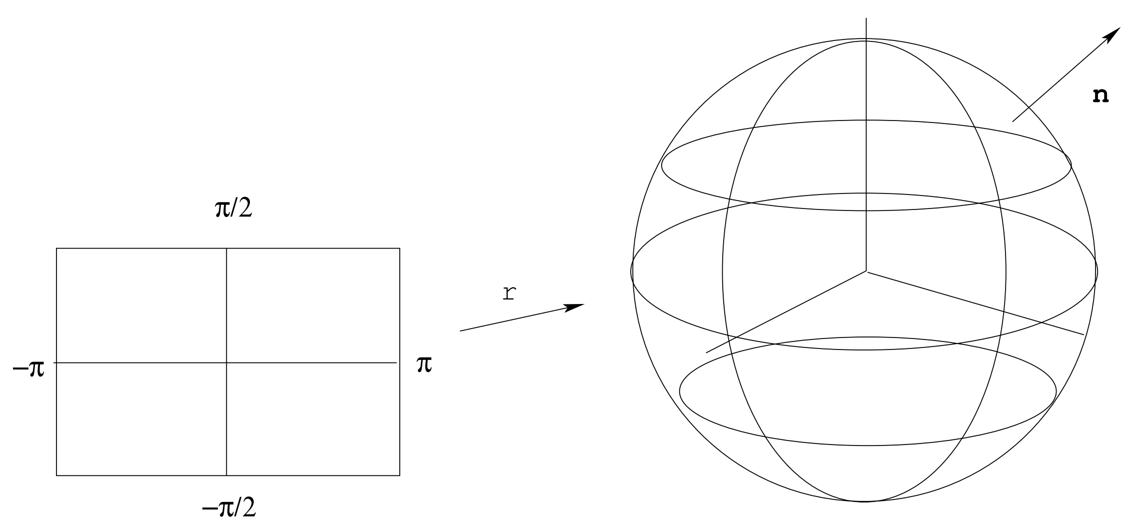
$$\int_S \mathbf{F} \cdot \mathbf{n}\, dS = \int_D \mathbf{F}(t, u) \cdot (\mathbf{r}_t \times \mathbf{r}_u)\, dA$$

Of course, you should check that the parameterization has the proper orientation. If not you can switch the parameters, that is replace $(t, u)$ with $(u, t)$ or change the sign of the result.

Let's take a quick look at the unit sphere centered at the origin in $\mathbb{R}^3$. We use the parameterization based on spherical coordinates, $r(\theta, \phi) = (x, y, z)$ where

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

for $-\pi \le \theta \le \pi$ and $-\pi/2 \le \phi \le \pi/2$.



We have

$$\begin{aligned}
\mathbf{r}_\theta \times \mathbf{r}_\phi &= \cos\phi\langle\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi\rangle \\
\mathbf{n} &= \langle\cos\phi\cos\theta, \cos\phi\sin\theta, \sin\phi\rangle
\end{aligned}$$

You might compare this parameterization with the one we had earlier in the chapter "Algebra and Geometry - seriously", the difference is that the order of the variables $\theta$ and $\phi$ is reversed and so is the direction of the normal.

### 7.3.3 Integration summary

We now have ten integrals for functions or vector fields. Here is a recap.

**Functions**

| | Integrating over a set of 1 dimension | 2 dimensions | 3 dimensions |
|---|---|---|---|
| $\mathbb{R}$ | $\int_a^b f(x)\,dx$ | | |
| $\mathbb{R}^2$ | $\int_C f(x,y)\,ds$ | $\int_D f(x,y)\,dA$ | |
| $\mathbb{R}^3$ | $\int_C f(x,y,z)\,ds$ | $\int_S f(x,y,z)\,dS$ | $\int_B f(x,y,z)\,dV$ |

**Vector fields**

| | Integrating over a set of 1 dimension | 2 dimensions |
|---|---|---|
| $\mathbb{R}^2$ | $\int_C \mathbf{F}\cdot\mathbf{T}\,ds$ | |
| | $\int_C \mathbf{F}\cdot\mathbf{n}\,ds$ | |
| $\mathbb{R}^3$ | $\int_C \mathbf{F}\cdot\mathbf{T}\,ds$ | $\int_S \mathbf{F}\cdot\mathbf{n}\,dS$ |

# Chapter 8

# The Fundamental Theorem of Calculus

The Fundamental Theorem of one-variable Calculus says
If $F$ is continuous on an open interval $I$, then

- for $a$ in $I$,
$$f(x) = \int_a^x F(t)dt$$
is an antiderivative for $F$ on $I$, that is, $f' = F$,

- if $g$ is any antiderivative of $F$ on $I$ then, for any $a$ and $b$ in $I$,
$$\int_a^b F(x)\,dx = g(b) - g(a)$$

So, the first part tells you how to get antiderivatives from integrals and the second, how to get integrals from antiderivatives. The first part is the most philosophically important and the second part is the most practically useful.

To some extent we can summarize the whole thing with
$$\int_a^b f'(x)\,dx = f(b) - f(a)$$

OK, what would the higher dimensional analog be?

## 8.1 Antiderivatives reloaded

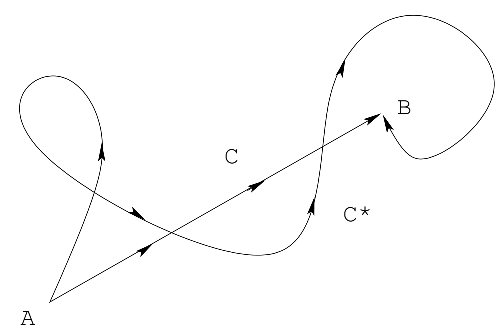
Suppose you have $f : \mathbb{R}^2 \to \mathbb{R}$, and $A$ and $B$ points in $\mathbb{R}^2$ and an oriented curve $C$ from $A$ to $B$. Integrating the "derivative" of $f$ along the curve would seem most naturally to mean to calculate
$$\int_C \nabla f \cdot \mathbf{T}\,ds$$

Parameterize the curve with $r(t) = (x(t), y(t))$, with $r(a) = A$ and $r(b) = B$ for some $a$ and $b$, and for convenience let $h(t) = f(x(t), y(t))$, then we have

$$\begin{aligned}
\int_C \nabla f \cdot \mathbf{T}\, ds &= \int_a^b \left( \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \right) dt \\
&= \int_a^b h'(t)\, dt \qquad \text{(Chain Rule)} \\
&= h(b) - h(a) \qquad \text{(Fundamental Theorem of Calculus)} \\
&= f(B) - f(A)
\end{aligned}$$

And there you have it, a version of the Fundamental Theorem of Calculus in $\mathbb{R}^2$. The same approach produces the same result in $\mathbb{R}^3$. Moreover, it wasn't that difficult. A curve is a one-dimensional set, we reduced the problem to a one variable calculation with the parameterization and applied the one variable version of the Fundamental Theorem of Calculus.

Wait a minute! This higher dimensional Fundamental Theorem of Calculus says something remarkable: the integral of the gradient of a function from $A$ to $B$ doesn't depend on how you get from $A$ to $B$ it is $f(B) - f(A)$, period.




We say that line integrals of the gradient are *independent of path* where that means curves between the same two points, not just any curves.

The physics terms that a vector field is conservative means it is somebody's gradient comes from the field representing forces. A conservative force is one for which the work done getting from one point to another does not depend on the path. It is a conservation of energy thing. The function whose gradient is the force is called a potential function, because the work is the change in potential energy in moving between the points.

Here is an example of perhaps the most important of these forces: gravity. According to Newton the force of gravity is proportional to $1/r^2$ where $r$ is the distance between two objects attracting each other. So, if a mass $M$ is at the origin and a mass $m$ is at $(x, y, z)$ the gravitational attraction has magnitude

$$|\mathbf{F}| \propto \frac{1}{r^2} = \frac{1}{x^2 + y^2 + z^2}$$

and acts in a direction pointing from $m$ to $M$,

$$\mathbf{u} = -\frac{\langle x, y, z\rangle}{|\langle x, y, z\rangle|}$$

so that with $G$ being the universal gravitational constant

$$\mathbf{F} = -\frac{mMG}{x^2+y^2+z^2}\frac{\langle x,y,z\rangle}{|\langle x,y,z\rangle|} = -\frac{mMG}{(x^2+y^2+z^2)^{\frac{3}{2}}}\langle x,y,z\rangle$$

and with a little work you can get the potential function

$$\phi(x,y,z) = \frac{mMG}{\sqrt{x^2+y^2+z^2}}$$

Let's go back to the general antiderivative problem in light of the first part of the Fundamental Theorem of Calculus and see if we can build one using integrals.

Suppose $\mathbf{F} = \langle P, Q\rangle$ is a vector field in $\mathbb{R}^2$. Do the following.

- Pick $(x_0, y_0)$ in $\mathbb{R}^2$.
- For each $(x, y)$ pick a curve from $(x_0, y_0)$ to $(x, y)$, call it $C_{(x,y)}$.
- Define
$$f(x,y) = \int_{C_{(x,y)}} \mathbf{F}\cdot\mathbf{T}\,ds$$

then $f$ should be an antiderivative of $\mathbf{F}$. But, not all vector fields have antiderivatives, so what can go wrong.

For one thing, $f$ may not be well-defined, that is $f(x, y)$ should depend only on $(x, y)$, but we had to choose a curve from $(x_0, y_0)$ to $(x, y)$. The value of the integral may depend not just on $(x, y)$ but on which curve we use. In other words, if the line integrals of $\mathbf{F}$ are not independent of path, this approach will definitely not work.

If you do have path independence, then it does work!

The problem is to show that $\nabla f = \mathbf{F}$, so look at

$$\begin{aligned} f_x(x,y) &= \lim_{h\to 0}\frac{1}{h}(f(x+h,y)-f(x,y)) \\ &= \lim_{h\to 0}\frac{1}{h}\left(\int_{C_{(x+h,y)}} \mathbf{F}\cdot\mathbf{T}\,ds - \int_{C_{(x,y)}} \mathbf{F}\cdot\mathbf{T}\,ds\right) \end{aligned}$$

There is another way to get from $(x_0, y_0)$ to $(x+h, y)$, namely go from $(x_0, y_0)$ to $(x, y)$, then from $(x, y)$ to $(x + h, y)$ using the parameteriztion $r(t) = (x + th, y)$ which draws the straight line segment between them for $0 \le t \le 1$, call it $C_h$.

Now, since line integrals of $\mathbf{F}$ are independent of path,

$$\int_{C_{(x+h,y)}} \mathbf{F}\cdot\mathbf{T}\,ds = \int_{C_{(x,y)}} \mathbf{F}\cdot\mathbf{T}\,ds + \int_{C_h} \mathbf{F}\cdot\mathbf{T}\,ds$$

so that

$$\begin{aligned} f_x(x,y) &= \lim_{h\to 0} \frac{1}{h} \int_{C_h} \mathbf{F}\cdot\mathbf{T}\, ds \\ &= \lim_{h\to 0} \int_0^1 P(x+th,y)\, dt \\ &= P(x,y) \end{aligned}$$

Similarly $f_y = Q$. Similarly in $\mathbb{R}^3$.

One thing that has happened is that we have a condition that tells us when we have an antiderivative, not just when we don't, namely

**F** has antiderivatives $\Longleftrightarrow$ line integrals of **F** are independent of path.

The only problem is how in the world would you know that line integrals of some vector field are independent of path?

We can simplify the problem a little. A *closed* curve is one whose initial and terminal points are the same - a loop. If line integrals of a vector field are independent of path then the integral of the vector field around a closed path will be zero, because it is the same as going directly from a point on the curve to itself by not moving at all. Moreover, if $C$ and $C^*$ are curves from $A$ to $B$, then $C$ followed by $-C^*$ is a closed curve, so that

$$\int_C \mathbf{F}\cdot\mathbf{T}\, ds - \int_{C^*} \mathbf{F}\cdot\mathbf{T}\, ds = 0$$

In other words,

The integrals of **F** around closed curves are zero $\Longleftrightarrow$ line integrals of **F** are path independent $\Longleftrightarrow$ **F** has antiderivatives.

Ok, we have reduced the problem from looking at all paths to looking only at closed paths - doesn't sound like much of improvement. But it is, as we are about to see.

## 8.2 Green's Theorem

Let's get to it.

**Green's Theorem[1]:** If $\mathbf{F} = \langle P, Q\rangle$ is defined on a region $D$ in $\mathbb{R}^2$, and $P$ and $Q$ have continuous partial derivatives on $D$, then

$$\int_D (Q_x - P_y)\, dA = \int_{\partial D} Pdx + Qdy$$

[1]George Green, 1793-1841

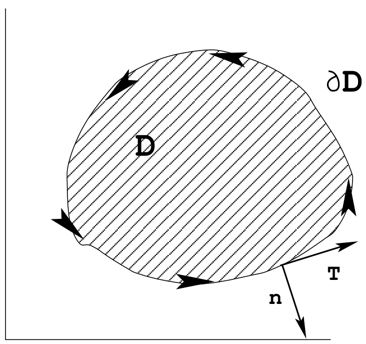
or

$$\int_D \operatorname{curl}(\hat{\mathbf{F}}) \cdot \mathbf{k}\, dA = \int_{\partial D} \mathbf{F} \cdot \mathbf{T}\, ds$$

and

$$\int_D \operatorname{div}(\mathbf{F})\, dA = \int_{\partial D} \mathbf{F} \cdot \mathbf{n}\, ds$$

where $\partial D$ is the boundary of $D$ traversed so that $D$ is on the left.




$\partial D$ is the geometrical boundary of $D$, oriented so that all the signs in the integrals are the way we want them. Recall also that $\mathbf{n}$ in terms of a parameterization was chosen to be $\langle y', -x' \rangle / |r'|$. The reason for that choice is so that the normal in this situation points outward from $D$, which, as before, ensures that all the signs in the integrals are the way we want them.

So, what is this all about? It is, in fact, another version of the Fundamental Theorem of Calculus. It says that integrals of derivatives of $\mathbf{F}$ over a set are completely determined by values of $\mathbf{F}$ on the boundary of the set, which is the basic philosophy of the Fundamental Theorem of Calculus.

I want to show you why this theorem is true, but perhaps before doing so, I should show you why anyone would care that it is true.

For one thing, it solves a problem we left at the end of the last section, at least in $\mathbb{R}^2$. Suppose you have a vector field $\mathbf{F} = \langle P, Q \rangle$ and $Q_x = P_y$. Suppose also you have a closed curve $C$, then $C$ encloses a region in $\mathbb{R}^2$, call it $D$ and either $C$ or $-C$ is $\partial D$, so that

$$\int_C P\, dx + Q\, dy = \pm \int_D (Q_x - P_y) dA = 0$$

That is, $Q_x = P_y$ implies that the integral around any closed curve of $\mathbf{F}$ is zero, so $\mathbf{F}$ has antiderivatives.

Almost, there is a catch. The partial derivatives of $P$ and $Q$ must be continuous on $D$ the interior of the curve $C$. Recall

$$\mathbf{F}(x, y) = \left\langle \frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right\rangle$$

then $Q_x = P_y$. The unit circle parameterized by $r(t) = (\cos t, \sin t)$ for $0 \leq t \leq 2\pi$ is a nice closed curve, and it doesn't take much to show that the integral of $\mathbf{F}$ around the circle is $2\pi$, not zero. Green's Theorem simply does not apply here. $P$ and $Q$ don't even exist at $(0, 0)$ much less have continuous derivatives, and $(0, 0)$ is in the region for which the circle is the boundary.

On the other hand, for $\mathbf{F} = \langle ye^{xy}, xe^{xy} \rangle$, we have $Q_x = e^{xy} + xye^{xy} = P_y$ and everything is continuous on all of $\mathbb{R}^2$, so $\mathbf{F}$ has antiderivatives.

Interestingly enough, the geometry of the set you want an antiderivative on plays a role in whether you can get one.

A set $S$ in $\mathbb{R}^2$ is *simply connected* means the region enclosed by any closed curve in $S$ is also in entirely $S$. Essentially, this simply means that $S$ has no holes in it.

Recall that for

$$\mathbf{F}(x, y) = \left\langle \frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right\rangle$$

We were able to find an antiderivative on the set $S = \{(x, y) : x > 0 \text{ or } y \neq 0\}$, $\mathbb{R}^2$ except for the negative $x$-axis. This set is simply connected. The domain of $\mathbf{F}$ is not.

So, our bottom line for antiderivatives in $\mathbb{R}^2$ is

For a vector field $\mathbf{F}$ whose components have continuous partial derivatives on a simply connected set $S$,

$\mathbf{F}$ has antiderivatives on $S \iff Q_x = P_y$ on $S$

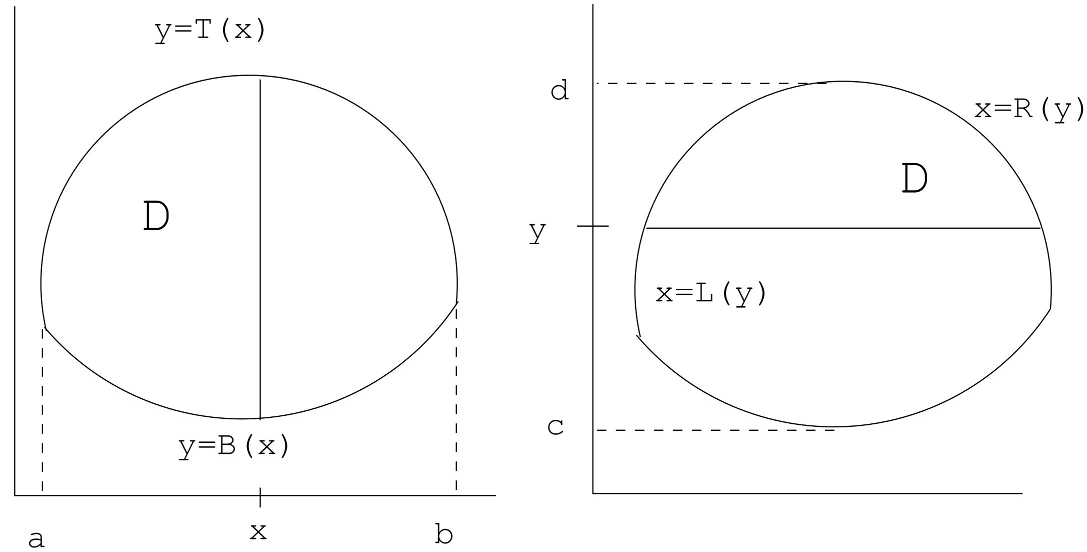
I will leave other justification for Green to the physicists.

Let me return to the theorem itself. First there appear to three versions. The second is just the first, with $Q_x - P_y$ disguised by

$$\text{curl}(\hat{\mathbf{F}}) \cdot \mathbf{k} = \text{curl}(\langle P, Q, 0 \rangle) \cdot \langle 0, 0, 1 \rangle$$

Third is just the first, where $-Q$ is in for $P$ and $P$ is in for $Q$, so that substituting into $(\quad)_x - (\quad)_y$ gives $P_x + Q_y = \text{div}(\mathbf{F})$.

Finally, let me illustrate why the theorem is true using a convex set, $D$. We can describe the top and bottom boundary as the graphs of functions, $T$ and $B$. these functions also parameterize the boundary of $D$. $(x, B(x))$ draws the bottom and $(x, T(x))$ the top. Note that as $x$ goes from left to right, $T$ draws the graph backwards from the correct orientation for $\partial D$, so a sign change will be needed. We could also describe the left and right sides using functions $L$ and $R$ and parameterize the boundary with $(L(y), y)$ which runs backwards and $(R(y), y)$.




So, we have

$$\begin{aligned}
-\int_D P_y(x,y)\,dA &= -\int_a^b \int_{B(x)}^{T(x)} P_y(x,y)\,dy\,dx \\
&= -\int_a^b (P(x,T(x)) - P(x,B(x)))\,dx \qquad \text{FTC!} \\
&= -\int_a^b P(x,T(x))\,dx + \int_a^b P(x,B(x))\,dx \\
&= \int_{\partial D} P\,dx
\end{aligned}$$

and

$$\begin{aligned}
\int_D Q_x(x,y)\,dA &= \int_c^d \int_{L(y)}^{R(y)} Q_x(x,y)\,dx\,dy \\
&= \int_c^d (Q(R(y),y) - Q(L(y),y))\,dy \qquad \text{FTC!} \\
&= \int_{\partial D} Q\,dy
\end{aligned}$$

combining the two gives Green's Theorem.

For non-convex regions, just chop them up into convex regions. When calculating the integrals for the pieces, you go across a cut in one direction on one piece and in the opposite direction on the other piece, so the contributions to the integrals cancel each other. Draw a picture.

This theorem is a Fundamental Theorem of Calculus because it says the integral of derivatives of **F** on a set is determined by the values of **F** on the boundary of the set, and because the one variable Fundamental Theorem is the key step in its proof.

## 8.3 Stokes' Theorem

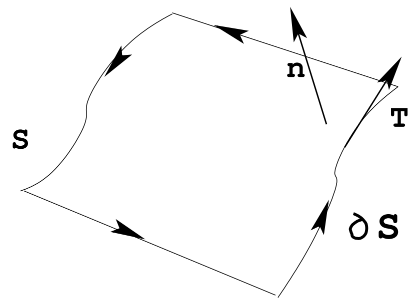
If you liked Green, you'll love Stokes.

**Stokes' Theorem**[2]**:** If $\mathbf{F}$ is a vector field on $\mathbb{R}^3$ whose components have continuous partial derivatives on a set containing a surface $S$, then

$$\int_S \nabla \times \mathbf{F} \cdot \mathbf{n}\, dS = \int_{\partial S} \mathbf{F} \cdot \mathbf{T}\, ds$$

Fundamental Theorem of Calculus philosophy is in place. The integral of derivatives of $\mathbf{F}$ on a set is determined entirely be the values of $\mathbf{F}$ on the boundary of the set.

What is $\partial S$? It is the geometrical boundary of $S$, a curve, traversed so that if you standing parallel to the normal vector to the surface the surface is on your left.




We can use this theorem on the antiderivative problem in $\mathbb{R}^3$ essentially the same way we used Green's Theorem in $\mathbb{R}^2$. Suppose $\mathbf{F}$ is a vector field with $\nabla \times \mathbf{F} = \mathbf{0}$. Suppose $C$ is a closed curve that $C$ or $-C$ is the boundary of a surface $S$. In general, it will happen, think of a wire loop dipped in soap. If $\mathbf{F}$ has continuous partial derivatives on and around the surface, then $\int_C \mathbf{F} \cdot \mathbf{T}\, ds = \pm \int_S \nabla \times \mathbf{F} \cdot \mathbf{n}\, dS = 0$.

A set $B$ in $\mathbb{R}^3$ where every closed curve in $B$ is the boundary of a surface in $B$ is *simply connected*. In $\mathbb{R}^3$ a simply connected set can have holes as long as they are bounded.

Again, we have

For a vector field $\mathbf{F}$ whose components have continuous partial derivatives on a simply connected set $B$ in $\mathbb{R}^3$,

$\mathbf{F}$ has antiderivatives on $B \iff \nabla \times \mathbf{F} = \mathbf{0}$ on $B$

Now, why is Stokes' Theorem true? It is not that difficult to believe. The second form of Green's Theorem is actually Stokes' Theorem with the set $D$ in $\mathbb{R}^2$, viewed as a surface in the $x$-$y$ plane in $\mathbb{R}^3$ with normal vector $\mathbf{k}$. To prove the theorem, you parameterize the surface and pull the problem into the parameter space in $\mathbb{R}^2$ and apply Green's Theorem. So, the FTC philosophy is embedded in the theorem in the form of Green's Theorem. Here's how it works.

Suppose, the set $D$ is the part of the domain of the parameterization $r(t,u) = (x,y,z)$ that draws the surface $S$. Parameterizing the boundary of $D$ will also parameterize the boundary of $S$.

[2] Sir George Gabriel Stokes, 1819-1903

So,

$$\begin{aligned}
\int_{\partial S} Pdx + Qdy + Rdz &= \int_{\partial D} P(x_t dt + x_u du) + Q(y_t dt + y_u du) + R(z_t dt + z_u du) \\
&= \int_{\partial D} (Px_t + Qy_t + Rz_t)dt + (Px_u + Qy_u + Rz_u)du \\
\text{(Green's Theorem)} &= \int_D ((Px_u + Qy_u + Rz_u)_t - (Px_t + Qy_t + Rz_t)_u)\, dA \\
&= \int_D (x_u P_t - x_t P_u + y_u Q_t - y_t Q_u + z_u R_t - z_t R_u)\, dA \\
\text{(Chain Rule)} &= \int_D ((P_x x_t + P_y y_t + P_z z_t)x_u - (P_x x_u + P_y y_u + P_z z_u)x_t + \\
&\quad (Q_x x_t + Q_y y_t + Q_z z_t)y_u - (Q_x x_u + Q_y y_u + Q_z z_u)y_t + \\
&\quad (R_x x_t + R_y y_t + R_z z_t)z_u - (R_x x_u + R_y y_u + R_z z_u)z_t)\, dA \\
&= \int_D (R_y - Q_z)(y_t z_u - y_u z_t) + (P_z - R_x)(x_u z_t - x_t z_u) + \\
&\quad (Q_x - P_y)(x_t y_u - x_u y_t)\, dA \\
&= \int_D \langle R_y - Q_z, P_z - R_x, Q_x - P_y \rangle \cdot \langle x_t, y_t, z_t \rangle \times \langle x_u, y_u, z_u \rangle\, dA \\
&= \int_S \nabla \times \mathbf{F} \cdot \mathbf{n}\, dS
\end{aligned}$$

So, Stokes' Theorem is done. The main reason for giving you the proof is to show how Stokes' Theorem, follows from Green's Theorem in two variables, which follows from the Fundamental Theorem of Calculus in one variable.

## 8.4 Gauss' Theorem

Also known as the Divergence Theorem.

**Gauss' Theorem[3]:** If $\mathbf{F}$ has continuous partial derivatives on a solid $B$, then

$$\int_B \operatorname{div}(\mathbf{F})\, dV = \int_{\partial B} \mathbf{F} \cdot \mathbf{n}\, dS$$

The $\partial B$ is the outer skin of the solid, oriented by the normal pointing away from the solid, the so-called outward pointing normal.

Besides propagating the Fundamental Theorem philosophy and being proved by it, which we will see in a bit, this theorem does things like relate the flow of a fluid across the boundary of a region to the divergence of the flow in the region.

[3]Karl Friedrich Gauss, 1777-1855

Mathematically speaking, Gauss' Theorem says things like, if $\mathbf{F} = \text{curl}(\mathbf{G})$, then the integral of $\mathbf{F}$ over a closed surface (no boundary, like the sphere) is zero.

It is also the $\mathbb{R}^3$ version of the third form of Green's Theorem.

To see why the theorem is true, suppose that the region $B$ is convex, then the top and bottom can be described explicitly by some functions, $z = T(x, y)$ and $z = B(x, y)$. Use these functions to define parameterizations, $(x, y, T(x, y))$ whose normal vector is chosen to be $\langle -T_x, -T_y, 1\rangle$ which is outward, and $(x, y, B(x, y))$ whose normal vector is $\langle -B_x, -B_y, 1\rangle$, which points inward, so change the sign.

Let me call $D = \{(x, y) : (x, y, z) \text{ is in } B \text{ for some } z\}$. We have

$$\begin{aligned}\int_B R_z\,dV &= \int_D \int_{B(x,y)}^{T(x,y)} R_z\,dz\,dA \\ &= \int_D (R(x,y,T(x,y)) - R(x,y,B(x,y)))\,dA \qquad \text{FTC} \\ &= \int_{\partial B} < 0, 0, R > \cdot \mathbf{n}\,dS\end{aligned}$$

Similarly, $\int_B Q_y\,dV = \int_{\partial B}\langle 0, Q, 0\rangle \cdot \mathbf{n}\,dS$ and $\int_B P_x\,dV = \int_{\partial B}\langle P, 0, 0\rangle \cdot \mathbf{n}\,dS$, so that adding them up finishes the job.

## 8.5 All of them

We have the following versions of the Fundamental Theorem of Calculus.

| | Integrating over a set of 1 dimension | 2 dimensions | 3 dimensions |
|---|---|---|---|
| $\mathbb{R}$ | $\int_a^b f'(x)\,dx = f(b) - f(a)$ | | |
| $\mathbb{R}^2$ | $\int_C \nabla f \cdot \mathbf{T}\,ds = f(B) - f(A)$ | $\int_D (Q_x - P_y)\,dA = \int_{\partial D} P dx + Q dy$ | |
| | | $\int_D \nabla \times \hat{\mathbf{F}} \cdot \mathbf{k}\,dA = \int_{\partial D} \mathbf{F} \cdot \mathbf{T}\,dS$ | |
| | | $\int_D \nabla \cdot \mathbf{F}\,dA = \int_{\partial D} \mathbf{F} \cdot \mathbf{n}\,ds$ | |
| $\mathbb{R}^3$ | $\int_C \nabla f \cdot \mathbf{T}\,ds = f(B) - f(A)$ | $\int_S \nabla \times \mathbf{F} \cdot \mathbf{n}\,dS = \int_{\partial S} \mathbf{F} \cdot \mathbf{T}\,ds$ | $\int_B \nabla \cdot \mathbf{F}\,dV = \int_{\partial B} \mathbf{F} \cdot \mathbf{n}\,dS$ |