Multivariable Calculus

George Cain & James Herod

Table of Contents

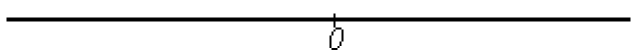
George Cain & James Herod
School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia 30332-0160

Chapter One

Euclidean Three-Space

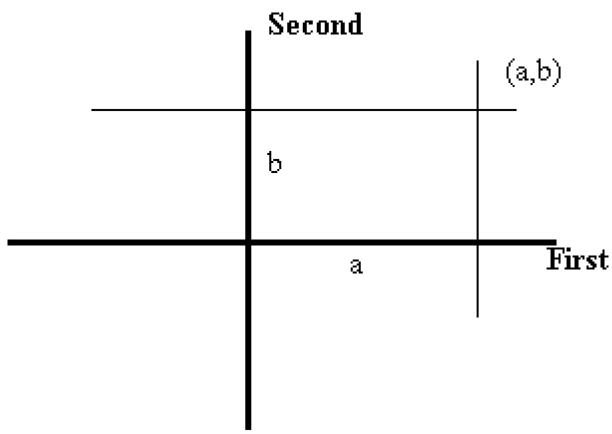
1.1 Introduction.

Let us briefly review the way in which we established a correspondence between the real numbers and the points on a line, and between ordered pairs of real numbers and the points in a plane. First, the line. We choose a point on a line and call it the ***origin***. We choose one direction from the origin and call it the ***positive*** direction. The opposite direction, not surprisingly, is called the ***negative*** direction. In a picture, we generally indicate the positive direction with an arrow or a plus sign:



Now we associate with each real number r a point on the line. First choose some unit of measurement on the line. For $r > 0$, associate with r the point on the line that is a distance r units from the origin in the positive direction. For $r < 0$, associate with r the point on the line that is a distance r units from the origin in the negative direction. The number 0 is associated with the origin. A moments reflection should convince you that this procedure establishes a so-called ***one-to-one correspondence*** between the real numbers and the points on a line. In other words, a real number determines exactly one point on a line, and, conversely, a point on the line determines exactly one real number. This line is called a ***real line***.

Next we establish a one-to-one correspondence between ordered pairs of real numbers and points in a plane. Take a real line, called the ***first axis***, and construct another real line, called the ***second axis***, perpendicular to it and passing through the origin of the first axis. Choose this point as the origin for the second axis. Now suppose we have an ordered pair (x_1, x_2) of reals. The point in the plane associated with this ordered pair is found by constructing a line parallel to the second axis through the point on the first axis corresponding to the real number x_1, and constructing a line parallel to the first axis through the point on the second axis corresponding to the real number x_2. The point at which these two lines intersect is the point associated with the ordered pair (x_1, x_2). A moments reflection here will convince you that there is exactly one point in the plane thus associated with an ordered pair (a, b), and each point in the plane is the point associated with some ordered pair (a, b):

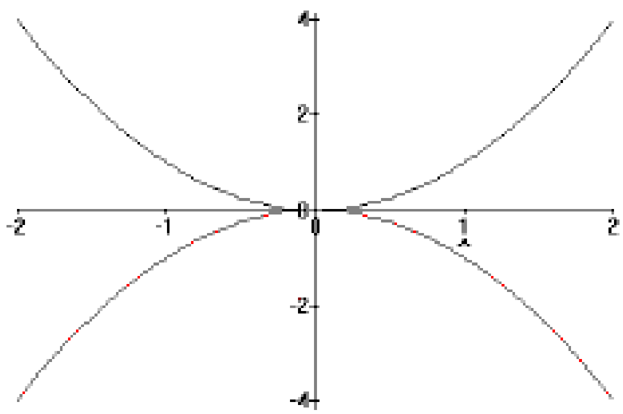


It is traditional to assume the point of view we have taken in this picture, in which the first axis is horizontal, the second axis is vertical, the positive direction on the first axis is to the right, and the positive direction on the second axis is up. We thus usually speak of the ***horizontal*** axis and the ***vertical*** axis, rather than the first axis and the second axis. We also frequently abuse the language by speaking of a *point* (x_1, x_2) when, of course, we actually mean the point associated with the ordered pair (x_1, x_2). The numbers x_1 and x_2 are called the ***coordinates*** of the point- x_1 is the first coordinate and x_2 is the second coordinate.

Given any collection of ordered pairs(A collection of ordered pairs is called a ***relation***.), a picture of the collection is obtained by simply looking at the set of points in the plane corresponding to the pairs in the given collection. Suppose we have an equation involving two variables, say x and y. Then this equation defines a collection of ordered pairs of numbers, namely all (x,y) that satisfy the equation. The corresponding picture in the plane is called the ***graph*** of the equation. For example, consider the equation $y^2 = x^4$. Let's take a look at the graph of this equation. A little algebra (very little, actually), convinces us that

$$\{(x,y): y^2 = x^4\} = \{(x,y): y = x^2\} \quad \{(x,y): y = -x^2\},$$

and we remember from the sixth grade that each of the sets on the right hand side of this equation is a parabola:



What do we do with all this? These constructions are, of course, the bases of analytic geometry, in which we join the subjects of algebra and geometry, to the benefit of both. A geometric figure (a subset of the plane) corresponds to a collection of ordered pairs of real numbers. Algebraic facts about the collection of ordered pairs of real are reflected by geometric facts about the subset of the plane, and, conversely, geometric facts about the plane subset are reflected by algebraic facts about the collection of pairs of reals.

Exercises

Draw a picture of the given relation:

1. $R = \{(x, y): 0 \le x \le 1, \text{ and } 1 \le y \le 4\}$

2. $R = \{(x, y): -4 \le x \le 4, \text{ and } -x^2 \le y \le x^2\}$

3. $R = \{(x, y): 1 \le y \le 2\} \cap \{(x, y): y \ge x^2\}$

4. $S = \{(x, y): x^2 + y^2 = 1, \text{ and } x \ge 0\}$

5. $S = \{(x, y): x^2 + y^2 \le 1\} \cap \{(x, y): y \le x^2\}$

6. $E = \{(r, s): |r| = |s|\}$

7. $T = \{(u, v): |u| + |v| = 1\}$

8. $R = \{(u, v): |u| + |v| \le 1\}$

9. $T = \{(x, y): x^2 = y^2\}$

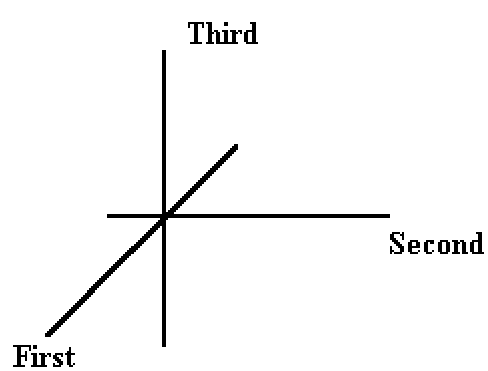
10. $A = \{(x, y): x^2 \le y^2\}$

11. $G = \{(s,t) : \max\{|s|,|t|\} = 1\}$

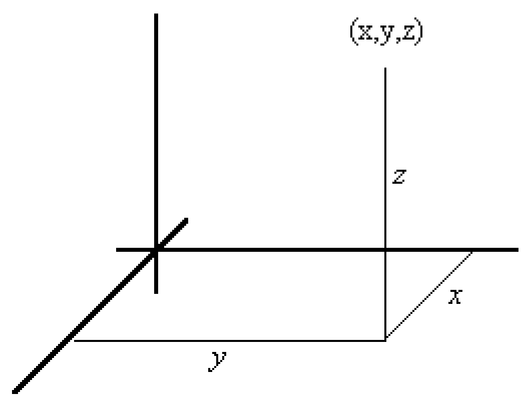
12. $B = \{(s,t) : \max\{|s|,|t|\} \le 1\}$

1.2 Coordinates in Three-Space

Now let's see what's doing in three dimensions. We shall associate with each ordered triple of real numbers a point in three space. We continue from where we left off in the previous section. Start with the plane constructed in the previous section, and construct a line perpendicular to both the first and second axes, and passing through the origin. This is the ***third***, axis. Now we must be careful about which direction on this third axis is chosen as the positive direction; it makes a difference. The positive direction is chosen to be the direction in which a right-hand threaded bolt would advance if the positive first axis is rotated to the positive second axis:



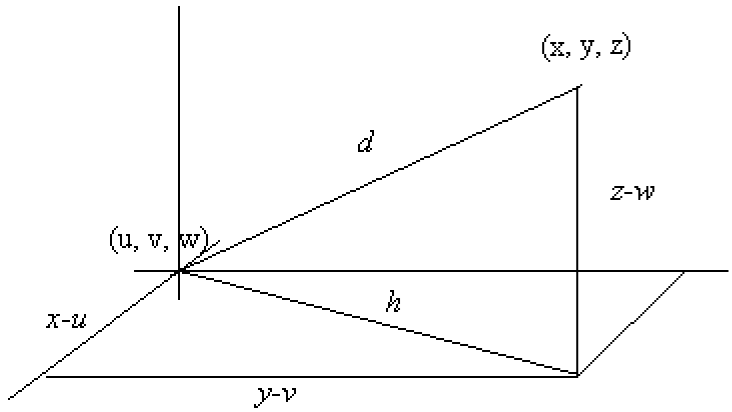
We now see how to define a one-to-one correspondence between ordered triples of real numbers (x_1, x_2, x_3) and the points in space. The association is a simple extension of the way in which we established a correspondence between ordered pair and points in a plane. Here's what we do. Construct a plane perpendicular to the first axis through the point x_1, a plane perpendicular to the second axis through x_2, and a plane perpendicular to the third axis through x_3. The point at which these three planes intersect is the point associated with the ordered triple (x_1, x_2, x_3). Some meditation on this construction should convince you that this procedure establishes a one-to-one correspondence between ordered triples of reals and points in space. As in the two dimensional, or plane, case, x_1 is called the first coordinate of the point, x_2 is called the second coordinate of the point, and x_3 is called the third coordinate of the point. Again, the point corresponding to $(0,0,0)$ is called the ***origin***, and we speak of the *point* (x_1, x_2, x_3), when we actually mean the point which corresponds to this ordered triple.



The three axes so defined is called a ***coordinate system*** for three space, and the three numbers x, y, and z, where (x, y, z) is the triple corresponding to the point P, are called the ***coordinates*** of P. The coordinate axes are sometimes given labels-most commonly, perhaps, the first axis is called the ***x*** axis, the second axis is called the ***y*** axis, and the third axis is called the ***z*** axis.

1.3 Some Geometry

Suppose P and Q are two points, and suppose space is endowed with a coordinate system such that $P=(x,y,z)$ and $Q=(u,v,w)$. How do we find the distance between P and Q? This simple enough; look at the picture:



We can see that $d^2 = h^2 + (z-w)^2$ and $h^2 = (x-u)^2 + (y-v)^2$. Thus we have

$$d^2 = (x-u)^2 + (y-v)^2 + (z-w)^2, \text{ or}$$

$$d = \sqrt{(x-u)^2 + (y-v)^2 + (z-w)^2}.$$

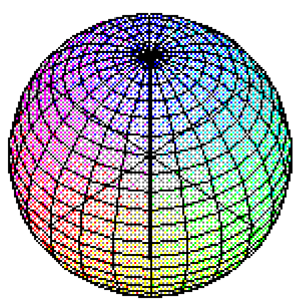
We saw that in the plane an equation in two variables defines in a natural way a collection of ordered pairs of numbers. The analogous situation obtains in three-space: an equation in three variables defines a collection of ordered triples. We thus speak of the collection of triples (x,y,z) which satisfy the equation

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

The collection of all such points is the ***graph*** of the equation. In this example, it is easy to see that the graph is precisely the set of all points at a distance of 1 from the origin-a sphere of radius 1 and center at the origin.

The graph of the equation $x = 0$ is simply the set of all points with first coordinate 0, and this is clearly the plane determined by the second axis and the third axis, or the *y* axis and the *z* axis. When the axes are labeled *x*, *y*, and *z*, this is known as ***the yz plane.*** . Similarly, the plane $y = 0$ is the ***xz plane***, and $z = 0$ is the ***xy plane***. These special planes are also called the ***coordinate planes.***

More often than not, it is difficult to see exactly what a graph of equation looks like, and even more difficult for most of us to draw it. Computers can help, but they usually draw rather poor pictures whose main application is in stimulating your own imagination sufficiently to allow you to see the graph in your mind's eye. An example:



This picture was drawn using *Maple.*

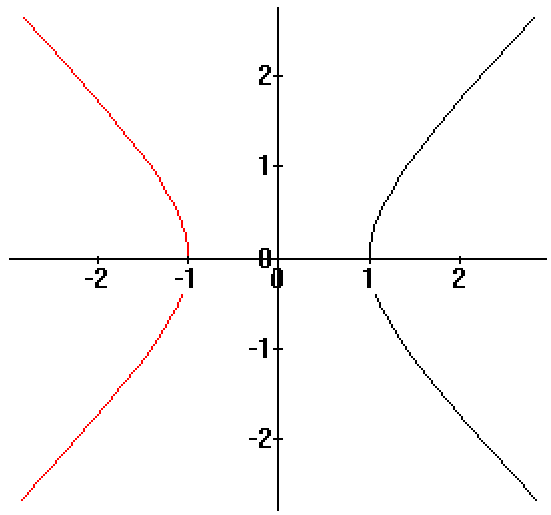
Let's look at a more complicated example. What does the graph of

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

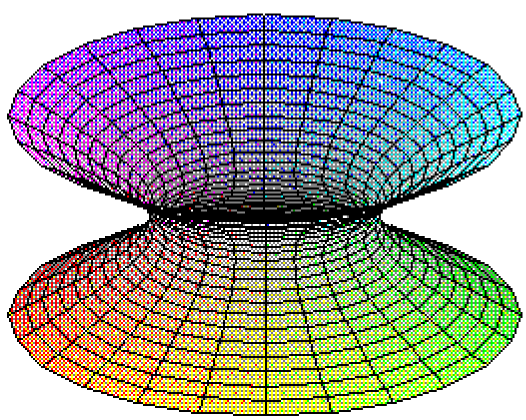
look like? We'll go after a picture of this one by slicing the graph with the coordinate planes. First, let's slice through it with the plane $z = 0$; then we see

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

a circle of radius 1 centered at the origin. Next, let's slice with the plane $y = 0$. Here we see $x^2 - z^2 = 1$, a hyperbola:



We, of course, see the same hyperbola when we slice the graph with the plane $x = 0$. What the graph looks like should be fairly clear by now:



This graph has a name; it is called a ***hyperboloid***.

Exercises

13. Describe the set of points $S = \{(x, y, z): x \geq 0,\ y \geq 0, \text{ and } z \geq 0\}$.

14. Describe the following sets

a) $S = \{(x, y, z):\ z \geq 0\}$

b) $S = \{(x, y, z):\ x \geq 5\}$

c) $R = \{(x, y, z):\ x^2 + y^2 \leq 1\}$

d) $T = \{(r, s, t):\ r^2 + s^2 + t^2 \leq 4\}$

15. Let G be the graph of the equation $x^2 + 4y^2 + 9z^2 = 36$.

a)Sketch the graphs of the curves sliced from G by the coordinate planes $x = 0$, $y = 0$, and $z = 0$.

b)Sketch G. (This graph is called an ***ellipsoid.***)

16. Let G be the graph of the equation $x^2 - 3y^2 + 4z^2 = 12$.

a)Sketch the graphs of the curves sliced from G by the coordinate planes $x = 0$, $y = 0$, and $z = 0$.

b)Sketch G. (Does this set look at all familiar to you?.)

1.4 Some More Geometry-Level Sets

The curves that result from slicing the graphs with the coordinate planes are special cases of what are called level sets of a set. Specifically, if S is a set, the intersection of S with a plane $z =$ constant is called a ***level set***. In case the level set is a curve, it is frequently called a ***level curve***. (The slices by planes x = constant, or y = constant are also level sets.) A family of level sets can provide a nice stimulant to your powers of visualization. Everyday examples of the use of level sets to describe a set are contour maps, in which the contours are, of course, just level curves ; and weather maps, in which, for instance, the isoclines on a 500mb chart are simply level curves for the 500mb surface. Let's illustrate with an example.

Let S be the graph of

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

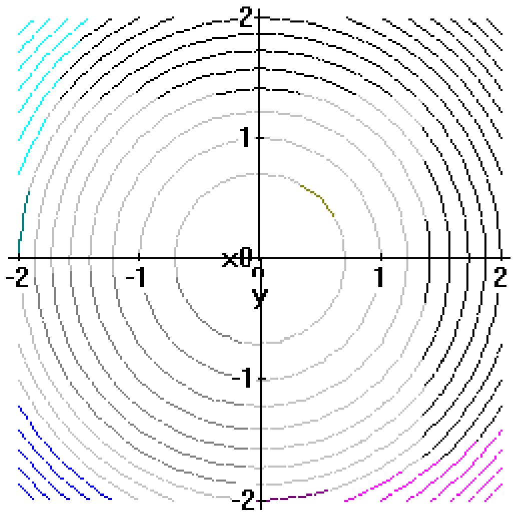
Now we look at the level set $z = c$:

$$c^2 - y^2 - x^2 = 1, \text{ or}$$

$$x^2 + y^2 = c^2 - 1.$$

Notice first that we have the same curve for $z = c$ and $z = -c$. The graph is symmetric about the plane $z = 0$. We shall thus look at just that part of the graph that is above the xy plane.

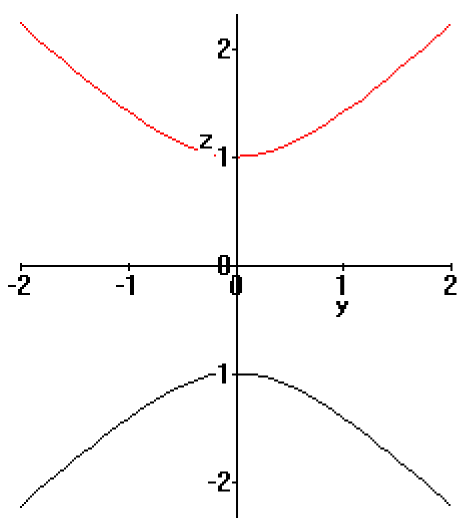
It is clear that these curves are concentric circles of radius $\sqrt{c^2 - 1}$ centered at the origin. There are no level sets for $|c| < 1$, and for $c = 1$ or -1, the level set is a single point, the origin.



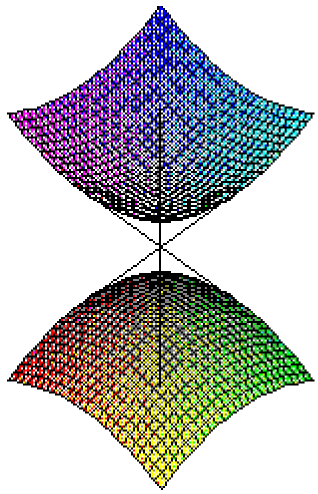
Next, slice with the planes $x = 0$ and $y = 0$ to get a better idea of what this thing looks like. For $x = 0$, we see

$$z^2 - y^2 = 1,$$

a hyperbola:



The slice by $y = 0$, of course, is the same. It is rather easy to visualize this graph. Here is a *Maple* drawn picture:



This also is called a hyperboloid. This is a hyperboloid of ***two sheets***, while the previously described hyperboloid is a hyperboloid of ***one sheet***.

Exercises

17. Let $S = \{(x_1, x_2, x_3) : |x_1| + |x_2| + |x_3| = 1\}$.

a)Sketch the coordinate plane slices of S.

b)Sketch the set S.

18. Let C be the graph of the equation $z^2 = 4(x^2 + y^2)$.

a)Sketch some level sets $z = c$.

b)Sketch the slices by the planes $x = 0$ and $y = 0$.

c)Sketch C. What does the man on the street call this set?

19. Using level sets, coordinate plane slices, and whatever, describe the graph of the equation $z = x^2 + y^2$. (This one has a name also; it is a ***paraboloid***.).

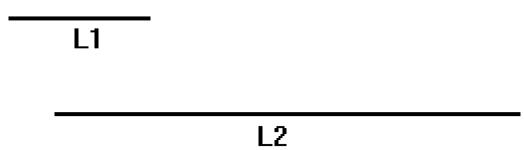
20. Using level sets, coordinate plane slices, and whatever, describe the graph of the equation $z = x^2 - y^2$.

Chapter Two

Vectors-Algebra and Geometry

2.1 Vectors

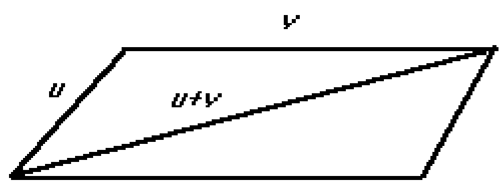
A ***directed line segment*** in space is a line segment together with a direction. Thus the directed line segment from the point ***P*** to the point ***Q*** is different from the directed line segment from ***Q*** to ***P***. We frequently denote the direction of a segment by drawing an arrow head on it pointing in its direction and thus think of a directed segment as a spear. We say that two segments have the same direction if they are parallel and their directions are the same:



Here the segments L1 and L2 have the same direction. We define two directed segments L and M to be ***equivalent*** (L M) if they have the same direction and have the same length. An ***equivalence class*** containing a segment L is the set of all directed segments equivalent with L. Convince yourself every segment in an equivalence class is equivalent with every other segment in that class, and two different equivalence classes must be disjoint. These equivalence classes of directed line segments are called ***vectors***. The members of a vector ***v*** are called ***representatives*** of ***v***. Given a directed segment u, the vector which contains u is called the vector ***determined*** by u. ***The length***, or ***magnitude***, of a vector ***v*** is defined to be the common length of the representatives of ***v***. It is generally designated by |***v***|. The ***angle*** between two vectors ***u*** and ***v*** is simply the angle between the directions of representatives of ***u*** and ***v.***

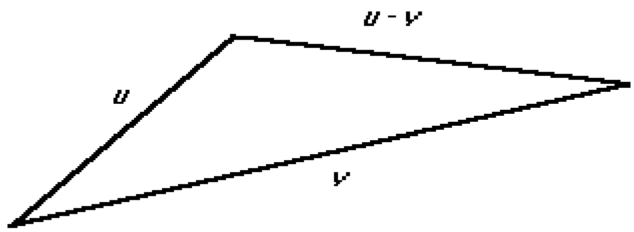
Vectors are just the right mathematical objects to describe certain concepts in physics. Velocity provides a ready example. Saying the car is traveling 50 miles/hour doesn't tell the whole story; you must specify in what direction the car is moving. Thus velocity is a vector-it has both magnitude and direction. Such physical concepts abound: force, displacement, acceleration, *etc.* The real numbers (or sometimes, the complex numbers) are frequently called ***scalars*** in order to distinguish them from vectors.

We now introduce an arithmetic, or algebra, of vectors. First, we define what we mean by the sum of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. Choose a spear u from $\boldsymbol{u}$ and a spear v from $\boldsymbol{v}$. Place the tail of v at the nose of u. The vector which contains the directed segment from the tail of u to the nose of v is defined to be $\boldsymbol{u}+\boldsymbol{v}$, the sum of $\boldsymbol{u}$ and $\boldsymbol{v}$. An easy consequence of elementary geometry is the fact that $|\boldsymbol{u} + \boldsymbol{v}| \leq |\boldsymbol{u}| + |\boldsymbol{v}|$. Look at the picture and convince yourself that the it does not matter which $\boldsymbol{u}$ spear or $\boldsymbol{v}$ spear you choose, and that $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$:



Convince yourself also that addition is associative: $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$. Since it does not matter where the parentheses occur, it is traditional to omit them and write simply $\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w}$.

Subtraction is defined as the inverse operation of addition. Thus the difference $\boldsymbol{u}$-$\boldsymbol{v}$ of two vectors is defined to be the vector you add to $\boldsymbol{v}$ to get $\boldsymbol{u}$. In pictures, if u is a representative of $\boldsymbol{u}$ and v is a representative of $\boldsymbol{v}$, and we put the tails of u and v together, the directed segment from the nose of v to the nose of u is a representative of $\boldsymbol{u} - \boldsymbol{v}$:



Now, what are we to make of ***u*** - ***u*** ? We define a special vector with 0 length, called the zero vector and denoted ***0.*** We may think of ***0*** as the collection of all degenerate line segments, or points. Note that the zero vector is special in that it has no direction (If you are going 0 miles/hour, the direction is not important!). To make our algebra of vectors nice , we make the zero vector behave as it should:

$$\boldsymbol{u} - \boldsymbol{u} = \boldsymbol{0} \text{ and } \boldsymbol{u} + \boldsymbol{0} = \boldsymbol{u}$$

for all vectors ***u***.

Next we define the product of a scalar r (*i.e.*, real number) with a vector ***u***. The product $r\boldsymbol{u}$ is defined to be the vector with length $|r||\boldsymbol{u}|$ and direction the same as the direction of ***u*** if $r > 0$, and direction opposite the direction of ***u*** if $r < 0$. Convince yourself that all the following nice properties of this multiplication hold:

$(r + s)\boldsymbol{u} = r\boldsymbol{u} + s\boldsymbol{u}$,

$r(\boldsymbol{u} + \boldsymbol{v}) = r\boldsymbol{u} + r\boldsymbol{v}$.

$0\boldsymbol{u} = \boldsymbol{0}$, and

$\boldsymbol{u} + (-1)\boldsymbol{v} = \boldsymbol{u} - \boldsymbol{v}$.

It is then perfectly safe to write $-\boldsymbol{u}$ to stand for $(-1)\boldsymbol{u}$.

Our next move is to define a one-to-one correspondence between vectors and points in space (This will, of course, also establish a one-to-one correspondence between vectors and ordered triples of real numbers.). The correspondence is quite easy; simply take a representative of the vector ***u*** and place its tail at the origin. The point at which is found the nose of this representative is the point associated with ***u.*** We handle the vector with no representatives by associating the origin with the zero vector. The fact that the point with coordinates (a, b, c) is associated with the vector ***u*** in this manner is shorthandedly indicated by writing ***u*** = (a, b, c). Strictly speaking this equation makes no sense; an equivalence class of directed line segments cannot possible be the same as a triple of real numbers, but this shorthand is usually clear and saves a lot of verbiage (The numbers a, b, and c are called the ***coordinates***, or ***components***, of ***u***.). Thus we frequently do not distinguish between points and vectors and indiscriminately speak of a *vector* (a,b,c) or of a point ***u.***

Suppose ***u*** = (a, b, c) and ***v*** = (x, y, z). Unleash your vast knowledge of elementary geometry and convince yourself of the truth of the following statements:

$|\boldsymbol{u}| = \sqrt{a^2 + b^2 + c^2}$,

$\boldsymbol{u} + \boldsymbol{v} = (a + x, b + y, c + d)$,

$\boldsymbol{u} - \boldsymbol{v} = (a - x, b - y, c - d)$, and

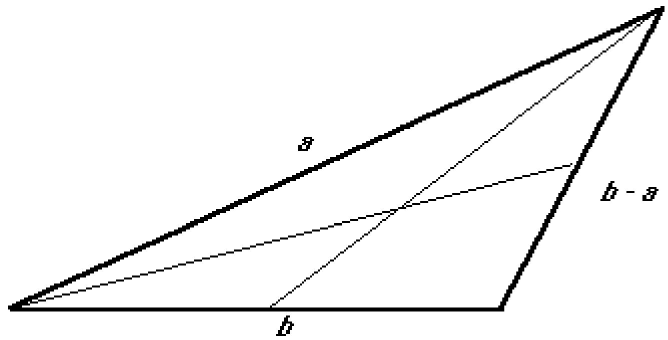
$r\boldsymbol{u} = (ra, rb, rc)$.

Let ***i*** be the vector corresponding to the point (1, 0, 0); let ***j*** be the vector corresponding to (0, 1, 0); and let ***k*** be the vector corresponding to (0, 0, 1). Any vector ***u*** can now be expressed as a linear combination of these special so-called ***coordinate*** vectors:

$$\boldsymbol{u} = (x, y, z) = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k} .$$

Example

Let's use our new-found knowledge of vectors to find where the medians of a triangle intersect. Look at the picture:



We shall find scalars s and t so that

$$\boldsymbol{a} + t\left(\frac{\boldsymbol{b}}{2} - \boldsymbol{a}\right) = s\left(\boldsymbol{a} + \frac{\boldsymbol{b} - \boldsymbol{a}}{2}\right).$$

Tidying this up gives us

$$\left(1 - t - \frac{s}{2}\right)\boldsymbol{a} = \left(\frac{s}{2} - \frac{t}{2}\right)\boldsymbol{b}.$$

This means that we must have

$$\frac{s}{2} - \frac{t}{2} = 0, \text{ and}$$

$$1 - t - \frac{s}{2} = 0.$$

Otherwise, $\boldsymbol{a}$ and $\boldsymbol{b}$ would be nonzero scalar multiples of one another, which would mean they have the same direction. It follows that

$$s = t = \frac{2}{3}.$$

This is, no doubt, the result you remember from Mrs. Turner's high school geometry class.

Exercises

1. Find the vector such that if its tail is at the point (x_1, y_1, z_1) its nose will be at the point (x_2, y_2, z_2).

2. Find the midpoint of the line segment joining the points (1, 5, 9) and (-3, 2, 3).

3. What is the distance between the points (x_1, y_1, z_1) and (x_2, y_2, z_2)?

4. Describe the set of points $L = \{t\boldsymbol{i}\colon\ -\infty < t < \infty\}$.

5. Let ***u*** = (2,3,8). Describe the set of points $L = \{t\boldsymbol{u}\colon\ -\infty < t < \infty\}$.

6. Describe the set of points $M = \{3\boldsymbol{k} + t\boldsymbol{i}\colon\ -\infty < t < \infty\}$.

7. Let ***u*** = (2,3,8).and ***v*** = (1,5,7). Describe the set of points $M = \{\boldsymbol{v} + t\boldsymbol{u}\colon\ -\infty < t < \infty\}$.

8. Describe the set $P = \{t\boldsymbol{i} + s\boldsymbol{j}\colon\ -\infty < t < \infty, \text{ and } -\infty < s < \infty\}$.

9. Describe the set $P = \{5\boldsymbol{k} + t\boldsymbol{i} + s\boldsymbol{j}\colon\ -\infty < t < \infty, \text{ and } -\infty < s < \infty\}$.

10. Let ***u*** = (2,-4,1) and ***v*** = (1,2,3). Describe the set

$$P = \{t\boldsymbol{u} + s\boldsymbol{v}\colon\ -\infty < t < \infty, \text{ and } -\infty < s < \infty\}.$$

11. Let ***u*** = (2,-4,1), ***v*** = (1,2,3), and ***w*** = (3,6,1). Describe the set

$$P = \{\boldsymbol{w} + t\boldsymbol{u} + s\boldsymbol{v}\colon\ -\infty < t < \infty, \text{ and } -\infty < s < \infty\}.$$

12. Describe the set $C = \{\cos t\,\boldsymbol{i} + \sin t\,\boldsymbol{j}\colon\ 0 \le t \le 2\pi\}$.

13. Describe the set $E = \{4\cos t\,\boldsymbol{i} + 3\sin t\,\boldsymbol{j}\text{: } 0 \le t \le 2\pi\}$.

14. Describe the set $P = \{t\boldsymbol{i} + t^2\boldsymbol{j}\text{: } -1 \le t \le 2\}$ **.**

15. Let **T** be the triangle with vertices (2, 5, 7), (-1, 2, 4), and (4, -2, -6). Find the point at which the medians intersect.

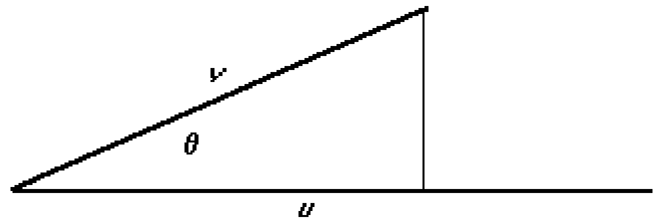
2.2 Scalar Product

You were perhaps puzzled when in grammar school you were first told that the work done by a force is the product of the force and the displacement since both force and displacement are, of course, vectors. We now introduce this product. It is a scalar and hence is called the ***scalar product.*** This scalar product $\boldsymbol{u}\ \boldsymbol{v}$ is defined by

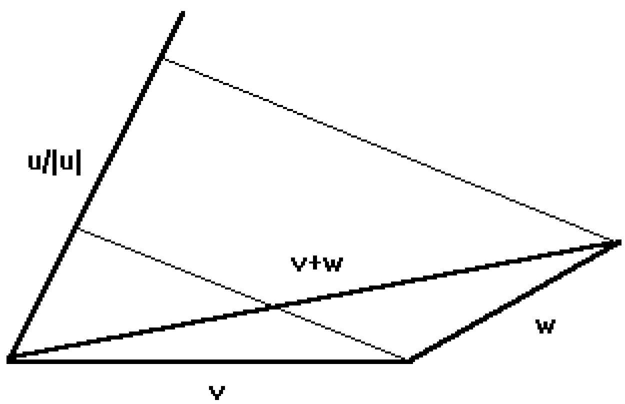
$$\boldsymbol{u}\ \boldsymbol{v} = |\boldsymbol{u}||\boldsymbol{v}|\cos\theta\ ,$$

where θ is the angle between ***u*** and ***v.*** The scalar product is frequently also called the ***dot product.*** Observe that $\boldsymbol{u}\ \boldsymbol{u} = |\boldsymbol{u}|^2$, and that $\boldsymbol{u}\ \boldsymbol{v} = 0$ if and only if ***u*** and ***v*** are perpendicular (or ***orthogonal***), or one or the other of the two is the zero vector. We avoid having to use the latter weasel words by defining the zero vector to be perpendicular to every vector; then we can say $\boldsymbol{u}\ \boldsymbol{v} = 0$ if and only if ***u*** and ***v*** are perpendicular.

Study the following picture to see that if |***u***| = 1, then $\boldsymbol{u}\ \boldsymbol{v}$ is the length of the projection of ***v*** onto ***u***. (More precisely, the length of the projection of a representative of ***v*** onto a representative of ***u***. Generally, where there is no danger of confusion, we omit mention of this, just as we speak of the length of vectors, the angle between vectors, *etc.*)



It is clear that $(a\boldsymbol{u}) \cdot (b\boldsymbol{v}) = (ab)\boldsymbol{u} \cdot \boldsymbol{v}$. Study the following picture until you believe that $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$ for any three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$.



Now let's get a recipe for the scalar product of $\boldsymbol{u} = (a, b, c)$ and $\boldsymbol{v} = (x, y, z)$:

$$\begin{aligned}\boldsymbol{u} \cdot \boldsymbol{v} &= (a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}) \cdot (x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}) \\ &= ax\boldsymbol{i} \cdot \boldsymbol{i} + ay\boldsymbol{i} \cdot \boldsymbol{j} + az\boldsymbol{i} \cdot \boldsymbol{k} + bx\boldsymbol{j} \cdot \boldsymbol{i} + by\boldsymbol{j} \cdot \boldsymbol{j} + bz\boldsymbol{j} \cdot \boldsymbol{k} + cx\boldsymbol{k} \cdot \boldsymbol{i} + cy\boldsymbol{k} \cdot \boldsymbol{j} + cz\boldsymbol{k} \cdot \boldsymbol{k} \\ &= ax + by + cz,\end{aligned}$$

since $\boldsymbol{i} \cdot \boldsymbol{i} = \boldsymbol{j} \cdot \boldsymbol{j} = \boldsymbol{k} \cdot \boldsymbol{k} = 1$ and $\boldsymbol{i} \cdot \boldsymbol{j} = \boldsymbol{i} \cdot \boldsymbol{k} = \boldsymbol{j} \cdot \boldsymbol{k} = 0$.

We thus see that it is remarkably simple to compute the scalar product of two vectors when we know their coordinates.

Example

Again, let's see how vectors can make geometry easy by using them to find the angle between a diagonal of a cube and the diagonal of a face of the cube.

Suppose the cube has edge length s. Introduce a coordinate system so that the faces are parallel to the coordinate planes, one vertex is the origin and the vertex at the other end of the diagonal from the origin is (s, s, s). The vector determined by this diagonal is thus $\mathbf{d} = s\mathbf{i} + s\mathbf{j} + s\mathbf{k}$ and the vector determined by the diagonal of the face in the horizontal coordinate plane is $\boldsymbol{f} = s\boldsymbol{i} + s\boldsymbol{j}$. Thus

$$\boldsymbol{d} \quad \boldsymbol{f} = |\boldsymbol{d}||\boldsymbol{f}|\cos\theta = s^2 + s^2,$$

where θ is the angle we seek. This gives us

$$\cos\theta = \frac{2s^2}{|\boldsymbol{d}||\boldsymbol{f}|} = \frac{2s^2}{\sqrt{3s^2}\sqrt{2s^2}} = \sqrt{\frac{2}{3}}.$$

Or,

$$\theta = Cos^{-1}\left(\sqrt{\frac{2}{3}}\right).$$

Exercises

16. Find the work done by the force $\boldsymbol{F} = 6\boldsymbol{i} - 3\boldsymbol{j} + 2\boldsymbol{k}$ in moving an object from the point (1, 4, -2) to the point (3, 2, 5).

17. Let ***L*** be the line passing through the origin and the point (2, 5), and let ***M*** be the line passing through the points (3, -2) and (5, 3). Find the smaller angle between ***L*** and ***M***.

18. Find an angle between the lines $3x + 2y = 1$ and $x - 2y = 3$.

19. Suppose ***L*** is the line passing through (1, 2) having slope -2, and suppose ***M*** is the line tangent to the curve $y = x^3$ at the point (1, 1).). Find the smaller angle between ***L*** and ***M***.

20. Find an angle between the diagonal and an adjoining edge of a cube.

21. Suppose the lengths of the sides of a triangle are *a*, *b*, and *c*; and suppose γ is the angle opposite the side having length *c*. Prove that

$$c^2 = a^2 + b^2 - 2ab\cos\gamma\ .$$

(This is, of course, the celebrated *Law of Cosines*.)

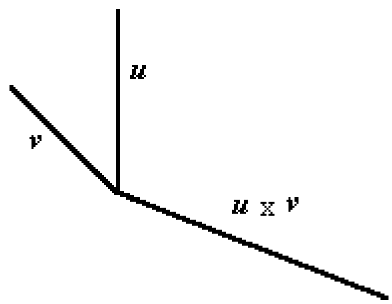
22. Let ***v*** = (1, 2, 5). . What is the graph of the equation $\boldsymbol{v}\ \ (x, y, z) = 0$?

2.3 Vector Product

Hark back to grammar school physics once again and recall what you were taught about the velocity of a point at a distance *r* from the axis of rotation; you were likely told that the velocity is $r\omega$, where ω is the rate at which the turntable is rotating-the so-called angular velocity. We now know that these quantities are actually vectors-ω is the angular *velocity*, and ***r*** is the *position* vector of the point in question. The grammar school quantities are the magnitudes of ω (the angular *speed*) and of ***r***. The *velocity* of the point is the so-called *vector product* of these two vectors. The ***vector product*** of vectors ***u*** and ***v*** is defined by

$$\boldsymbol{u} \times \boldsymbol{v} = |\boldsymbol{u}||\boldsymbol{v}|\sin\theta\ \ |\ \ \boldsymbol{n}\ ,$$

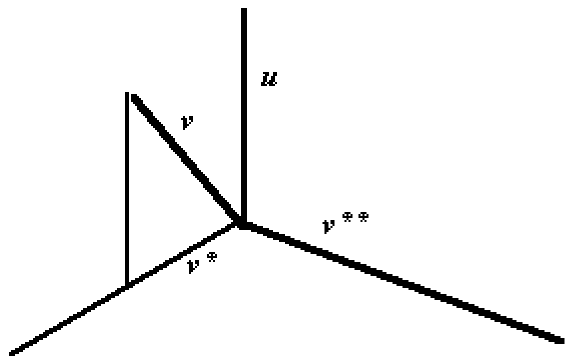
where θ is the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ and $\boldsymbol{n}$ is a vector of length 1 (such vectors are called ***unit vectors***) which is orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{v}$ and which points in the direction a right-hand threaded bolt would advance if $\boldsymbol{u}$ were rotated into the direction of $\boldsymbol{v}$.



Note first that this is a somewhat more exciting product than you might be used to:

the *order* of the factors makes a difference. Thus $\boldsymbol{u} \times \boldsymbol{v} = -\boldsymbol{v} \times \boldsymbol{u}$.

Now let's find a geometric construction of the vector product $\boldsymbol{u} \times \boldsymbol{v}$. Proceed as follows. Let *P* be a plane perpendicular to $\boldsymbol{u}$. Now project $\boldsymbol{v}$ onto this plane, giving us a vector $\boldsymbol{v}^*$ perpendicular to $\boldsymbol{u}$ and having length $|\boldsymbol{v}||\sin\theta\,|$. Now rotate this vector $\boldsymbol{v}^*$ 90 degrees around $\boldsymbol{u}$ in the "positive direction." (By the positive direction of rotation about a vector $\boldsymbol{a}$, we mean the diction that would cause a right-hand threaded bolt to advance in the direction of $\boldsymbol{a}$.) This gives a vector $\boldsymbol{v}^{**}$ having the same length as $\boldsymbol{v}^*$ and having the direction of $\boldsymbol{u} \times \boldsymbol{v}$. Thus $\boldsymbol{u} \times \boldsymbol{v} = |\boldsymbol{u}|\boldsymbol{v}^{**}$:



Now, why did we go to all this trouble to construct $\boldsymbol{u} \times \boldsymbol{v}$ in this fashion? Simple. It makes it much easier to see that for any three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$, we have

$$\boldsymbol{u} \times (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \times \boldsymbol{v} + \boldsymbol{u} \times \boldsymbol{w} .$$

(Draw a picture!)

We shall see how to compute this vector product $\boldsymbol{u} \times \boldsymbol{v}$ for

$$\boldsymbol{u} = (a, b, c) = a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k} \text{ and } \boldsymbol{v} = (x, y, z) = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k} .$$

We have

$$\begin{aligned} \boldsymbol{u} \times \boldsymbol{v} &= (a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}) \times (x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}) \\ &= ax(\boldsymbol{i} \times \boldsymbol{i}) + ay(\boldsymbol{i} \times \boldsymbol{j}) + az(\boldsymbol{i} \times \boldsymbol{k}) + \\ &\qquad bx(\boldsymbol{j} \times \boldsymbol{i}) + by(\boldsymbol{j} \times \boldsymbol{j}) + bz(\boldsymbol{j} \times \boldsymbol{k}) + \\ &\qquad cx(\boldsymbol{k} \times \boldsymbol{i}) + cy(\boldsymbol{k} \times \boldsymbol{j}) + cz(\boldsymbol{k} \times \boldsymbol{k}) \end{aligned}$$

This looks like a terrible mess, until we note that

$$\boldsymbol{i} \times \boldsymbol{i} = \boldsymbol{j} \times \boldsymbol{j} = \boldsymbol{k} \times \boldsymbol{k} = \boldsymbol{0} ,$$

$$\boldsymbol{i} \times \boldsymbol{j} = -(\boldsymbol{j} \times \boldsymbol{i}) = \boldsymbol{k} ,$$

$$\boldsymbol{j} \times \boldsymbol{k} = -(\boldsymbol{k} \times \boldsymbol{j}) = \boldsymbol{i} \text{, and}$$

$$\boldsymbol{k} \times \boldsymbol{i} = -(\boldsymbol{i} \times \boldsymbol{k}) = \boldsymbol{j} .$$

Making these substitutions in the above equation for $\boldsymbol{u} \times \boldsymbol{v}$ gives us

$$\boldsymbol{u}\times\boldsymbol{v} = (bz - cy)\boldsymbol{i} + (cx - az)\boldsymbol{j} + (ay - bx)\boldsymbol{k}\ .$$

This is not particularly hard to remember, but there is a nice memory device using determinants:

$$\boldsymbol{u}\times\boldsymbol{v} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a & b & c \\ x & y & z \end{vmatrix}.$$

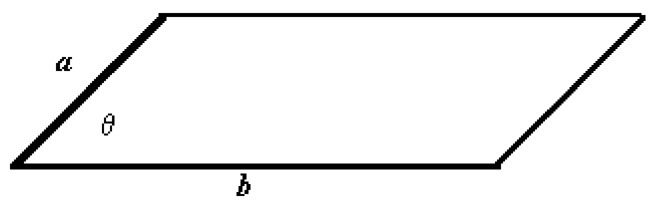
Example

Let’s find the velocity of a point on the surface of the Earth relative to a coordinate system whose origin is fixed at its center-we thus shall consider only motion due to the Earth’s rotation, and neglect its motion about the sun, *etc.* For our point on the Earth, choose Room 254, Skiles Classroom Building at Georgia Tech. The latitude of the room is about 33.75 degrees (North, of course.), and it is about 3960 miles from the center of the Earth. As we said, the origin of our coordinate system is the center of the Earth. We choose the third axis to point through the North Pole; In other words, the coordinate vector $\boldsymbol{k}$ points through the North Pole. The velocity of our room, is of course, not a constant, but changes as the Earth rotates. We find the velocity at the instant our room is in the coordinate plane determined by the vectors $\boldsymbol{i}$ and $\boldsymbol{k}$.

The Earth makes one complete revolution every 24 hours, and so its angular velocity ω is $\omega = \frac{2\pi}{24}\boldsymbol{k} \approx 02618\boldsymbol{k}$ radians/hour. The position vector $\boldsymbol{r}$ of our room is $\boldsymbol{r} = 3960(\cos(33.75)\boldsymbol{i} + \sin(33.75)\boldsymbol{k}) \approx 3292.6\boldsymbol{i} + 2200.1\boldsymbol{k}$ miles. Our velocity is thus

$$\omega\times\boldsymbol{r}=\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ 0 & 0 & 02618\\ 3292.6 & 0 & 2200.1\end{vmatrix}\approx 862\boldsymbol{j}\ \text{miles/hour.}$$

Suppose we want to find the area of a parallelogram, the non-parallel sides of which are representatives of the vectors ***a*** and ***b***:



The area A is clearly $A=|\boldsymbol{a}||\boldsymbol{b}|\sin\theta=|\boldsymbol{a}\times\boldsymbol{b}|$.

Example

Find the are of the parallelogram with a vertex (1,4,-2) and the vertices at the other ends of the sides adjoining this vertex are (4, 7, 8), and (6, 10, 20). This is easy. This is just as in the above picture with $\boldsymbol{a}=(4-1)\boldsymbol{i}+(7-4)\boldsymbol{j}+(8-(-2))\boldsymbol{k}=3\boldsymbol{i}+3\boldsymbol{j}+10\boldsymbol{k}$ and $\boldsymbol{b}=(6-1)\boldsymbol{i}+(10-4)\boldsymbol{j}+(20-(-2))\boldsymbol{k}=5\boldsymbol{i}+6\boldsymbol{j}+22\boldsymbol{k}$. So we have

$$\boldsymbol{a}\times\boldsymbol{b}=\begin{vmatrix}\boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k}\\ 3 & 3 & 10\\ 5 & 6 & 22\end{vmatrix}=6\boldsymbol{i}-16\boldsymbol{j}+3\boldsymbol{k},$$

and so,

$$\text{Area}=|\boldsymbol{a}\times\boldsymbol{b}|=\sqrt{6^2+16^2+3^2}=\sqrt{301}.$$

Exercises

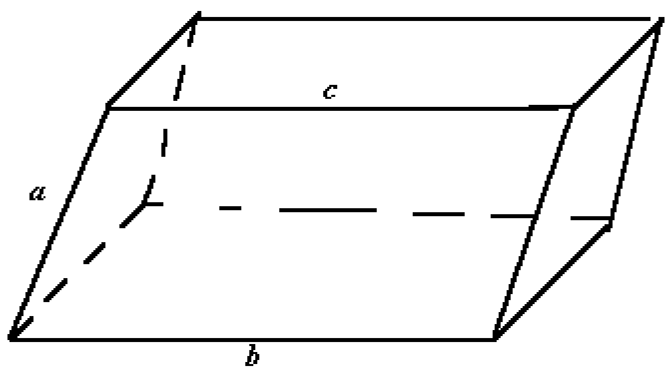
23. Find a a vector perpendicular to the plane containing the points (1,4,6), (-1,2,-7), and (-3,6,10).

24. Are the points (0,4,7), (2, 6, 8), and (5, 10, 20) collinear? Explain how you know?

25. Find the torque created by the force $\boldsymbol{f} = 3\boldsymbol{i} + 2\boldsymbol{j} - 3\boldsymbol{k}$ acting at the point $\boldsymbol{a} = \boldsymbol{i} - 2\boldsymbol{j} - 7\boldsymbol{k}$.

26. Find the area of the triangle whose vertices are (0,0,0), (1,2,3), and (4,7,12).

27. Find the volume of the parallelepiped



Chapter Three

Vector Functions

3.1 Relations and Functions

We begin with a review of the idea of a function. Suppose A and B are sets. The ***Cartesian product*** $A \times B$ of these sets is the collection of all ordered pairs (a,b) such that $a \quad A$ and $b \quad B$. A ***relation*** R is simply a subset of $A \times B$. The ***domain*** of R is the set $dom\, R = \{a \quad A : (a,b) \quad R\}$. In case $A = B$ and the domain of R is all of A, we call R a ***relation on A***. A relation $R \quad A \times B$ such that $(a,b) \quad R$ and $(a,c) \quad R$ only if $b = c$ is called a ***function***. In other words, if R is a function, and $a \quad dom\, R$, there is exactly one ordered pair $(a,b) \quad R$. The second "coordinate" b is thus uniquely determined by a. It is usually denoted $R(a)$. If $R \quad A \times B$ is a relation, the ***inverse*** of R is the relation $R^{-1} \quad B \times A$ defined by $R^{-1} = \{(b,a) : (a,b) \quad R\}$.

Example

Let A be the set of all people who have ever lived and let $S \quad A \times A$ be the relation defined by $S = \{(a,b) : b \text{ is the mother of } a\}$. The S is a relation on A, and is, in fact, a function. The relation S^{-1} is not a function, and $dom S^{-1} \neq A$.

The fact that $f \quad A \times B$ is a function with $dom\, f = A$ is frequently indicated by writing $f : A \to B$, and we say f is a function ***from*** A to B. Very often a function f is defined by specifying the domain, and giving a recipe for finding *f(a)*. Thus we may define the function f from the interval [0,1] to the real numbers by $f(x) = x^2$. This says that f is the collection of all ordered pairs (x, x^2) in which $x \quad [0,1]$.

Exercises

1. Let A be the set of all Georgia Tech students, and let B be the set of real numbers. Define the relation $W \subseteq A \times B$ by $W = \{(a,b) : b$ is the weight (in pounds) of $a\}$. Is W a function? Is W^{-1} a function? Explain.

2. Let X be set of all states of the U. S., and let Y be the set of all U. S. municipalities. Define the relation $c \subseteq X \times Y$ by $c = \{(x,y) : y$ is the capital of $x\}$. Explain why c is a function, and find c(Nevada), c(Missouri), and c(Kentucky).

3. With X, Y as in Exercise **2**, let b be the function $b = \{(x,y) : y$ is the largest city in $x\}$.

a)What is b(South Carolina)?

b)What is b(California)?

c)Let $f = c \cap b$, where c is the function defined in Exercise **2**. Find $dom\, f$.

4. Suppose $f \subseteq X \times Y$ and $g \subseteq X \times Y$. If f is a function, is it necessarily true that $f \cap g$ is a function? Prove your answer.

5. Suppose $f \subseteq X \times Y$ and $g \subseteq X \times Y$. If f and g are both functions, is it necessarily true that $f \cup g$ is a function? Prove your answer.

6. Suppose $f : X \to Y$ is a function and the inverse f^{-1} is also a function.

a)What is $f^{-1}(f(x))$? Explain.

b)If $y \in dom\, f^{-1}$, what is $f(f^{-1}(y))$? Explain.

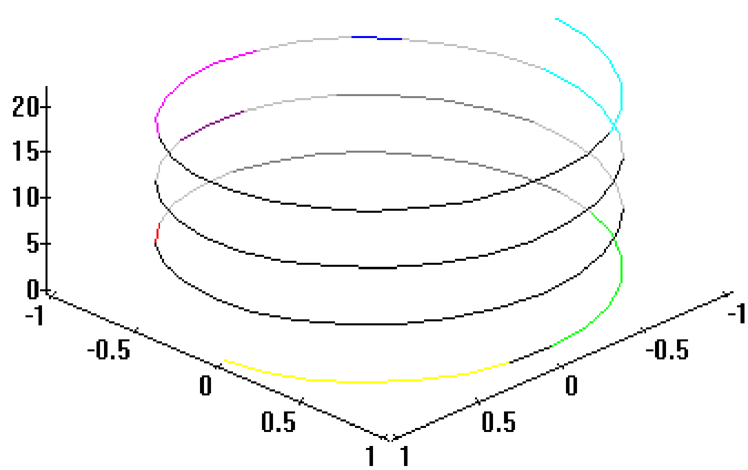
3.2 Vector Functions

Our interest now will be focused on functions $f : X \to Y$ in which Y is a set of vectors. These are called ***vector functions***, or sometimes, ***vector-valued*** functions. Initially, we shall be solely interested in the special case in which X is a "nice" set of real numbers, such as an interval. As the drama unfolds, we shall see that such functions provide just the right tool for describing curves in space.

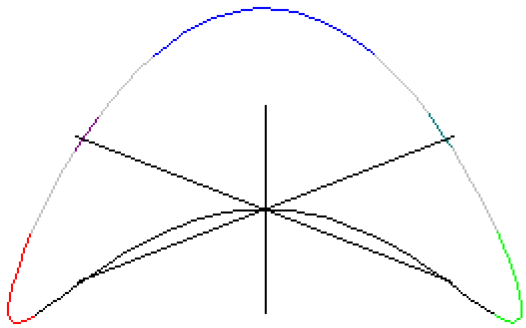
Let's begin with a simple example. Let X be the entire real line and let the function $\boldsymbol{f}$ be defined by $\boldsymbol{f}(t) = t\boldsymbol{i} + t^2\boldsymbol{j}$. It should be reasonably clear that if we place the tail of

$\boldsymbol{f}(t)$ (actually,. a representative of $\boldsymbol{f}(t)$) at the origin, the nose will lie on the curve $y = x^2$. In fact, as t varies over the reals, the nose traces out this curve. The function $\boldsymbol{f}$ is called a ***vector description*** of the curve. Let's look at another example. This time, let $\boldsymbol{g}(t) = \cos t\boldsymbol{i} + \sin t\boldsymbol{j}$ for $0 \le t \le 4\pi$. What is the curve described by this function? First, note that for all t, we have $|\boldsymbol{g}(t)| = 1$. The nose of $\boldsymbol{g}$ thus always lies on the circle of radius one centered at the origin. It's not difficult to see that, in fact, as t varies from 0 to 2π, the nose moves around the circle once, and as t varies on from 2π to 4π, the nose traces out the circle again.

The real usefulness of vector descriptions is most evident when we consider curves in space. Let $\boldsymbol{f}(t) = \cos t\boldsymbol{i} + \sin t\boldsymbol{j} + t\boldsymbol{k}$, for all $t \ge 0$. Now, what curve is followed by the nose of $\boldsymbol{f}(t)$? Notice first that if we look down on this curve from someplace up the positive third axis (In other words, $\boldsymbol{k}$ is pointing directly at us.), we see the circle described by $\cos t\boldsymbol{i} + \sin t\boldsymbol{j}$. As t increases, we run around this circle and the third component of our position increases linearly. Convince yourself now that this curve looks like this:



This curve is called a ***helix***, or more precisely, a ***right circular helix***. The picture was drawn by *Maple*. Let's draw another. How about the curve described by the vector function $\boldsymbol{g}(t) = \cos t\boldsymbol{i} + \sin t\boldsymbol{j} + \sin(2t)\boldsymbol{k}$? This one is just a bit more exciting. Here's a computer drawn picture:



(This time we put the axes where they are "supposed to be.")

Observe that in giving a vector description, we are in effect specifying the three coordinates of points on the curves as ordinary real valued functions defined on a subset of the reals. Assuming the axes are labeled x, y, and z, the curve described by the vector function

$$\mathbf{r}(t) = f(t)\mathbf{i} + g(t)\mathbf{j} + h(t)\mathbf{k}$$

is equivalently described by the equations

$$x = f(t)$$
$$y = g(t)$$
$$z = h(t)$$

These are called ***parametric equations*** of the curve (The variable t is called the ***parameter***.).

Exercises

7. Sketch or otherwise describe the curve given by $\mathbf{f}(t) = t\mathbf{i} + t^3\mathbf{k}$ for $-1 \le t \le 3$.

8. Sketch or otherwise describe the curve given by $\mathbf{f}(t) = (2t-3)\mathbf{i} + (3t+1)\mathbf{j}$.
[Hint: Find an equation in x and y, the graph of which is the given curve.]

9. Sketch or otherwise describe the curve given by $\boldsymbol{c}(t) = \cos t\boldsymbol{i} + \sin t\boldsymbol{j} + 7\boldsymbol{k}$.

10. Sketch or otherwise describe the curve given by $\boldsymbol{c}(t) = \cos(t^2)\boldsymbol{i} + \sin(t^2)\boldsymbol{j} + 7\boldsymbol{k}$.

11. Find an equation in x and y, the graph of which is the curve $\boldsymbol{g}(t) = 3\cos t\boldsymbol{i} + 4\sin t\boldsymbol{j}$.

12. a)Find a vector equation for the graph of $y = x^3 + 2x^2 + x + 5$.

b)Find a vector equation for the graph of $x = y^3 + 2y^2 + y + 5$.

13. Find a vector equation for the graph of $x^{\frac{2}{3}} + y^{\frac{2}{3}} = 1$.

14. a)Sketch or otherwise describe the curve given by the function $\boldsymbol{r}(t) = \boldsymbol{a} + t\boldsymbol{b}$, where $\boldsymbol{a} = 2\boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k}$ and $\boldsymbol{b} = \boldsymbol{i} + 3\boldsymbol{j} - 5\boldsymbol{k}$.

b)Express $\boldsymbol{r}(t)$ in the form $\boldsymbol{r}(t) = f(t)\boldsymbol{i} + g(t)\boldsymbol{j} + h(t)\boldsymbol{k}$.

15. Describe the curve given by $\boldsymbol{L}(t) = (3t+1)\boldsymbol{i} + (1-t)\boldsymbol{j} + 2t\boldsymbol{k}$.

16. Find a vector function for the straight line passing through the point (1,4,-2) in the direction of the vector $\boldsymbol{v} = \boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$.

17. a)Find a vector function for the straight line passing through the points (1,2,4) and (3,1,5).

b)Find a vector function for the line segment joining the points (1,2,4) and (3,1,5).

18. Let L be the line through the points (1,5,-2) and (2,2,4); and let M be the line through the points (2,4,6) and (-3,1,-2). Find a vector description of the line which passes through the point (1,1,2) and is perpendicular to both L and M.

3.3 Limits and Continuity

Recall from grammar school what we mean when we say the limit at t_0 of a real-valued, or scalar, function f is L. The definition for vector functions is essentially the same. Specifically, suppose $\boldsymbol{f}$ is a vector valued function, t_0 is a real number, and $\boldsymbol{L}$ is a vector such that for every real number $\varepsilon > 0$, there is a $\delta > 0$ such that $|\boldsymbol{f}(t) - \boldsymbol{L}| < \varepsilon$ whenever $0 < |t - t_0| < \delta$ and t is in the domain of $\boldsymbol{f}$. This is traditionally written

$$\lim_{t \to t_0} \boldsymbol{f}(t) = \boldsymbol{L}.$$

The vector $\boldsymbol{L}$ is called a ***limit of f at a***.

Suppose $\alpha(t)$ is a scalar function for which $\lim_{t \to t_0} \alpha(t) = a$, and $\boldsymbol{f}$ is a vector function for which $\lim_{t \to t_0} \boldsymbol{f}(t) = \boldsymbol{L}$. It is but a modest exercise to show that

$$\lim_{t \to t_0} (\alpha(t)\boldsymbol{f}(t)) = a\boldsymbol{L}.$$

To see this, we use the "behold!" method. Let $\varepsilon > 0$ be given. Choose $\delta_1, \delta_2, \delta_3$, and δ_4 so that

$$|\boldsymbol{f}(t) - \boldsymbol{L}| < \frac{\varepsilon}{3(1+|a|)} \text{ for } 0 < |t - t_0| < \delta_1;$$

$$|\boldsymbol{f}(t) - \boldsymbol{L}| < \sqrt{\frac{\varepsilon}{3}} \text{ for } 0 < |t - t_0| < \delta_2;$$

$$|\alpha(t) - a| < \frac{\varepsilon}{3(1+|\boldsymbol{L}|)} \text{ for } 0 < |t - t_0| < \delta_3; \text{ and}$$

$$|\alpha(t) - a| < \sqrt{\frac{\varepsilon}{3}} \text{ for } 0 < |t - t_0| < \delta_4.$$

Now let $\delta = \min\{\delta_1, \delta_2, \delta_3, \delta_4\}$ suppose t is such that $0 < |t - t_0| < \delta$. Then

$$\begin{aligned} |\alpha(t)\boldsymbol{f}(t) - a\boldsymbol{L}| &= |a(\boldsymbol{f}(t) - \boldsymbol{L}) + \boldsymbol{L}(\alpha(t) - a) + (\alpha(t) - a)(\boldsymbol{f}(t) - \boldsymbol{L})| \\ &\le |a(\boldsymbol{f}(t) - \boldsymbol{L})| + |\boldsymbol{L}(\alpha(t) - a)| + |(\alpha(t) - a)||(\boldsymbol{f}(t) - \boldsymbol{L})| \\ &< \frac{|a|}{3(1+|a|)}\varepsilon + \frac{|\boldsymbol{L}|}{3(1+|\boldsymbol{L}|)}\varepsilon + \sqrt{\frac{\varepsilon}{3}}\sqrt{\frac{\varepsilon}{3}} < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \end{aligned}$$

Or, in other words,

$$\lim_{t \to t_0}(\alpha(t)\,\boldsymbol{f}(t)) = a\boldsymbol{L}\ ,$$

which is what we set out to show.

Now suppose $\boldsymbol{f}(t) = x(t)\boldsymbol{i} + y(t)\boldsymbol{j} + z(t)\boldsymbol{k}$ and $\boldsymbol{L} = a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k}$. Then we see that $\lim\limits_{t \to t_0} \boldsymbol{f}(t) = \boldsymbol{L}$ if and only if

$$\lim_{t \to t_0} x(t) = a,$$

$$\lim_{t \to t_0} y(t) = b, \text{ and}$$

$$\lim_{t \to t_0} z(t) = c.$$

It is now easy to show that all the usual nice properties of limits are valid for vector functions:

$$\lim_{t \to t_0}(\boldsymbol{f}(t) + \boldsymbol{g}(t)) = \lim_{t \to t_0} \boldsymbol{f}(t) + \lim_{t \to t_0} \boldsymbol{g}(t).$$

$$\lim_{t \to t_0}(\boldsymbol{f}(t)\ \ \boldsymbol{g}(t)) = (\lim_{t \to t_0} \boldsymbol{f}(t))\ \ (\lim_{t \to t_0} \boldsymbol{g}(t))\,.$$

$$\lim_{t \to t_0}(\boldsymbol{f}(t) \times \boldsymbol{g}(t)) = (\lim_{t \to t_0} \boldsymbol{f}(t)) \times (\lim_{t \to t_0} \boldsymbol{g}(t))\,.$$

We are now ready to say what we mean by a vector function's being continuous at a point of its domain. Suppose t_0 is in the domain of the vector function $\boldsymbol{f}$. Then we say $\boldsymbol{f}$ ***is continuous at*** $\boldsymbol{t_0}$ if it is true that $\lim\limits_{t \to t_0} \boldsymbol{f}(t) = \boldsymbol{f}(t_0)$. It is easy to see that if

$$\boldsymbol{f}(t) = x(t)\boldsymbol{i} + y(t)\boldsymbol{j} + z(t)\boldsymbol{k}\ ,$$

then $\boldsymbol{f}$ is continuous at t_0 if and only if each of the everyday scalar functions $x(t), y(t),$ and $z(t)$ is continuous at t_0. This shows there is nothing particularly mysterious or exotic about continuity of vector functions.

If $\boldsymbol{f}$ is continuous at each point of its domain, then we say simply that $\boldsymbol{f}$ is ***continuous***,

Exercises

19. Is it possible for a function $\boldsymbol{f}$ to have more than one limit at $t = t_0$? Prove your answer.

20. Suppose m is a continuous real-valued function and $\boldsymbol{f}$ is a continuous vector-valued function. Is the vector function $\boldsymbol{h}$ defined by $\boldsymbol{h}(t) = m(t)\,\boldsymbol{f}(t)$ also continuous? Explain.

21. Let $\boldsymbol{f}$ and $\boldsymbol{g}$ be continuous at $t = t_0$. Is the function $\boldsymbol{h}$ defined by $\boldsymbol{h}(t) = \boldsymbol{f}(t) \times \boldsymbol{g}(t)$ continuous? Explain. How about the function $r(t) = \boldsymbol{f}(t)\ \ \boldsymbol{g}(t)$?

22. Let $\boldsymbol{r}(t) = t\boldsymbol{i} + t^2\boldsymbol{j} + \frac{1}{t}\boldsymbol{k}$. Is $\boldsymbol{r}$ a continuous function? Explain.

23. Suppose $\boldsymbol{r}$ is a continuous function. Explain how you know that the length function $n(t) = |\boldsymbol{r}(t)|$ is continuous.

Chapter Four

Derivatives

4.1 Derivatives

Suppose $\boldsymbol{f}$ is a vector function and t_0 is a point in the interior of the domain of $\boldsymbol{f}$ (t_0 in the interior of a set S of real numbers means there is an interval centered at t_0 that is a subset of S.). The derivative is defined just as it is for a plain old everyday real valued function, except, of course, the derivative is a vector. Specifically, we say that ***f* is differentiable at t_0** if there is a vector $\boldsymbol{v}$ such that

$$\lim_{t \to t_0} \frac{1}{h}[\boldsymbol{f}(t_0 + h) - \boldsymbol{f}(t_0)] = \boldsymbol{v}.$$

The vector $\boldsymbol{v}$ is called the ***derivative of f at t_0***.

Now, how would we find such a thing? Suppose $\boldsymbol{f}(t) = a(t)\boldsymbol{i} + b(t)\boldsymbol{j} + c(t)\boldsymbol{k}$. Then

$$\frac{1}{h}[\boldsymbol{f}(t_0 + h) - \boldsymbol{f}(t_0)] = \left(\frac{a(t_0 + h) - a(t_0)}{h}\right)\boldsymbol{i} + \left(\frac{b(t_0 + h) - b(t_0)}{h}\right)\boldsymbol{j} + \left(\frac{c(t_0 + h) - c(t_0)}{h}\right)\boldsymbol{k}.$$

It should now be clear that the vector function $\boldsymbol{f}$ is differentiable at t_0 if and only if each of the coordinate functions $a(t), b(t),$ and $c(t)$ is. Moreover, the vector derivative $\boldsymbol{v}$ is $\boldsymbol{v} = a'(t)\boldsymbol{i} + b'(t)\boldsymbol{j} + c'(t)\boldsymbol{k}$.

Now we "know" what the derivative of a vector function is, and we know how to compute it, but what is it, really? Let's see. Let $\boldsymbol{f}(t) = t\boldsymbol{i} + t^3\boldsymbol{j}$. This is, of course, a vector function which describes the graph of the function $y = x^3$. Let's look at the derivative of $\boldsymbol{f}$ at t_0: $\boldsymbol{v} = \boldsymbol{i} + 3t_0^2\boldsymbol{j}$. Convince yourself that the direction of the vector $\boldsymbol{v}$ is the direction tangent to the graph of $y = x^3$ at the point (t_0, t_0^3). It is not so clear what we should define to be the tangent to a curve other than a plane curve. Again, vectors come to our rescue. If $\boldsymbol{f}$ is a vector description of a space curve, the direction of the derivative $\boldsymbol{f}'(t)$ vector is the ***tangent*** direction at the point $\boldsymbol{f}(t)$ -the derivative $\boldsymbol{f}'(t)$ is said to be tangent to the curve at $\boldsymbol{f}(t)$.

If $\boldsymbol{f}(t)$ specifies the position of a particle at time *t*, then, of course, the derivative is the ***velocity*** of the particle, and its length $|\boldsymbol{f}'(t)|$ is the ***speed***. Thus the distance the particle travels from time $t = a$ to time $t = b$ is given by the integral of the speed:

$$d = \int_a^b |\boldsymbol{f}'(t)|dt.$$

If the particle behaves nicely, this distance is precisely the length of the arc of the curve from $\boldsymbol{f}(a)$ to $\boldsymbol{f}(b)$. It should be clear what we mean by "behaves nicely". . For the distance traveled by the particle to be the same as the length of its path, there must be no "backtracking", or reversing direction. This means we must not allow the velocity to be zero for any *t* between *a* and *b*.

Example

Consider the function $\boldsymbol{r}(t) = \cos t\boldsymbol{i} + \sin t\boldsymbol{j}$. Then the derivative, or velocity, is $\boldsymbol{r}'(t) = -\sin t\boldsymbol{i} + \cos t\boldsymbol{j}$. This vector is indeed tangent to the curve described by $\boldsymbol{r}$ (which we already know to be a circle of radius 1 centered at the origin.) at $\boldsymbol{r}(t)$. Note that the scalar product $\boldsymbol{r}(t)\ \boldsymbol{r}'(t) = -\sin t\cos t + \sin t\cos t = 0$, and so the tangent vector and the vector from the center of the circle to the point on the circle are perpendicular-a well-known fact you learned from Mrs. Turner in 4th grade. Note that the derivative is never zero-there is no value of t for which both cos$t$ and sin$t$ vanish. The length of a piece of the curve can thus be found by integrating the speed:

$$p = \int_0^{2\pi} |\mathbf{r}'(t)|dt = \int_0^{2\pi} \sqrt{\sin^2 t + \cos^2 t}dt = \int_0^{2\pi} dt = 2\pi .$$

No surprise here.

Exercises

1. a)Find a vector tangent to the curve $\boldsymbol{f}(t) = t^2\boldsymbol{i} + t^3\boldsymbol{j} + (1-t)\boldsymbol{k}$ at the point (1, 1, 0).

b)Find a vector equation for the line tangent to this same curve at the point (1, 1, 0).

2. The position of a particle is given by $\boldsymbol{r}(t) = \cos(t^3)\boldsymbol{i} + \sin(t^3)\boldsymbol{j}$.

a)Find the velocity of the particle.

b)Find the speed of the particle.

c)Describe the path of the particle, and find its length.

3. Let $\boldsymbol{L}$ be the line tangent to the curve $\boldsymbol{g}(t) = 10\cos t\boldsymbol{i} + 10\sin t\boldsymbol{j} + 16t\boldsymbol{k}$ at the point $(\frac{10}{\sqrt{2}}, \frac{10}{\sqrt{2}}, 4\pi)$. Find the point at which $\boldsymbol{L}$ intersects the ***i-j*** plane.

4. Let $\boldsymbol{L}$ be the straight line passing through the point (5, 0, 3) in the direction of the vector $\boldsymbol{a} = \boldsymbol{i} + 2\boldsymbol{j} - \boldsymbol{k}$, and let $\boldsymbol{M}$ be the straight line passing through the point (0, 0, 6) in the direction of $\boldsymbol{b} = \boldsymbol{i} - 3\boldsymbol{j} + 2\boldsymbol{k}$.

a)Are $\boldsymbol{L}$ and $\boldsymbol{M}$ parallel? Explain.

b)Do $\boldsymbol{L}$ and $\boldsymbol{M}$ intersect? Explain.

5. Let $\boldsymbol{L}$ be the straight line passing through the point (1, 1, 3) in the direction of the vector $\boldsymbol{a} = 2\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{k}$, and let $\boldsymbol{M}$ be the straight line passing through the point (0, 1, 5) in the direction of $\boldsymbol{b} = 3\boldsymbol{i} - \boldsymbol{j} + 2\boldsymbol{k}$. Find the distance between $\boldsymbol{L}$ and $\boldsymbol{M}$.

6. Find the length of the arc of the curve $\boldsymbol{R}(t) = 3\cos t\boldsymbol{i} + 3\sin t\boldsymbol{j} + 4t\boldsymbol{k}$ between the points (3, 0, 0) and (3, 0, 16π).

7. Find an integral the value of which is the length of the curve $y = x^2$ between the points (-1, 1) and (1, 1).

4.2 Geometry of Space Curves-Curvature

Let $\boldsymbol{R}(t)$ be a vector description of a curve. Then the distance $s(t)$ along the curve from the point $\boldsymbol{R}(t_0)$ to the point $\boldsymbol{R}(t)$ is, as we have seen, simply

$$s(t) = \int_{t_0}^{t} |\boldsymbol{R}'(\xi)| d\xi ;$$

assuming, of course, that $\boldsymbol{R}'(t) \neq 0$. The speed is

$$\frac{ds}{dt} = |\boldsymbol{R}'(t)| .$$

Now then the vector

$$\boldsymbol{T} = \frac{\boldsymbol{R}'(t)}{|\boldsymbol{R}'(t)|} = \frac{\boldsymbol{R}'(t)}{ds/dt} = \boldsymbol{R}'(t)\frac{dt}{ds} = \frac{d\boldsymbol{R}}{ds}$$

is tangent to $\boldsymbol{R}$ and has length one. It is called the ***unit tangent vector***.

Consider next the derivative

$$\frac{d}{ds}\boldsymbol{T}\;\;\boldsymbol{T} = \boldsymbol{T}\;\;\frac{d\boldsymbol{T}}{ds} + \frac{d\boldsymbol{T}}{ds}\;\;\boldsymbol{T} = 2\boldsymbol{T}\;\;\frac{d\boldsymbol{T}}{ds}\,.$$

But we know that $\boldsymbol{T}\;\;\boldsymbol{T} = |\boldsymbol{T}|^2 = 1$. Thus $\boldsymbol{T}\;\;\frac{d\boldsymbol{T}}{ds} = 0$, which means that the vector $\frac{d\boldsymbol{T}}{ds}$ is perpendicular, or orthogonal, or normal, to the tangent vector ***T***. The length of this vector is called the ***curvature*** and is usually denoted by the letter κ. Thus

$$\kappa = \left|\frac{d\boldsymbol{T}}{ds}\right|.$$

The unit vector

$$\boldsymbol{N} = \frac{1}{\kappa}\frac{d\boldsymbol{T}}{ds}$$

is called the ***principal unit normal vector***, and its direction is sometimes called the ***principal normal direction.***

Example

Consider the circle of radius a and center at the origin: $\boldsymbol{R}(t) = a\cos t\boldsymbol{i} + a\sin t\boldsymbol{j}$.

Then $\boldsymbol{R}'(t) = -a\sin t\boldsymbol{i} + a\cos t\boldsymbol{j}$, and $\frac{ds}{dt} = |\boldsymbol{R}'(t)| = \sqrt{a^2\sin^2 t + a^2\cos^2 t} = \sqrt{a^2} = |a| = a$.

Thus

$$\boldsymbol{T} = \frac{1}{a}\boldsymbol{R}'(t) = -\sin t\boldsymbol{i} + \cos t\boldsymbol{j}\,.$$

Let’s not stop now.

$$\frac{d\boldsymbol{T}}{ds} = \frac{d\boldsymbol{T}}{dt}\frac{dt}{ds} = \frac{1}{a}\frac{d\boldsymbol{T}}{dt} = \frac{1}{a}(-\cos t\boldsymbol{i} - \sin t\boldsymbol{j})\,.$$

Thus $\kappa = \left|\frac{d\boldsymbol{T}}{ds}\right| = \frac{1}{a}$, and $\boldsymbol{N} = -(\cos t\boldsymbol{i} + \sin t\boldsymbol{j})$. So the curvature is the reciprocal of the radius and the principal normal vector points back toward the center of the circle.

Another Example

This time let $\boldsymbol{R} = (t+1)\boldsymbol{i} + 2t\boldsymbol{j} + t^2\boldsymbol{k}$. First, $\boldsymbol{R}'(t) = \boldsymbol{i} + 2\boldsymbol{j} + 2t\boldsymbol{k}$, and so $\frac{ds}{dt} = |\boldsymbol{R}'(t)| = \sqrt{5+4t^2}$. The unit tangent is then

$$\boldsymbol{T} = \frac{1}{\sqrt{5+4t^2}}(\boldsymbol{i} + 2\boldsymbol{j} + 2t\boldsymbol{k}).$$

It's a bit of a chore now to find the curvature and the principal normal, so let's use a computer algebra system; *viz., Maple*:

First, let's enter the unit tangent vector ***T***:

$$T := t \to \frac{[1, 2, 2t]}{\mathrm{sqrt}(5+4t^2)}$$

See if we got it right:

T(t);

$$\frac{[1, 2, 2t]}{\sqrt{5+4t^2}}$$

Fine. Now differentiate:

$$A := t \to \frac{\mathrm{simplify}(\mathrm{diff}(\mathrm{T}(t), t))}{\mathrm{sqrt}(5+4t^2)}$$

A(t);

$$\frac{5\,[0, 0, 2] + 4\,[0, 0, 2]\,t^2 - 4\,[1, 2, 2t]\,t}{(5+4t^2)^2}$$

We need to tidy this up:

$$B := t \to \frac{\text{evalm}(\text{numer}(A(t)))}{\text{denom}(A(t))}$$

B(t);

$$\frac{[-4\,t \quad -8\,t \quad 10]}{(5+4\,t^2)^2}$$

This vector is, of course, the normal $\frac{d\boldsymbol{T}}{ds}$. We continue and find the curvature κ and the principal normal ***N***.

```
kappa:=t->simplify(sqrt(dotprod(B(t),B(t))));
kappa(t);
```

$$2\,\frac{\sqrt{5}}{(5+4\,t^2)^{3/2}}$$

$$N := t \to \frac{B(t)}{\kappa(t)}$$

N(t);

$$\frac{1}{10}\,\frac{[-4\,t \quad -8\,t \quad 10]\sqrt{5}}{\sqrt{5+4\,t^2}}$$

So there we have at last the speed $\frac{ds}{dt}$, the unit tangent ***T***, the curvature κ., and the principal normal ***N***.

Exercises

8. Find a line tangent to the curve $\boldsymbol{R}(t) = (t^2 + t)\boldsymbol{i} + (t+1)\boldsymbol{j} - (t^3 + 5)\boldsymbol{k}$ and passing through the point (5, -2, 15), or show there is no such line.

9. Find the unit tangent ***T***, the principal normal ***N*,** and the curvature κ, for the curves:

a) $\boldsymbol{R}(t) = 5\cos(t)\boldsymbol{i} + 5\sin(t)\boldsymbol{j} + 2t\boldsymbol{k}$

b) $\boldsymbol{R}(t) = (2t+3)\boldsymbol{i} + (5-t^2)\boldsymbol{j}$

c) $\boldsymbol{R}(t) = e^t \cos t\boldsymbol{i} + e^t \sin t\boldsymbol{j} + 6\boldsymbol{k}$

10. Find the curvature of the curve $y = f(x)$ at $(x_0, f(x_0))$.

11. Find the curvature of $\boldsymbol{R}(t) = t\boldsymbol{i} + t^2\boldsymbol{j}$. At what point on the curve is the curvature the largest? smallest?

12. Find the curvature of $\boldsymbol{R}(t) = t\boldsymbol{i} + t^3\boldsymbol{j}$. At what point on the curve is the curvature the largest? smallest?

4.3 Geometry of Space Curves-Torsion

Let $\boldsymbol{R}(t)$ be a vector description of a curve. If ***T*** is the unit tangent and ***N*** is the principal unit normal, the unit vector $\boldsymbol{B} = \boldsymbol{T} \times \boldsymbol{N}$ is called the ***binormal***. Note that the binormal is orthogonal to both ***T*** and ***N***. Let's see about its derivative $\frac{d\boldsymbol{B}}{ds}$ with respect to arclength s. First, note that $\boldsymbol{B} \ \ \boldsymbol{B} = |\boldsymbol{B}|^2 = 1$, and so $\boldsymbol{B} \ \ \frac{d\boldsymbol{B}}{ds} = 0$, which means that being orthogonal to ***B***, the derivative $\frac{d\boldsymbol{B}}{ds}$ is in the plane of ***T*** and ***N***. Next, note that ***B is*** perpendicular to the tangent vector ***T***, and so $\boldsymbol{B} \ \ \boldsymbol{T} = 0$. Thus $\frac{d\boldsymbol{B}}{ds} \ \ \boldsymbol{T} = 0$. So what have we here? The vector $\frac{d\mathbf{B}}{ds}$ is perpendicular to both $\mathbf{B}$ and $\mathbf{T}$, and so must have the direction of $\mathbf{N}$ (or, of course, - $\mathbf{N}$). This means

$$\frac{d\mathbf{B}}{ds} = -\tau \mathbf{N} .$$

The scalar τ is called the ***torsion***.

Example

Let's find the torsion of the helix $\mathbf{R}(t) = a\cos t\mathbf{i} + a\sin t\mathbf{j} + bt\mathbf{k}$. Here we go! $\mathbf{R}'(t) = -a\sin t\mathbf{i} + a\cos t\mathbf{j} + b\mathbf{k}$. Thus $\frac{ds}{dt} = |\mathbf{R}'(t)| = \sqrt{a^2 + b^2}$, and we have

$$\mathbf{T} = \frac{1}{\sqrt{a^2 + b^2}}(-a\sin t\mathbf{i} + a\cos t\mathbf{j} + b\mathbf{k}) .$$

Now then

$$\frac{d\mathbf{T}}{ds} = \frac{d\mathbf{T}}{dt}\frac{dt}{ds} = \frac{-a}{(a^2 + b^2)}(\cos t\mathbf{i} + \sin t\mathbf{j}) .$$

Therefore,

$$\kappa = \frac{a}{(a^2 + b^2)} \text{ and } \mathbf{N} = -(\cos t\mathbf{i} + \sin t\mathbf{j}) .$$

Let's don't stop now:

$$\boldsymbol{B} = \boldsymbol{T} \times \boldsymbol{N} = \frac{1}{\sqrt{a^2 + b^2}} \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ -a\sin t & a\cos t & b \\ -\cos t & -\sin t & 0 \end{vmatrix} = \frac{1}{\sqrt{a^2 + b^2}} (b\sin t\boldsymbol{i} - b\cos t\boldsymbol{j} + a\boldsymbol{k});$$

and

$$\frac{d\boldsymbol{B}}{ds} = \frac{d\boldsymbol{B}}{dt}\frac{dt}{ds} = \frac{b}{(a^2 + b^2)}(\cos t\boldsymbol{i} + \sin t\boldsymbol{j}) = \frac{-b}{(a^2 + b^2)}\boldsymbol{N}.$$

The torsion, at last:

$$\tau = \frac{b}{a^2 + b^2}.$$

Suppose the curve $\boldsymbol{R}(t)$ is such that the torsion is zero for all values of *t*. In other words, $\frac{d\boldsymbol{B}}{ds} \equiv 0$. Look at

$$\frac{d}{ds}[(\boldsymbol{R}(t) - \boldsymbol{R}(t_0)) \;\; \boldsymbol{B}] = \frac{d\boldsymbol{R}}{ds} \;\; \boldsymbol{B} + (\boldsymbol{R}(t) - \boldsymbol{R}(t_0)) \;\; \frac{d\boldsymbol{B}}{ds} = 0.$$

Thus the scalar product $(\boldsymbol{R}(t) - \boldsymbol{R}(t_0)) \;\; \boldsymbol{B}$ is constant. It is 0 at t_0, and hence it is 0 for all values of *t*. This means that $\boldsymbol{R}(t) - \boldsymbol{R}(t_0)$ and $\boldsymbol{B}$ are perpendicular for all t, and so $\boldsymbol{R}(t) - \boldsymbol{R}(t_0)$ lies in a plane perpendicular to $\boldsymbol{B}$. In other words, the curve described by $\boldsymbol{R}(t)$ is a plane curve.

Exercises

13. Find the binormal and torsion for the curve $\boldsymbol{R}(t) = 4\cos t\boldsymbol{i} + 3\sin t\boldsymbol{k}$.

14. Find the binormal and torsion for the curve $\boldsymbol{R}(t) = \frac{\sin t}{\sqrt{2}}\boldsymbol{i} + \cos t\boldsymbol{j} + \frac{\sin t}{\sqrt{2}}\boldsymbol{k}$.

15. Find the curvature and torsion for $\boldsymbol{R}(t) = t\boldsymbol{i} + t^2\boldsymbol{j} + t^3\boldsymbol{k}$.

16. Show that the curve $\boldsymbol{R}(t) = t\boldsymbol{i} + \frac{1+t}{t}\boldsymbol{j} + \frac{1-t^2}{t}\boldsymbol{k}$ lies in a plane.

17. What is the vector $\boldsymbol{B} \times \boldsymbol{T}$? How about $\boldsymbol{N} \times \boldsymbol{T}$?

4.4 Motion

Suppose t is time and $\boldsymbol{R}(t)$ is the position vector of a body. Then the curve described by $\boldsymbol{R}(t)$ is the path, or trajectory, of the body, $\boldsymbol{v}(t) = \frac{d\boldsymbol{R}}{dt}$ is the velocity, and $\boldsymbol{a}(t) = \frac{d\boldsymbol{v}}{dt}$ is the acceleration. We know that $\boldsymbol{v}(t) = \frac{ds}{dt}\boldsymbol{T}$, and so the direction of the velocity is the unit tangent $\boldsymbol{T}$. Let's see about the direction of the acceleration:

$$\begin{aligned}\boldsymbol{a}(t) &= \frac{d\boldsymbol{v}}{dt} = \frac{d^2 s}{dt^2}\boldsymbol{T} + \frac{ds}{dt}\frac{d\boldsymbol{T}}{dt} \\ &= \frac{d^2 s}{dt^2}\boldsymbol{T} + \left(\frac{ds}{dt}\right)^2 \kappa \boldsymbol{N}\end{aligned},$$

since $\frac{d\boldsymbol{T}}{dt} = \frac{ds}{dt}\kappa\boldsymbol{N}$. This tells us that the acceleration is always in the plane of the vectors $\boldsymbol{T}$ and $\boldsymbol{N}$. The derivative of the speed $\frac{d^2s}{dt^2}$ is the ***tangential component*** of the acceleration, and $\kappa\left(\frac{ds}{dt}\right)^2$ is the ***normal component*** of the acceleration.

Example

Suppose a person who weighs 160 pounds moves around a circle having radius 20 feet at a constant speed of 60 miles/hour. What is the magnitude of the force on this person at any time?

First, we know the force $\boldsymbol{f}$ is the mass times the acceleration: $\boldsymbol{f}(t) = m\boldsymbol{a}(t)$. Thus

$$\boldsymbol{f} = m\frac{d^2s}{dt^2}\boldsymbol{T} + m\kappa\left(\frac{ds}{dt}\right)^2\boldsymbol{N}$$

also have The speed is a constant 60 miles/hour, or 88 feet/second; in other words, $\frac{ds}{dt} = 88$ and $\frac{d^2s}{dt} = 0$. Hence,

$$|\boldsymbol{f}| = |m\kappa\left(\frac{ds}{dt}\right)^2\boldsymbol{N}| = m\kappa\left(\frac{ds}{dt}\right)^2.$$

The mass $m = \frac{160}{32} = 5$ slugs, and the curvature $\kappa = \frac{1}{20}$. The magnitude of the force is thus $|\mathbf{f}| = \frac{5\ 88^2}{20} = 1936$ pounds.

Exercises

18. The position of an object at time t is given by $\boldsymbol{r}(t) = t\boldsymbol{i} + (t^3 - 2)\boldsymbol{j} + 2t\boldsymbol{k}$. Find the velocity, the speed, and the tangential and normal components of the acceleration.

19. A force $\boldsymbol{f}(t) = t^2\boldsymbol{i} + (t-1)\boldsymbol{j} + \boldsymbol{k}$ newtons is applied to an object of mass 2 kilograms. At time t = 0, the object is at the origin. Find its position at time t.

20. A projectile of weight w is fired from the origin with an initial speed v_0 in the direction of the vector $\cos\theta\,\boldsymbol{i} + \sin\theta\,\boldsymbol{j}$, and the only force acting on the projectile is $\boldsymbol{f} = -w\boldsymbol{j}$.

a)Find a vector description of the trajectory of the projectile.

b)Find an equation the graph of which is the trajectory.

21. A 16 lb. bowling ball is rolled along a track with a circular vertical loop of radius a feet. What must the speed of the ball be in order for it not to fall from the track? What must the speed of an 8 lb. ball be in order for it not to fall?

Chapter Five

More Dimensions

5.1 The Space $\boldsymbol{R}^n$

We are now prepared to move on to spaces of dimension greater than three. These spaces are a straightforward generalization of our Euclidean space of three dimensions. Let n be a positive integer. *The* ***n-dimensional Euclidean space*** $\boldsymbol{R}^n$ is simply the set of all ordered n-tuples of real numbers $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. Thus $\boldsymbol{R}^1$ is simply the real numbers, $\boldsymbol{R}^2$ is the plane, and $\boldsymbol{R}^3$ is Euclidean three-space. These ordered n-tuples are called ***points***, or ***vectors***. This definition does not contradict our previous definition of a vector in case $n = 3$ in that we identified each vector with an ordered triple (x_1, x_2, x_3) and spoke of the triple as being a vector.

We now define various arithmetic operations on $\boldsymbol{R}^n$ in the obvious way. If we have vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ in $\boldsymbol{R}^n$, the sum $\boldsymbol{x} + \boldsymbol{y}$ is defined by

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n),$$

and multiplication of the vector $\boldsymbol{x}$ by a scalar a is defined by

$$a\boldsymbol{x} = (ax_1, ax_2, \ldots, ax_n).$$

It is easy to verify that $a(\boldsymbol{x} + \boldsymbol{y}) = a\boldsymbol{x} + a\boldsymbol{y}$ and $(a + b)\boldsymbol{x} = a\boldsymbol{x} + b\boldsymbol{x}$.

Again we see that these definitions are entirely consistent with what we have done in dimensions 1, 2, and 3-there is nothing to unlearn. Continuing, we define the ***length***, or ***norm*** of a vector $\boldsymbol{x}$ in the obvious manner

$$|\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}.$$

The ***scalar product*** of $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\boldsymbol{x} \quad \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i \; .$$

It is again easy to verify the nice properties:

$$|\boldsymbol{x}|^2 = \boldsymbol{x} \quad \boldsymbol{x} \geq 0,$$

$$|a\boldsymbol{x}| = |a||\boldsymbol{x}|,$$

$$\boldsymbol{x} \quad \boldsymbol{y} = \boldsymbol{y} \quad \boldsymbol{x} \, ,$$

$$\boldsymbol{x} \quad (\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{x} \quad \boldsymbol{y} + \boldsymbol{x} \quad \boldsymbol{z} \text{, and}$$

$$(a\boldsymbol{x}) \quad \boldsymbol{y} = a(\boldsymbol{x} \quad \boldsymbol{y}) \, .$$

The geometric language of the three dimensional setting is retained in higher dimensions; thus we speak of the “length” of an *n*-tuple of numbers. In fact, we also speak of $d(\boldsymbol{x}, \boldsymbol{y}) = |\boldsymbol{x} - \boldsymbol{y}|$ as the ***distance*** between $\boldsymbol{x}$ and $\boldsymbol{y}$. We can, of course, no longer rely on our vast knowledge of Euclidean geometry in our reasoning about $\boldsymbol{R}^{\boldsymbol{n}}$ when $n > 3$. Thus for $n \leq 3$, the fact that $|\boldsymbol{x} + \boldsymbol{y}| \leq |\boldsymbol{x}| + |\boldsymbol{y}|$ for any vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ was a simple consequence of the fact that the sum of the lengths of two sides of a triangle is at least as big as the length of the third side. This inequality remains true in higher dimensions, and, in fact, is called the ***triangle inequality***, but requires an essentially algebraic proof. Let’s see if we can prove it.

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$. Then if a is a scalar, we have

$$|a\boldsymbol{x} + \boldsymbol{y}|^2 = (a\boldsymbol{x} + \boldsymbol{y}) \quad (a\boldsymbol{x} + \boldsymbol{y}) \geq 0 \, .$$

Thus,

$$(a\boldsymbol{x} + \boldsymbol{y}) \quad (a\boldsymbol{x} + \boldsymbol{y}) = a^2 \boldsymbol{x} \quad \boldsymbol{x} + 2a\boldsymbol{x} \quad \boldsymbol{y} + \boldsymbol{y} \quad \boldsymbol{y} \geq 0.$$

This is a quadratic function in a and is never negative; it must therefore be true that

$$4(\boldsymbol{x} \quad \boldsymbol{y})^2 - 4(\boldsymbol{x} \quad \boldsymbol{x})(\boldsymbol{y} \quad \boldsymbol{y}) \leq 0 \text{, or}$$

$$|\boldsymbol{x} \quad \boldsymbol{y}| \leq |\boldsymbol{x}||\boldsymbol{y}| \, .$$

This last inequality is the celebrated ***Cauchy-Schwarz-Buniakowsky inequality***. It is exactly the ingredient we need to prove the triangle inequality.

$$|\boldsymbol{x}+\boldsymbol{y}|^2 = (\boldsymbol{x}+\boldsymbol{y}) \cdot (\boldsymbol{x}+\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{x} + 2\boldsymbol{x} \cdot \boldsymbol{y} + \boldsymbol{y} \cdot \boldsymbol{y}\,.$$

Applying the ***C-S-B*** inequality, we have

$$|\boldsymbol{x}+\boldsymbol{y}|^2 \le |\boldsymbol{x}|^2 + 2|\boldsymbol{x}||\boldsymbol{y}| + |\boldsymbol{y}|^2 = (|\boldsymbol{x}|+|\boldsymbol{y}|)^2, \text{ or}$$

$$|\boldsymbol{x}+\boldsymbol{y}| \le |\boldsymbol{x}| + |\boldsymbol{y}|.$$

Corresponding to the coordinate vectors $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$, the coordinate vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are defined in $\boldsymbol{R}^n$ by

$$\begin{aligned}
\boldsymbol{e}_1 &= (1,0,0,0,\ldots,0) \\
\boldsymbol{e}_2 &= (0,1,0,0,\ldots,0) \\
\boldsymbol{e}_3 &= (0,0,1,0,\ldots,0)\,, \\
&\vdots \\
\boldsymbol{e}_n &= (0,0,0,\ldots,0,1)
\end{aligned}$$

Thus each vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ may be written in terms of these coordinate vectors:

$$\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i\,.$$

Exercises

1. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two vectors in $\boldsymbol{R}^n$. Prove that $|\boldsymbol{x}+\boldsymbol{y}|^2 = |\boldsymbol{x}|^2 + |\boldsymbol{y}|^2$ if and only if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$. (Adopting more geometric language from three space, we say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are ***perpendicular*** or ***orthogonal*** if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$.)

2. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two vectors in $\boldsymbol{R}^n$. Prove

a) $|\boldsymbol{x}+\boldsymbol{y}|^2 - |\boldsymbol{x}-\boldsymbol{y}|^2 = 4\boldsymbol{x} \cdot \boldsymbol{y}$.

b) $|\boldsymbol{x}+\boldsymbol{y}|^2 + |\boldsymbol{x}-\boldsymbol{y}|^2 = 2[|\boldsymbol{x}|^2 + |\boldsymbol{y}|^2]$.

3. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two vectors in $\boldsymbol{R}^n$. Prove that $\left|\,|\boldsymbol{x}| - |\boldsymbol{y}|\,\right| \leq |\boldsymbol{x}| + |\boldsymbol{y}|$.

4. Let $P \quad \boldsymbol{R}^4$ be the set of all vectors $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ such that

$$3x_1 + 5x_2 - 2x_3 + x_4 = 15.$$

Find vectors $\boldsymbol{n}$ and $\boldsymbol{a}$ such that $P = \{\boldsymbol{x} \quad \boldsymbol{R}^4 : \boldsymbol{n} \;\; (\boldsymbol{x} - \boldsymbol{a}) = 0\}$.

5. Let $\boldsymbol{n}$ and $\boldsymbol{a}$ be vectors in $\boldsymbol{R}^n$, and let $P = \{\boldsymbol{x} \quad \boldsymbol{R}^n : \boldsymbol{n} \;\; (\boldsymbol{x} - \boldsymbol{a}) = 0$.

a)Find an equation in $x_1, x_2, \ldots,$ and x_n such that $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \quad P$ if and only if the coordinates of $\boldsymbol{x}$ satisfy the equation.

b)Describe the set P be in case $n = 3$. Describe it in case $n = 2$.

[The set P is called a ***hyperplane through a***.]

5.2 Functions

We now consider functions $F: \boldsymbol{R}^n \to \boldsymbol{R}^p$. Note that when $n = p = 1$, we have the usual grammar school calculus functions, and when $n = 1$ and $p = 2$ or 3, we have the vector valued functions of the previous chapter. Note also that except for very special circumstances, graphs of functions will not play a big role in our understanding. The set of points $(\boldsymbol{x}, F(\boldsymbol{x}))$ resides in $\boldsymbol{R}^{n+p}$ since $\boldsymbol{x} \quad \boldsymbol{R}^n$ and $F(\boldsymbol{x}) \quad \boldsymbol{R}^p$; this is difficult to "see" unless $n + p \leq 3$.

We begin with a very special kind of functions, the so-called linear functions. A function $F: \boldsymbol{R}^n \to \boldsymbol{R}^p$ is said to be a ***linear*** function if

i) $F(\boldsymbol{x} + \boldsymbol{y}) = F(\boldsymbol{x}) + F(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \quad \boldsymbol{R}^n$, and

ii) $F(a\boldsymbol{x}) = aF(\boldsymbol{x})$ for all scalars a and $\boldsymbol{x} \quad \boldsymbol{R}^n$.

Example

Let $n = p = 1$, and define F by $F(x) = 3x$. Then

$$F(x + y) = 3(x + y) = 3x + 3y = F(x) + F(y) \text{ and}$$

$$F(ax) = 3(ax) = a3x = aF(x).$$

This F is a linear function.

Another Example

Let $\boldsymbol{F}:\boldsymbol{R} \to \boldsymbol{R^3}$ be defined by $\boldsymbol{F}(t) = t\boldsymbol{i} + 2t\boldsymbol{j} - 7t\boldsymbol{k} = (t, 2t, -7t)$. Then

$$\begin{aligned}\boldsymbol{F}(t+s) &= (t+s)\boldsymbol{i} + 2(t+s)\boldsymbol{j} - 7(t+s)\boldsymbol{k} \\ &= [t\boldsymbol{i} + 2t\boldsymbol{j} - 7t\boldsymbol{k}] + [s\boldsymbol{i} + 2s\boldsymbol{j} - 7s\boldsymbol{k}] \\ &= \boldsymbol{F}(t) + \boldsymbol{F}(s)\end{aligned}$$

Also,

$$\begin{aligned}\boldsymbol{F}(at) &= at\boldsymbol{i} + 2at\boldsymbol{j} - 7at\boldsymbol{k} \\ &= a[t\boldsymbol{i} + 2t\boldsymbol{j} - 7t\boldsymbol{k}] = a\boldsymbol{F}(t)\end{aligned}$$

We see yet another linear function.

One More Example

Let $F:\boldsymbol{R^3} \to \boldsymbol{R^4}$ be defined by

$$F((x_1, x_2, x_3)) = (2x_1 - x_2 + 3x_3,\ x_1 + 4x_2 - 5x_3,\ -x_1 + 2x_2 + x_3,\ x_1 + x_3).$$

It is easy to verify that F is indeed a linear function.

A ***translation*** is a function $T:\boldsymbol{R^p} \to \boldsymbol{R^p}$ such that $T(\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{x}$, where $\boldsymbol{a}$ is a fixed vector in $\boldsymbol{R^n}$. A function that is the composition of a linear function followed by a translation is called an ***affine*** function. An affine function F thus has the form $F(\boldsymbol{x}) = \boldsymbol{a} + L(\boldsymbol{x})$, where L is a linear function.

Example

Let $F:\boldsymbol{R} \to \boldsymbol{R^3}$ be defined by $F(t) = (2+t,\ 4t-3,\ t)$. Then F is affine. Let $\boldsymbol{a} = (2,4,0)$ and $L(t) = (t,\ 4t,\ t)$. Clearly $F(t) = \boldsymbol{a} + L(t)$.

Exercises

6. Which of the following functions are linear? Explain your answers.

a) $f(x) = -7x$

b) $g(x) = 2x - 5$

c) $F(x_1, x_2) = (2x_1 + x_2,\ x_1 - x_2,\ 3x_1,\ 5x_1 - 2x_2,\ x_1)$

d) $G(x_1, x_2, x_3) = x_1x_2 + x_3$

e) $F(t) = (2t,\ t,\ 0,\ -2t)$

f) $h(x_1, x_2, x_3, x_4) = (1,\ 0,\ 0)$

g) $f(x) = \sin x$

7. a)Describe the graph of a linear function from $\boldsymbol{R}$ to $\boldsymbol{R}$.
b)Describe the graph of an affine function from $\boldsymbol{R}$ to $\boldsymbol{R}$.

Chapter Six

Linear Functions and Matrices

6.1 Matrices

Suppose $f:\boldsymbol{R}^n \to \boldsymbol{R}^p$ be a linear function. Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ be the coordinate vectors for $\boldsymbol{R}^n$. For any $\boldsymbol{x}$ $\boldsymbol{R}^n$, we have $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \ldots + x_n\boldsymbol{e}_n$. Thus

$$f(\boldsymbol{x}) = f(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \ldots + x_n\boldsymbol{e}_n) = x_1 f(\boldsymbol{e}_1) + x_2 f(\boldsymbol{e}_2) + \ldots + x_n f(\boldsymbol{e}_n).$$

Meditate on this; it says that a linear function is entirely determined by its values $f(\boldsymbol{e}_1), f(\boldsymbol{e}_2), \ldots, f(\boldsymbol{e}_n)$. Specifically, suppose

$$\begin{aligned}
&f(\boldsymbol{e}_1) = (a_{11}, a_{21}, \ldots, a_{p1}),\\
&f(\boldsymbol{e}_2) = (a_{12}, a_{22}, \ldots, a_{p2}),\\
&\vdots\\
&f(\boldsymbol{e}_n) = (a_{1n}, a_{2n}, \ldots, a_{pn}).
\end{aligned}$$

Then

$$\begin{aligned}
f(\boldsymbol{x}) = (&a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n,\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n, \ldots,\\
&a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pn}x_n).
\end{aligned}$$

The numbers a_{ij} thus tell us everything about the linear function *f*. . To avoid labeling these numbers, we arrange them in a rectangular array, called a ***matrix***:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{bmatrix}$$

The matrix is said to ***represent*** the linear function *f*.

For example, suppose $f: \boldsymbol{R^2} \to \boldsymbol{R^3}$ is given by the receipt

$$f(x_1, x_2) = (2x_1 - x_2,\ x_1 + 5x_2,\ 3x_1 - 2x_2).$$

Then $f(\boldsymbol{e}_1) = f(1,0) = (2,1,3)$, and $f(\boldsymbol{e}_2) = f(0,1) = (-1,5,-2)$. The matrix representing *f* is thus

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

Given the matrix of a linear function, we can use the matrix to compute $f(\boldsymbol{x})$ for any $\boldsymbol{x}$. This calculation is systematized by introducing an arithmetic of matrices. First, we need some jargon. For the matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pn} \end{bmatrix},$$

the matrices $[a_{i1}, a_{i2}, \ldots, a_{in}]$ are called ***rows*** of *A*, and the matrices $\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{pj} \end{bmatrix}$ are called ***columns*** of *A*. Thus *A* has *p* rows and *n* columns; the ***size*** of *A* is said to be $\boldsymbol{p} \times \boldsymbol{n}$. A vector in $\boldsymbol{R}^n$ can be displayed as a matrix in the obvious way, either as a $1 \times n$ matrix, in which case the matrix is called a ***row vector***, or as a $\boldsymbol{n} \times 1$ matrix, called a ***column vector***. Thus the matrix representation of *f* is simply the matrix whose columns are the column vectors $f(\boldsymbol{e}_1), f(\boldsymbol{e}_1), \ldots, f(\boldsymbol{e}_n)$.

Example

Suppose $f: \boldsymbol{R}^3 \to \boldsymbol{R}^2$ is defined by

$$f(x_1, x_2, x_3) = (2x_1 - 3x_2 + x_3,\ \ -x_1 + 2x_2 - 5x_3).$$

So $f(\boldsymbol{e}_1) = f(1,0,0) = (2,-1)$, $f(\boldsymbol{e}_2) = f(0,1,0) = (-3,2)$, and $f(\boldsymbol{e}_3) = f(0,0,1) = (1,-5)$. The matrix which represents *f* is thus

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

Now the recipe for computing *f(x)* can be systematized by defining the product of a matrix $\boldsymbol{A}$ and a column vector $\boldsymbol{x}$. Suppose $\boldsymbol{A}$ is a $p \times n$ matrix and $\boldsymbol{x}$ is a $n \times 1$ column vector. For each $i = 1, 2, \ldots, p,$ let $\boldsymbol{r}_i$ denote the i^{th} row of $\boldsymbol{A}$. We define the product $\boldsymbol{Ax}$ to be the $p \times 1$ column vector given by

$$\boldsymbol{Ax} = \begin{bmatrix} \boldsymbol{r}_1 \cdot \boldsymbol{x} \\ \boldsymbol{r}_2 \cdot \boldsymbol{x} \\ \vdots \\ \boldsymbol{r}_p \cdot \boldsymbol{x} \end{bmatrix}.$$

If we consider all vectors to be represented by column vectors, then a linear function $f : \boldsymbol{R}^n \to \boldsymbol{R}^p$ is given by $f(\boldsymbol{x}) = \boldsymbol{Ax}$, where, of course, $\boldsymbol{A}$ is the matrix representation of f.

Example

Consider the preceding example:

$$f(x_1, x_2, x_3) = (2x_1 - 3x_2 + x_3, \; -x_1 + 2x_2 - 5x_3).$$

We found the matrix representing f to be

$$\boldsymbol{A} = \begin{bmatrix} 2 & -3 & 1 \\ -1 & 2 & -5 \end{bmatrix}.$$

Then

$$\boldsymbol{Ax} = \begin{bmatrix} 2 & -3 & 1 \\ -1 & 2 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 3x_2 + x_3 \\ -x_1 + 2x_2 - 5x_3 \end{bmatrix} = f(\boldsymbol{x})$$

Exercises

1. Find the matrix representation of each of the following linear functions:

a) $f(x_1, x_2) = (2x_1 - x_2,\ x_1 + 4x_2,\ -7x_1,\ 3x_1 + 5x_2)$.

b) $\boldsymbol{R}(t) = 4t\boldsymbol{i} - 5t\boldsymbol{j} - 2t\boldsymbol{k}$.

c) $L(x) = 6x$.

2. Let g be define by $g(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A} = \begin{bmatrix} 2 & -1 \\ -2 & 1 \\ 0 & -3 \\ 3 & 5 \end{bmatrix}$. Find $g(3,-9)$.

3. Let $f: \boldsymbol{R}^2 \rightarrow \boldsymbol{R}^2$ be the function in which $f(\mathbf{x})$ is the vector that results from rotating the vector $\boldsymbol{x}$ about the origin $\frac{\pi}{4}$ in the counterclockwise direction.

a)Explain why *f* is a linear function.

b)Find the matrix representation for *f*.

d)Find $f(4,-9)$.

4. Let $f: \boldsymbol{R}^2 \rightarrow \boldsymbol{R}^2$ be the function in which $f(\mathbf{x})$ is the vector that results from rotating the vector $\boldsymbol{x}$ about the origin θ in the counterclockwise direction. Find the matrix representation for *f*.

5. Suppose $g: \boldsymbol{R}^2 \rightarrow \boldsymbol{R}^2$ is a linear function such that $g(1,2) = (4,7)$ and $g(-2,1) = (2,2)$.

Find the matrix representation of *g*.

6. Suppose $f: \boldsymbol{R^n} \to \boldsymbol{R^p}$ and $g: \boldsymbol{R^p} \to \boldsymbol{R^q}$ are linear functions. Prove that the composition $g \circ f: \boldsymbol{R^n} \to \boldsymbol{R^q}$ is a linear function.

7. Suppose $f: \boldsymbol{R^n} \to \boldsymbol{R^p}$ and $g: \boldsymbol{R^n} \to \boldsymbol{R^p}$ are linear functions. Prove that the function $f + g: \boldsymbol{R^n} \to \boldsymbol{R^p}$ defined by $(f+g)(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$ is a linear function.

6.2 Matrix Algebra

Let us consider the composition $h = g \circ f$ of two linear functions $f: \boldsymbol{R^n} \to \boldsymbol{R^p}$ and $g: \boldsymbol{R^p} \to \boldsymbol{R^q}$. Suppose $\boldsymbol{A}$ is the matrix of *f* and $\boldsymbol{B}$ is the matrix of *g*. Let's see about the matrix $\boldsymbol{C}$ of *h*. We know the columns of $\boldsymbol{C}$ are the vectors $g(f(\boldsymbol{e}_j)), j = 1,2,\ldots,n$, where, of course, the vectors $\boldsymbol{e}_j$ are the coordinate vectors for $\boldsymbol{R^n}$. Now the columns of $\boldsymbol{A}$ are just the vectors $f(\boldsymbol{e}_j), j = 1,2,\ldots,n$. Thus the vectors $g(f(\boldsymbol{e}_j))$ are simply the products $\boldsymbol{B}f(\boldsymbol{e}_j)$. In other words, if we denote the columns of $\boldsymbol{A}$ by $\boldsymbol{k}_i$, $i = 1,2,\ldots,n$, so that $\boldsymbol{A} = [\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_n]$, then the columns of $\boldsymbol{C}$ are $\boldsymbol{Bk}_1, \boldsymbol{Bk}_2, \ldots, \boldsymbol{Bk}_n$, or in other words, $\boldsymbol{C} = [\boldsymbol{Bk}_1, \boldsymbol{Bk}_2, \ldots, \boldsymbol{Bk}_n]$.

Example

Let the matrix $\boldsymbol{B}$ of g be given by $\boldsymbol{B} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & -5 & 8 \\ 2 & 7 & -3 \\ 2 & -2 & 1 \end{bmatrix}$ and let the matrix $\boldsymbol{A}$ of f be given by $\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ -4 & -3 \end{bmatrix}$. Thus $f:\boldsymbol{R^2} \to \boldsymbol{R^3}$ and $g:\boldsymbol{R^3} \to \boldsymbol{R^4}$ (Note that for the composition $h = g \circ f$ to be defined, it must be true that the number of columns of $\boldsymbol{B}$ be the same as the number of rows of $\boldsymbol{A}$.). Now, $\boldsymbol{k}_1 = \begin{bmatrix} 3 \\ 1 \\ -4 \end{bmatrix}$ and $\boldsymbol{k}_2 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$, and so

$\mathbf{Bk}_1 = \begin{bmatrix} -5 \\ -40 \\ 25 \\ 0 \end{bmatrix}$ and $\mathbf{Bk}_2 = \begin{bmatrix} -5 \\ -35 \\ 25 \\ -3 \end{bmatrix}$. The matrix $\boldsymbol{C}$ of the composition is thus

$$\mathbf{C} = \begin{bmatrix} -5 & -5 \\ -40 & -35 \\ 25 & 25 \\ 0 & -3 \end{bmatrix}.$$

These results inspire us to define a product of matrices. Thus, if $\boldsymbol{B}$ is an $n \times p$ matrix, and $\boldsymbol{A}$ is a $p \times q$ matrix, the ***product BA*** of these matrices is defined to be the $n \times q$ matrix whose columns are the column vectors $\boldsymbol{Bk}_j$, where $\boldsymbol{k}_j$ is the j^{th} column of $\boldsymbol{A}$. Now we can simply say that the matrix representation of the composition of two linear functions is the product of the matrices representing the two functions.

There are several interesting and important things to note regarding matrix products. First and foremost is the fact that in general $\boldsymbol{BA} \neq \boldsymbol{AB}$, even when both products are defined (The product ***BA*** obviously defined only when the number of columns of ***B*** is the same as the number of rows of ***A***.). Next, note that it follows directly from the fact that $h \circ (f \circ g) = (h \circ f) \circ g$ that for ***C***(***BA)*** = (***CB***)***A.*** Since it does not matter where we insert the parentheses in a product of three or more matrices, we usually omit them entirely.

It should be clear that if *f* and *g* are both functions from $\boldsymbol{R}^n$ to $\boldsymbol{R}^p$, then the matrix representation for the sum $f + g\colon \boldsymbol{R}^n \to \boldsymbol{R}^p$ is simply the matrix

$$\boldsymbol{A}+\boldsymbol{B}=\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & & & \\ a_{p1}+b_{p1} & a_{p2}+b_{p2} & \dots & a_{pn}+b_{pn} \end{bmatrix},$$

where

$$\boldsymbol{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix}$$

is the matrix of *f*, and

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & & & \\ b_{p1} & b_{p2} & \cdots & b_{pn} \end{bmatrix}$$

is the matrix of g. Meditating on the properties of linear functions should convince you that for any three matrices (of the appropriate sizes) $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, it is true that

$$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}.$$

Similarly, for appropriately sized matrices, we have $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC}+\boldsymbol{BC}$.

Exercises

8. Find the products:

a) $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ b) $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix}$

c) $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}$ d) $\begin{bmatrix} 1 & -3 & 2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 5 \\ -2 & 3 \\ 0 & 2 \\ -3 & 4 \end{bmatrix}$

9. Find a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ b) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

10. Let $\boldsymbol{A}(\theta)$ be the 2×2 matrix for the linear function that rotates the plane θ counterclockwise. Compute the product $\boldsymbol{A}(\theta)\boldsymbol{A}(\eta)$, and use the result to give identities for $\cos(\theta + \eta)$ and $\sin(\theta + \eta)$ in terms of $\cos\theta$, $\cos\eta$, $\sin\theta$, and $\sin\eta$.

11. a)Find the matrix for the linear function that rotates $\boldsymbol{R}^3$ about the coordinate vector $\boldsymbol{j}$ by $\frac{\pi}{4}$ (In the positive direction, according to the usual "right hand rule" for rotation.).

b)Find a vector description for the curve that results from applying the linear transformation in a) to the curve $\boldsymbol{R}(t) = \cos t\boldsymbol{i} + \sin t\boldsymbol{j} + t\boldsymbol{k}$.

12. Suppose $f:\boldsymbol{R}^2 \to \boldsymbol{R}^2$ is linear. Let C be the circle of radius 1 and center at the origin. Find a vector description for the curve $f(C)$.

13. Suppose $g:\boldsymbol{R}^2 \to \boldsymbol{R}^n$ is linear. Suppose moreover that $g(1,1) = (2,3)$ and $g(-1,1) = (4,-5)$. Find the matrix of g.

Chapter Seven

Continuity, Derivatives, and All That

7.1 Limits and Continuity

Let $\boldsymbol{x_0} \in \boldsymbol{R}^n$ and $r > 0$. The set $B(\boldsymbol{a};r) = \{\boldsymbol{x} \in \boldsymbol{R}^n : |\boldsymbol{x} - \boldsymbol{a}| < r\}$ is called the ***open ball of radius r centered at*** $\boldsymbol{x_0}$. The ***closed ball of radius r centered at*** $\boldsymbol{x_0}$ is the set $\overline{B}(\boldsymbol{a};r) = \{\boldsymbol{x} \in \boldsymbol{R}^n : |\boldsymbol{x} - \boldsymbol{a}| \le r\}$. Now suppose $\boldsymbol{D} \subset \boldsymbol{R}^n$. A point $\boldsymbol{a} \in \boldsymbol{D}$ is called an ***interior point of D*** if there is an open ball $B(\boldsymbol{a};r) \subset \boldsymbol{D}$. The collection of all interior points of $\boldsymbol{D}$ is called the ***interior*** of D, and is usually denoted *int* $\boldsymbol{D}$. A set $\boldsymbol{U}$ is said to be ***open*** if $\boldsymbol{U}$ = *int* $\boldsymbol{U}$.

Suppose $f : \boldsymbol{D} \to \boldsymbol{R}^p$, where $\boldsymbol{D} \subset \boldsymbol{R}^n$ and suppose $\boldsymbol{a} \in \boldsymbol{R}^n$ is such that every open ball centered at $\boldsymbol{a}$ meets the domain $\boldsymbol{D}$. If $\boldsymbol{y} \in \boldsymbol{R}^p$ is such that for every $\varepsilon > 0$, there is a $\delta > 0$ so that $|f(\boldsymbol{x}) - \boldsymbol{y}| < \varepsilon$ whenever $0 < |\boldsymbol{x} - \boldsymbol{a}| < \delta$, then we say that $\boldsymbol{y}$ is the limit of f at $\boldsymbol{a}$. This is written

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = \boldsymbol{y},$$

and $\boldsymbol{y}$ is called the ***limit of f at a***.

Notice that this agrees with our previous definitions in case $\boldsymbol{n} = 1$ and $\boldsymbol{p} = 1,2$, or 3. The usual properties of limits are relatively easy to establish:

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}}(f(\boldsymbol{x}) + g(\boldsymbol{x})) = \lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) + \lim_{\boldsymbol{x} \to \boldsymbol{a}} g(\boldsymbol{x}), \text{ and}$$

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} af(\boldsymbol{x}) = a \lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}).$$

Now we are ready to say what we mean by a continuous function $f : \boldsymbol{D} \to \boldsymbol{R}^p$, where $\boldsymbol{D} \subset \boldsymbol{R}^n$. Again this definition will not contradict our previous lower dimensional definitions. Specifically, we say that f is ***continuous at*** $\boldsymbol{a} \in \boldsymbol{D}$ if $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = f(\boldsymbol{a})$. If f is continuous at each point of its domain $\boldsymbol{D}$, we say simply that f is ***continuous***.

Example

Every linear function is continuous. To see this, suppose $f:\boldsymbol{R}^n \to \boldsymbol{R}^p$ is linear and $\boldsymbol{a} \in \boldsymbol{R}^n$. Let $\varepsilon > 0$. Now let $M = \max\{|f(\boldsymbol{e}_1)|,|f(\boldsymbol{e}_2)|,\ldots,|f(\boldsymbol{e}_n)|\}$ and let $\delta = \dfrac{\varepsilon}{nM}$. Then for $\boldsymbol{x}$ such that $0 < |\boldsymbol{x}-\boldsymbol{a}| < \delta$, we have

$$\begin{aligned}
|f(\boldsymbol{x}) - f(\boldsymbol{a})| &= |f(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \ldots + x_n\boldsymbol{e}_n) - f(a_1\boldsymbol{e}_1 + a_2\boldsymbol{e}_2 + \ldots + a_n\boldsymbol{e}_n)| \\
&= |(x_1 - a_1) f(\boldsymbol{e}_1) + (x_2 - a_2) f(\boldsymbol{e}_2) + \ldots + (x_n - a_n) f(\boldsymbol{e}_n)| \\
&\le |x_1 - a_1||f(\boldsymbol{e}_1)| + |x_2 - a_2||f(\boldsymbol{e}_2)| + \ldots + |x_n - a_n||f(\boldsymbol{e}_n)| \\
&\le (|x_1 - a_1| + |x_2 - a_2| + \ldots + |x_n - a_n|)M \\
&\le n|\boldsymbol{x}-\boldsymbol{a}|M \\
&< \varepsilon
\end{aligned}$$

Thus $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = f(\boldsymbol{a})$ and so f is continuous.

Another Example

Let $f:\boldsymbol{R}^2 \to \boldsymbol{R}$ be defined by $f(\boldsymbol{x}) = f(x_1, x_2) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2}, & \text{for } x_1^2 + x_2^2 \neq 0 \\ 0, & \text{otherwise} \end{cases}$.

Let's see about $\lim_{\boldsymbol{x}\to(0,0)} f(\boldsymbol{x})$. Let $\boldsymbol{x} = \alpha(1,1)$. Then for all $\alpha \neq 0$, we have

$$f(\boldsymbol{x}) = f(\alpha,\alpha) = \frac{\alpha^2}{\alpha^2 + \alpha^2} = \frac{1}{2}.$$

Now. let $\boldsymbol{x} = \alpha(1,0) = (\alpha, 0)$. It follows that all $\alpha \neq 0$, $f(\boldsymbol{x}) = 0$. What does this tell us? It tells us that for any $\delta > 0$, there are vectors $\boldsymbol{x}$ with $0 < |\boldsymbol{x} - (0,0)| < \delta$ such that $f(\boldsymbol{x}) = \frac{1}{2}$ and such that $f(\boldsymbol{x}) = 0$. This, of course, means that $\lim_{\boldsymbol{x} \to (0,0)} f(\boldsymbol{x})$ does not exist.

7.2 Derivatives

Let $f : \boldsymbol{D} \to \boldsymbol{R}^p$, where $\boldsymbol{D} \subset \boldsymbol{R}^n$, and let $\boldsymbol{x_0} \in \text{int}\, \boldsymbol{D}$. Then f is ***differentiable at*** $\boldsymbol{x_0}$ if there is a linear function $\boldsymbol{L}$ such that

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{1}{|\boldsymbol{h}|}[f(\boldsymbol{x_0} + \boldsymbol{h}) - f(\boldsymbol{x_0}) - \boldsymbol{L}(\boldsymbol{h})] = \boldsymbol{0}.$$

The linear function $\boldsymbol{L}$ is called the ***derivative of f at*** $\boldsymbol{x_0}$. It is usual to identify the linear function $\boldsymbol{L}$ with its matrix representation and think of the derivative at a $p \times n$ matrix. Note that in case $n = p = 1$, the matrix $\boldsymbol{L}$ is simply the 1×1 matrix whose sole entry is the every day grammar school derivative of f.

Now, how do find the derivative of f? Suppose f has a derivative at $\boldsymbol{x_0}$. First, let $\boldsymbol{h} = t\boldsymbol{e}_j = (0,0,\ldots,0,t,0,\ldots,0)$. Then

$$f(\boldsymbol{x} + \boldsymbol{h}) = f(x_1, x_2, \ldots, x_j + t, \ldots, x_n) = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_j + t, \ldots, x_n) \\ f_2(x_1, x_2, \ldots, x_j + t, \ldots, x_n) \\ \vdots \\ f_p(x_1, x_2, \ldots, x_j + t, \ldots, x_n) \end{bmatrix},$$

and

$$\boldsymbol{Lh} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & & & \\ m_{p1} & m_{p2} & \cdots & m_{pn} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ t \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} m_{1j}t \\ m_{2j}t \\ \vdots \\ m_{pj}t \end{bmatrix},$$

where $\boldsymbol{x_0} = (x_1, x_2, \ldots, x_n)$, *etc.*

Now then,

$$\frac{1}{|\mathbf{h}|}[f(\mathbf{x_0} + \mathbf{h}) - f(\mathbf{x_0}) - \mathbf{L}(\mathbf{h})]$$

$$= \frac{1}{t}\begin{bmatrix} f_1(x_1, x_2, \ldots, x_j + t, \ldots, x_n) - f_1(x_1, x_2, \ldots, x_n) - m_{1j}t \\ f_2(x_1, x_2, \ldots, x_j + t, \ldots, x_n) - f_2(x_1, x_2, \ldots, x_n) - m_{2j}t \\ \vdots \\ f_p(x_1, x_2, \ldots, x_j + t, \ldots, x_n) - f_p(x_1, x_2, \ldots, x_n) - m_{pj}t \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{f_1(x_1, x_2, \ldots, x_j + t, \ldots, x_n) - f_1(x_1, x_2, \ldots, x_n)}{t} - m_{1j} \\ \dfrac{f_2(x_1, x_2, \ldots, x_j + t, \ldots, x_n) - f_2(x_1, x_2, \ldots, x_n)}{t} - m_{2j} \\ \vdots \\ \dfrac{f_p(x_1, x_2, \ldots, x_j + t, \ldots, x_n) - f_p(x_1, x_2, \ldots, x_n)}{t} - m_{pj} \end{bmatrix}$$

Meditate on this vector. For each component,

$$\lim_{t \to 0} \frac{f_i(x_1, x_2, \ldots, x_j + t, \ldots, x_n) - f_i(x_1, x_2, \ldots, x_n)}{t}$$

$$= \frac{d}{ds} f_i(x_1, x_2, \ldots, s, \ldots, x_n)\Big|_{s = x_j}$$

This derivative has a name. It is called the ***partial derivative of*** f_i ***with respect to the*** j^{th} ***variable***. There are many different notations for the partial derivatives of a function $g(x_1, x_2, \ldots, x_n)$. The two most common are:

$$g_{,j}(x_1, x_2, \ldots, x_n)$$
$$\frac{\partial}{\partial x_j} g(x_1, x_2, \ldots, x_n)$$

The requirement that $\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{1}{|\boldsymbol{h}|}[f(\boldsymbol{x_0} + \boldsymbol{h}) - f(\boldsymbol{x_0}) - \boldsymbol{L}(\boldsymbol{h})] = \boldsymbol{0}$ now translates into

$$m_{ij} = \frac{\partial f_i}{\partial x_j},$$

and, *mirabile dictu*, we have found the matrix $\boldsymbol{L}$!

Example

Let $f : \boldsymbol{R^2} \to \boldsymbol{R^2}$ be given by $f(x_1. x_2) = \begin{bmatrix} 3x_1 \sin x_2 \\ x_1^3 + x_1 x_2^2 \end{bmatrix}$. Assume f is differentiable and let's find the derivative (more precisely, the matrix of the derivative. This matrix will, of course, be 2×2: $\boldsymbol{L} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$. Now

$$f_1(x_1, x_2) = 3x_1 \sin x_2, \text{ and}$$
$$f_2(x_1, x_2) = x_1^3 + x_1 x_2^2$$

Compute the partial derivatives:

$$\frac{\partial f_1}{\partial x_1} = 3 \sin x_2$$
$$\frac{\partial f_2}{\partial x_1} = 3x_1^2 + x_2^2,$$

and

$$\frac{\partial f_1}{\partial x_2} = 3x_1 \cos x_2$$
$$\frac{\partial f_2}{\partial x_2} = 2x_1 x_2 .$$

The derivative is thus

$$\boldsymbol{L} = \begin{bmatrix} 3\sin x_2 & 3x_1 \cos x_2 \\ 3x_1^2 + x_2^2 & 2x_1 x_2 \end{bmatrix}.$$

We now know how to find the derivative of f at $\boldsymbol{x}$ if we know the derivative exists; but how do we know when there is a derivative? The function f is differentiable at $\boldsymbol{x}$ if the partial derivatives exist *and are continuous.* It should be noted that it is not sufficient just for the partial derivatives to exist.

Exercises

1. Find all partial derivatives of the given functions:

a) $f(x, y) = x^2 y^3$

b) $f(x, y, z) = x^2 yz + z\cos(xy)$

c) $g(x_1, x_2, x_3) = x_1 x_2 x_3 + x_2$

d) $h(x_1, x_2, x_3, x_4) = \dfrac{x_3 \sin(e^{x_1})}{x_2 + x_4}$

2. Find the derivative of the linear function whose matrix is $\begin{bmatrix} 1 & 3 & 2 \\ -2 & 7 & 0 \end{bmatrix}$.

3. What is the derivative a linear function whose matrix is $\boldsymbol{A}$?

4. Find the derivative of $\boldsymbol{R}(t) = \cos t\boldsymbol{i} + \sin t\boldsymbol{j} + t\boldsymbol{k}$.

5. Find the derivative of $f(x, y) = x^2 y^3$.

6. Find the derivative of

$$f(x_1, x_2, x_3) = \begin{bmatrix} x_1 x_3 + e^{x_2} \\ \dfrac{x_3 \log(x_1 + x_2^2)}{x_2} \\ x_1 x_3^2 + 5 \end{bmatrix}.$$

7.3 The Chain Rule

Recall from elementary one dimensional calculus that if a function is differentiable at a point, it is also continuous there. The same is true here in the more general setting of functions $f : \boldsymbol{R}^n \to \boldsymbol{R}^p$. Let's see why this is so. Suppose f is differentiable at $\boldsymbol{a}$ with derivative $\boldsymbol{L}$. Let $\boldsymbol{h} = \boldsymbol{x} - \boldsymbol{a}$. Then $\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = \lim_{\boldsymbol{h} \to \boldsymbol{0}} f(\boldsymbol{a} + \boldsymbol{h})$. Now,

$$f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a}) = |\boldsymbol{h}| \left[\frac{f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a}) - \boldsymbol{L}(h)}{|\boldsymbol{h}|} \right] - \boldsymbol{L}(\boldsymbol{h})$$

Now look at the limit of this as $|\boldsymbol{h}| \to 0$:

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \left[\frac{f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a}) - \boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|} \right] = \boldsymbol{0}$$

because f is differentiable at $\boldsymbol{a}$, and $\lim_{\boldsymbol{h}\to\boldsymbol{0}} \boldsymbol{L}(\boldsymbol{h}) = \boldsymbol{L}(\boldsymbol{0}) = \boldsymbol{0}$ because the linear function $\boldsymbol{L}$ is continuous. Thus $\lim_{\boldsymbol{h}\to\boldsymbol{0}}(f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a})) = \boldsymbol{0}$, or $\lim_{\boldsymbol{h}\to\boldsymbol{0}} f(\boldsymbol{a}+\boldsymbol{h}) = f(\boldsymbol{a})$, which means f is continuous at $\boldsymbol{a}$.

Next, let's see what the celebrated chain rule looks like in higher dimensions. Let $f:\boldsymbol{R}^n \to \boldsymbol{R}^p$ and $g:\boldsymbol{R}^p \to \boldsymbol{R}^q$. Suppose the derivative of f at $\boldsymbol{a}$ is $\boldsymbol{L}$ and the derivative of g at $f(\boldsymbol{a})$ is $\boldsymbol{M}$. We go on a quest for the derivative of the composition $g \circ f:\boldsymbol{R}^n \to \boldsymbol{R}^q$ at $\boldsymbol{a}$. Let $r = g \circ f$, and look at $r(\boldsymbol{a}+\boldsymbol{h}) - r(\boldsymbol{a}) = g(f(\boldsymbol{a}+\boldsymbol{h})) - g(f(\boldsymbol{a}))$. Next, let $\boldsymbol{k} = f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a})$. Then we may write

$$
\begin{aligned}
r(\boldsymbol{a}+\boldsymbol{h}) - r(\boldsymbol{a}) - \boldsymbol{ML}(\boldsymbol{h}) &= g(f(\boldsymbol{a}+\boldsymbol{h})) - g(f(\boldsymbol{a})) - \boldsymbol{ML}(\boldsymbol{h}) \\
&= g(f(\boldsymbol{a})+\boldsymbol{k}) - g(f(\boldsymbol{a})) - \boldsymbol{M}(\boldsymbol{k}) + \boldsymbol{M}(\boldsymbol{k}) - \boldsymbol{ML}(\boldsymbol{h}). \\
&= g(f(\boldsymbol{a})+\boldsymbol{k}) - g(f(\boldsymbol{a})) - \boldsymbol{M}(\boldsymbol{k}) + \boldsymbol{M}(\boldsymbol{k} - \boldsymbol{L}(\boldsymbol{h}))
\end{aligned}
$$

Thus,

$$
\frac{r(\boldsymbol{a}+\boldsymbol{h}) - r(\boldsymbol{a}) - \boldsymbol{ML}(\boldsymbol{h})}{|\boldsymbol{h}|} = \frac{g(f(\boldsymbol{a})+\boldsymbol{k}) - g(f(\boldsymbol{a})) - \boldsymbol{M}(\boldsymbol{k})}{|\boldsymbol{h}|} + \boldsymbol{M}\left(\frac{\boldsymbol{k} - \boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right)
$$

Now we are ready to see what happens as $|\boldsymbol{h}| \to 0$. look at the second term first:

$$
\begin{aligned}
\lim_{\boldsymbol{h}\to\boldsymbol{0}} \boldsymbol{M}\left(\frac{\boldsymbol{k} - \boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right) &= \lim_{\boldsymbol{h}\to\boldsymbol{0}} \boldsymbol{M}\left(\frac{f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a}) - \boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right) = \boldsymbol{M}\left(\lim_{\boldsymbol{h}\to\boldsymbol{0}}\left(\frac{f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a}) - \boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right)\right) \\
&= \boldsymbol{M}(\boldsymbol{0}) = \boldsymbol{0}
\end{aligned}
$$

since $\boldsymbol{L}$ is the derivative of f at $\boldsymbol{a}$ and $\boldsymbol{M}$ is linear, and hence continuous.

Now we need to see what happens to the term

$$
\lim_{\boldsymbol{h}\to\boldsymbol{0}} \left(\frac{g(f(\boldsymbol{a})+\boldsymbol{k}) - g(f(\boldsymbol{a})) - \boldsymbol{M}(\boldsymbol{k})}{|\boldsymbol{h}|}\right).
$$

This is a bit tricky. Note first that because *f* is differentiable at $\boldsymbol{a}$, we know that

$$\frac{|\boldsymbol{k}|}{|\boldsymbol{h}|} = \frac{|f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a})|}{|\boldsymbol{h}|}$$

behaves nicely as $|\boldsymbol{h}| \to 0$. Next,

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \left(\frac{g(f(\boldsymbol{a})+\boldsymbol{k}) - g(f(\boldsymbol{a})) - \boldsymbol{M}(\boldsymbol{k})}{|\boldsymbol{h}|} \; \frac{|\boldsymbol{k}|}{|\boldsymbol{k}|} \right)$$
$$= \lim_{\boldsymbol{h}\to\boldsymbol{0}} \left(\frac{g(f(\boldsymbol{a})+\boldsymbol{k}) - g(f(\boldsymbol{a})) - \boldsymbol{M}(\boldsymbol{k})}{|\boldsymbol{k}|} \right) \frac{|\boldsymbol{k}|}{|\boldsymbol{h}|} = \boldsymbol{0}$$

since the derivative of g at $f(\boldsymbol{a})$ is $\boldsymbol{M}$, and $\frac{|\boldsymbol{k}|}{|\boldsymbol{h}|}$ is well-behaved. Finally at last, we have shown that

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \left(\frac{r(\boldsymbol{a}+\boldsymbol{h}) - r(\boldsymbol{a}) - \boldsymbol{ML}(\boldsymbol{h})}{|\boldsymbol{h}|} \right) = \boldsymbol{0},$$

which means the derivative of the composition $r = g \circ f$ is simply the composition, or matrix product, of the derivatives. What could be more pleasing from an esthetic point of view!

Example

Let $f(t) = (t^2, 1+t^3)$ and $g(x_1, x_2) = (2x_1 - x_2)^3$, and let $r = g \circ f$. First, we shall find the derivative of r at $t = 2$ using the Chain Rule. The derivative of *f* is

$$\boldsymbol{L} = \begin{bmatrix} 2t \\ 3t^2 \end{bmatrix},$$

and the derivative of g is

$$\boldsymbol{M} = \begin{bmatrix} 6(2x_1 - x_2)^2 & -3(2x_1 - x_2)^2 \end{bmatrix}.$$

At $t = 2$, $\boldsymbol{L} = \begin{bmatrix} 4 \\ 12 \end{bmatrix}$; and at $g(f(2)) = g(4,9)$, $\boldsymbol{M} = \begin{bmatrix} 6 & -3 \end{bmatrix}$. Thus the derivative of the composition is $\boldsymbol{ML} = \begin{bmatrix} 6 & -3 \end{bmatrix}\begin{bmatrix} 4 \\ 12 \end{bmatrix} = [-12] = -12$.

Now for fun, let's find an explicit recipe for r and differentiate: $r(t) = g(f(t)) = g(t^2, 1 + t^3) = (2t^2 - 1 - t^3)^3$. Thus $r'(t) = 3(2t^2 - 1 - t^3)^2(4t - 3t^2)$, and so $r'(2) = 3(1)(8 - 12) = -12$. It is, of course, very comforting to get the same answer as before.

There are several different notations for the matrix of the derivative of $f: \boldsymbol{R}^n \to \boldsymbol{R}^p$ at $\boldsymbol{x} \in \boldsymbol{R}^n$ The most usual is simply $f'(\boldsymbol{x})$.

Exercises

7. Let $g(x_1, x_2, x_3) = (x_1 x_3, x_2 x_3 + 1)$ and $f(x_1, x_2) = (x_1 x_2 \sin x_1, x_1 + 3x_2, x_2 - 2x_1^2)$.
Find the derivative of $g \circ f$ at (2,-4).

8. Let $u(x, y, z) = (x + y^2, 2xy, x \sin y, x^3 y^2)$ and $v(r, s, t, q) = (r + s - q^3, (r - t)e^s)$.
a)Which, if either, of the composition functions $u \circ v$ or $v \circ u$ is defined? Explain.
b)Find the derivative of your answer to part a).

9. Let $f(x, y) = (e^{(x+y)}, e^{(x-y)})$ and $g(x, y) = (x - y^3, x^2 + y)$.

a)Find the derivative of $f \circ g$ at the point (1,-2).

b)Find the derivative of $g \circ f$ at the point (1,-2).

c) Find the derivative of $f \circ f$ at the point (1,-2).

d) Find the derivative of $g \circ g$ at the point (1,-2).

10. Suppose $r = t^2 \cos t$ and $t = x^2 - 3y^2$. Find the partial derivatives $\frac{\partial r}{\partial x}$ and $\frac{\partial r}{\partial y}$.

7.4 More Chain Rule Stuff

In the everyday cruel world, we seldom compute the derivative of the composition of two functions by explicitly multiplying the two derivative matrices. Suppose, as usual, we have $r = g \circ f : \boldsymbol{R}^n \to \boldsymbol{R}^q$. The the derivative is, as we now know,

$$r'(\boldsymbol{x}) = r'(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \cdots & \frac{\partial r_1}{\partial x_n} \\ \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} & \cdots & \frac{\partial r_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial r_p}{\partial x_1} & \frac{\partial r_p}{\partial x_2} & \cdots & \frac{\partial r_p}{\partial x_n} \end{bmatrix}.$$

We can thus find the derivative using the Chain Rule only in the very special case in which the compsite function is real valued. Specifically, suppose $g : \boldsymbol{R}^p \to \boldsymbol{R}$ and $f : \boldsymbol{R}^n \to \boldsymbol{R}^p$. Let $r = g \circ f$. Then r is simply a real-valued function of $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. Let's use the Chain Rule to find the partial derivatives.

$$r'(\mathbf{x}) = \begin{bmatrix} \frac{\partial r}{\partial x_1} & \frac{\partial r}{\partial x_2} & \cdots & \frac{\partial r}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial g}{\partial y_1} & \frac{\partial g}{\partial y_2} & \cdots & \frac{\partial g}{\partial y_p} \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial f_p}{\partial x_1} & \frac{\partial f_p}{\partial x_2} & \cdots & \frac{\partial f_p}{\partial x_n} \end{bmatrix}$$

Thus makes it clear that

$$\frac{\partial r}{\partial x_j} = \frac{\partial g}{\partial y_1}\frac{\partial f_1}{\partial x_j} + \frac{\partial g}{\partial y_2}\frac{\partial f_2}{\partial x_j} + \cdots + \frac{\partial g}{\partial y_p}\frac{\partial f_p}{\partial x_j}.$$

Frequently, engineers and other malefactors do not use a different name for the composition $g \circ f$, and simply use the name g to denote both the composition $g \circ f(x_1, x_2, \ldots, x_n) = g(f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_p(x_1, x_2, \ldots, x_n))$ and the function g given by $g(\mathbf{y}) = g(y_1, y_2, \ldots, y_p)$. Since $y_j = f_j(x_1, x_2, \ldots, x_n)$, these same folks also frequently just use y_j to denote the function f_j. The Chain Rule given above then looks even nicer:

$$\frac{\partial g}{\partial x_j} = \frac{\partial g}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial g}{\partial y_2}\frac{\partial y_2}{\partial x_j} + \cdots + \frac{\partial g}{\partial y_p}\frac{\partial y_p}{\partial x_j}.$$

Example

Suppose $g(x, y, z) = x^2 y + ye^z$ and $x = s + t$, $y = st^3$, and $z = s^2 + 3t^2$. Let us find the partial derivatives $\frac{\partial g}{\partial r}$ and $\frac{\partial g}{\partial t}$. We know that

$$\frac{\partial g}{\partial s} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial s}$$
$$= 2xy(1) + (x^2 + e^z)t^3 + ye^z(2s)$$
$$= 2xy + (x^2 + e^z)t^3 + 2sye^z$$

Similarly,

$$\frac{\partial g}{\partial t} = \frac{\partial g}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial g}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial g}{\partial z}\frac{\partial z}{\partial t}$$
$$= 2xy(1) + (x^2 + e^z)3st^2 + ye^z(6t)$$
$$= 2xy + 3(x^2 + e^z)st^2 + 6tye^z$$

These notational shortcuts are fine and everyone uses them; you should, however, be aware that it is a practice sometimes fraught with peril. Suppose, for instance, you have $g(x,y,z) = x^2 + y^2 + z^2$, and $x = t + z$, $y = t^2 + 2z$, and $z = t^3$. Now it is not at all clear what is meant by the symbol $\frac{\partial g}{\partial z}$. Meditate on this.

Exercises

11. Suppose $g(x,y) = f(x-y, y-s)$. Find $\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}$.

12. Suppose the temperature T at the point (x,y,z) in space is given by the function $T(x,y,z) = x^2 + xyz - zy^2$. Find the derivative with respect to t of a particle moving along the curve described by $\boldsymbol{r}(t) = \cos t\boldsymbol{i} + \sin t\boldsymbol{j} + 3t\boldsymbol{k}$.

13. Suppose the temperature T at the point (x, y, z) in space is given by the function $T(x,y,z) = x^2 + y^2 + z^2$. A particle moves along the curve described by $\boldsymbol{r}(t) = \sin \pi t\boldsymbol{i} + \cos \pi t\boldsymbol{j} + (t^2 - 2t + 2)\boldsymbol{k}$. Find the coldest point on the trajectory.

14. Let $r(x, y) = f(x)g(y)$, and suppose $x = t$ and $y = t$. Use the Chain Rule to find $\dfrac{dr}{dt}$.

Chapter Eight

$f: \boldsymbol{R}^n \to \boldsymbol{R}$

8.1 Introduction

We shall now turn our attention to the very important special case of functions that are real, or scalar, valued. These are sometimes called ***scalar fields***. In the very, but important, special subcase in which the dimension of the domain space is 2, we can actually look at the graph of a function. Specifically, suppose $f: \boldsymbol{R}^n \to \boldsymbol{R}$. The collection $S = \{(x_1, x_2, x_3) \in \boldsymbol{R}^3 : f(x_1, x_2) = x_3\}$ is called the ***graph of f***. If f is a reasonably nice function, then S is what we call a surface. We shall see more of this later. Let us now return to the general case of a function $f: \boldsymbol{R}^n \to \boldsymbol{R}$. The derivative of f is just a row vector $f'(\boldsymbol{x}) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \cdots \quad \frac{\partial f}{\partial x_n}\right]$. It is frequently called the ***gradient of f*** and denoted *grad* f or ∇f.

8.2 The Directional Derivative

In the applications of scalar fields it is of interest to talk of the rate of change of the function in a specified direction. Suppose, for instance, the function $T(x, y, z)$ gives the temperature at points (x, y, z) in space, and we might want to know the rate at which the temperature changes as we move in a specified direction. Let $f: \boldsymbol{R}^n \to \boldsymbol{R}$, let $\boldsymbol{a} \in \boldsymbol{R}^n$, and let $\boldsymbol{u} \in \boldsymbol{R}^n$ be a vector such that $|\boldsymbol{u}| = 1$. Then the ***directional derivative of f at a in the direction of the vector u*** is defined to be

$$D_{\boldsymbol{u}} f(\boldsymbol{a}) = \frac{d}{dt} f(\boldsymbol{a} + t\boldsymbol{u})\Big|_{t=0}.$$

Now that we are experts on the Chain Rule, we know at once how to compute such a thing. It is simply

$$D_{\boldsymbol{u}} f(\boldsymbol{a}) = \frac{d}{dt} f(\boldsymbol{a} + t\boldsymbol{u})\Big|_{t=0} = \quad f \quad \boldsymbol{u}\,.$$

Example

The surface of a mountain is the graph of $f(x, y) = 700 - x^2 - 5y^2$. In other words, at the point (x, y), the height is $f(x, y)$. The positive y-axis points North, and, of course, then the positive x-axis points East. You are on the mountain side above the point (2, 4) and begin to walk Southeast. What is the slope of the path at the starting point? Are you going uphill or downhill? (Which!?).

The answers to these questions call for the directional derivative. We know we are at the point $\boldsymbol{a} = (2,4)$, but we need a unit vector $\boldsymbol{u}$ in the direction we are walking. This is, of course, just $\boldsymbol{u} = \frac{1}{\sqrt{2}}(1,-1)$. Next we compute the gradient $\quad f(x,y) = [-2x, -10y]$. At the point $\boldsymbol{a}$ this becomes $\quad f(2,4) = [-2,-40]$, and at last we have $\quad f \quad \boldsymbol{u} = \frac{-2+40}{\sqrt{2}} = \frac{38}{\sqrt{2}}$. This gives us the slope of the path; it is positive so we are going uphill. Can you tell in which direction the path will be level?

Another Example

The temperature in space is given by $T(x, y, z) = x^2 y + yz^3$. From the point (1,1,1), in which direction does the temperature increase most rapidly?

We clearly need the direction in which the directional derivative is largest. The directional derivative is simply $\quad T \quad \boldsymbol{u} = | \quad T|\cos\theta$, where θ is the angle between $\quad T$ and $\boldsymbol{u}$. Anyone can see that this will be largest when $\theta = 0$. Thus T in creases most rapidly in the direction of the gradient of T. Here that direction is $[2xy, x^2 + z^3, 3yz^2]$. At (1,1,1), this becomes [2, 2, 3].

Exercises

1. Find the derivative of $f(x, y, z) = x \log z + 2xy$ at (1, 2, 1) in the direction of the vector [1, 2 ,2].

2. Find the derivative of $f(x, y, z) = x \cos y + 3z^3 - xz$ at (1, π, 1) in the direction of the vector [3, -2 ,2].

3. Find the directions in which $g(x, y) = x^2 y + e^{xy} \sin y$ increases and decreases most rapidly from the point (1, 0).

4. The surface of a hill is the graph of the equation $z = 1000 + x^2 - x^4 - y^2$. You stand on the hill above the point (5,3) and pour out a glass of water. In which direct will it begin to run? Explain.

5. The position of a particle at time t is given by $\boldsymbol{r}(t) = 3(t^2 - \sin t)\boldsymbol{i} + t\boldsymbol{j} - \cos t\boldsymbol{k}$, and the position of another particle is $\boldsymbol{R}(t) = t^2\boldsymbol{i} + (t^3 + t)\boldsymbol{j} + \sin t\boldsymbol{k}$. At time $t = \pi$, what is the rate of change of the distance between the two particles? Are they getting closer to one another, or are they getting farther apart? (Which!) Explain.

8.3 Surface Normals

Let $f : \boldsymbol{R^3} \to \boldsymbol{R}$ be a function and let c be some constant. Recall that the set $S = \{(x, y, z) \quad \boldsymbol{R^3} : f(x, y, z) = c\}$ is called a ***level set***, or ***level surface***, of the function f. Suppose $\boldsymbol{r}(t) = x(t)\boldsymbol{i} + y(t)\boldsymbol{j} + z(t)\boldsymbol{k}$ describes a curve in $\boldsymbol{R^3}$ that lies on the surface S. This means, of course, that $f(\boldsymbol{r}(t)) = f(x(t), y(t), z(t)) = c$. Now look at the derivative with respect to t of this equation:

$$\frac{d}{dt} f(\boldsymbol{r}(t)) = \quad f \quad \boldsymbol{r}'(t) = 0\,.$$

In other words, the gradient of f and the tangent to the curve are perpendicular. Note there was nothing special about our choice of $\boldsymbol{r}(t)$; it is *any* curve on the surface. The gradient $\;f$ is thus ***perpendicular***, or ***normal*** to the surface $f(x, y, z) = c$.

Example

Suppose we want to find an equation of the plane tangent to the surface

$$x^2 + 3y^2 + 2z^2 = 12$$

at the point (1, -1, 2). For an equation of a plane, we need a point $\boldsymbol{a}$ on the plane and a vector $\boldsymbol{N}$ normal to the plane. Then the equation we seek is simply $\boldsymbol{N} \quad (\boldsymbol{x} - \boldsymbol{a}) = 0$, where $\boldsymbol{x} = (x, y, z)$. In the case at hand, we have a point on the plane: $\boldsymbol{a}$ = (1, -1, 2). Let's find a normal vector $\boldsymbol{N}$. We have just learned that the gradient of $f(x, y, z) = x^2 + 3y^2 + 2z^2$ does the job.

$$f(x, y, z) = [2x, 6y, 4z],$$

and so $\boldsymbol{N} = \quad f(1,-1,2) = [2,-6,8]$. The tangent plane is thus given by the equation $\boldsymbol{N} \quad (\boldsymbol{x} - \boldsymbol{a}) = 0$, which in this case is

$$2(x-1) - 6(y+1) + 8(z-2) = 0 .$$

You should note that the discussion here didn't depend on the dimension of the domain. Thus if $f : \boldsymbol{R^2} \rightarrow \boldsymbol{R}$, then the set $\{(x,y) \quad \boldsymbol{R^2} : f(x,y) = c\}$ is a level curve of *f*, and the gradient of *f* is normal to such a curve.

Combining these results with what we know about the directional derivative, we see that at a point the value of a function increases most rapidly in a direction normal to the level set passing through that point. On a contour map of a portion of the Earth's surface, for example, the steepest path is in the direction normal to the contour lines.

Exercises

6. Find an equation for the plane tangent to the surface $z = x^2 + 2y^2$ at the point (1,1,3).

7. Find an equation for the plane tangent to the surface $z = \log(x^2 + y^2)$ at the point $(1,0,0)$.

8. Find an equation for the plane tangent to the surface $\cos \pi x - x^2 y + e^{xz} + yz = 4$ at the point (0,1,2).

9. Find an equation of the straight line tangent to the curve of intersection of the surfaces $x^3 + 3x^2y^2 + y^3 + 4xy - z^2 = 0$ and $x^2 + y^2 + z^2 = 11$ at the point (1, 1, 3).

8.4 Maxima and Minima

Let $f : \boldsymbol{R}^n \to \boldsymbol{R}$. A point $\boldsymbol{a}$ in the domain of f is called a ***local minimum*** if there is an open ball $B(\boldsymbol{a};r)$ centered at $\boldsymbol{a}$ such that $f(\boldsymbol{x}) - f(\boldsymbol{a}) \geq 0$ for all $\boldsymbol{x} \quad B(\boldsymbol{a};r)$. If f is a nice function, then this means the directional derivative $D_{\boldsymbol{u}} f(\boldsymbol{a}) \geq 0$ for all unit vectors $\boldsymbol{u}$. In other words, $\quad f(\boldsymbol{a}) \ \boldsymbol{u} \geq 0$. Then it must be true that both $\quad f(\boldsymbol{a}) \ \boldsymbol{u} \geq 0$ and $- \ f(\boldsymbol{a}) \ \boldsymbol{u} = \ f(\boldsymbol{a}) \ (-\boldsymbol{u}) \geq 0$. This can be so for every $\boldsymbol{u}$ only of $\quad f(\boldsymbol{a}) = \boldsymbol{0}$. Thus f has a local minimum at a point at which it has a derivative only if the derivative is zero there.

You should guess the definition of a local maximum and see why it must be true that the gradient is zero at such a point. Thus if $\boldsymbol{a}$ is a local minimum or a local maximum of f, and if f has a derivative at $\boldsymbol{a}$, then the derivative $\quad f(\boldsymbol{a}) = 0$. You should be aware of the fact that here, just as in Mrs. Turner's elementary calculus class, the converse is not necessarily true. We may have $\quad f(\boldsymbol{a}) = \boldsymbol{0}$ without $\boldsymbol{a}$ being either a local minimum or a local maximum.

Example

Let us find all local maxima and local minima of the function

$$f(x,y) = x^2 + xy + y^2 + 3x - 3y + 4\,.$$

Meditate on just how should proceed. This function clearly has a derivative everywhere, so at any local maximum or minimum, this derivative, or gradient, must be zero. So let's begin by finding all points at which $\quad f(\boldsymbol{a}) = \boldsymbol{0}$. In other words, we want (x, y) at which $\dfrac{\partial f}{\partial x} = 0$ and $\dfrac{\partial f}{\partial y} = 0$:

$$\frac{\partial f}{\partial x} = 2x + y + 3 = 0$$

$$\frac{\partial f}{\partial y} = x + 2y - 3 = 0$$

We are thus faced with the border-line trivial problem of solving the system of equations

$$\begin{aligned} 2x + y &= -3 \\ x + 2y &= 3 \end{aligned} .$$

There is just one solution: $(x, y) = (-3,\ 3)$. Now let us reflect on what we have here. What we have actually found is all the points that *cannot possibly be local minima or maxima*. These are all points *except* (-3, 3).. All we know right now is that this point is the only possible candidate. Let's find out what we have by the hammer and tongs method of examining the quantity $f(-3+x,\ 3+y) - f(-3,3)$:

$$\begin{aligned} f(-3+x,\ 3+y) - f(-3,3) &= f(-3+x,\ 3+y) - (-5) \\ &= (-3+x)^2 + (-3+x)(3+y) + (3+y)^2 + 3(-3+x) - 3(3+y) + 9 \\ &= x^2 + xy + y^2 = \left(x + \frac{y}{2}\right)^2 + \frac{3y^2}{4} \end{aligned}$$

It is therefore clear that $f(-3+x,\ 3+y) - f(-3,3) \geq 0$, which means that (-3, 3) is a local minimum.

Exercises

In each of the following, find all local maxima and minima:

10. $f(x,y) = x^2 + 3xy + 3y^2 - 6x + 3y - 6$

11. $f(x,y) = x^2 + xy + 3x + 2y + 5$

12. $f(x,y) = 2xy - 5x^2 - 2y^2 + 4x - 4$

13. $f(x,y) = x^2 + 2xy$

14. $f(x,y) = y - x^2$

8.5 Least Squares

We shall next look at some very simple, yet important, applications in which the location of a minimum value of a function is sought.

Suppose we have a set of *n* points in the plane, say $(x_1, y_1),(x_2, y_2),\ldots,(x_n, y_n)$, and we seek the straight line that "best" fits this collection of points. We first decide what we mean by "best". Let's say we mean the line that minimizes the sum of the squares of the vertical distances from the points to the line. We can describe all nonvertical lines in the world by means of two variables, traditionally called m and b. Thus every such line has the form $y = mx + b$. Our quest is thus for the values of m and b at which the function

$$f(m,b) = \sum_{i=1}^{n} (mx_i + b - y_i)^2$$

has its minimum value. Knowing these values will give us our line.

We simply apply our vast and growing knowledge of calculus and find where the gradient of f is 0:

$$\nabla f = \left(\frac{\partial f}{\partial m}, \frac{\partial f}{\partial b}\right) = \mathbf{0} \quad .$$

Now,

$$\frac{\partial f}{\partial m} = \sum_{i=1}^{n} 2x_i(mx_i + b - y_i) = 2[m\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i], \text{ and}$$

$$\frac{\partial f}{\partial b} = \sum_{i=1}^{n} 2(mx_i + b - y_i) = 2[m\sum_{i=1}^{n} x_i + nb - \sum_{i=1}^{n} y_i].$$

We are thus faced with solving the 2 x 2 linear system

$$m\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$
$$m\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i$$

Meditate sufficiently to convince yourself that there is always exactly one solution to this system, and continue meditating sufficiently to convince yourself that there must be an honest-to-goodness minimum of the original function at this solution.

Let's have a go at an example. Suppose we have the following table of values:

x	**y**
0	1
1	2
2	4
3	3.5
4	5

5	4
7	7
8	9
9	12
10	18
12	21
15	29

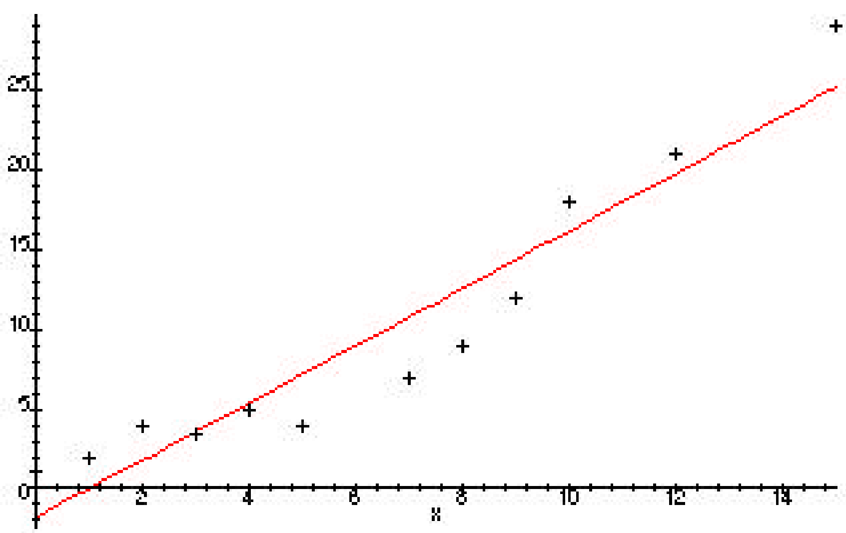
The linear system for m and b is

$$\begin{aligned} 718m + 76b &= 1156.5 \\ 76m + 12b &= 115.5 \end{aligned}$$

Solving this system gives us $m = \dfrac{255}{142}$ and $b = -\dfrac{993}{568}$. In other words, the line that best fits the data in the "sense of least squares" is

$$y = \frac{255}{142}x - \frac{993}{568}$$

Here's a picture of this line together with the data points:



Looks pretty good!

Exercises

15. Here is a table of Köchel numbers versus year of composition for the compositions of W. A. Mozart. Find the "least squares" straight line approximation to this table and use it to estimate the year in which Mozart's *Sinfonia Concertante in E-flat major* was composed.

Köchel Number	Year composed
1	1761
75	1771
155	1772
219	1775
271	1777
351	1780
425	1783
503	1786
575	1789
626	1791

[This problem is taken from *Calculus and Analytic Geometry* (*8th Edition*), by Thomas & Finney.]

16. Find some data somewhere (The *Statistical Abstract of the United States* is a good source of interesting data.), find the least squares linear approximation to the data, and say something intelligent about your results.

8.6 More Maxima and Minima

In real life, one is most likely interested in finding the places at which the largest and smallest values of a function $f:\boldsymbol{D} \to \boldsymbol{R}$ occur, rather than in simply finding local maxima and minima. (Here $\boldsymbol{D}$ is a subset of $\boldsymbol{R}^n$.). To begin, let's think a moment about how we can tell if there *is* a maximum or minimum value of f on $\boldsymbol{D}$. First, we suppose that f is continuous—otherwise, anything can happen! Next, what properties of $\boldsymbol{D}$ will insure the existence of a biggest and smallest value of f ? The answer is fairly simple. Certainly $\boldsymbol{D}$ must be a closed subset of $\boldsymbol{R}^n$; consider, for example the function $f:(0,1) \to \boldsymbol{R}$ given simply by $f(x) = x$, which has neither a maximum nor a minimum on $\boldsymbol{D} = (0,1)$. Having the domain be closed, however, is not sufficient to guarantee the existence of a maximum and minimum. Consider, for example $f:\boldsymbol{R} \to \boldsymbol{R}$ again with $f:(0,1) \to \boldsymbol{R}$ given by $f(x) = x$. The domain $\boldsymbol{R}$ is certainly closed, but f has neither a maximum nor a minimum. We need also to have the domain be *bounded.* It turns out that for continuous f, if the domain $\boldsymbol{D}$ is both ***closed*** and ***bounded***, then there must necessarily be a maximum and a minimum value for f on $\boldsymbol{D}$. Let's think a moment about what the candidates for such points are. If the biggest or smallest value of f occurs in the interior of $\boldsymbol{D}$, then surely the point at which it occurs is a local maximum (or minimum). If f has a gradient there, then the gradient must be 0. The points at which the largest or smallest values occur must therefore be either i)points in the interior of $\boldsymbol{D}$ at which the gradient of f vanishes, ii)points in the interior at which the gradient of f does not exist, or iii)points in $\boldsymbol{D}$ but not in the interior of $\boldsymbol{D}$ (that is, points on the boundary of $\boldsymbol{D}$).

Hark back to Mrs. Turner's third grade calculus class. How did you find the maximum value of a function f whose domain $\boldsymbol{D}$ is a closed interval $[a,b] \quad \boldsymbol{R}$? Recall found all points in the interior (that is, in the open interval (a,b)) at which the derivative vanishes. You then simply evaluated f at these points, evaluated f at any points in (a,b) at which there is no derivative, evaluated f at the two end points of the interval (in this one dimensional case, the boundary of $\boldsymbol{D}$ is particularly simple.), and then picked out the biggest and smallest numbers you computed. The situation in higher dimensions is a bit more complicated, mostly because the boundary of even a nice domain $\boldsymbol{D}$ is *not* a nice finite set as in the case of an interval, but is an infinite set. Let's look at an example.

Example

A flat circular plate has the shape of the region $\{(x,y) \quad \boldsymbol{R^2} : x^2 + y^2 \leq 1\}$. The temperature at the point (x,y) on the plate is given by $T(x,y) = x^2 + 2y^2 - x$. Our assignment is to find the hottest and coldest points on the plate. According to our previous discussion, candidates for the hottest and coldest points are all points inside the circular boundary at which the gradient of T is 0 and all points on the boundary. (Note that T has a gradient at all points inside the circle.) First, let's find where among all points (x,y) such that $x^2 + y^2 < 1$, the ones at which $T = (2x-1, 4y) = \boldsymbol{0}$. This is easy; it should be clear there is just one such point: $(\frac{1}{2}, 0)$. Now for the more difficult part, finding the candidates on the boundary. Note that the boundary may be described by the vector equation

$$\boldsymbol{r}(t) = \cos t\, \boldsymbol{i} + \sin t\, \boldsymbol{j} \text{ , where } 0 \leq t \leq 2\pi .$$

The temperature on this set is then given by

$$T(t) = T(\boldsymbol{r}(t)),\ 0 \leq t \leq 2\pi$$

[Here we are abusing the notation, as we have done before, by using the same name for the function $T(x,y)$ and the composition $T(\boldsymbol{r}(t))$.] We are now faced with the one dimensional problem of finding the maximum and minimum values of a nice differentiable function of one variable on a closed interval. First, we know the endpoints of the interval are candidates: $t = 0$, and $t = 2\pi$. We have at this point added one more point to our list of candidates: $\boldsymbol{r}(0) = \boldsymbol{r}(2\pi) = (1,0)$. Now for candidates inside the interval, we seek places at which the derivative $\frac{dT}{dt} = 0$. From the Chain Rule, we know

$$\frac{dT}{dt} = \quad T(\boldsymbol{r}(t)) \;\; \boldsymbol{r}'(t) = (2\cos t - 1, 4\sin t) \;\; (-\sin t, \cos t) = 2\cos t \sin t + \sin t \;.$$

The equation $\frac{dT}{dt} = 0$ now becomes

$$2\cos t \sin t + \sin t = 0, \text{ or}$$
$$\sin t(2\cos t + 1) = 0$$

Thus $\sin t = 0$, or $2\cos t + 1 = 0$. We have, in other words, $y = 0$, or $x = -\frac{1}{2}$. When $y = 0$, then $x = 1$ or $x = -1$; and when $x = -\frac{1}{2}$, then $y = \frac{\sqrt{3}}{2}$ or $y = -\frac{\sqrt{3}}{2}$. Thus our new candidates are $(1,0)$, $(-1,0)$, $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$, and $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})$. These together with the one we have already found, $(\frac{1}{2}, 0)$, make up our entire list of possibilities for the hottest and coldest points on the plate. All we need do now is to compute the temperature at each of these points:

$$T(\frac{1}{2}, 0) = \frac{1}{4} - \frac{1}{2} = -\frac{1}{4}.$$
$$T(1,0) = 1 - 1 = 0$$
$$T(-1,0) = 1 + 1 = 2$$
$$T(-\frac{1}{2}, \frac{\sqrt{3}}{2}) = T(-\frac{1}{2}, -\frac{\sqrt{3}}{2}) = \frac{1}{4} + \frac{3}{2} + \frac{1}{2} = \frac{9}{4}$$

Finally, we have our answer. The coldest point is $(\frac{1}{2}, 0)$, and the hottest points are $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$ and $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})$.

Exercises

17. Find the maximum and minimum value of $f(x, y) = x^2 - xy + y^2 + 4$ on the closed area in the first quadrant bounded by the triangle formed by the lines $x = 0$, $y = 4$, and $y = x$.

18. Find the maximum and minimum values of $f(x, y) = (4y - y^2)\cos x$ on the closed area bounded by the rectangle $1 \le y \le 3$, $-\frac{\pi}{4} \le x \le \frac{\pi}{4}$.

8.7 Even More Maxima and Minima

It should be clear now that the really troublesome part of finding maxima and minima is in dealing with the constrained problem; that is, the problem of finding the maxima and minima of a given function on a set of lower dimension than the domain of the function. In the problems of the previous section, we were fortunate in that it was easy to find parametric representations of the these sets; in general, this, of course, could be quite difficult. Let's see what we might do about this difficulty.

Suppose we are faced with the problem of finding the maximum or minimum value of the function $f : \boldsymbol{D} \to \boldsymbol{R}$, where $\boldsymbol{D} = \{(x, y) \quad \boldsymbol{R}^2 : g(x, y) = 0\}$, where g is a nice function. (In other words, $\boldsymbol{D}$ is a level curve of g.) Suppose $\boldsymbol{r}(t)$ is a vector description of the curve $\boldsymbol{D}$. Now then, we are seeking a maximum or minimum of the function $F(t) = f(\boldsymbol{r}(t))$. At a maximum or minimum, we must have $\frac{dF}{dt} = 0$. (Here g is sufficiently nice to insure that $g(x, y) = 0$ is a closed curve, and so there are no endpoints to worry about.) The Chain Rule tells us that $\frac{dF}{dt} = \quad f \quad \boldsymbol{r}' = 0$. Thus at a maximum or minimum, the gradient of f must be perpendicular to the tangent to $g(x, y) = 0$. But if $\quad f$ is perpendicular to the tangent to the level curve $g(x, y) = 0$, then it must have the same direction as the normal to this curve. This is just what we need to know, for the gradient of g is normal to this curve. Thus at a maximum or minimum, f and g must "line up". Thus $f = \lambda \; g$, and there is no need actually to know a vector representation $\boldsymbol{r}$ for $g(x,y) = 0$.

Let's see this idea in action. Suppose we wish to find the largest and smallest values of $f(x,y) = x^2 + y^2$ on the curve $x^2 - 2x + y^2 - 4y = 0$.

Here, we may take $g(x,y) = x^2 - 2x + y^2 - 4y$. Then $f = 2x\boldsymbol{i} + 2y\boldsymbol{j}$, and $g = (2x-2)\boldsymbol{i} + (2y-4)\boldsymbol{j}$, and our equation $f = \lambda \; g$ becomes

$$2x = \lambda\,(2x-2)$$
$$2y = \lambda\,(2y-4)$$

We obtain a third equation from the requirement that the point (x,y) be on the curve $g(x,y) = 0$. In other words, we need to find all solutions to the system of equations

$$2x = \lambda\,(2x-2)$$
$$2y = \lambda\,(2y-4)$$
$$x^2 - 2x + y^2 - 4y = 0$$

The first two equations become

$$x(\lambda - 1) = \lambda$$
$$y(\lambda - 1) = 2\lambda$$

Thus $x = \dfrac{\lambda}{\lambda - 1}$ and $y = \dfrac{2\lambda}{\lambda - 1}$. (What about the possibility that $\lambda - 1 = 0$?). The last equation then becomes $\dfrac{\lambda^2}{(\lambda-1)^2} - \dfrac{2\lambda}{\lambda - 1} + \dfrac{4\lambda^2}{(\lambda-1)^2} - \dfrac{8\lambda}{\lambda - 1} = 0$; or,

$$\lambda^2 - 2\lambda(\lambda - 1) = 0,$$
$$\lambda^2 - 2\lambda = 0$$

We have two solutions: $\lambda = 0$ and $\lambda = 2$. What do you make of the solution $\lambda = 0$? These values of λ give us two candidates for places at which extrema occur: $x = 0$ and $y = 0$; and $x = 2$ and $y = 4$. Now then $f(0,0) = 0$, and $f(2,4) = 4 + 16 = 20$. There we have them—the minimum value is 0 and it occurs at (0,0); and the maximum value is 20, and it occurs at (2,4).

This method for finding "constrained" extrema is generally called the method of ***Lagrange Multipliers***. (The variable λ is called a ***Lagrange multiplier***.)

Exercises

19. Use the method of Lagrange multipliers to find the largest and smallest values of $f(x, y) = 4x + 3y$ on the circle $x^2 + y^2 = 1$.

20. Find the points on the ellipse $x^2 + 2y^2 = 1$ at which $f(x, y) = xy$ has its extreme values.

21. Find the points on the curve $x^2 + xy + y^2 = 1$ that are nearest to and farthest from the origin.

Chapter Nine

The Taylor Polynomial

9.1 Introduction

Let f be a function and let $\boldsymbol{F}$ be a collection of "nice" functions. The approximation problem is simply to find a function $g \quad \boldsymbol{F}$ that is "close" to the given function f. There are two issues immediately. How is the collection $\boldsymbol{F}$ selected, and what do we mean by "close"? The answers depend on the problem at hand. Presumably we want to do something to f that is difficult or impossible (This might be something as simple as finding $f(x)$ for some x.). The collection $\boldsymbol{F}$ would thus consist of functions to which it is easy to do that which we wish to do to f. Our measure of how close one function is to another would try to reflect the closeness of the results of our operations. Now, what are we talking about here. Suppose, for example, we wish to find $f(x)$. Our collection $\boldsymbol{F}$ of functions should include functions that are easy to evaluate at x, and two function would be "close" simply if there values are close. We might, for instance, want to evaluate $\sin x$ for all x is some interval I. The collection $\boldsymbol{F}$ could be a collection of second degree polynomials. The approximation problem is then to find elements of $\boldsymbol{F}$ that make the "distance" $\max\{|\sin x - p(x)| : x \quad I\}$ as small as possible. Similarly, we might want to find the integral of some function f over an interval I. Here we would want $\boldsymbol{F}$ to consist of functions easily integrated and measure the distance between functions by the difference of their integrals over I. In the previous chapter, we found the "best" straight line approximation to a set of data points. In that case, the collection $\boldsymbol{F}$ consisted of all nonvertical straight lines, and we measured the distance between functions by the sum of the squares of their differences on a specified set of points $\{x_1, x_2, \ldots, x_n\}$. You can imagine many other examples.

9.2 The Taylor Polynomial

We look first at a simple but useful problem: Given a nice function $f : \boldsymbol{D} \quad \boldsymbol{R} \to \boldsymbol{R}$, a point a in the interior of the domain $\boldsymbol{D}$, and an integer n, find a polynomial p of degree $\leq n$ such that

$$
\begin{aligned}
&p(a) = f(a) \\
&p'(a) = f'(a) \\
&p''(a) = f''(a) \\
&\vdots \\
&p^{(n)}(a) = f^{(n)}(a)
\end{aligned}
$$

We solve the problem by the Behold Method. Simply verify that

$$p(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \ldots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

does the job! It is also fairly easy to see that this polynomial is the *only* polynomial of degree $\leq n$ that does the job. Suppose q is also a polynomial with degree $g \leq n$ such that

$$
\begin{aligned}
&p(a) = f(a) \\
&p'(a) = f'(a) \\
&p''(a) = f''(a) \\
&\vdots \\
&p^{(n)}(a) = f^{(n)}(a)
\end{aligned}
$$

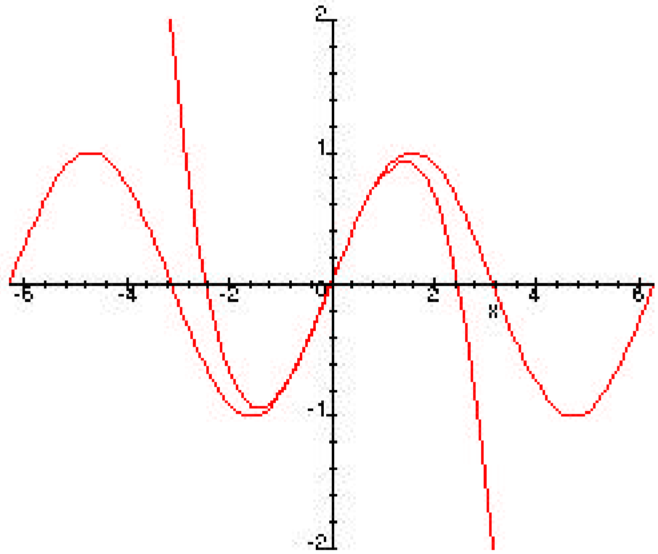
and consider the function $r = p - q$. Note that r is also a polynomial of degree $\leq n$. But

$$r(a) = r'(a) = r''(a) = \ldots = r^{(n)}(a) = 0 .$$

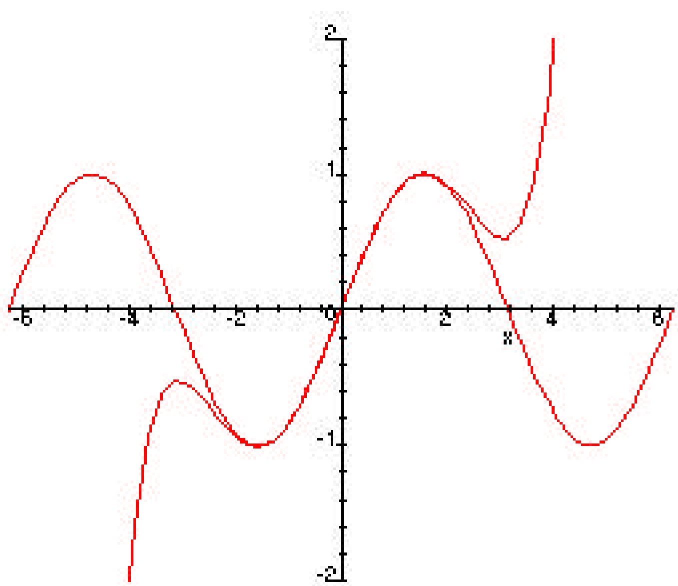
Or, in other words, r has a zero of order $n + 1$, and the only way this can happen is if $r(x) \equiv 0$ for *all* x. That is, $p(x) \equiv q(x)$ identically.

Example

Let $f(x) = \sin x$ and let $a = 0$. Let's find the Taylor polynomial for a few different values of n. For $n = 1$, we have simply $p_1(x) = f(a) + f'(a)(x-a) = \sin 0 + \cos 0(x) = x$. Note that for $n = 2$, we have $p_2(x) = \sin 0 + \cos 0(x) - \sin 0(x^2) = x$, also. Let's take a look at the next Taylor polynomial. Here $p_3(x) = x - \frac{x^3}{6}$. Let's draw some pictures; we'll look at the graph of p_3 and f. We shall use *Maple.*



What we see is that the Taylor polynomial looks like a pretty good approximation as long as we don't get too far away from $a = 0$. Let us continue. Convince yourself that $p_4 = p_3$, and $p_5(x) = x - \frac{x^3}{6} + \frac{x^5}{120}$. Another picture:



Exercises

1. Find the Taylor polynomial of degree n for $f(x) = e^x$ at $a = 0$.

2. Find the Taylor polynomial of degree n for $f(x) = x^3$ at $a = 1$.

3. Find the Taylor polynomial of degree 3 for $f(x) = \log x$ at $a = 1$.

4. Find the Taylor polynomial of degree n for $f(x) = \sin x$ at $a = 0$.

5. Find the Taylor polynomial of degree 3 for $f(x) = \sqrt{x}$ at $a = 4$.

9.3 Error

Let's see how close the Taylor polynomial is to the function f. To do this, suppose p is the Taylor polynomial of degree $\leq n$ for the function f at a, and consider the function

$$g(t) = f(t) - p(t) - \frac{(t-a)^{n+1}}{(x-a)^{n+1}}(f(x) - p(x)).$$

(We assume $x \neq a$.) Note that $g(a) = g(x) = 0$. Now, from the Mean Value Theorem (or Rolle's Theorem, or whatever.) we know that $g'(\xi_1) = 0$ for some ξ_1 between a and x. But note also that $g'(a) = f'(a) - p'(a) - \frac{(n+1)(a-a)^n}{(x-a)^{n+1}}(f(x) - p(x)) = 0$. It thus follows from the Mean Value Theorem that the derivative of g' is zero at some ξ_2 between a and ξ_1. Also, $g''(a) = f''(a) - p''(a) - \frac{(n+1)n(a-a)^{n-1}}{(x-a)^{n+1}}(f(x) - p(x)) = 0$. Once again, from the celebrated Mean Value Theorem, we conclude that $g'''(\xi_3) = 0$ for some ξ_3 between a and ξ_2. Continuing in this fashion, we are finally able to conclude that $g^{(n+1)}(\xi) = 0$ for some ξ. Let's see what this looks like.

$$g^{(n+1)}(t) = f^{(n+1)}(t) - p^{(n+1)}(t) - \frac{(n+1)!}{(x-a)^{n+1}}(f(x) - p(x))$$

and so $g^{(n+1)}(\xi)=0$ becomes

$$f^{(n+1)}(\xi)-\frac{(n+1)!}{(x-a)^{n+1}}(f(x)-p(x))=0\ .$$

(Remember, p is a polynomial of degree $\le n$, and so $p^{(n+1)}(t)\equiv 0$. From this we obtain an expression for the difference between *f* and the Taylor polynomial g :

$$f(x)-p(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}\ .$$

Example

Remember when in 7th grade physics class, Mr. Crews replaced the sine of a "small" angle θ by θ itself ? He assured us that for small angles this was just fine. Well, what was going on here? Let's see if our new-found knowledge of Taylor polynomials will help. Observe that $p(\theta)=\theta$ is simply the Taylor polynomial of degree ≤ 2 for $f(\theta)=\sin\theta$ at $a=0$. Using the result just derived, we have that

$$\sin\theta-\theta=\frac{-\sin\xi}{6}\theta^3\ .$$

Now, we don't know what ξ is, but we do know that $|\sin\xi\}\le 1$; thus

$$|\sin\theta-\theta|\le\frac{\theta^3}{6}\ ,$$

and we have a precise estimate of the error incurred by substituting θ for $\sin\theta$. Suppose, for example, that $\theta=10^\circ$; then what? Well, $\theta=\frac{10}{360}2\pi=\frac{\pi}{18}$. Then the error we get when we use $\frac{\pi}{18}$ instead of $\sin\frac{\pi}{18}$ is estimated by

$$\left|\sin\frac{\pi}{18}-\frac{\pi}{18}\right|\le\frac{1}{6}\left(\frac{\theta}{18}\right)^3\le 0.008862.$$

Now we know exactly what "pretty close" means. For 10 degrees, I guess that's "not too bad."

Exercises

6. a)Find the Taylor polynomial of degree ≤ 2 for $f(x) = e^x$ at $a = 0$.

b)Use the result of part a) to find an approximation for $\sqrt{e}$.

c)Find as small an upper bound as you can for the difference between your approximation found in part b) and $\sqrt{e}$.

7. Use the Taylor polynomial found in Exercise 3 to approximate $\log(1.1)$ and find an upper bound for the magnitude of the difference between your approximation and $\log(1.1)$.

8. For what values of x can you replace $\sin x$ by $x - \dfrac{x^3}{6}$ with an error of magnitude no greater than 3×10^{-4} ?

9. Calculate e with an error of less than 10^{-6} .

Chapter Ten

Sequences, Series, and All That

10.1 Introduction

Suppose we want to compute an approximation of the number e by using the Taylor polynomial p_n for $f(x) = e^x$ at a =0. This polynomial is easily seen to be

$$p_n(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \ldots + \frac{x^n}{n!}.$$

We could now use $p_n(1)$ as an approximation to e. We know from the previous chapter that the error is given by

$$e - p_n(1) = \frac{e^{\xi}}{(n+1)!} 1^{n+1},$$

where $0 < \xi < 1$. Assume we know that e <3, and we have the estimate

$$0 \le e - p_n(1) \le \frac{3}{(n+1)!}.$$

Meditate on this error estimate. It tells us that we can make this error as small as we like by choosing n sufficiently large. This is expressed formally by saying that the *limit* of $p_n(1)$ as *n becomes infinite* is e. This is the idea we shall study in this chapter.

10.2 Sequences

A ***sequence*** of real numbers is simply a function from a subset of the nonnegative integers into the reals. If the domain is infinite, we say the sequence is an ***infinite sequence.*** (Guess what a *finite* sequence is.) We shall be concerned only with infinite sequences, and so the modifier will usually be omitted. We shall also almost always consider sequences in which the domain is either the entire set of nonnegative or positive integers.

There are several notational conventions involved in writing and talking about sequences. If $f: Z_+ \to \boldsymbol{R}$, it is customary to denote $f(n)$ by f_n, and the sequence itself by (f_n). (Here Z_+ denotes the positive integers.) Thus, for example, $\left(\frac{1}{n}\right)$ is the sequence

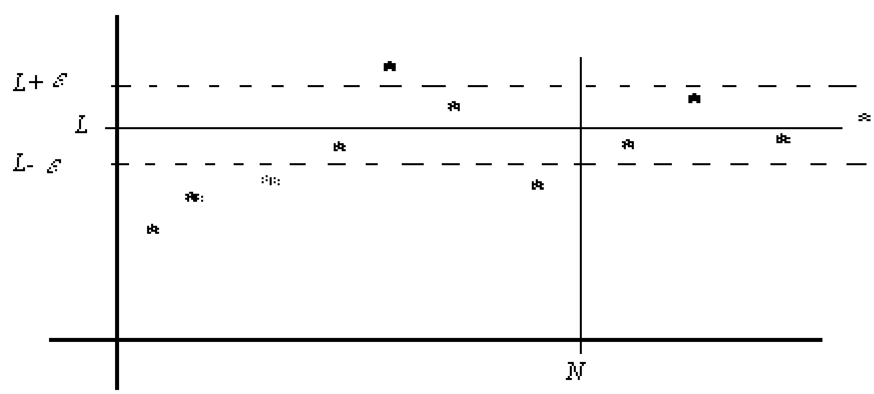
f defined by $f(n) = \frac{1}{n}$. The function values f_n are called ***terms*** of the sequence. Frequently one sees a sequence described by writing something like

$$1, 4, 9, \ldots, n^2, \ldots .$$

This is simply another way of describing the sequence (n^2).

Let (a_n) be a sequence and suppose there is a number L such that for any $\varepsilon > 0$, there is an integer N such that $|a_n - L| < \varepsilon$ for all $n > N$. Then L is said to be a ***limit*** of the sequence, and (a_n) is said to ***converge*** to L. This is usually written $\lim_{n \to \infty} a_n = L$. Now, what does this really mean? It says simply that as n gets big, the terms of the sequence get close to L. I hope it is clear that 0 is a limit of the sequence $\left(\frac{1}{n}\right)$. From the discussion in the Introduction to this chapter, it should be reasonably clear that a limit of the sequence $\left(1 + \frac{1}{2} + \frac{1}{6} + \ldots + \frac{1}{n!}\right)$ is e.

The graph of a sequence is pretty dreary compared with the graph of a function whose domain is an interval of reals, but nevertheless, a look at some pictures can help understand some of these definitions. Suppose the sequence (a_n) converges to L. Look at the graph of (a_n):



The fact that L is a limit of the sequence means that for any $\varepsilon > 0$, there is an N so that to the right of N, all the spots are in the strip of width 2ε centered at L.

Exercises

1. Prove that a sequence can have at most one limit (We may thus speak of ***the*** limit of a sequence.).

2. Give an example of a sequence that does not have a limit. Explain.

3. Suppose the sequence $(a_n) = a_0, a_1, a_2, \ldots$ converges to L. Explain how you know that the sequence $(a_{n+5}) = a_5, a_6, a_7, \ldots$ also converges to L.

4. Find the limit of the sequence $\left(\frac{3}{n^2}\right)$, or explain why it does not converge.

5. Find the limit of the sequence $\left(\frac{3n^2 + 2n - 7}{n^2}\right)$, or explain why it does not converge.

6. Find the limit of the sequence $\left(\frac{5n^3 - n^2 + 7n + 2}{3n^3 + n^2 - n + 10}\right)$, or explain why it does not converge.

7. Find the limit of the sequence $\left(\frac{\log n}{n}\right)$, or explain why it does not converge.

10.3 Series

Suppose (a_n) is a sequence. The sequence $(a_0 + a_1 + \ldots + a_n)$ is called a ***series.*** It is a little neater to write if we use the usual summation notation: $\left(\sum_{k=0}^{n} a_k\right)$. We have seen an example of such a thing previously; *viz.*,

$$\left(1 + \frac{1}{2} + \frac{1}{6} + \ldots + \frac{1}{n!}\right) = \left(\sum_{k=0}^{n} \frac{1}{k!}\right).$$

It is usual to replace $\lim_{n\to\infty} \sum_{k=0}^{n} a_k$ by $\sum_{k=0}^{\infty} a_k$. Thus, one would, for example, write

$$e = \sum_{k=0}^{\infty} \frac{1}{k!}.$$

One also frequently sees the limit $\sum_{k=0}^{\infty} a_k$ written as $a_0 + a_1 + \ldots + a_n + \ldots$. And one more word of warning. Some poor misguided souls also use $\sum_{k=0}^{\infty} a_k$ to stand simply for the series $\left(\sum_{k=0}^{n} a_k\right)$. It is usually clear whether the series or the limit of the series is meant, but it is nevertheless an offensive practice that should be ruthlessly and brutally suppressed.

Example

Let's consider the series $\left(\sum_{k=0}^{n} \frac{1}{2^k}\right) = \left(1 + \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^n}\right)$. Let

$$S_n = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^n}. \text{ Then}$$

$$\frac{1}{2} S_n = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^n} + \frac{1}{2^{n+1}}.$$

Thus

$$\frac{S_n}{2} = S_n - \frac{1}{2} S_n = 1 - \frac{1}{2^{n+1}}.$$

This makes it quite easy to see that $\lim_{n \to \infty} \frac{S_n}{2} = 1$, or $\lim_{n \to \infty} S_n = 2$. In other words,

$$\sum_{k=0}^{\infty} \frac{1}{2^k} = 2.$$

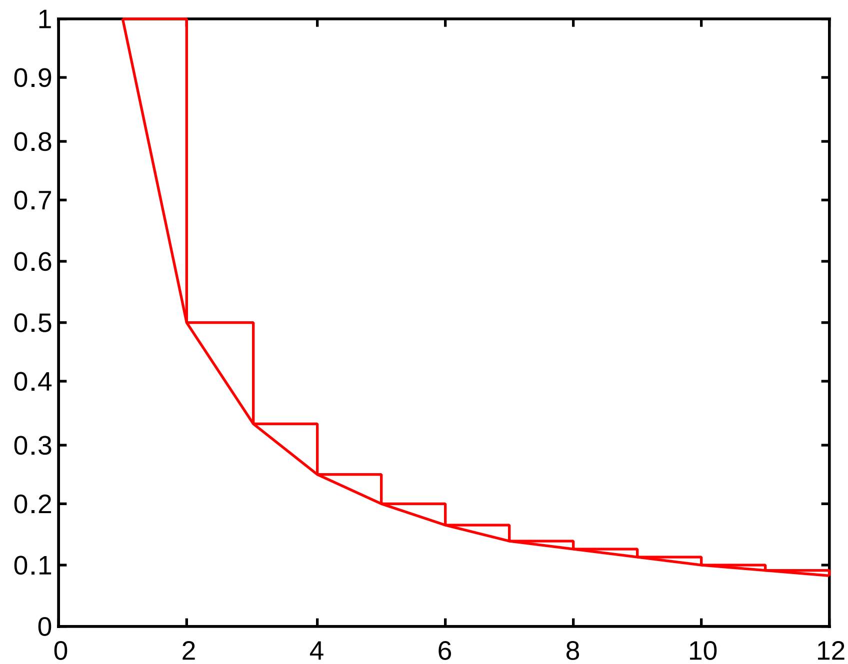
Observe that for series $\left(\sum_{k=0}^{n} a_k\right)$ to converge, it must be true that $\lim_{n \to \infty} a_n = 0$. To see this, suppose $L = \sum_{k=0}^{\infty} a_k$, and observe that $a_n = \sum_{k=0}^{n} a_k - \sum_{k=0}^{n-1} a_k$. Thus,

$$\lim_{n \to \infty} a_n = \lim_{n \to \infty} \left(\sum_{k=0}^{n} a_k - \sum_{k=0}^{n-1} a_k\right) = \lim_{n \to \infty} \sum_{k=0}^{n} a_k - \lim_{n \to \infty} \sum_{k=0}^{n-1} a_k$$
$$= L - L = 0.$$

In other words, if $\lim_{n\to\infty} a_n \neq 0$, then the series $\left(\sum_{k=0}^{n} a_k\right)$ does not have a limit.

Another Example

Consider the series $\left(\sum_{k=1}^{n} \frac{1}{k}\right)$. First, note that $\lim_{n\to\infty} \frac{1}{k} = 0$. Thus we do not know that the series does not converge; that is, we still don't know anything. Look at the following picture:



The curve is the graph of $y = \frac{1}{x}$. Observe that the area under the "stairs" is simply $\sum_{k=1}^{11} \frac{1}{k}$.

Now convince yourself that $\sum_{k=1}^{n} \frac{1}{k}$ is larger than the area under the curve $y = \frac{1}{x}$ from $x = 1$ to $x = n+1$. In other words,

$$\sum_{k=1}^{n} \frac{1}{k} > \int_{1}^{n+1} \frac{1}{x}\,dx = \log(n+1)\,.$$

We know that $\log(n+1)$ can be made as large as we wish by choosing n sufficiently large. Thus $\sum_{k=1}^{n} \frac{1}{k}$ can be made as large as we wish by choosing n sufficiently large. From this it follows that the series $\left(\sum_{k=1}^{n} \frac{1}{k}\right)$ does *not* have a limit. (This series has a name. It is called the ***harmonic series***.)

The method we used to show that the harmonic series does not converge can be used on many other series. We simply consider a picture like the one above. Suppose we have a series $\left(\sum_{k=1}^{n} a_k\right)$ such that $a_k > 0$ for all k . Suppose f is a decreasing function such that $f(k) = a_k$ for all k . Then if the limit $\lim_{R\to\infty} \int_1^R f(x)dx$ does not exist, the series is divergent.

Exercises

8. Find the limit of the series $\left(\sum_{k=0}^{n} \frac{1}{3^n}\right)$, or explain why it does not converge.

9. Find the limit of the series $\left(\sum_{k=0}^{n} \frac{5}{\sqrt{n+3}}\right)$, or explain why it does not converge.

10. Find a value of n that will insure that $1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n} > 10^6$.

11. Let $0 \le \theta \le 1$. Prove that $\sin\theta = \sum_{k=0}^{\infty} (-1)^k \frac{\theta^{2k+1}}{(2k+1)!}$.

[Hint: $p_{2n+1}(\theta) = \sum_{k=0}^{n} (-1)^k \frac{\theta^{2k+1}}{(2k+1)!}$ is the Taylor polynomial of degree $\le 2n+1$ for the function $f(\theta) = \sin\theta$ at $a = 0$.]

12. Suppose we have a series $\left(\sum_{k=1}^{n} a_k\right)$ such that $a_k > 0$ for all k, and suppose f is a decreasing function such that $f(k) = a_k$ for all k. Show that if the limit $\lim_{R\to\infty}\int_1^R f(x)dx$ exists, then the series is convergent.

13. a) Find all p for which the series $\left(\sum_{k=1}^{n} \frac{1}{k^p}\right)$ converges.

b) Find all p for which the series in a) diverges.

10.4 More Series

Consider a series $\left(\sum_{k=0}^{n} a_k\right)$ in which $a_k \geq 0$ for all k. This is called a ***positive series.*** Let $\left(\sum_{k=0}^{n} b_k\right)$ be another positive series. Suppose that $b_k \leq a_k$ for all $k \geq N$, where N is simply some integer. Now suppose further that we know that $\left(\sum_{k=0}^{n} a_k\right)$ converges. This tells us all about the series $\left(\sum_{k=0}^{n} b_k\right)$. Specifically, it tells us that this series also converges. Let's see why that is. First note the obvious: $\left(\sum_{k=0}^{n} b_k\right)$ converges if and only if $\left(\sum_{k=N}^{n} b_k\right)$ converges. Next, observe that for all n, we have $\sum_{k=N}^{n} b_k \leq \sum_{k=N}^{n} a_k$, from which it follows at once that $\lim_{n\to\infty}\sum_{k=N}^{n} b_k$ exists.

Example

What about the convergence of the series $\left(\sum_{k=1}^{n} \frac{1}{n^3+3n^2+n+4}\right)$? Observe first that $\frac{1}{n^3+3n^2+n+4} < \frac{1}{n^3}$. Then observe that the series $\left(\sum_{k=1}^{n} \frac{1}{n^3}\right)$ converges because $\lim_{R\to\infty}\int_1^R \frac{1}{x^3}dx = \lim_{R\to\infty}\left(\frac{-1}{3R^2}+\frac{1}{3}\right)=\frac{1}{3}$. Thus $\left(\sum_{k=1}^{n} \frac{1}{n^3+3n^2+n+4}\right)$ converges.

Suppose that, as before, $\left(\sum_{k=0}^{n} a_k\right)$ and $\left(\sum_{k=0}^{n} b_k\right)$ are positive series, and $b_k \le a_k$ for all $k \geq N$, where N is some number. This time, suppose we know that $\left(\sum_{k=0}^{n} b_k\right)$ is divergent. Then it should not be too hard for you to convince yourself that $\left(\sum_{k=0}^{n} a_k\right)$ must be divergent, also.

Exercises

Which of the following series are convergent and which are divergent? Explain your answers.

14. $\left(\sum_{k=0}^{n} \frac{1}{2e^k+k}\right)$

15. $\left(\sum_{k=0}^{n} \frac{1}{2k+1}\right)$

16. $\left(\sum_{k=2}^{n} \frac{1}{\log k}\right)$

17. $\left(\sum_{k=0}^{n} \frac{1}{k^2+k-1}\right)$

10.5 Even More Series

We look at one more very nice way to help us determine if a positive series has a limit. Consider a series $\left(\sum_{k=0}^{n} a_k\right)$, and suppose $a_k > 0$ for all k. Next suppose the sequence $\left(\frac{a_{k+1}}{a_k}\right)$ is convergent, and let

$$r = \lim_{k\to\infty} \frac{a_{k+1}}{a_k}.$$

The number r tells us almost everything about the convergence of the series $\left(\sum_{k=0}^{n} a_k\right)$. Let's see about it.

First, suppose that $r < 1$. Then the number $\rho = r + \frac{1-r}{2}$ is positive and less than 1. For all sufficiently large k, we know that $\frac{a_{k+1}}{a_k} \le \rho$. In other words, there is an N so that $a_{k+1} \le a_k \rho$ for all $k \ge N$. Thus

$$a_{k+1} \le a_k \rho \le a_{k-1}\rho^2 \le a_{k-2}\rho^3 \le \ldots \le a_N \rho^{k+1-N}.$$

Look now at the series

$$\left(\sum_{k=N}^{n} a_N \rho^{k-N}\right) = \left(a_N(1 + \rho + \rho^2 + \ldots \rho^{n-N})\right).$$

This one converges because the Geometric series $\left(\sum_{k=0}^{n} \rho^k\right)$ converges (Recall that $0 < \rho < 1$.). It now follows from the previous section that our original series $\left(\sum_{k=0}^{n} a_k\right)$ has a limit.

A similar argument should convince you that if $r > 1$, then the series $\left(\sum_{k=0}^{n} a_k\right)$ does not have a limit.

The "method" of the previous section is usually called the ***Comparison Test***, while that of this section is usually called the ***Ratio Test***.

Exercises

Which of the following series are convergent and which are divergent? Explain your answers.

18. $\left(\sum_{k=0}^{n} \frac{10^k}{k!}\right)$

19. $\left(\sum_{k=0}^{n} \frac{3^{2k+1}}{5^k}\right)$

20. $\left(\sum_{k=0}^{n} \frac{3^{2k+1}}{10^k}\right)$

21. $\left(\sum_{k=1}^{n} \frac{3^k}{5^k (k^4 + k + 1)}\right)$

22. $\left(\sum_{k=1}^{n} \frac{3^k (k^4 + k + 1)}{5^k}\right)$

10.6 A Final Remark

The "tests" for convergence of series that we have seen so far all depended on the series having positive terms. We need to say a word about the situations in which this is not necessarily the case. First, if the terms of a series $\left(\sum_{k=0}^{n} a_k\right)$ alternate in sign, and if it is true that $|a_{k+1}| \le |a_k|$ for all k, then $\lim_{k \to \infty} a_k = 0$ is sufficient to insure convergence of the series. This is not too hard to see—meditate on it for a while.

The second result is a bit harder to see, and we'll just put out the result as the word, asking that you accept it on faith. It says simply that if the series $\left(\sum_{k=0}^{n} |a_k|\right)$ converges, then so also does the series $\left(\sum_{k=0}^{n} a_k\right)$. Thus, faced with an arbitrary series $\left(\sum_{k=0}^{n} a_k\right)$, we may unleash out arsenal of tests on the series $\left(\sum_{k=0}^{n} |a_k|\right)$. If we find this one to be convergent, then the original series is also convergent. If, of course, this series turns out not to be convergent, then we still do not know about the original series.

Chapter Eleven

Taylor Series

11.1 Power Series

Now that we are knowledgeable about series, we can return to the problem of investigating the approximation of functions by Taylor polynomials of higher and higher degree. We begin with the idea of a so-called power series. A ***power series*** is a series of the form

$$\left(\sum_{k=0}^{n} c_k (x-a)^k\right).$$

A power series is thus a sequence of special polynomials: each term is obtained from the previous one by adding a constant multiple of the next higher power of $(x-a)$. Clearly the question of convergence will depend on x, as will the limit where there is one. The k th term of the series is $c_k(x-a)^k$ so the Ratio Test calculation looks like

$$r(x) = \lim_{k\to\infty}\left|\frac{c_{k+1}(x-a)^{k+1}}{c_k(x-a)^k}\right| = |x-a| \lim_{k\to\infty}\left|\frac{c_{k+1}}{c_k}\right|.$$

Recall that our series converges for $r(x) < 1$ and diverges for $r(x) > 1$. Thus this series converges absolutely for all values of x if the number $\lim_{k\to\infty}\left|\frac{c_{k+1}}{c_k}\right| = 0$. Otherwise, we have absolute convergence for $|x-a| < \lim_{k\to\infty}\left|\frac{c_k}{c_{k+1}}\right|$ and divergence for $|x-a| > \lim_{k\to\infty}\left|\frac{c_k}{c_{k+1}}\right|$. The number $R = \lim_{k\to\infty}\left|\frac{c_k}{c_{k+1}}\right|$ is called the ***radius of convergence,*** and the interval $|x-a| < R$ is called the ***interval of convergence.*** There are thus exactly three possibilities for the convergence of our power series $\left(\sum_{k=0}^{n} c_k (x-a)^k\right)$:

(i)The series converges for no value of x except $x = a$; or (ii)The series converges for all values of x; or

(iii)There is a positive number R so that the series converges for $|x-a|< R$ and diverges for $|x-a| > R$.

Note that the Ratio Test tells us nothing about the convergence or divergence of the series at the two points where $|x-a| = R$.

Example

Consider the series $\left(\sum_{k=0}^{n} k!x^k\right)$. Then $R = \lim_{k\to\infty}\left|\frac{c_k}{c_{k+1}}\right| = \lim_{k\to\infty}\frac{k!}{(k+1)!} = \lim_{k\to\infty}\frac{1}{k+1} = 0$.

Thus this series converges only when $x = 0$.

Another Example

Now look at the series $\left(\sum_{k=0}^{n} 3^k(x-1)^k\right)$. Here $R = \lim_{k\to\infty}\left|\frac{c_k}{c_{k+1}}\right| = \lim_{k\to\infty}\frac{3^k}{3^{k+1}} = \lim_{k\to\infty}\frac{1}{3} = \frac{1}{3}$.

Thus, this one converges for $|x-1| < \frac{1}{3}$ and diverges for $|x-1| > \frac{1}{3}$.

Exercises

Find the interval of convergence for each of the following power series:

1. $\left(\sum_{k=0}^{n} (x+5)^k\right)$

2. $\left(\sum_{k=0}^{n} \frac{1}{k}(x-1)^k\right)$

3. $\left(\sum_{k=0}^{n} \frac{k}{3k+1}(x-4)^k\right)$

4. $\left(\sum_{k=0}^{n} \frac{3^k}{k!}(x+1)^k\right)$

5. $\left(\sum_{k=0}^{n} \frac{k!}{7(k^2+1)}(x-9)^k\right)$

11.2 Limit of a Power Series

If the interval of convergence of the power series $\left(\sum_{k=0}^{n} c_k(x-a)^k\right)$ is $|x-a| < R$, then, of course, the limit of the series defines a function f:

$$f(x) = \sum_{k=0}^{\infty} c_k(x-a)^k \text{, for } |x-a| < R.$$

It is known that this function has a derivative, and this derivative is the limit of the derivative of the series. Moreover, the differentiated series has the same interval of convergence as that of the series defining f. Thus for all x in the interval of convergence, we have

$$f'(x) = \sum_{k=1}^{\infty} kc_k(x-a)^{k-1}.$$

We can now apply this result to the power series for the derivative and conclude that f has all derivatives, and they are given by

$$f^{(p)}(x) = \sum_{k=p}^{\infty} k(k-1)\ldots(k-p+1)c_k(x-a)^{k-p}.$$

Example

We know that $\dfrac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ for $|x|<1$. It follows that

$$\frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1} = 1 + 2x + 3x^2 + 4x^3 + \ldots$$

for $|x|<1$.

It is, miraculously enough, also true that the limit of a power series can be integrated, and the integral of the limit is the limit of the integral. Once again, the interval of convergence of the integrated series remains the same as that of the original series:

$$\int_a^x f(t)dt = \sum_{k=0}^{\infty} \frac{c_k}{k+1}(x-a)^{k+1}.$$

Example

We may simply integrate the Geometric series to get

$$\log(1-x) = -\sum_{k=0}^{\infty} \frac{x^{k+1}}{k+1}, \text{ for } -1 < x < 1, \text{ or } 0 < 1-x < 2.$$

It is also valid to perform all the usual arithmetic operations on power series. Thus if $f(x) = \sum_{k=0}^{\infty} c_k x^k$ and $g(x) = \sum_{k=0}^{\infty} d_k x^k$ for $|x|<r$, then

$$f(x) \pm g(x) = \sum_{k=0}^{\infty} (c_k \pm d_k) x^k, \text{ for } |x|<r.$$

Also,

$$f(x)g(x) = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} c_i d_{k-i} \right) c_k x^k, \text{ for } |x|<r.$$

The essence of the story is that power series behave as if they were “infinite degree” polynomials—the limits of power series are just about the nicest functions in the world.

Exercises

6. What is the limit of the series $\left(\sum_{k=0}^{n} x^{2k}\right)$? What is its interval of convergence?

7. What is the limit of the series $\left(\sum_{k=1}^{n} 2(-1)^k kx^{2k-1}\right)$? What is its interval of convergence?

8. Find a power series that converges to $\tan^{-1} x$ on some nontrivial interval.

9. Suppose $f(x) = \sum_{k=0}^{\infty} c_k (x-a)^k$. What is $f^{(p)}(a)$?

11.3 Taylor Series

Our major interest in finding a power series that converges to a given function. The obvious candidate for such a series is simply the sequence of Taylor polynomials of increasing degree. Thus if *f* is a given function, and *a* is a point in the interior of the domain of *f*, the ***Taylor Series for f*** *at a* is the series

$$\left(\sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k\right).$$

The Taylor Series is thus an "infinite degree" Taylor Polynomial>

In general, the Taylor series for a function may not converge on any nontrivial interval to *f*, but, mercifully, for many sufficiently nice functions it does. In such cases, we are provided with the nice answer to the question proposed back in Chapter Nine: Can we approximate the function *f* as well as we like by a Taylor Polynomial for sufficiently large degree?

Example

The Taylor series for $f(x) = \sin x$ at $x = a$ is simply $\left(\sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!}\right)$. An easy calculation shows us that the radius of convergence is infinite, or in other words, this power series converges for all x. But is the limit $\sin x$? That's easy to decide. From Section 9.3, we know that

$$\left|\sin x - \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!}\right| \leq \frac{|x|^{2n+3}}{(2n+3)!},$$

and we know that

$$\lim_{n\to\infty} \frac{|x|^{2n+3}}{(2n+3)!} = 0,$$

no mater what x is. Thus we have

$$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}, \text{ for all } x.$$

Exercises

10. Find the Taylor Series at $a = 0$ for $f(x) = e^x$. Find the interval of convergence and show that the series converges to f on this interval.

11. Find the Taylor Series at $a = 0$ for $f(x) = \cos x$. Find the interval of convergence and show that the series converges to f on this interval.

12. Find the derivative of the cosine function by differentiating the Taylor Series you found in Problem #11.

13. Find the Taylor Series at $a = 1$ for $f(x) = \log x$. Find the interval of convergence and show that the series converges to f on this interval.

14. Let the function f be defined by

$$f(x) = \begin{cases} 0, \text{ for } x = 0 \\ e^{-1/x^2}, \text{ for } x \neq 0 \end{cases}.$$

Find the Taylor Series at $a = 0$ for f. Find the interval of convergence and the limit of the series.

Chapter Twelve

Integration

12.1 Introduction

We now turn our attention to the idea of an integral in dimensions higher than one. Consider a real-valued function $f : \boldsymbol{D} \to \boldsymbol{R}$, where the domain $\boldsymbol{D}$ is a nice closed subset of Euclidean n-space $\boldsymbol{R}^n$. We shall begin by seeing what we mean by the integral of f over the set $\boldsymbol{D}$; then later we shall see just what such an abstract thing might be good for in real life. Mrs. Turner taught us all about the case $n = 1$. As it was in extending the definition of a derivative to higher dimensions, our definition of the integral in higher dimensions will include the definition for dimension 1 we learned in grammar school—as always, there will be nothing to unlearn. Let us again hark back to our youth and review what we know about the integral of $f : \boldsymbol{D} \to \boldsymbol{R}$ in case $\boldsymbol{D}$ is a nice connected piece of the real line $\boldsymbol{R}$. First, in this context, the only nice closed pieces of $\boldsymbol{R}$ are the closed intervals; we thus have $\boldsymbol{D}$ is a set $[a,b]$, where $b > a$. Recall that we defined a ***partition*** P of the interval to be simply a finite subset $\{x_0, x_1, \ldots, x_n\}$ of $[a,b]$ with $a = x_0 < x_1 < x_2 < \ldots < x_n = b$. The ***mesh*** of a partition is $\max\{|x_i - x_{i-1}| : i = 1,2\ldots,n\}$. We then defined a ***Riemann sum*** $S(P)$ for this partition to be a sum

$$S(P) = \sum_{i=1}^{n} f(x_i^*)\Delta x_i ,$$

where $\Delta x_i = x_i - x_{i-1}$ is simply the length of the subinterval $[x_{i-1}, x_i]$ and x_i^* is any point in this subinterval. (Thus there is not just one Riemann sum for a partition P; the sum obviously also depends on the choices of the points x_i^*. This is not reflected in the notation.)

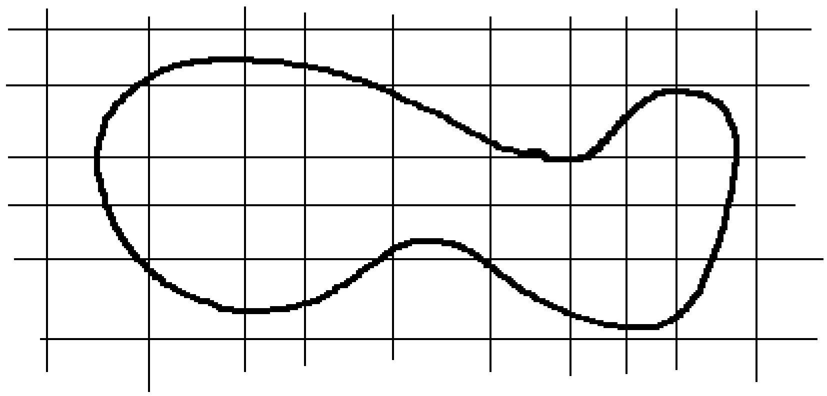
Now, if there is a number L such that we can make all Riemann sums as close as we like to L by choosing the mesh of the partition sufficiently small, then f is said to be

integrable over the interval, and the number L is called the ***integral of f over* [a, b].** This number L is almost always denoted $\int_a^b f(x)dx$. More formally, we say that L is the integral of f over $[a,b]$ if for every $\varepsilon > 0$, there is a δ so that $|S(P) - L| < \varepsilon$ for every partition P having mesh < δ. You no doubt remember from your first encounter with this integral that it initially seemed like an impossible thing to compute in any reasonable situation, but then some version of the Fundamental Theorem of Calculus came to the rescue.

12.2 Two Dimensions

Let us begin our study of higher dimensional integrals with the two dimensional case. As we have seen so often in the past, in extending calculus ideas to higher dimensions, most of the excitement occurs in taking the step from one dimension to two dimensions—seldom is the step from 97 to 98 dimensions very interesting. We shall thus begin by looking at the integral of $f: \boldsymbol{D} \to \boldsymbol{R}$ for the case in which $\boldsymbol{D}$ is a nice closed subset of the plane. Complications appear at once. On the real line, nice closed sets are simply closed intervals; in the plane, nice closed sets are considerably more interesting:

A moment's reflection convinces us that the domain ***D*** can, even in just two dimensions, be considerably more complicated than it is in one dimension. First, capture ***D*** inside a rectangle with sides parallel to the coordinate axes; and then divide this rectangle into subrectangles by partitioning each of its sides:



Now, label the subrectangles that meet ***D***, say with subscripts $i = 1, 2, \ldots, n$. The largest area of all such rectangles is called the mesh of the subdivision. In each such rectangle, choose a point (x_i^*, y_i^*) in ***D***. A Riemann sum S now looks like

$$S = \sum_{i=1}^{n} f(x_i^*, y_i^*)\Delta A_i \ ,$$

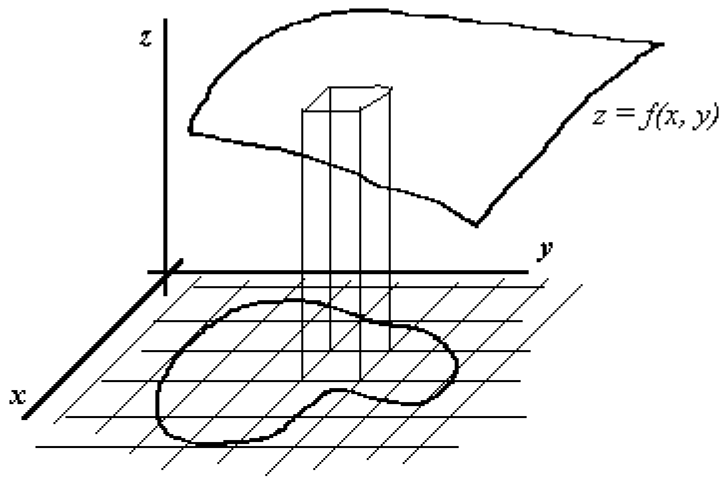
where ΔA_i is the area of the rectangle from which (x_i^*, y_i^*) is chosen. Now if there is a number L such that we can get as close to L as we like by choosing the mesh of the subdivision sufficiently small, then f is said to be integrable over ***D***, and the number L is the integral of f over ***D***. The number L is usually written with two snake signs:

$$\iint_{\boldsymbol{D}} f(x, y)\,dA \ .$$

Such integrals over two dimensional domains are frequently referred to as ***double integrals***.

I hope the definition of the integral in case ***D*** is a nice subset of $\boldsymbol{R}^3$ is evident. We capture ***D*** inside a box, and subdivide the box into boxes, *etc.* , *etc*. There will be more of the higher dimensional stuff later.

Let's look a bit at some geometry. For the purpose of drawing a reasonable picture, let us suppose that $f(x, y) \geq 0$ everywhere on ***D***.



Each term $f(x_i^*, y_i^*)\Delta A_i$ is the volume of a box with base the rectangle A_i and height $f(x_i^*, y_i^*)$. The top of the box thus meets the surface $z = f(x, y)$. The Riemann sum is thus the total volume of all such boxes. Convince yourself that as the size of the bases of the boxes goes to 0, the boxes "fill up" the solid bounded below by the *x*-*y* plane, above by the surface $z = f(x, y)$, and on the sides by the cylinder determined by the region ***D***. The integral $\iint_D f(x, y)\,dA$ is thus equal to the volume of this solid. If $f(x, y) \leq 0$, then, of course, we get the negative of the volume bounded below by the surface $z = f(x, y)$, above by the *x*-*y* plane, *etc*.

Suppose a and b are constants, and $\boldsymbol{D} = \boldsymbol{E} \quad \boldsymbol{F}$, where ***E*** and ***F*** are nice domains whose interiors do not meet. The following important properties of the double integral should be evident:

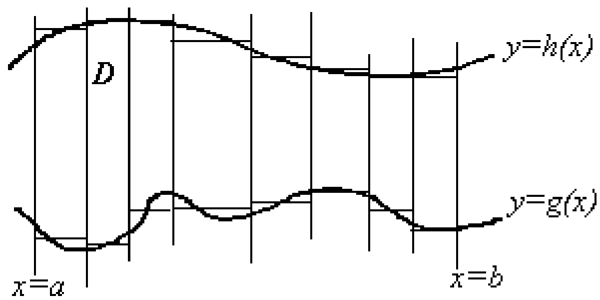
$$\iint_D [af(x,y) + bg(x,y)]dA = a\iint_D f(x,y)dA + b\iint_D g(x,y)dA, \text{ and}$$

$$\iint_D f(x,y)dA = \iint_E f(x,y)dA + \iint_F f(x,y)dA.$$

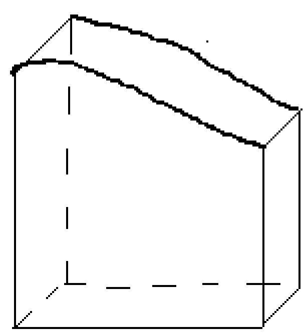
Now, how on Earth do we ever find an integral $\iint_D f(x,y)dA$? Let's see. Again, we shall look at a picture, and again we shall draw our picture as if $f(x,y) \geq 0$. It should be clear what happens if this is not the case.

We assume our domain $\boldsymbol{D}$ has a special form; specifically, we suppose it to be bounded above by the curve $y = h(x)$, below by $y = g(x)$, on the left by $x = a$, and on the right by $x = b$:

It is convenient for us to think of the integral $\iint_D f(x,y)dA$ as the volume of the blob bounded below by $\boldsymbol{D}$ in the x-y plane and above by the surface $z = f(x,y)$. Think of finding this volume by dividing the blob into slices parallel to the y-axis and adding up the volumes of the slices. To approximate the volumes of these slices, we use slabs:



We partition the x interval $[a, b]$: $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$. In each subinterval $[x_{i-1}, x_i]$ choose a point x_i^*. Our approximating slab has as its base the rectangle of "width" $\Delta x_i = x_i - x_{i-1}$ and height $h(x_i^*) - g(x_i^*)$; the roof is $z = f(x_i^*, y)$. The volume of the slab is the cross section area times the thickness, or $[\int_{g(x_i^*)}^{h(x_i^*)} f(x_i^*, y)dy]\Delta x_i$.



The sum of the volumes of the approximating slabs is thus

$$S = \sum_{i=1}^{n} [\int_{g(x_i^*)}^{h(x_i^*)} f(x_i^*, y)dy]\Delta x_i \ .$$

The double integral we seek is just the "limit" of these as we take thinner and thinner slabs; or finer and finer partitions of the interval $[a, b]$. But Lo! The above sums are

Riemann sums for the ordinary one dimensional integral of the function $F(x) = \int_{g(x)}^{h(x)} f(x,y)dy$, and so the double integral is given by

$$\iint_D f(x,y)dA = \int_a^b F(x)dx$$

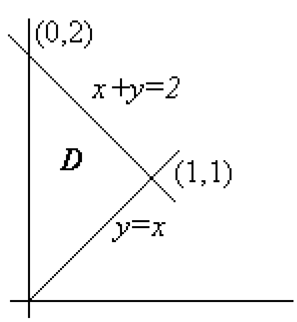
$$= \int_a^b [\int_{g(x)}^{h(x)} f(x,y)dy]dx$$

The double integral is thus equal to an integral of an integral, usually called an ***iterated integral***. It is traditional to omit the brackets and write the iterated integral simply as

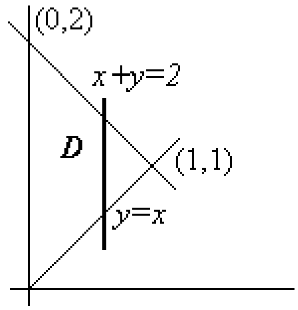
$\int_a^b \int_{g(x)}^{h(x)} f(x,y)dydx$.

Example

Let's find the double integral $\iint_D [x^2 + y^2]dA$, where ***D*** is the area enclosed by the lines $y = x$, $x = 0$, and $x + y = 2$. The first item of business here is to draw a picture of ***D*** (We *always* need a picture of the domain of integration.):



It should be clear from the picture that in the language of our discussion, $g(x) = x$, $h(x) = 2 - x$, $a = 0$, and $b = 1$. So slice parallel to the y axis:



The lower end of the slice is at $y = x$ and the upper end is at $y = 2 - x$. The "volume" is thus

$$\int_x^{2-x}[x^2 + y^2]dy = x^2 y + \frac{y^3}{3}\Bigg|_{y=x}^{y=2-x} = x^2(2-x) + \frac{(2-x)^3}{3} - x^3 - \frac{x^3}{3} = 2x^2 + \frac{(2-x)^3}{3} - \frac{7}{3}x^3,$$

and we have such a slice for all x from $x = 0$ to $x = 1$. Thus

$$\begin{aligned}\iint_D [x^2 + y^2]dA &= \int_0^1 [2x^2 + \frac{(2-x)^3}{3} - \frac{7}{3}x^3]dx \\ &= \frac{2x^3}{3} - \frac{(2-x)^4}{12} - \frac{7x^4}{12}\Bigg|_0^1 \\ &= \frac{16}{12} = \frac{4}{3}\end{aligned}$$

Exercises

1. Find $\iint_D x^2 dA$, where $\boldsymbol{D}$ is the domain bounded by the curves $y = 4 - x^2$ and $y = 3x$.

2. Find $\iint_D (x^2 - y)dA$, where ***D*** is the area in the first quadrant enclosed by the coordinate axes and the line $2x + y = 4$.

3. Use double integration to find the area of the region enclosed by the curves $x - y = 2$ and $y = -x^2$.

4. Find the volume of the solid cut from the first octant by the surface $z = 4 - x^2 - y$.

5. Sketch the domain of integration and evaluate the iterated integral:

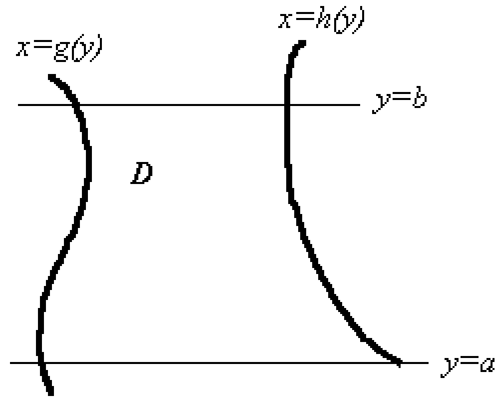
$$\int_0^1 \int_x^1 y^2 e^{xy} dy dx .$$

6. Sketch the domain of integration and evaluate the iterated integral:

$$\int_1^{\log 8} \int_0^{\log x} e^{x+y} dy dx .$$

7. Find the volume of the wedge cut from the first octant by the cylinder $z = 12 - 3y^2$ and the plane $x + y = 2$.

8. Suppose you have a double integral $\iint_D f(x, y)dA$ in which the domain ***D*** is bounded on the left by the curve $x = g(y)$, on the right by $x = h(y)$, below by $y = a$, and above by $y = b$.



Give an iterated integral for the double integral in which the first integration is with respect to x , and explain what's going on.

9. Give a double integral for the area of the region bounded by $x = y^2$ and $x = 2y - y^2$, and evaluate the integral.

Chapter Thirteen

More Integration

13.1 Some Applications

Think now for a moment back to elementary school physics. Suppose we have a system of point masses and forces acting on the masses. Specifically, suppose that for each $i = 1, 2, \ldots, n$ we have a point mass m_i whose position in space at time t is given by the vector $\boldsymbol{r}_i$ .. Assume moreover that there is a force $\boldsymbol{f}_i$ acting on this mass. Thus according to Sir Isaac Newton, we have

$$\boldsymbol{f}_i = m_i \frac{d^2 \boldsymbol{r}_i}{dt^2}$$

for each i. Now sum these equations to get

$$\boldsymbol{F} = \sum_{i=1}^{n} \boldsymbol{f}_i = \sum_{i=1}^{n} m_i \frac{d^2 \boldsymbol{r}_i}{dt^2}, \text{ or}$$

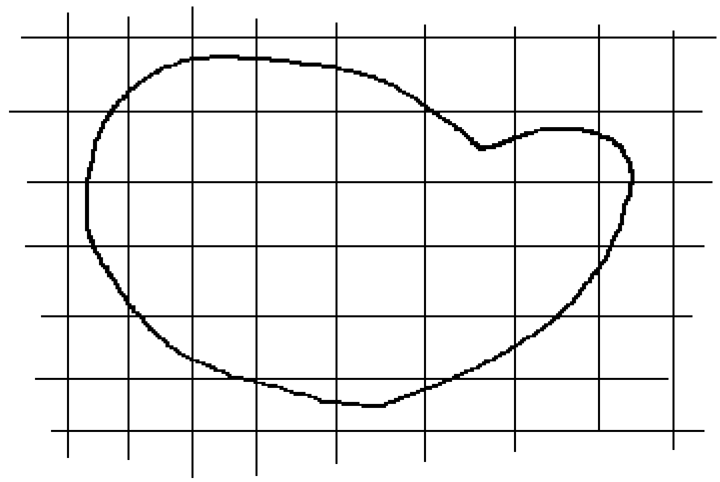
$$\boldsymbol{F} = M \frac{d^2}{dt} \left(\frac{\sum_{i=1}^{n} m_i \boldsymbol{r}_i}{\sum_{i=1}^{n} m_i} \right),$$

where $M = \sum_{i=1}^{n} m_i$. Reflect for a moment on this equation. If we define $\boldsymbol{R}$ by

$\boldsymbol{R} = \frac{\sum_{i=1}^{n} m_i \boldsymbol{r}_i}{\sum_{i=1}^{n} m_i}$, then the equation becomes $\boldsymbol{F} = M \frac{d^2 \boldsymbol{R}}{dt^2}$. Thus the sum of the external forces on the system of masses is the total mass times the acceleration of the mystical point $\boldsymbol{R}$. This point $\boldsymbol{R}$ is called the ***center of mass*** of the system.

In case the total mass is continuously distributed in space, the "sum" in the equation for $\boldsymbol{R}$ becomes an integral. Let's look at what this means in two dimensions.

Suppose we have a plate and the mass density of the plate at (x,y) is given by $\rho(x,y)$. To find the center of mass of the plate, we approximate its location by chopping it into a bunch of small pieces and treating each of these pieces as a point mass.



Now choose a point $\boldsymbol{r}_i = x_i^*\boldsymbol{i} + y_i^*\boldsymbol{j}$ in each rectangle. The mass of this rectangle will be approximately $\rho(x_i^*, y_i^*)\Delta A_i$, where ΔA_i is the area of the rectangle. The equation for the center of mass of this system of rectangles is then

$$\tilde{\boldsymbol{R}} = \frac{\sum_{i=1}^{n} m_i \boldsymbol{r}_i}{\sum_{i=1}^{n} m_i} = \frac{\sum_{i=1}^{n} \rho(x_i^*, y_i^*)\boldsymbol{r}_i \Delta A_i}{\sum_{i=1}^{n} \rho(x_i^*, y_i^*)\Delta A_i}$$

$$= \frac{1}{\sum_{i=1}^{n} \rho(x_i^*, y_i^*)\Delta A_i} \left\{ \left[\sum_{i=1}^{n} \rho(x_i^*, y_i^*) x_i^* \Delta A_i \right] \boldsymbol{i} + \left[\sum_{i=1}^{n} \rho(x_i^*, y_i^*) y_i^* \Delta A_i \right] \boldsymbol{j} \right\}$$

The three sums in the previous line are Riemann sums for two dimensional integrals! Thus as we take smaller and smaller rectangles, *etc*., we obtain for $\boldsymbol{R}$, the location of the center of mass

$$\boldsymbol{R} = \frac{1}{\iint\limits_P \rho(x,y)\,dA}\left\{\left[\iint\limits_P x\rho(x,y)\,dA\right]\boldsymbol{i} + \left[\iint\limits_P y\rho(x,y)\,dA\right]\boldsymbol{j}\right\}$$

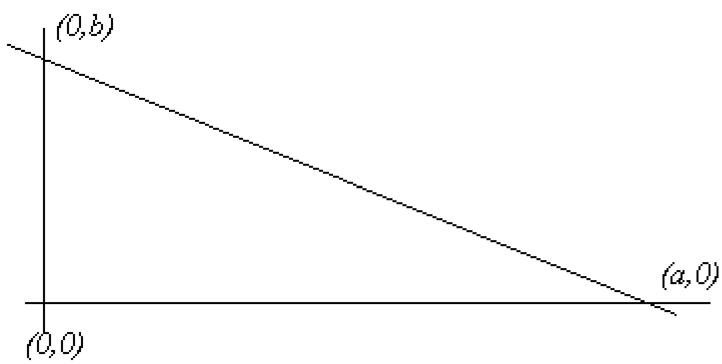
In other words, the coordinates $(\bar{x},\bar{y})$ of the center of mass of P are given by

$$\bar{x} = \frac{\left[\iint\limits_P x\rho(x,y)dA\right]}{M}, \text{ and } \bar{y} = \frac{\left[\iint\limits_P y\rho(x,y)dA\right]}{M},$$

where $M = \iint\limits_P \rho(x,y)dA$ is the total mass of the plate.

Example

Let's find the center of mass of a plate having the shape of the plane region enclosed by the triangle



and having constant density (In this case, we say the mass is ***uniformly distributed*** over the region. Suppose $\rho(x,y) = k$. First,

$$\iint\limits_T x\rho(x,y)dA = k\int_0^a\int_0^{b(1-x/a)} xdydx = k\int_0^a xb(1-x/a)dx = k\frac{a^2b}{6}, \text{ and then}$$

$$\iint\limits_T y\rho(x,y)\,dA = k\int_0^a\int_0^{b(1-x/a)} ydydx = \frac{kb^2}{2}\int_0^a (1-x/a)^2 dx = k\frac{ab^2}{6}.$$

Also, $M = \iint_T k\,dA = k\iint_T dA = k\frac{ab}{2}$. Thus,

$$\bar{x} = \frac{a}{3}, \text{ and } \bar{y} = \frac{b}{3}.$$

Meditate on the fact that the location of the center of mass does not depend on the value of the constant k. Note that in general, if the density is constant, then the constant slips out through the integral signs and cancels top and bottom in the recipe for the coordinates $(\bar{x}, \bar{y})$. This is what most of our intuitions tell us, I believe. It is, nevertheless, comforting to see this fact come out in the mathematical wash. In this case of constant density, the center of mass thus depends only on the geometry of the plate; it is thus a geometric property of the region. It is called the ***centroid*** of the region. One must never confuse the two concepts; intimately related though they be, they are different. The center of mass is something a physical body has, while the centroid is an abstract mathematical something.

Exercises

1. Find the center of mass of a plate of density $\rho(x, y) = y + 1$ having the shape of the area bounded by the line $y = 1$ and the parabola $y = x^2$.

2. Find the center of mass of the smaller of the two regions cut from the elliptical region $x^2 + 4y^2 = 12$ by the parabola $x = 4y^2$ if the density $\rho(x, y) = 5x$.

3. Find the centroid of the semicircular region $\{(x, y) \quad \boldsymbol{R^2} : x^2 + y^2 \le a^2, \text{ and } y \ge 0\}$.

4. Find the centroid of the region bounded by the horizontal axis and one arch of the sine curve. (That is, the region between $x = 0$ and $x = \pi$ bounded above by $y = \sin x$ and below by $y = 0$.)

5. Find the centroid of the region bounded by the curves $y^2 = 2x$, $x + y = 4$, and $y = 0$.

6. The area of a region A is $\int_0^2 \int_{x^2-4}^0 dydx + \int_0^4 \int_0^{\sqrt{x}} dydx$. Draw a picture of the region.

7. Let $f : \boldsymbol{D} \to \boldsymbol{R}$ be a function defined on a nice subset $\boldsymbol{D}$ $\quad \boldsymbol{R}^2$. The ***average value*** A of f on $\boldsymbol{D}$ is defined to be $A = \dfrac{1}{\text{area of } \boldsymbol{D}} \iint_{\boldsymbol{D}} f(x, y) dA$.

a)Find the average depth of a bowl having the shape of the bottom half of the sphere $x^2 + y^2 + z^2 = 1$.

b)Find the average depth of a bowl having the shape of the part of the paraboloid $z = x^2 + y^2 - 1$ below the x-y plane.

8. Let $\boldsymbol{D}$ be the region inside the circle $x^2 + (y - a)^2 = a^2$ that lies below the line $y = a$.

a)Find the centroid of $\boldsymbol{D}$.

b)Find the point on the semicircular boundar of $\boldsymbol{D}$ that is closest to the centroid.

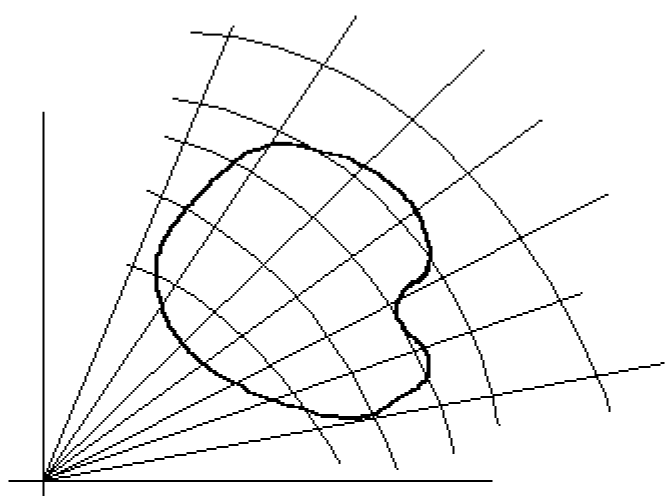
13.2 Polar Coordinates

Now we shall see what happens when we express a double integral as an iterated integral in some coordinate system other than the usual rectangular, or Cartesian, coordinate system. We shall see more of this later; right now, let's look at what happens in ***polar coordinates***.

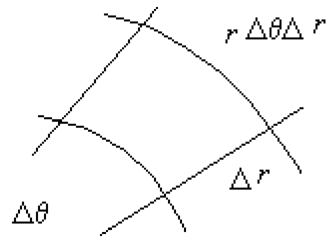
Suppose we have the integral $\iint_D f(x, y)\,dA$. In polar coordinates, we know that we must substitute

$$x = r\cos\theta, \text{ and}$$
$$y = r\sin\theta.$$

There is, however, more to it than this. When we divided the plane into regions formed by the curves $x =$ constant and $y =$ constant, we got rectangles, *etc.*, *etc.* Now we divide the plane into regions formed by the curves $r =$ constant and $\theta =$ constant, where r and θ are the usual polar coordinates. This results in funny shaped regions:



Now, a typical region looks like



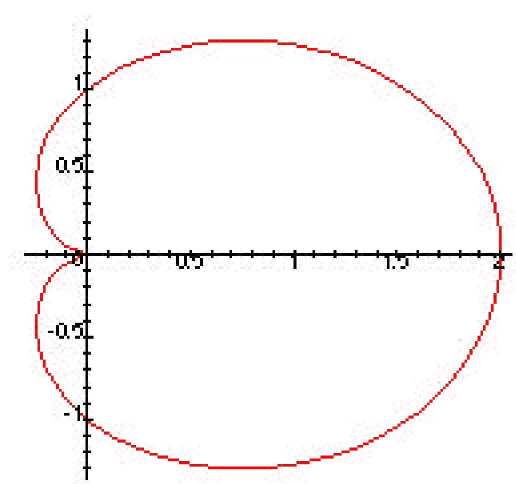
The area of this region is thus something like $\Delta A \approx r\Delta r\Delta\theta$, and our iterated integral looks like

$$\iint_D f(x,y)dA = \iint f(r\cos\theta, r\sin\theta)rdrd\theta$$

together with the appropriate limits of integration. (We may, of course, integrate first with respect to θ and then with respect to r if this is convenient.) We desperately need to see an example.

Example

Let's find the centroid of the region enclosed by the curve whose equation in polar coordinates is $r = 1 + \cos\theta$. Here is a picture drawn by *Maple*:



The centroid $(\bar{x}, \bar{y})$ is given by

$$\bar{x} = \frac{\iint_D x\,dA}{\iint_D dA}, \text{ and } \bar{y} = \frac{\iint_D y\,dA}{\iint_D dA}.$$

First. let's find the integral $\iint_D x\,dA$. Now, when we hold θ fixed and integrate first with respect to r, the lower limit is independent of θ and is always $r = 0$, while the upper limit depends, of course on θ and is $r = 1 + \cos\theta$. We have a slice for each value of θ from $\theta = 0$ to $\theta = 2\pi$, and so our iterated integral looks like

$$\iint_D x\,dA = \int_0^{2\pi}\int_0^{1+\cos\theta} r\cos\theta\; r\,dr\,d\theta = \int_0^{2\pi}\int_0^{1+\cos\theta} r^2\cos\theta\; dr\,d\theta.$$

It is downhill all the way now:

$$\int_0^{2\pi}\int_0^{1+\cos\theta} r^2\cos\theta\; dr\,d\theta = \frac{1}{3}\int_0^{2\pi}(1+\cos\theta)^3\cos\theta\,d\theta$$
$$= \frac{1}{3}\int_0^{2\pi}[\cos\theta + 3\cos^2\theta + 3\cos^3\theta + \cos^4\theta]d\theta$$
$$= \frac{1}{3}[0 + \frac{3}{2}\int_0^{2\pi}(1+\cos 2\theta)d\theta + 0 + \frac{1}{4}\int_0^{2\pi}(1+\cos 2\theta)^2 d\theta]$$
$$= \pi + \frac{\pi}{6} + \frac{1}{12}\int_0^{2\pi}\cos^2 2\theta\,d\theta = \pi + \frac{\pi}{6} + \frac{\pi}{12} = \frac{15\pi}{12} = \frac{5}{4}\pi$$

Now for the other integrals.

It should be clear that $\iint_D y\,dA = \int_0^{2\pi}\int_0^{1+\cos\theta} r^2\sin\theta\; dr\,d\theta = 0$. Finally,

$$\iint_D dA = \int_0^{2\pi}\int_0^{1+\cos\theta} r\,dr\,d\theta = \frac{1}{2}\int_0^{2\pi}(1+\cos\theta)^2\,d\theta$$

$$= \pi + \frac{1}{4}\int_0^{2\pi}(1+\cos 2\theta)\,d\theta$$

$$= \pi + \frac{\pi}{2} = \frac{3}{2}\pi$$

We are, at last, done.

$$\bar{x} = \frac{\frac{5}{4}\pi}{\frac{3}{2}\pi} = \frac{5}{6}, \text{ and } \bar{y} = 0.$$

Exercises

9. Find the area of the region enclosed by the curve with polar equation $r = \sin 2\theta$.

10. Evaluate the integral $\iint_D (x+y)\,dA$, where $\boldsymbol{D}$ is the region in the first quadrant inside the circle $x^2 + y^2 = a^2$ and below the line $y = x\sqrt{3}$.

11. Find the centroid of the region in the first quadrant inside the circle $r = a$ and between the rays $\theta = 0$ and $\theta = \alpha$, where $0 \le \alpha \le \frac{\pi}{2}$. What is the limiting position of the centroid as $\alpha \to 0$?

12. Evaluate $\iint_R e^{x^2+y^2}\,dA$, where R is the semicircular region bounded above by $y = \sqrt{1-x^2}$ and below by the x axis.

13. Find the area enclosed by one leaf of the rose $r = \cos 3\theta$.

14. Find the area of the region inside $r = 1 + \cos\theta$ and outside $r = 1$.

13. 3 Three Dimensions

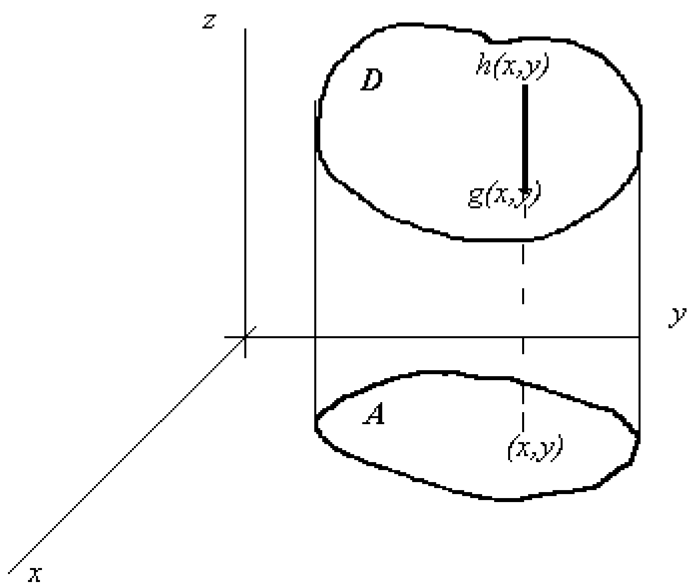
We move along to integrals in three dimensions. The idea is quite simple. Suppose we have a function $f : \boldsymbol{D} \to \boldsymbol{R}$, where $\boldsymbol{D}$ is a nice subset of $\boldsymbol{R}^3$. Capture $\boldsymbol{D}$ inside a big box (*i.e.*, a rectangular parallelepiped). Now subdivide this box by partitioning each of its sides. The volume of the largest such box is called the ***mesh*** of the subdivision. In each box that meets $\boldsymbol{D}$, choose a point (x_i^*, y_i^*, z_i^*) in $\boldsymbol{D}$. A Riemann sum S now looks like

$$S = \sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*)\Delta V_i ,$$

where ΔV_i is the volume of the box from which (x_i^*, y_i^*, z_i^*) was chosen. (The summation is over all boxes that meet $\boldsymbol{D}$.) If there is a number L such that $|S - L|$ can be made arbitrarily small by choosing a subdivision of sufficiently small mesh, then we say that f is ***integrable*** over $\boldsymbol{D}$, and the number L is called the ***integral of f over D***. This integral is usually written with three snake signs:

$$\iiint_{\boldsymbol{D}} f(x, y, z)dV .$$

Let's see how to evaluate such a thing by considering iterated integrals. Here's what we do. First, project $\boldsymbol{D}$ onto a coordinate plane. (We choose the x-y plane as an example.)

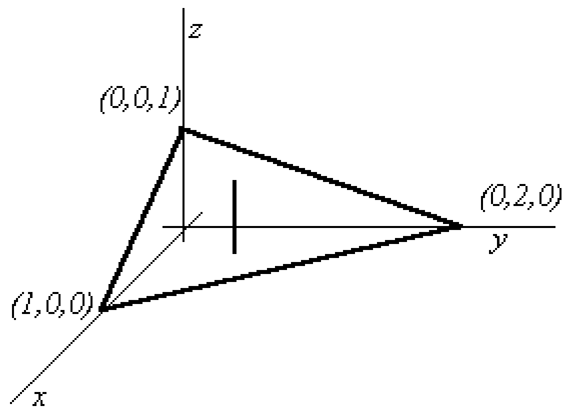


Let ***A*** be the region in the x-y plane onto which ***D*** projects. Assume that a vertical line through a point $(x,y) \in$ ***A*** enters ***D*** through the surface $z = g(x,y)$ and exits through the surface $z = h(x,y)$. In other words, the blob ***D*** is the solid above the region ***A*** between the surfaces $z = g(x,y)$ and $z = h(x,y)$. Now we simply integrate the integral $\int_{g(x,y)}^{h(x,y)} f(x,y,z)dz$ over the region ***A***:

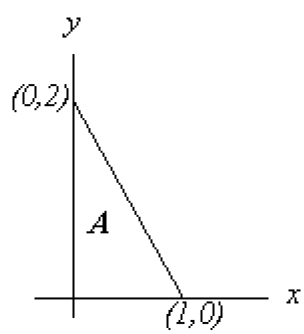
$$\iiint_D f(x,y,z)dV = \iint_A \left(\int_{g(x,y)}^{h(x,y)} f(x,y,z)dz \right) dA.$$

Example

Let's find the integral $\iiint_D (x + 2y + z)dV$, where ***D*** is the tetrahedron with vertices (0,0,0), (1,0,0), (0,2,0), and (0,0,1).



When we project ***D*** onto the *x*-*y* plane, the bottom of ***D*** is the surface $z = 0$ and the top of ***D*** is $x + \frac{y}{2} + z = 1$, or $z = 1 - x - \frac{y}{2}$. The projection is simply the triangle



Our iterated integral is thus simply $\iint_A \left(\int_0^{1-x-y/2} (x + 2y + z)dz \right) dA$. We now write the double integral over ***A*** as an iterated integral, and we have

$$\iiint_D (x + 2y + z)dV = \iint_A \left(\int_0^{1-x-y/2} (x + 2y + z)dz \right) dA$$

$$= \int_0^1 \int_0^{2(1-x)} \int_0^{1-x-y/2} (x + 2y + z)dzdydx .$$

Again, it is traditional to omit the parentheses in the iterated integral. All we need do now is integrate three times. Let's use *Maple* for the calculations, but look at the intermediate steps, rather than just use one statement. Here we go.

For the first integration, we want $\int_0^{1-x-y/2}(x+2y+z)dz$:

int(x+2*y+z,z=0..(1-x-y/2));

$$-\frac{1}{2}x^2-2xy+\frac{3}{2}y-\frac{7}{8}y^2+\frac{1}{2}$$

Thus,

$$\int_0^{1-x-y/2}(x+2y+z)dz=-\frac{1}{2}x^2-2xy+\frac{3}{2}y-\frac{7}{8}y^2+\frac{1}{2},$$

and our next integral is

$$\int_0^{2(1-x)}\int_0^{1-x-y/2}(1+2y+z)dzdy=\int_0^{2(1-x)}(-\frac{1}{2}x^2-2xy+\frac{3}{2}y-\frac{7}{8}y^2+\frac{1}{2})dy.$$

Maple again:

int(-(x^2)/2-2*x*y+(3/2)*y-(7/8)*y^2+1/2,y=0..2*(1-x));

$$-4x-\frac{2}{3}x^3+3x^2+\frac{5}{3}$$

Thus,

$$\int_0^{2(1-x)}(-\frac{1}{2}x^2-2xy+\frac{3}{2}y-\frac{7}{8}y^2+\frac{1}{2})dy=-4x-\frac{2}{3}x^3+3x^2+\frac{5}{3},$$

and finally,

int(-4*x-(2/3)*x^3+3*x^2+(5/3),x=0..1);

$$\frac{1}{2}$$

At last!

$$\int_0^1 \int_0^{2(1-x)} \int_0^{1-x-y/2} (x+2y+z)\,dz\,dy\,dx = \frac{1}{2}.$$

We make a few obvious observations. First, if $\boldsymbol{S}$ is a solid, the volume V of the solid is simply $V = \iiint_S dV$. If the mass density of a blob having the shape of $\boldsymbol{S}$ is $\rho(x,y,z)$, then the mass M of the blob is $M = \iiint_S \rho(x,y,z)\,dV$, and the location $(\bar{x},\bar{y},\bar{z})$ of the center of mass is given by

$$\bar{x} = \frac{\iiint_S x\rho(x,y,z)\,dV}{M}$$

$$\bar{y} = \frac{\iiint_S y\rho(x,y,z)\,dV}{M}$$

$$\bar{z} = \frac{\iiint_S z\rho(x,y,z)\,dV}{M}$$

Exercises

15. Find the volume of the tetrahedron having vertices (0,0,0), (*a*,0,0),(0,*b*,0), and (0,0,*c*).

16. Find the centroid of the tetrahedron in the previous exercise.

17. Evaluate $\iiint_S (xy + z^2)\,dV$, where S is the set $S = \{(x,y,z): 0 \le z \le 1-|x|-|y|\}$.

18. Find the volume of the region in the first octant bounded by the coordinate planes and the surface $z = 4 - x^2 - y$.

19. Write six different iterated integrals for the volume of the tetrahedron cut from the first octant by the plane $12x + 4y + 3z = 12$.

20. A solid is bounded below by the surface $z = 4y^2$, above by the surface $z = 4$, and on the ends by the surfaces $x = 1$ and $x = -1$. Find the centroid.

21. Find the volume of the region common to the interiors of the cylinders $x^2 + y^2 = 1$ and $x^2 + z^2 = 1$.

Chapter Fourteen

One Dimension Again

14.1 Scalar Line Integrals

Now we again consider the idea of the integral in one dimension. When we were introduced to the integral back in elementary school, we considered only functions defined on nice subsets of the real line. The notion of an integral of a function $f:\boldsymbol{D} \to \boldsymbol{R}$ in which $\boldsymbol{D}$ is a nice one dimensional set, but is *not* a subset of the reals is our next object of study. To get some idea of why one might care about such a thing, consider the simple problem of finding the mass of a piece of wire having the shape of an arc of a space curve $\boldsymbol{C}$ and having a given density $\rho(\boldsymbol{r})$. How might we approach such a problem? Simple enough! We subdivide, or partition, the curve with a finite set of points, say $\{\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_n\}$. On the subarc joining $\boldsymbol{r}_{i-1}$ to $\boldsymbol{r}_i$, we choose a point, say $\boldsymbol{r}_i^*$, and evaluate the function $\rho(\boldsymbol{r}_i^*)$. Now we multiply this times the length of the line segment joining the points $\boldsymbol{r}_{i-1}$ and $\boldsymbol{r}_i$ for an approximation to the mass of this arc of our curve. Then sum these to obtain an approximation for the total mass:

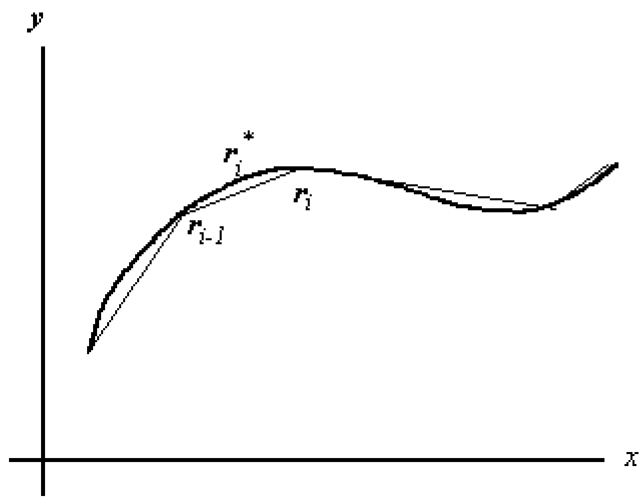
$$S = \sum_{i=1}^{n} \rho(\boldsymbol{r}_i^*)|\boldsymbol{r}_i - \boldsymbol{r}_{i-1}|.$$

Then we all believe that the "limit" of these sums as we choose finer and finer partitions of the curve should be the actual, honest-to-goodness mass of the wire.

Let's abstract the essence of the discussion. Suppose $f:\boldsymbol{C} \to \boldsymbol{R}$ is a function whose domain $\boldsymbol{C}$ is a curve (in $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$, or wherever). We subdivide the curve as in the preceding discussion and choose a point $\boldsymbol{r}_i^*$ on the subarc joining $\boldsymbol{r}_{i-1}$ to $\boldsymbol{r}_i$. The sum

$$S = \sum_{i=1}^{n} f(\boldsymbol{r}_i^*)|\boldsymbol{r}_i - \boldsymbol{r}_{i-1}|$$

again is called a ***Riemann sum***. If there is a number L such that all Riemann sums are arbitrarily close to L for sufficiently fine partitions, then we say f is ***integrable*** on ***C***, and the number L is called the ***integral of f on C*** and is denoted $\int_C f(\boldsymbol{r})d\boldsymbol{r}$. This integral is also frequently referred to as a ***line integral***.



This is wonderful, but how do find such an integral? It is remarkably simple and easy. Suppose we have a vector description of the curve ***C***; say $\boldsymbol{r}(t)$, for $a \le t \le b$. We partition the curve by partitioning the interval $[a,b]$: If $\{a = t_0, t_1, \ldots, t_n = b\}$ is a partition of the interval, then the points $\{\boldsymbol{r}(t_0), \boldsymbol{r}(t_1), \ldots, \boldsymbol{r}(t_n)\}$ partition the curve ***C***. We obtain the point $\boldsymbol{r}_i^*$ on the subarc joining $\boldsymbol{r}(t_{i-1})$ to $\boldsymbol{r}(t_i)$ by choosing $t_i^* \quad [t_{i-1}, t_i]$ and letting $\boldsymbol{r}_i^* = \boldsymbol{r}(t_i^*)$. Our Riemann sum now looks like

$$S = \sum_{i=1}^{n} f(\boldsymbol{r}(t_i^*)|\boldsymbol{r}(t_i) - \boldsymbol{r}(t_{i-1})|.$$

Next, multiply the terms on the right by one, but one disguised as $\frac{\Delta t_i}{\Delta t_i}$, where, of course, $\Delta t_i = t_i - t_{i-1}$. Then we see

$$S = \sum_{i=1}^{n} f(\boldsymbol{r}(t_i^*))\left|\frac{\boldsymbol{r}(t_i) - \boldsymbol{r}(t_{i-1})}{\Delta t_i}\right|\Delta t_i .$$

We know that $\lim_{\Delta t \to 0}\left|\frac{\boldsymbol{r}(t_i) - \boldsymbol{r}(t_{i-1})}{\Delta t_i}\right| = \left|\frac{d\boldsymbol{r}}{dt}\right|$, and so it is not hard to convince oneself that the "limiting" value of the Riemann sums is

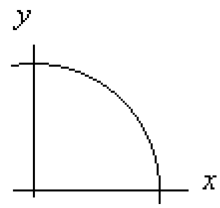
$$\int_a^b f(\boldsymbol{r}(t))\left|\frac{d\boldsymbol{r}(t)}{dt}\right|dt .$$

We have thus turned the problem into one we know how to solve—a plain old everyday elementary calculus integral. Hence,

$$\int_C f(\boldsymbol{r})d\boldsymbol{r} = \int_a^b f(\boldsymbol{r}(t))\left|\frac{d\boldsymbol{r}(t)}{dt}\right|dt .$$

Example

Suppose we have a wire in the shape of a quarter circle of radius 2, and the density of the wire is given by $\rho(x, y) = y$. What is the mass of the wire? Well, we know the mass is simply the integral $\int_C y d\boldsymbol{r}$, where $\boldsymbol{C}$ is the quarter circle:



A vector description of the curve is $\boldsymbol{r}(t) = 2\cos t\boldsymbol{i} + 2\sin t\boldsymbol{j}$, for $0 \le t \le \frac{\pi}{4}$. Thus we have $\left|\frac{d\boldsymbol{r}}{dt}\right| = |-2\sin t\boldsymbol{i} + 2\cos t\boldsymbol{j}| = 2$, and the integral becomes simply

$$\int_C y\,d\boldsymbol{r} = \int_0^{\pi/4} 4\sin t\,dt = 4.$$

Let's see what happens if we use a different vector description of the curve, say $\boldsymbol{r}(t) = t\boldsymbol{i} + \sqrt{4-t^2}\,\boldsymbol{j}$ for $0 \le t \le 2$. We have $\left|\frac{d\boldsymbol{r}}{dt}\right| = \left|\boldsymbol{i} - \frac{t}{\sqrt{4-t^2}}\boldsymbol{j}\right| = \frac{2}{\sqrt{4-t^2}}$. Hence

$$\int_C y\,d\boldsymbol{r} = \int_0^2 \sqrt{4-t^2}\left(\frac{2}{\sqrt{4-t^2}}\right)dt = \int_0^2 2\,dt = 4.$$

We get, as we must, the same answer.

Exercises

1. Evaluate the integral $\int_C (x - y + z + 2)\,d\boldsymbol{r}$, where $\boldsymbol{C}$ is the curve $\boldsymbol{r}(t) = t\boldsymbol{i} + (1-t)\boldsymbol{j} + \boldsymbol{k}$, $0 \le t \le 1$.

2. Evaluate the integral $\int_C \sqrt{x^2+y^2}\,d\boldsymbol{r}$, where $\boldsymbol{C}$ is the curve $\boldsymbol{r}(t) = 4\cos t\boldsymbol{i} + 4\sin t\boldsymbol{j} + 3t\boldsymbol{k}$,

$-2\pi \le t \le 2\pi$.

3. Find the centroid of a semicircle of radius a.

4. Find the mass of a wire having the shape of the curve $\boldsymbol{r}(t) = (t^2-1)\boldsymbol{j} + 2t\boldsymbol{k},\ 0 \le t \le 1$ if the density is $\rho(t) = \frac{3}{2}t$.

5. Find the center of mass of a wire having the shape of the curve

$$\boldsymbol{r}(t) = t\boldsymbol{i} + \frac{2\sqrt{2}}{3}t^{3/2}\,\boldsymbol{j} + \frac{t^2}{2}\boldsymbol{k},\ 0 \le t \le 2,$$

if the density is $\rho(t) = \frac{1}{t+1}$.

6. What is $\int_C d\boldsymbol{r}$?

14.2 Vector Line Integrals

Now we are introduce something perhaps a little different from what we have seen to now—integrals with vector valued integrands. Specifically, suppose $\boldsymbol{C}$ is a space curve and $\boldsymbol{f}:\boldsymbol{C} \to \boldsymbol{R}^3$ is a function from $\boldsymbol{C}$ into the Euclidean space $\boldsymbol{R}^3$. We are going to define an integral $\int_C \boldsymbol{f}(\boldsymbol{r})\ d\boldsymbol{r}$. Why should we care about such a thing? Again, let's think about a physical model. You learned in fifth grade physics that the work done by a force F acting through a distance d is simply the product Fd. The force F and the displacement d are, of course, really vectors, and we saw earlier in life that the "product" of the two is actually the scalar, or dot, product of the two vectors. Now, in general, neither of these quantities will be constant, and we will have a variable force $\boldsymbol{F}(\boldsymbol{r})$ acting along a curve $\boldsymbol{C}$ in space. How do we compute the work done in this situation? Let's see. Once more, we partition the curve by choosing a sequence of points $\{\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_n\}$ on the curve, with $\boldsymbol{r}_0$ being the initial point and $\boldsymbol{r}_n$ being the final point. Now, of course, there is an *orientation,* or direction, specified on the curve. One may think of specifying an orientation by simply putting an arrow on the curve—it thus makes sense to speak of the initial point and the terminal point of the curve. Exactly as in the scalar integrand case, we choose a point $\boldsymbol{r}_i^*$ on the subarc joining $\boldsymbol{r}_{i-1}$ to $\boldsymbol{r}_i$, and evaluate $\boldsymbol{F}(\boldsymbol{r}_i^*)$. Now then, the work done in going from $\boldsymbol{r}_{i-1}$ to $\boldsymbol{r}_i$ is approximately the scalar product $\boldsymbol{F}(\boldsymbol{r}_i^*) \cdot (\boldsymbol{r}_i - \boldsymbol{r}_{i-1})$. Add all these up for an approximation to the total work done:

$$S = \sum_{i=1}^{n} \boldsymbol{F}(\boldsymbol{r}_i^*) \cdot (\boldsymbol{r}_i - \boldsymbol{r}_{i-1}) .$$

The course should be obvious now; we take finer and finer partitions, and the limiting value of the sums is the integral

$$\int_{\boldsymbol{C}} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} .$$

This integral too is called a line integral. To prevent confusion, we sometimes speak of ***scalar*** line integrals and ***vector*** line integrals. How to find such a vector integral should be clear from the discussion of scalar line integrals. We let $\boldsymbol{r}(t)$, $a \le t \le b$, be a vector description of $\boldsymbol{C}$. (Here $\boldsymbol{r}(a)$ is the initial point and $\boldsymbol{r}(b)$ is the terminal point.) The discussion proceeds almost exactly as it did in the previous section and we get

$$\int_C \boldsymbol{F}(\boldsymbol{r})\, d\boldsymbol{r} = \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \frac{d\boldsymbol{r}}{dt}\, dt \,.$$

Example

Find $\int_C [(xy + z^2)\boldsymbol{i} + (x + z)\boldsymbol{j} + 2yz\boldsymbol{k}]\, d\boldsymbol{r}$, where $\boldsymbol{C}$ is the straight line from the origin to the point (1,2,3). The line $\boldsymbol{C}$ has a vector description $\boldsymbol{r}(t) = t\boldsymbol{i} + 2t\boldsymbol{j} + 3t\boldsymbol{k}$. Thus, $\frac{d\boldsymbol{r}}{dt} = \boldsymbol{i} + 2\boldsymbol{j} + 3\boldsymbol{k}$, and so

$$\int_C [(xy + z^2)\boldsymbol{i} + (x + z)\boldsymbol{j} + 2yz\boldsymbol{k}]\, d\boldsymbol{r} = \int_0^1 [(2t^2 + 9t^2)\boldsymbol{i} + (t + 3t)\boldsymbol{j} + 12t^2\boldsymbol{k}]\, (\boldsymbol{i} + 2\boldsymbol{j} + 3\boldsymbol{k})dt$$

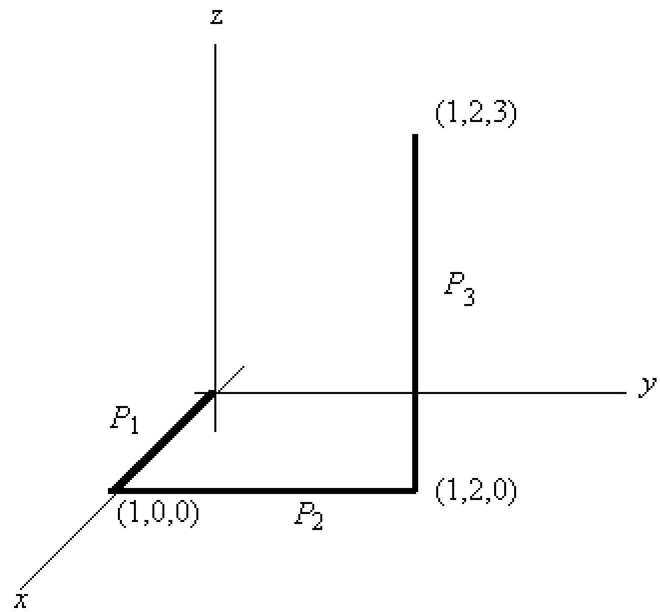
$$= \int_0^1 (2t^2 + 8t + 36t^2)dt = \int_0^1 (38t^2 + 8t)dt$$

$$= \frac{38}{3}t^3 + 4t^2 \Big|_0^1 = \frac{50}{3}.$$

Nothing to it.

Another Example

Now let's integrate the same function from (0,0,0) t0 (1,2,3), but this time along the path P in the picture:



Here the path P is the union of the three nice curves, P_1, P_2, and P_3, so our integral is the sum of three integrals:

$$\int_P \boldsymbol{F}(x,y,x)\ d\boldsymbol{r} = \int_{P_1} \boldsymbol{F}(x,y,x)\ d\boldsymbol{r} + \int_{P_2} \boldsymbol{F}(x,y,x)\ d\boldsymbol{r} + \int_{P_3} \boldsymbol{F}(x,y,x)\ d\boldsymbol{r},$$

where

$$\boldsymbol{F}(x,y,z) = (xy + z^2)\boldsymbol{i} + (x+z)\boldsymbol{j} + 2yz\boldsymbol{k}.$$

A vector description of P_1 is simply $\boldsymbol{r}(t) = t\boldsymbol{i}$, $0 \le t \le 1$. Thus

$$\int_{P_1} \boldsymbol{F}(x,y,z)\ d\boldsymbol{r} = \int_0^1 \boldsymbol{F}(t,0,0) \cdot \boldsymbol{i}\, dt = \int_0^1 t\boldsymbol{j} \cdot \boldsymbol{i}\, dt = 0.$$

For P_2, we have $\boldsymbol{r}(t) = \boldsymbol{i} + t\boldsymbol{j}$, $0 \le \mathrm{t} \le 2$. This gives us

$$\int_{P_2} \boldsymbol{F}(x,y,z)\ d\boldsymbol{r} = \int_0^2 \boldsymbol{F}(1,t,0) \cdot \boldsymbol{j}\, dt = \int_0^2 (t\boldsymbol{i} + \boldsymbol{j}) \cdot \boldsymbol{j}\, dt = \int_0^2 dt = 2.$$

Finally, for P_3, there is $\boldsymbol{r}(t) = \boldsymbol{i} + 2\boldsymbol{j} + t\boldsymbol{k},\ 0 \le t \le 3$; and so

$$\int_{P_3} \boldsymbol{F}(x,y,z)\ d\boldsymbol{r} = \int_0^3 \boldsymbol{F}(1,2,t)\ \boldsymbol{k}\ dt = \int_0^3 [(2+t^2)\boldsymbol{i} + (1+t)\boldsymbol{j} + 4t\boldsymbol{k}]\ \boldsymbol{k}\ dt$$

$$= \int_0^3 4t\ dt = 18\,.$$

At last, we have then $\int_P \boldsymbol{F}(x,y,z)\ d\boldsymbol{r} = 0 + 2 + 18 = 20$.

Exercises

7. Evaluate $\int_C [xy\boldsymbol{i} + x^2\boldsymbol{j}]\ d\boldsymbol{r}$, where C is the arc of the curve $y = x^2$ from (0,0) to (1,1).

8. Evaluate $\int_C (\cos x\ \boldsymbol{i} - y\boldsymbol{j})\ d\boldsymbol{r}$ where C the part of the curve $y = \sin x$ from (0,0) to (π,0).

9. Evaluate the line integral of $\boldsymbol{F}(x,y,z) = xyz\boldsymbol{i} + (xy + yz)\boldsymbol{j} + z^2\boldsymbol{k}$ from (0,0,0) to (-1,1,2) along the line segment joining these two points.

10. Evaluate the line integral of $\boldsymbol{F}(x,y,z) = (x-z)\boldsymbol{i} + (y-z)\boldsymbol{j} - (x+y)\boldsymbol{k}$ along the polygonal path from (0,0,0) to (1,0,0) to (1,1,0) to (1,1,1).

11. Integrate $\boldsymbol{F}(x,y) = \dfrac{1}{x^2 + y^2}(-y\boldsymbol{i} + x\boldsymbol{j})$ one time around the circle $x^2 + y^2 = a^2$ in the counterclockwise direction.

14.3 Path Independence

Suppose we evaluate the vector line integral $\int_C \boldsymbol{F}(\boldsymbol{r}) \ d\boldsymbol{r}$, where C is a curve from the point $\boldsymbol{p}$ to the point $\boldsymbol{q}$. Let $\boldsymbol{r}(t)$, $a \le t \le b$, be a vector description of C. Then, of course, we have $\boldsymbol{r}(a) = \boldsymbol{p}$ and $\boldsymbol{r}(b) = \boldsymbol{q}$. As we have already seen,

$$\int_C \boldsymbol{F}(\boldsymbol{r}) \ d\boldsymbol{r} = \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \ \frac{d\boldsymbol{r}}{dt} \, dt \, .$$

Now let us make the very special assumption that there exists a real-valued (or scalar) function $g\colon \boldsymbol{R^3} \to \boldsymbol{R}$ such that the derivative, or gradient, of g is the integrand $\boldsymbol{F}$:

$$g = \boldsymbol{F} \, .$$

Next let's use the Chain Rule to compute the derivative of the composition $h(t) = g(\boldsymbol{r}(t))$:

$$h'(t) = \quad g \ \frac{d\boldsymbol{r}}{dt} = \boldsymbol{F}(\boldsymbol{r}(t)) \ \frac{d\boldsymbol{r}}{dt} \, .$$

This is, *mirabile dictu*, precisely the integrand in our line integral:

$$\int_C \boldsymbol{F}(\boldsymbol{r}) \ d\boldsymbol{r} = \int_a^b \boldsymbol{F}(\boldsymbol{r}(t)) \ \frac{d\boldsymbol{r}}{dt} \, dt = \int_a^b h'(t) dt = h(b) - h(a) = g(\boldsymbol{p}) - g(\boldsymbol{q}) \, .$$

This is a very exciting result and calls for some meditation. Note that the curve C has completely disappeared from the answer. The value of the integral depends only on the values of the function g at the endpoints; the path from $\boldsymbol{p}$ to $\boldsymbol{q}$ does not affect the answer. The line integral is ***path independent***. The result is esthetically pleasing and is clearly the lineal descendant of the fundamental theorem of calculus we learned so many years ago.

A moment's reflection on the examples we have seen should convince us that a lot of integrals are *not* path independent, thus many very nice functions $\boldsymbol{F}$ (or ***vector fields***) are not the gradient of any function. A function $\boldsymbol{F}$ that is the gradient of a function g is said to be ***conservative*** and the function g is said to be a ***potential*** function for $\boldsymbol{F}$.

Let's suppose the domain $\boldsymbol{D}$ of the function $\boldsymbol{F}\colon \boldsymbol{D} \to \boldsymbol{R}^3$ is open and connected (Thus any two points in $\boldsymbol{D}$ may be joined by a nice path.) We have just seen that if there exists a function $g\colon \boldsymbol{D} \to \boldsymbol{R}$ such that $\boldsymbol{F} = \quad g$, then the integral of $\boldsymbol{F}$ between any two points of $\boldsymbol{D}$ does not depend on the path between the two points. It turns out, as we shall see, that the converse of this is true. Specifically, if every integral of $\boldsymbol{F}$ in $\boldsymbol{D}$ is path independent, then there is a function g such that $\boldsymbol{F} = \quad g$. Let's see why this is so.

Choose a point $\boldsymbol{p} = (x_0, y_0, z_0) \quad \boldsymbol{D}$. Now define $g(\boldsymbol{s}) = g(x, y, z)$ to be the integral from $\boldsymbol{p}$ to $\boldsymbol{s}$ along any curve joining these points. We are assuming path independence of the integral, so it matters not what curve we choose. Okay, now we compute the partial derivative $\dfrac{\partial g}{\partial x}$. The domain $\boldsymbol{D}$ is open and hence includes an open ball centered at $\boldsymbol{s} = (x, y, z) \quad \boldsymbol{D}$. Choose a point $\boldsymbol{q} = (x_1, y, z)$ in such an open ball, and let $\boldsymbol{L}$ be the straight line segment from $\boldsymbol{s}$ to $\boldsymbol{q}$. Then, of course, $\boldsymbol{L}$ lies in $\boldsymbol{D}$. Now let's integrate $\boldsymbol{F}$ from $\boldsymbol{p}$ to $\boldsymbol{s}$ by going along any curve $\boldsymbol{C}$ from $\boldsymbol{p}$ to $\boldsymbol{q}$ and then along $\boldsymbol{L}$ from $\boldsymbol{q}$ to $\boldsymbol{s}$:

$$g(\boldsymbol{s}) = g(x, y, z) = \int_C \boldsymbol{F}(\boldsymbol{r}) \quad d\boldsymbol{r} + \int_L \boldsymbol{F}(\boldsymbol{r}) \quad d\boldsymbol{r}.$$

The first integral on the right does not depend on x, and so $\dfrac{\partial}{\partial x}\displaystyle\int_C \boldsymbol{F}(\boldsymbol{r}) \quad d\boldsymbol{r} = 0$. Thus

$$\frac{\partial g}{\partial x} = \frac{\partial}{\partial x}\int_L \boldsymbol{F}(\boldsymbol{r}) \quad d\boldsymbol{r}.$$

We clearly need to find $\int_L \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r}$. This is easy. Suppose

$$\boldsymbol{F}(\boldsymbol{r}) = f_1(\boldsymbol{r})\boldsymbol{i} + f_2(\boldsymbol{r})\boldsymbol{j} + f_3(\boldsymbol{r})\boldsymbol{k}.$$

A vector description of $\boldsymbol{L}$ is simply $\boldsymbol{r}(t) = t\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$, $x_1 \le t \le x$. Thus $\frac{d\boldsymbol{r}}{dt} = \boldsymbol{i}$, and our line integral becomes simply $\int_L \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \int_{x_1}^{x} f_1(t,y,z)dt$. We are almost done, for note that now

$$\frac{\partial}{\partial x}\int_L \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} = \frac{\partial}{\partial x}\int_{x_1}^{x} f_1(t,y,z)dt = f_1(x,y,z).$$

Hence

$$\frac{\partial g}{\partial x} = f_1.$$

It should be clear to one and all how to show that $\frac{\partial g}{\partial y} = f_2$ and $\frac{\partial g}{\partial z} = f_3$, thus giving us the desired result: $\boldsymbol{F} = \quad g$.

Exercises

12. Prove that $\frac{\partial g}{\partial y} = f_2$, where g and f_2 are as in the preceding discussion.

13. Prove that if $\boldsymbol{F}: \boldsymbol{D} \to \boldsymbol{R^3}$, where $\boldsymbol{D}$ is open and connected, and every $\int_C \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r}$ is path independent, then $\oint_P \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{r} = 0$ for every closed path in $\boldsymbol{D}$.(A ***closed*** path, or curve, is one with no endpoints.) [Physicists and others like to use a snake sign with a little circle superimposed on it $\oint$ to indicate that the path of integration is closed.]

14. Prove that if $\boldsymbol{F}:\boldsymbol{D} \to \boldsymbol{R}^3$, where $\boldsymbol{D}$ is open and connected, and $\oint_{P} \boldsymbol{F}(\boldsymbol{r})\ d\boldsymbol{r} = 0$ for every closed path in $\boldsymbol{D}$, then every $\int_{C} \boldsymbol{F}(\boldsymbol{r})\ d\boldsymbol{r}$ is path independent .

15. a)Find a potential function g for the function $\boldsymbol{F}(\boldsymbol{r}) = yz\boldsymbol{i} + xz\boldsymbol{j} + xy\boldsymbol{k}$.

b)Evaluate the line integral $\int_{C} \boldsymbol{F}(\boldsymbol{r})\ d\boldsymbol{r}$, where $\boldsymbol{C}$ is the curve

$$\boldsymbol{r}(t) = (e^t \sin t)\boldsymbol{i} + t^2 e^{3t}\boldsymbol{j} + \cos^3 t\boldsymbol{k},\ 0 \le t \le 1 .$$

16. a)Find a potential function g for the function $\boldsymbol{F}(\boldsymbol{r}) = e^{y+2z}(\boldsymbol{i} + x\boldsymbol{j} + 2x\boldsymbol{k})$.

b)Find another potential function for $\boldsymbol{F}$ in part a).

b)Evaluate the line integral $\int_{C} \boldsymbol{F}(\boldsymbol{r})\ d\boldsymbol{r}$, where $\boldsymbol{C}$ is the curve

$$\boldsymbol{r}(t) = t\cos 2t^2\boldsymbol{i} + 4t\boldsymbol{j} + e^{2t}\boldsymbol{k},\ 0 \le t \le \sqrt{\pi} .$$

17. Evaluate $\oint_{E} [(e^x \sin y + 3y)\boldsymbol{i} + (e^x \cos y + 2x - 2y)\boldsymbol{j}]\ d\boldsymbol{r}$ where $\boldsymbol{E}$ is the ellipse $4x^2 + y^2 = 4$ oriented clockwise.

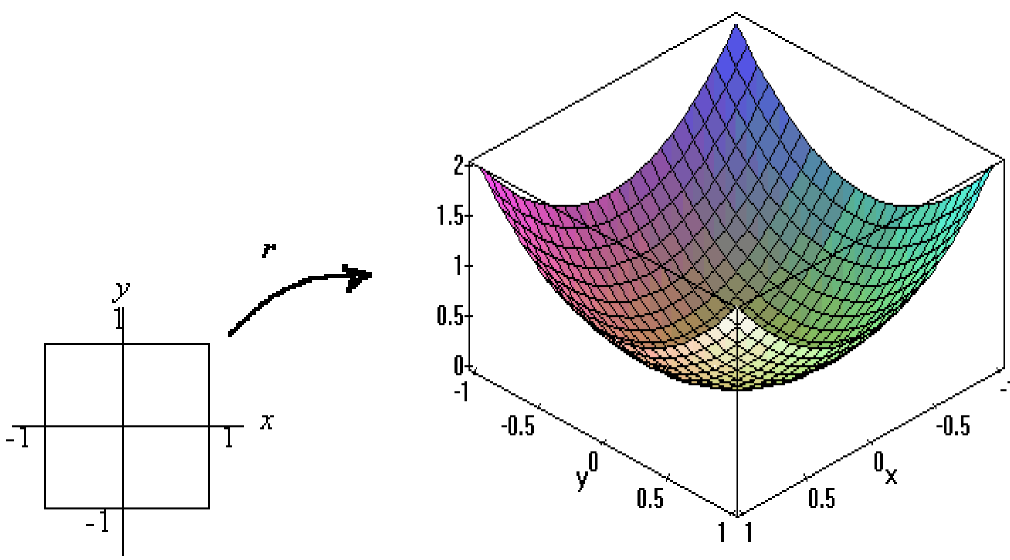
[Really good hint: Find the gradient of $g(x,y,z) = e^x \sin y + 2xy - y^2$.]

Chapter Fifiteen

Surfaces Revisited

15.1 Vector Description of Surfaces

We look now at the very special case of functions $\boldsymbol{r}: \boldsymbol{D} \to \boldsymbol{R}^3$, where $\boldsymbol{D} \ \ \boldsymbol{R}^2$ is a nice subset of the plane. We suppose $\boldsymbol{r}$ is a nice function. As the point $(s,t) \ \ \boldsymbol{D}$ moves around in $\boldsymbol{D}$, if we place the tail of the vector $\boldsymbol{r}(s,t)$ at the origin, the nose of this vector will trace out a surface in three-space. Look, for example at the function $\boldsymbol{r}: \boldsymbol{D} \to \boldsymbol{R}^3$, where $\boldsymbol{r}(s,t) = s\boldsymbol{i} + t\boldsymbol{j} + (s^2 + t^2)\boldsymbol{k}$, and $\boldsymbol{D} = \{(s,t) \ \ \boldsymbol{R}^2 : -1 \le s,t \le 1\}$. It shouldn't be difficult to convince yourself that if the tail of $\boldsymbol{r}(s,t)$ is at the origin, then the nose will be on the paraboloid $z = x^2 + y^2$, and for all $(s,t) \ \ \boldsymbol{D}$, we get the part of the paraboloid above the square $-1 \le x, y \le 1$. It is sometimes helpful to think of the function $\boldsymbol{r}$ as providing a map from the region $\boldsymbol{D}$ to the surface.



The vector function $\boldsymbol{r}$ is called a ***vector description*** of the surface. This is, of course, exactly the two dimensional analogue of the vector description of a curve.

For a curve, $\boldsymbol{r}$ is a function from a nice piece of the real line into three space; and for a surface, $\boldsymbol{r}$ is a function from a nice piece of the plane into three space.

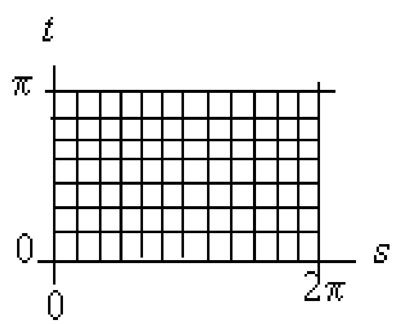
Let's look at another example. Here, let

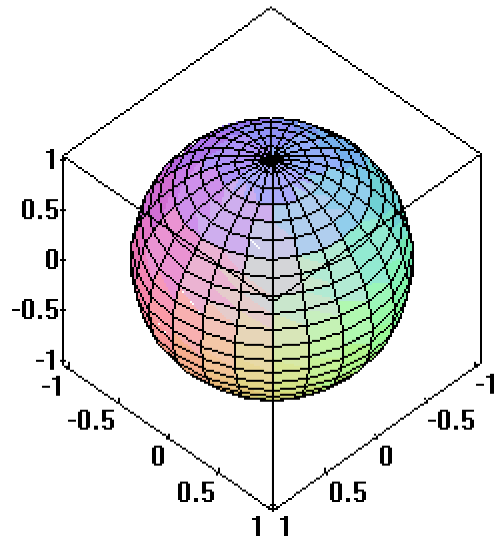
$$\boldsymbol{r}(s,t) = \cos s\ \sin t\ \boldsymbol{i} + \sin s\ \sin t\ \boldsymbol{j} + \cos t\ \boldsymbol{k}\ ,$$

for $0 \le t \le \pi$ and $0 \le s \le 2\pi$. What have we here? First, notice that

$$\begin{aligned}|\boldsymbol{r}(s,t)|^2 &= (\cos s\ \sin t)^2 + (\sin s\ \sin t)^2 + (\cos t)^2 \\ &= \sin^2 t\ (\cos^2 s + \sin^2 s) + \cos^2 t \\ &= \sin^2 t + \cos^2 t = 1\end{aligned}$$

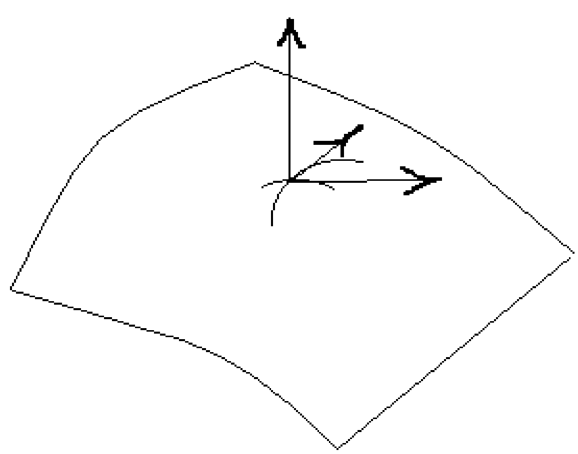
Thus the nose of $\boldsymbol{r}$ is always on the sphere of radius one and centered at the origin. Notice next, that the variable, or parameter, s is the longitude of $\boldsymbol{r}(s,t)$; and the variable t is the latitude of $\boldsymbol{r}(s,t)$. (More precisely, t is co-latitude.) A moment's reflection on this will convince you that as $\boldsymbol{r}$ is a description of the entire sphere. We have a map of the sphere on the rectangle





Observe that the entire lower edge of the rectangle (the line from $(0,0)$ to $(2\pi,0)$) is mapped by $\boldsymbol{r}$ onto the North Pole, while the upper edge is mapped onto the South Pole.

Let $\boldsymbol{r}(s,t),\ (s,t) \in \boldsymbol{D}$ be a vector description of a surface S, and let $\boldsymbol{p} = \boldsymbol{r}(\bar{s},\bar{t})$ be a point on S. Now, $\boldsymbol{c}(s) = \boldsymbol{r}(s,\bar{t})$ is a curve on the surface that passes through he point $\boldsymbol{p}$. Thus the vector $\frac{d\boldsymbol{c}}{ds} = \frac{\partial \boldsymbol{r}}{\partial s}(\bar{s},\bar{t})$ is tangent to this curve at the point $\boldsymbol{p}$. We see in the same way that the vector $\frac{\partial \boldsymbol{r}}{\partial t}(\bar{s},\bar{t})$ is tangent to the curve $\boldsymbol{r}(\bar{s},t)$ at $\boldsymbol{p}$.



At the point $\boldsymbol{p} = \boldsymbol{r}(\bar{s}, \bar{t})$ on the surface S, the vectors $\dfrac{\partial \boldsymbol{r}}{\partial s}$ and $\dfrac{\partial \boldsymbol{r}}{\partial t}$ are thus tangent to S.

Hence the vector $\dfrac{\partial \boldsymbol{r}}{\partial s} \times \dfrac{\partial \boldsymbol{r}}{\partial t}$ is normal to S.

Example

Let's find a vector normal to the surface given by the vector description $\boldsymbol{r}(s,t) = s\boldsymbol{i} + t\boldsymbol{j} + (s^2 + t^2)\boldsymbol{k}$ at a point. We need to find the partial derivatives $\dfrac{\partial \boldsymbol{r}}{\partial s}$ and $\dfrac{\partial \boldsymbol{r}}{\partial s}$:

$$\frac{\partial \boldsymbol{r}}{\partial s} = \boldsymbol{i} + 2s\boldsymbol{k}\text{, and } \frac{\partial \boldsymbol{r}}{\partial t} = \boldsymbol{j} + 2t\boldsymbol{k}\,.$$

The normal $\boldsymbol{N}$ is

$$\boldsymbol{N} = \frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 1 & 0 & 2s \\ 0 & 1 & 2t \end{vmatrix} = -2s\boldsymbol{i} - 2t\boldsymbol{j} + \boldsymbol{k}\,.$$

Meditate on the geometry here and convince yourself that this result is at least reasonable.

Exercises

1. Give a vector description for the surface $z = \sqrt{x + 2y^2}$, $x, y \geq 0$.

2. Give a vector description for the ellipsoid $4x^2 + y^2 + 8z^2 = 16$.

3. Give a vector description for the cylinder $x^2 + y^2 = 1$.

4. Describe the surface given by $\boldsymbol{r}(s,t) = s\cos t\boldsymbol{i} + s\sin t\boldsymbol{j} + s\boldsymbol{k}$, $0 \le t \le 2\pi$, $-1 \le s \le 1$.

5. Describe the surface given by $\boldsymbol{r}(s,t) = s\cos t\boldsymbol{i} + s\sin t\boldsymbol{j} + s^2\boldsymbol{k}$, $0 \le t \le 2\pi$, $1 \le s \le 2$.

6. Give a vector description for the sphere having radius 3 and centered at the point (1,2,3).

7. Find an equation (*I.e.*, a vector description) of the line normal to the sphere $x^2 + y^2 + z^2 = a^2$ at the point $(\frac{a}{\sqrt{3}}, \frac{a}{\sqrt{3}}, -\frac{a}{\sqrt{3}})$.

8. Find a scalar equation (*I.e.*, of the form $f(x,y,z) = 0$) of the plane tangent to the sphere $x^2 + y^2 + z^2 = a^2$ at the point $(\frac{a}{\sqrt{3}}, \frac{a}{\sqrt{3}}, -\frac{a}{\sqrt{3}})$.

9. Find all points on the surface $\boldsymbol{r}(s,t) = (s^2 + t^2)\boldsymbol{i} + (s + 3t)\boldsymbol{j} - st\boldsymbol{k}$ at which the tangent plane is parallel to the plane $5x - 6y + 2z = 7$, or show there are no such points.

10. Find an equation of the plane that contains the point (1,-2,3) and is parallel to the plane tangent to the surface $\boldsymbol{r}(s,t) = (s+t)\boldsymbol{i} + s^2\boldsymbol{j} - 2t^2\boldsymbol{k}$ at the point (1, 4,-18).

15.2 Integration

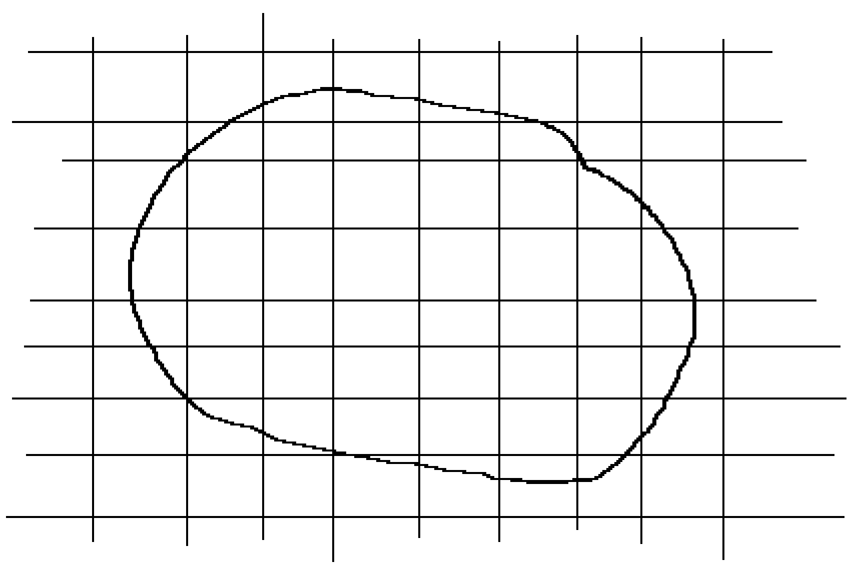
Suppose we have a nice surface S and a function $f : S \to \boldsymbol{R}$ defined on the surface. We want to define an integral of f on S as the limit of some sort of Riemann sum in the way in which we have already defined various integrals. Here we have a slight problem in that we really are not sure at this point exactly what we might mean by the *area* of a small piece of surface. We assume the surface is sufficiently smooth to allow us to approximate the area of a small piece of it by a small planar region, and then add up these approximations to get a Riemann sum, *etc.*, *etc.* Let's be specific.

We subdivide S into a number of small pieces $S_1, S_2, \ldots, S_n$ each having area ΔA_i, select points $\boldsymbol{r}_i^* = (x_i^*, y_i^*, z_i^*) \quad S_i$, and form the Riemann sum

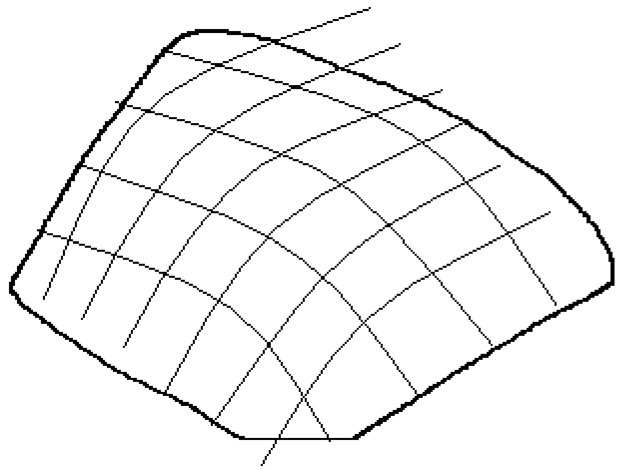
$$R = \sum_{i=1}^{n} f(\boldsymbol{r}_i^*)\Delta A_i \ .$$

Then, of course, we take finer and finer subdivisions, and if the corresponding Riemann sums have a limit, this limit is the thing we call the integral of f on S: $\iint_S f(\boldsymbol{r})\,dS$.

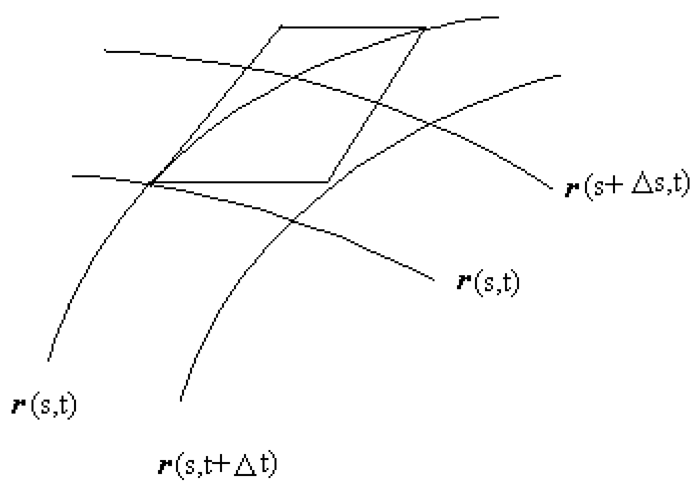
Now, how do find such a thing. We need a vector description of S, say $\boldsymbol{r}: \boldsymbol{D} \to \boldsymbol{r}(\boldsymbol{D}) = S$. The surface S is subdivided by subdividing the region $\boldsymbol{D} \quad \boldsymbol{R}^2$ into rectangles in the usual way:



The images of the vertical lines, s = constant, form a family of "parallel" curves on the surface, and the images of the horizontal lines t = constant, also form a family of such curves:



Let's look closely at one of the subdivisions:



We paste a parallelogram tangent to the surface at the point $\boldsymbol{r}(s_i, t_i)$ as shown. The lengths of the sides of this parallelogram are $\left|\frac{\partial \boldsymbol{r}}{\partial s}(s_i, t_i)\Delta s_i\right|$ and $\left|\frac{\partial \boldsymbol{r}}{\partial t}(s_i, t_i)\Delta t_i\right|$. The area is then $\left|\left(\frac{\partial \boldsymbol{r}}{\partial s}(s_i, t_i)\Delta s_i\right) \times \left(\frac{\partial \boldsymbol{r}}{\partial t}(s_i, t_i)\Delta t_i\right)\right|$, and we use the approximation

$$\Delta A_i \approx \left|\left(\frac{\partial \boldsymbol{r}}{\partial s}(s_i, t_i)\right) \times \left(\frac{\partial \boldsymbol{r}}{\partial t}(s_i, t_i)\right)\right| \Delta s_i \Delta t_i$$

in the Riemann sums:

$$R = \sum_{i=1}^{n} f(\boldsymbol{r}(s_i, t_i)) \left|\left(\frac{\partial \boldsymbol{r}}{\partial s}(s_i, t_i)\right) \times \left(\frac{\partial \boldsymbol{r}}{\partial t}(s_i, t_i)\right)\right| \Delta s_i \Delta t_i .$$

These are just the Riemann sums for the usual old time double integral of the function

$$F(s,t) = \sum_{i=1}^{n} f(\boldsymbol{r}(s_i, t_i)) \left|\left(\frac{\partial \boldsymbol{r}}{\partial s}(s_i, t_i)\right) \times \left(\frac{\partial \boldsymbol{r}}{\partial t}(s_i, t_i)\right)\right|$$

over the plane region $\boldsymbol{D}$. Thus,

$$\iint_S f(\boldsymbol{r})\,dS = \iint_{\boldsymbol{D}} f(\boldsymbol{r}(s,t)) \left|\frac{\partial \boldsymbol{r}}{\partial s}(s,t) \times \frac{\partial \boldsymbol{r}}{\partial t}(s,t)\right| dA .$$

Example

Let's use our new-found knowledge to find the area of a sphere of radius a. Observe that the area of a surface S is simply the integral $\iint_S dS$. In the previous section, we found a vector description of the sphere:

$$\boldsymbol{r}(s,t) = a\cos s\ \sin t\,\boldsymbol{i} + a\sin s\ \sin t\,\boldsymbol{j} + a\cos t\,\boldsymbol{k}\ ,$$

$0 \le t \le \pi$ and $0 \le s \le 2\pi$. Compute the partial derivatives:

$$\frac{\partial\,\boldsymbol{r}}{\partial\,s} = -a\sin s\ \sin t\,\boldsymbol{i} + a\cos s\ \sin t\,\boldsymbol{j},\ \text{and}$$

$$\frac{\partial\,\boldsymbol{r}}{\partial\,t} = a\cos s\ \cos t\,\boldsymbol{i} + a\sin s\ \cos t\,\boldsymbol{j} - a\sin t\,\boldsymbol{k}$$

Then

$$\frac{\partial\,\boldsymbol{r}}{\partial\,s} \times \frac{\partial\,\boldsymbol{r}}{\partial\,t} = a^2 \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ -\sin s\ \sin t & \cos s\ \sin t & 0 \\ \cos s\ \cos t & \sin s \cos t & \text{-sint} \end{vmatrix}$$

$$= a^2[-\cos s\ \sin^2 t\,\boldsymbol{i} - \sin s\ \sin^2 t\,\boldsymbol{j} - \sin t\cos t\,\boldsymbol{k}]$$

Next we need to find the length of this vector:

$$\left|\frac{\partial\,\boldsymbol{r}}{\partial\,s} \times \frac{\partial\,\boldsymbol{r}}{\partial\,t}\right| = a^2[\cos^2 s\ \sin^4 t + \sin^2 s\ \sin^4 t + \sin^2 t\cos^2 t]^{1/2}$$
$$= a^2[\sin^4 t + \sin^2 t\cos^2 t]^{1/2} = a^2[\sin^2 t(\sin^2 t + \cos^2 t)]^{1/2}$$
$$= a^2|\sin t|$$

Hence,

$$\text{Area} = \iint_S dS = \iint_D \left|\frac{\partial\,\boldsymbol{r}}{\partial\,s} \times \frac{\partial\,\boldsymbol{r}}{\partial\,t}\right| dA = \iint_D a^2|\sin t|dA$$

$$= a^2 \int_0^{\pi} \int_0^{2\pi} |\sin t| \, ds dt$$

$$= 2\pi a^2 \int_0^{\pi} \sin t dt = 4\pi a^2$$

Another Example

Let's find the centroid of a hemispherical shell H of radius a. Choose our coordinate system so that the shell is the surface $x^2 + y^2 + z^2 = a^2$, $z \geq 0$. The centroid $(\bar{x}, \bar{y}, \bar{z})$ is given by

$$\bar{x} = \frac{\iint_H x dS}{\iint_H dS} \qquad \bar{y} = \frac{\iint_H y dS}{\iint_H dS} \quad \text{and} \quad \bar{z} = \frac{\iint_H z ds}{\iint_H dS}.$$

First, note from the symmetry of the shell that $\bar{x} = \bar{y} = 0$. Second, it should be clear from the precious example that $\iint_H dS = 2\pi a^2$. This leaves us with just integral to evaluate:

$\iint_H z dS$. Most of the work was done in the example before this one. This hemisphere has the same vector description as the sphere, except for the fact that the domain of $\boldsymbol{r}$ is the rectangle $0 \leq s \leq 2\pi$, $0 \leq t \leq \frac{\pi}{2}$. Thus

$$\iint_H z\,dS = a^2 \int_0^{\pi/2}\int_0^{2\pi} a\cos t \left| \frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t} \right| ds\,dt$$

$$= a^3 \int_0^{\pi/2}\int_0^{2\pi} \cos t \ \sin t \, ds\,dt = 2\pi a^3 \int_0^{\pi/2} \cos t \ \sin t \, dt$$

$$= \pi a^3 \ \sin^2 t \Big|_0^{\pi/2} = \pi a^3$$

And so we have $\bar{z} = \dfrac{\pi a^3}{2\pi a^2} = \dfrac{a}{2}$. Is this the result you expected?

Yet One More Example

Our new definition of a surface integral certainly includes the old one for plane surfaces. Look at the "surface" described by the vector function

$$\boldsymbol{r}(\theta, r) = r\cos\theta\,\boldsymbol{i} + r\sin\theta\,\boldsymbol{j}\,,$$

with $\boldsymbol{r}$ defined on some subset D of the $\theta - r$ plane. For what we hope will be obvious reasons, we are using the letters θ and r instead of s and t. Now consider an integral

$$\iint_S f(x, y)\,dS$$

over the surface S described by $\boldsymbol{r}$. We know this integral to be given by

$$\iint_S f(x, y)\,dS = \iint_D f(r\cos\theta\,, r\sin\theta\,) \left| \frac{\partial \boldsymbol{r}}{\partial \theta} \times \frac{\partial \boldsymbol{r}}{\partial r} \right| dA\,.$$

Let's find the partial derivatives:

$$\frac{\partial \boldsymbol{r}}{\partial \theta} = -r\sin\theta\, \boldsymbol{i} + r\cos\theta\, \boldsymbol{j}\text{, and}$$

$$\frac{\partial \boldsymbol{r}}{\partial r} = \cos\theta\, \boldsymbol{i} + \sin\theta\, \boldsymbol{j}\,.$$

Thus,

$$\frac{\partial \boldsymbol{r}}{\partial \theta} \times \frac{\partial \boldsymbol{r}}{\partial r} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ -r\sin\theta & r\cos\theta & 0 \\ \cos\theta & \sin\theta & 0 \end{vmatrix} = -r\boldsymbol{k}\,,$$

and we have $\left|\frac{\partial \boldsymbol{r}}{\partial \theta} \times \frac{\partial \boldsymbol{r}}{\partial r}\right| = r$. Hence,

$$\iint_S f(x,y)\,dS = \iint_D f(r\cos\theta, r\sin\theta)\left|\frac{\partial \boldsymbol{r}}{\partial \theta} \times \frac{\partial \boldsymbol{r}}{\partial r}\right| dA = \iint_D f(r\cos\theta, r\sin\theta)\, r\,dA\,.$$

This should look familiar!

Exercises

11. Find the area of that part of the surface $z = x^2 + y^2$ that lies between the planes $z = 1$ and $z = 2$.

12. Find the centroid of the surface given in Problem **11**.

13. Find the area of that part of the Earth that lies North of latitude 45°. (Assume the surface of the Earth is a sphere.)

14. A spherical shell of radius a is centered at the origin. Find the centroid of that part of it which is in the first octant.

15. a)Find the centroid of the solid right circular cone having base radius a and altitude h.

b)Find the centroid of the lateral surface of the cone in part a).

16. Find the area of the ellipse cut from the plane $z = 2x$ by the cylinder $x^2 + y^2 = 1$.

17. Evaluate $\iint_S (x + y + z)dS$, where S is the surface of the cube cut from the first octant by the planes $x = a$, $y = a$, and $z = a$.

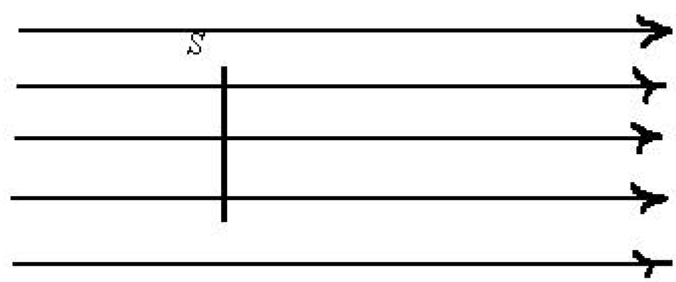
18. Evaluate $\iint_S x\sqrt{y^2 + 1}\, dS$, where S is the surface cut from the paraboloid $y^2 + 4z = 16$ by the planes $x = 0$, $x = 1$, and $z = 0$.

Chapter Sixteen

Integrating Vector Functions

16.1 Introduction

Suppose water (or some other incompressible fluid) flows at a constant velocity $\boldsymbol{v}$ in space (through a pipe, for instance), and we wish to know the rate at which the water flows across a rectangular surface S that is normal to the stream lines:

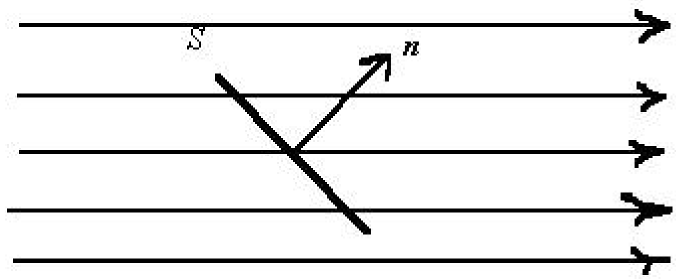


What is the rate at which the fluid flows through S? Let $M(t)$ denote the total volume of fluid that has passed through the surface at time t. The amount of fluid that flows through during the time between t and $t + \Delta t$ is simply

$$M(t + \Delta t) - M(t) = |\boldsymbol{v}| a \Delta t \ ,$$

where a is the area of S. Thus, the rate of flow through S is $\frac{dM}{dt} = |\boldsymbol{v}| a$.

The result is slightly more complicated when various exciting changes are made. Clearly there is nothing special about the surface's being a rectangle. But suppose that S is placed at an angle to the stream lines instead of being placed normal to the them. Then we have $\frac{dM}{dt} = \boldsymbol{v} \ \ \boldsymbol{n} \ a$, where $\boldsymbol{n}$ is a unit normal to the surface S.



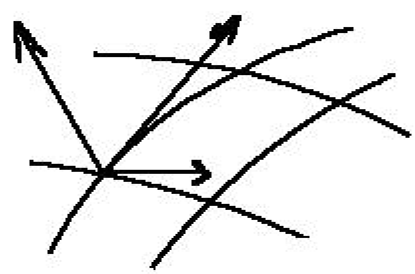
Observe that matters which unit normal to the plane surface we choose. If we choose the other normal ($-\boldsymbol{n}$), then our rate will be the negative of this one. We must thus specify an ***orientation*** of the surface. We are computing the rate of flow from one side of the surface to the other, and so we have to specify the "sides", so to speak.

16.2 Flux

Now, let's look at the general situation. The surface is not restricted to being a plane surface, and the velocity of the flow is not restricted to being constant in space; it may vary with position as well as time. Specifically, suppose S is a surface, together with an orientation—that is, some means of specifying two "sides"—and suppose $\boldsymbol{F}(\boldsymbol{r})$ is a function $\boldsymbol{F}: \boldsymbol{R^3} \to \boldsymbol{R^3}$, which is the velocity of the incompressible fluid. How do we find the rate of flow through the surface S from one side to the other?

First, let's come to grips with the problem of specifying an ***orientation*** for S. We say that an ***orientation*** for S is a continuous function $\boldsymbol{n}: S \to \boldsymbol{R^3}$ such that $\boldsymbol{n}(\boldsymbol{r})$ is normal to S and $|\boldsymbol{n}(\boldsymbol{r})| = 1$ for all $\boldsymbol{r}$ S. A surface together with an orientation is called an ***oriented surface***. At first blush this looks simple enough, and the unsophisticated might guess that every surface has an orientation (or may be ***oriented***, as we sometimes say). But this is not so! There are many surfaces for which an orientation does not exist. You may recall from grammar school a simple example of such a surface, the so-called Möbius band, or strip. Here is my feeble attempt to draw one:

Now we see about finding the rate of flow through the oriented surface S. The strategy should be old-hat by now. We subdivide S and look at "small" parallelograms tangent to the surface:



As we have done so often, we suppose the subdivisions are small and approximate the rate of flow, or ***flux***, through the subdivision by the rate of flow through the tangent parallelogram.

$$\Delta S_i = \boldsymbol{F}(\boldsymbol{r}_i^*) \quad \boldsymbol{n}\Delta A_i \,,$$

and then add them to obtain yet another type of Riemann sum $R = \sum_{i=1}^{n} \boldsymbol{F}(\boldsymbol{r}_i^*) \quad \boldsymbol{n}\Delta A_i$. If these sums have a limiting value as the size of the subdivisions go to zero, this is what we call the integral of $\boldsymbol{F}$ over the oriented surface S:

$$\iint_S \boldsymbol{F}(\boldsymbol{r}) \quad d\boldsymbol{S} \,.$$

It should be clear now what we do to evaluate such an integral. As usual, we consider a vector description of the surface S: $\boldsymbol{r}\colon \boldsymbol{D} \to S \quad \boldsymbol{R}^3$, where $\boldsymbol{D} \quad \boldsymbol{R}^2$. We subdivide S by subdividing the region $\boldsymbol{D}$ into rectangles formed by lines s = constant and t= constant, and looking at the curves $\boldsymbol{r}(s,\bar{t})$ and $\boldsymbol{r}(\bar{s},t)$ on the surface, exactly as we did in integrating a scalar function over a surface S. Most conveniently now, the vector product $\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t}$ gives us not only a vector such that $\left|\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t}\right| \Delta s \Delta t$ is the area of the approximating parallelogram, but also one which is normal to the surface. There is the slight problem of the orientation of S. Thus $\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t}$ may not point in the direction of the specified orientation, in which case, of course, we simply replace $\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t}$ by its negative, $\frac{\partial \boldsymbol{r}}{\partial t} \times \frac{\partial \boldsymbol{r}}{\partial s}$. (We may think of just reversing the roles of s and t.) We have in the Riemann sums,

$$R = \sum_{i=1}^{n} \boldsymbol{F}(\boldsymbol{r}_i^*) \cdot \left(\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t}\right) \Delta s_i \Delta t_i ,$$

and, as before, we obtain

$$\iint_S \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \iint_D \boldsymbol{F}(\boldsymbol{r}(s,t)) \cdot \left(\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t}\right) dA .$$

The concept we have developed here is purely mathematical and is done independent of any physical interpretation, such as our fluid flow interpretation. What we have is just an integral of a vector function $\boldsymbol{F}$ (or field) over an oriented surface S. This is generally called the ***flux*** of $\boldsymbol{F}$ over S. There are many physical interpretations of this concept; you have perhaps seen some of them in elementary school physics. There is electric flux, the flux of an electric field; magnetic flux; gravitational flux, *etc., etc.*

Example

Let S be the sphere of radius a oriented so that the normal points "out" of the sphere, and let $\boldsymbol{F}(\boldsymbol{r}) = \frac{c}{|\boldsymbol{r}|^3}\boldsymbol{r}$, where c is a constant. Let's find $\iint_S \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S}$. Use the vector description of S we used in the first Example of the previous section:

$$\boldsymbol{r}(s,t) = a\cos s\sin t\boldsymbol{i} + a\sin s\sin t\boldsymbol{j} + a\cos t\boldsymbol{k},$$

$0 \le s \le 2\pi,\ 0 \le t \le \pi$. We have already found that

$$\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t} = a^2 \sin t[-\cos s\sin t\boldsymbol{i} - \sin s\sin t\boldsymbol{j} - \cos t\boldsymbol{k}].$$

Modest meditation should convince you that this normal points into the sphere, and is thus the negative of the one we need for the specified orientation of S.

Next, the integrand is given by

$$\boldsymbol{F}(\boldsymbol{r}) = \frac{c}{|\boldsymbol{r}|^3}\boldsymbol{r} = \frac{c}{a^3}a[\cos s\sin t\boldsymbol{i} + \sin s\sin t\boldsymbol{j} + \cos t\boldsymbol{k}],$$

and our integral becomes

$$\iint_S \boldsymbol{F}(\boldsymbol{r})\ d\boldsymbol{S} = \int_0^{\pi}\int_0^{2\pi} \frac{c}{a^2}[\cos s\sin t\boldsymbol{i} + \sin s\sin t\boldsymbol{j} + \cos t\boldsymbol{k}]\ a^2 \sin t[\cos s\sin t\boldsymbol{i} + \sin s\sin t\boldsymbol{j} + \cos t\boldsymbol{k}]dsdt$$

$$= c\int_0^{\pi}\int_0^{2\pi} \sin t[\cos^2 s\sin^2 t + \sin^2 s\sin^2 t + \cos^2 t]dsdt$$

$$= c\int_0^{\pi}\int_0^{2\pi} \sin t\ dsdt = 2\pi c\int_0^{\pi} \sin t\ dt = 4\pi c.$$

Note that the radius a of the sphere has disappeared—the value of the integral is independent of the radius of the sphere.

Exercises

1. Find $\iint_S [z\boldsymbol{i} + x^2\boldsymbol{k}]\ d\boldsymbol{S}$, where S is that part of the surface $z = x^2 + y^2$ that lies above the square $\{(x, y): -1 \le x \le 1, \text{ and } -1 \le y \le 1\}$, oriented so that the normal points upward.

2. Find the flux of $\boldsymbol{F}(x, y, z) = x\boldsymbol{i} + z\boldsymbol{j}$ out of the tetrahedron bounded by the coordinate planes and the plane $x + 2y + 3z = 6$.

3. Find the flux of $\boldsymbol{F}(\boldsymbol{r}) = \dfrac{c}{|\boldsymbol{r}|^3}\boldsymbol{r}$ out of the surface of the cube $-a \le x, y, z \le a$, where c and a are positive constants.

4. Find the flux of the function $\boldsymbol{F}(x, y, z) = 4x\boldsymbol{i} + 4y\boldsymbol{j} + 2\boldsymbol{k}$ outward through the surface cut from the bottom of the paraboloid $z = x^2 + y^2$ by the plane $z = 1$.

5. Find the flux of the function $\boldsymbol{F}(x, y, z) = z^2\boldsymbol{i} + x\boldsymbol{j} - 3z\boldsymbol{k}$ upward throught the surface cut from the cylinder $z = 4 - y^2$ by the planes $x = 0,\ x = 1,$ and $z = 0$.

6. Let S be the surface defined by

$$y = \log x\ ,\ 1 \le x \le e\ ,\ 0 \le z \le 1\ ,$$

and let $\boldsymbol{n}$ be the orientation of S such that $\boldsymbol{n}(\boldsymbol{r})\quad \boldsymbol{j} > 0$ for all $\boldsymbol{r}\quad S$. Find the flux

$$\iint_S [2y\boldsymbol{j} + z\boldsymbol{k}]\ d\boldsymbol{S}\ .$$

Chapter Seventeen

Gauss and Green

17.1 Gauss's Theorem

Let $\boldsymbol{B}$ be the box, or rectangular parallelepiped, given by

$$\boldsymbol{B} = \{(x,y,z): x_0 \le x \le x_1,\ y_0 \le y \le y_1,\ z_0 \le z \le z_1\}\ ;$$

and let $\boldsymbol{S}$ be the surface of $\boldsymbol{B}$ with the orientation that points out of $\boldsymbol{B}$. Let $\boldsymbol{F}: \boldsymbol{B} \to \boldsymbol{R}^3$ be a nice function, or field. For reasons that will become apparent as the drama unfolds, let's compute the flux

$$\iint_S \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S}\ .$$

We shall do this by computing the surface integral over each of the six sides of $\boldsymbol{B}$ and adding the results. Let S_1 be the side in the plane $x = x_1$; let S_2 be the side in the plane $x = x_0$; let S_3 be the side in the plane $y = y_1$; let S_4 be the side in the plane $y = y_0$; and let S_5 and S_6 be the obvious things. We begin by computing the integral

$$\iint_{S_1} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S}\ .$$

A vector description of S_1 is quite easy to come by; it is, of course, simply

$$\boldsymbol{r}(s,t) = x_1\boldsymbol{i} + s\boldsymbol{j} + t\boldsymbol{k}\ ,$$

where $y_0 \le s \le y_1$ and $z_0 \le t \le z_1$. (Obviously, s is simply y, and t is z .) Then

$$\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t} = \boldsymbol{j} \times \boldsymbol{k} = \boldsymbol{i}\ .$$

It is clear this is the specified orientation. If $\boldsymbol{F}(\boldsymbol{r}) = p(x,y,z)\boldsymbol{i} + q(x,y,z)\boldsymbol{j} + r(x,y,z)\boldsymbol{k}$, then

$$\iint_{S_1} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \int_{y_0}^{y_1}\int_{z_0}^{z_1} \boldsymbol{F}(x_1,s,t) \cdot \boldsymbol{i}\ dtds$$

$$= \int_{y_0}^{y_1}\int_{z_0}^{z_1} p(x_1,s,t)\ dtds$$

A vector description for the opposite side, $x = x_0$, is just

$$\boldsymbol{r}(s,t) = x_0\boldsymbol{i} + s\boldsymbol{j} + t\boldsymbol{k} \, ,$$

and we have

$$\iint_{S_2} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \int_{y_0}^{y_1}\int_{z_0}^{z_1} \boldsymbol{F}(x_0, s, t) \cdot (-\boldsymbol{i}) \; dtds$$

$$= \int_{y_0}^{y_1}\int_{z_0}^{z_1} -p(x_0, s, t) \; dtds$$

The sum of these two is then

$$\iint_{S_1} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} + \iint_{S_2} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \int_{y_0}^{y_1}\int_{z_0}^{z_1} [\mathrm{p}(\mathrm{x}_1, \mathrm{s}, \mathrm{t}) - \mathrm{p}(\mathrm{x}_0, \mathrm{s}, \mathrm{t})] \; dtds \, .$$

Observe that

$$p(x_1, s, t) - p(x_0, s, t) = \int_{x_0}^{x_1} \frac{\partial p}{\partial x}(x, s, t) dx \, .$$

Substitution of this into the previous equation gives us

$$\iint_{S_{10}} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} + \iint_{S_{21}} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \int_{y_0}^{y_1}\int_{z_0}^{z_1}\int_{x_0}^{x_1} \frac{\partial p}{\partial x}(x, s, t) dx dt ds$$

$$= \iiint_B \frac{\partial p}{\partial x} dV$$

and we have turned the sum of the two surface integrals into a plain ol' volume integral .

It should be clear how we also obtain

$$\iint_{S_3} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} + \iint_{S_4} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \iiint_B \frac{\partial q}{\partial y} dV \, , \text{ and}$$

$$\iint_{S_5} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} + \iint_{S_6} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \iiint_B \frac{\partial r}{\partial z} dV \, .$$

The flux over the entire surface S is thus the sum of these:

$$
(⌘)\qquad
\begin{aligned}
\iint_S \boldsymbol{F}(\boldsymbol{r})\cdot d\boldsymbol{S} &= \iiint_B \frac{\partial p}{\partial x}dV + \iiint_B \frac{\partial q}{\partial y}dV + \iiint_B \frac{\partial r}{\partial z}dV \\
&= \iiint_B \left[\frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} + \frac{\partial r}{\partial z}\right]dV
\end{aligned}
$$

We have now found the surface integral, or flux, in terms of an ordinary volume integral.

Now, suppose we have an "arbitrary" solid region $\boldsymbol{B}$ bounded by a surface $\boldsymbol{S}$, together with a function $\boldsymbol{F}(\boldsymbol{r}) = p(x,y,z)\boldsymbol{i} + q(x,y,z)\boldsymbol{j} + r(x,y,z)\boldsymbol{k}$ defined on $\boldsymbol{B}$. Trap $\boldsymbol{B}$ in a box and subdivide the box into parallelepipeds. Consider those parallelepipeds $\{B_i : i = 1,2,\ldots,n\}$ that meet $\boldsymbol{B}$. The surface that bounds B_i will be called S_i, and oriented so that the normal points out. The union $\boldsymbol{P}_n = \cup\{B_i\}$ of all the B_i is thus an approximation to the original solid $\boldsymbol{B}$.

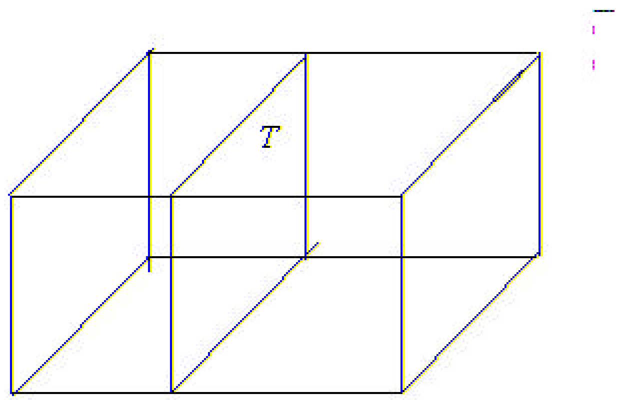
Apply the equation (⌘) to each of these and sum the equations:

$$\sum_i \iint_{S_i} \boldsymbol{F}(\boldsymbol{r})\cdot d\boldsymbol{S} = \sum_i \iiint_{B_i} \left[\frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} + \frac{\partial r}{\partial z}\right]dV\ .$$

The sum on the right hand side is just the integral over $\boldsymbol{P}_n$:

$$\sum_i \iint_{S_i} \boldsymbol{F}(\boldsymbol{r})\cdot d\boldsymbol{S} = \iiint_{P_{ni}} \left[\frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} + \frac{\partial r}{\partial z}\right]dV\ .$$

Take a closer look at the sum of the surface integrals on the left hand side of this equation. Suppose parallelepipeds B_j and B_k are adjacent, and call the common side T :



In the sum of surface integrals, the integral over the common side *T* appears twice, once from the integral over S_j, the surface of B_j and once from the integral over S_k, the surface of B_k. These integrals, will, however, have opposite signs because the orientation of *T* has one direction as a part of the surface of B_j and the opposite direction as a part of the surface of B_k. These two terms thus sum to zero and cancel each other. In the sum of all the surface integrals, we are therefore left with only the integrals over sides that are not adjacent to another box. A moments reflection, and you see that what is left is precisely the integral over the boundary $\boldsymbol{S}_n$ of $\boldsymbol{P}_n$ with the outward pointing orientation. *Mirabile dictu*, this is precisely the equation (⌘):

$$\iint_{S_n} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \iiint_{P_n} \left[\frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} + \frac{\partial r}{\partial z} \right] dV .$$

Now, as everyone can see coming, we look at the limit of this equation as we take smaller and smaller subdivisions. Then $\boldsymbol{P}_n \to \boldsymbol{B}$ and $\boldsymbol{S}_n \to \boldsymbol{S}$, giving us precisely the same result for the arbitrary region $\boldsymbol{B}$:

$$\iint_{S} \boldsymbol{F}(\boldsymbol{r}) \cdot d\boldsymbol{S} = \iiint_{B} \left[\frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} + \frac{\partial r}{\partial z} \right] dV .$$

This is really a big deal—such a big deal that it has its own name. This is called ***Gauss's Theorem***, or the ***Divergence Theorem***.

The integrand in the volume integral also has a name; it is called the ***divergence*** of the function $\boldsymbol{F}$. It is usually designated either $div\ \boldsymbol{F}$, or $\nabla \cdot \boldsymbol{F}$. Thus,

$$div\,\boldsymbol{F} = \nabla \cdot \boldsymbol{F} = \frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} + \frac{\partial r}{\partial z}.$$

With this new definition, Gauss's Theorem looks like

$$\iint_S \mathbf{F}(\mathbf{r}) \cdot d\mathbf{S} = \iiint \nabla \cdot \mathbf{F}(\mathbf{r})\, dV$$

Example

Let's find the divergence of $\boldsymbol{F}(\boldsymbol{r}) = \frac{c}{|\boldsymbol{r}|^3}\boldsymbol{r}$. First we need to see $\boldsymbol{F}$ in the form

$$\boldsymbol{F}(x,y,z) = p(x,y,z)\boldsymbol{i} + q(x,y,z)\boldsymbol{j} + r(x,y,z)\boldsymbol{k}\,.$$

That's easy:

$$\boldsymbol{F} = \frac{c}{(x^2+y^2+z^2)^{3/2}}[x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}],$$

and so

$$p = \frac{cx}{(x^2+y^2+z^2)^{3/2}},$$

$$q = \frac{cy}{(x^2+y^2+z^2)^{3/2}},$$

$$r = \frac{cz}{(x^2+y^2+z^2)^{3/2}}.$$

A bit of elementary school calculus (remember Mrs. Turner!), and we have

$$\frac{\partial p}{\partial x} = c\frac{x^2 + y^2 + z^2 - 3x^2}{(x^2 + y^2 + z^2)^{5/2}},$$

$$\frac{\partial q}{\partial y} = c\frac{x^2 + y^2 + z^2 - 3y^2}{(x^2 + y^2 + z^2)^{5/2}},$$

$$\frac{\partial r}{\partial z} = c\frac{x^2 + y^2 + z^2 - 3z^2}{(x^2 + y^2 + z^2)^{5/2}}.$$

Hence, $\nabla \cdot \boldsymbol{F} = 0$ everywhere (except, of course, for $\boldsymbol{r} = 0$, where $\boldsymbol{F}$ is not defined.).

Gauss's Theorem now tells us that the integral of $\boldsymbol{F}$ over any closed surface that does not enclose $\boldsymbol{r} = 0$ must be zero. This might be the ho-hum of the week save for the fact that the function $\boldsymbol{F}$ is a common one. It is the gravitational field of a point mass fixed at the origin, or the electric intensity field for a point charge fixed at the origin, or any field in which the magnitude is inversely proportional to the distance from the origin and which points in the direction of the origin.

Exercises

1. Find the outward flux of the function $\boldsymbol{F} = (y - x)\boldsymbol{i} + (z - y)\boldsymbol{j} + (y - x)\boldsymbol{k}$ across the boundary of the cube bounded by the planes $x = \pm 4$, $y = \pm 4$, and $z = \pm 4$.

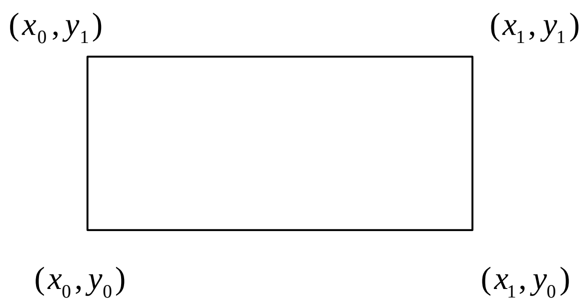
2. Find $\iint_S [y\boldsymbol{i} + xy\boldsymbol{j} - z\boldsymbol{k}] \cdot d\boldsymbol{S}$, where S is the boundary of the solid inside the cylinder $x^2 + y^2 \le 1$ between $z = 0$ and $z = x^2 + y^2$, with the outward pointing orientation.

3. Find $\iint_S [\log(x^2 + y^2)\boldsymbol{i} + \left(\frac{2z}{x}\tan^{-1}\frac{y}{x}\right)\boldsymbol{j} + z\sqrt{x^2 + y^2}\boldsymbol{k}] \cdot d\boldsymbol{S}$, where S is the boundary of the solid $\{(x, y, z): 1 \le x^2 + y^2 \le 2,\ -1 \le z \le 2\}$.

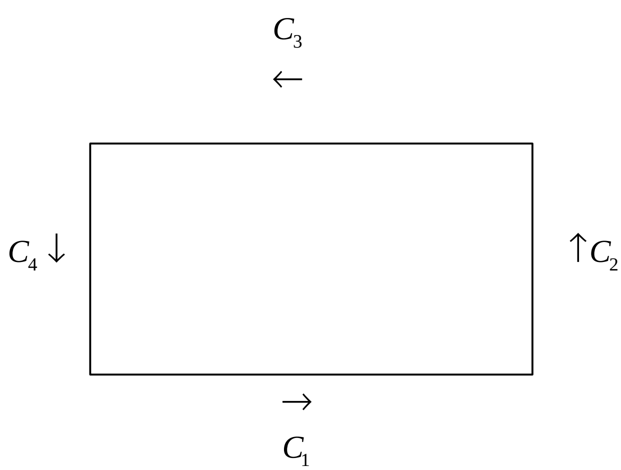
4. Let $\boldsymbol{B}$ a region in $\boldsymbol{R^3}$, and let $f:\boldsymbol{B}\to\boldsymbol{R}$ be a function such that $\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}+\frac{\partial^2 f}{\partial z^2}=0$ in $\boldsymbol{B}$ (Such a function f is said to be ***harmonic*** in $\boldsymbol{B}$.). Let S be the boundary of $\boldsymbol{B}$. Show that $\iint_S \nabla f\cdot d\boldsymbol{S}=0$.

17.2 Green's Theorem

Let R be the rectangular region in the plane bounded by the rectangle with vertices $(x_0,y_0),(x_1,y_0),(x_1,y_1),$ and (x_0,y_1).



Suppose $\boldsymbol{F}:R\to\boldsymbol{R^2}$ is a vector function given by $\boldsymbol{F}(x,y)=p(x,y)\boldsymbol{i}+q(x,y)\boldsymbol{j}$. Now, let's compute the vector line integral of $\boldsymbol{F}$ around the rectangular boundary C in the counterclockwise direction. We shall compute the integral in four parts: the integrals along each of the straight line segments making up the boundary.



Thus,

$$\int_C \boldsymbol{F} \cdot d\boldsymbol{r} = \int_{C_1} \boldsymbol{F} \cdot d\boldsymbol{r} + \int_{C_2} \boldsymbol{F} \cdot d\boldsymbol{r} + \int_{C_3} \boldsymbol{F} \cdot d\boldsymbol{r} + \int_{C_4} \boldsymbol{F} \cdot d\boldsymbol{r} .$$

We shall work out the evaluation of one of these in some painful detail; it should then be rather obvious how to do the others. Start with a vector description of C_1:

$$\boldsymbol{r}(t) = t\boldsymbol{i} + y_0\boldsymbol{j} , \quad x_0 \le t \le x_1 .$$

Then, of course, $\dfrac{d\boldsymbol{r}}{dt} = \boldsymbol{i}$, and our line integral becomes

$$\int_{C_1} \boldsymbol{F} \cdot d\boldsymbol{r} = \int_{x_0}^{x_1} [\, p(t, y_0)\boldsymbol{i} + q(t, y_0)\,\boldsymbol{j}] \cdot \boldsymbol{i}\; dt = \int_{x_0}^{x_1} p(t, y_0) dt .$$

In a similar fashion, we get

$$\int_{C_3} \boldsymbol{F} \cdot d\boldsymbol{r} = \int_{x_0}^{x_1} -p(t, y_1) dt .$$

Thus,

$$\begin{aligned}\int_{C_1} \boldsymbol{F} \cdot d\boldsymbol{r} + \int_{C_3} \boldsymbol{F} \cdot d\boldsymbol{r} &= \int_{x_0}^{x_1} -[\, p(t, y_1) - p(t, y_0)]dt \\ &= \int_{x_0}^{x_1}\int_{y_0}^{y_1} -\frac{\partial p}{\partial y}(t, s)ds dt \\ &= \iint_R -\frac{\partial p}{\partial y} dA\end{aligned}$$

In essentially the same manner, we find that

$$\int_{C_2} \boldsymbol{F} \cdot d\boldsymbol{r} + \int_{C_4} \boldsymbol{F} \cdot d\boldsymbol{r} = \iint_R \frac{\partial q}{\partial x} dA .$$

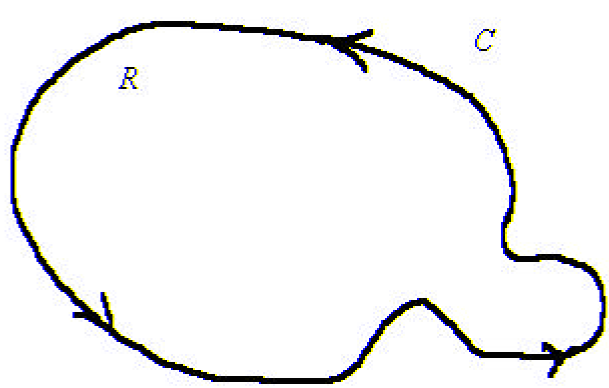
Thus

$$\int_C \boldsymbol{F} \cdot d\boldsymbol{r} = \int_{C_1} \boldsymbol{F} \cdot d\boldsymbol{r} + \int_{C_2} \boldsymbol{F} \cdot d\boldsymbol{r} + \int_{C_3} \boldsymbol{F} \cdot d\boldsymbol{r} + \int_{C_4} \boldsymbol{F} \cdot d\boldsymbol{r}$$

$$= \iint_R \left[\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y} \right] dA$$

We have turned a one dimensional vector integral into a double integral, similar to the way in which in the previous section we turned a two dimensional vector integral into a triple integral.

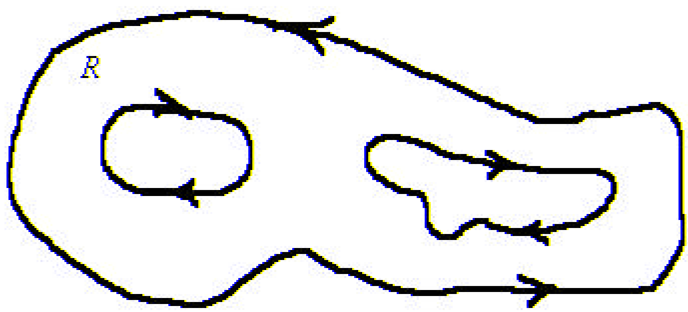
Now suppose we have a reasonable region *R* bounded by a reasonable curve *C* with a counterclockwise orientation:



Now cover this region with rectangles, and apply the above recipe to each rectangle, and add all the equations, *etc.*, *etc.*, just as we did with the parallelepipeds in deriving Gauss's Theorem. When the dust settles, we have the same result:

$$\int_C \boldsymbol{F} \cdot d\boldsymbol{r} = \iint_R \left[\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y} \right] dA .$$

This is called ***Green's Theorem***. You should note that the same equation is valid even if the region *R* is bounded by more than one closed curve.



Here the boundary C consists of three curves with the orientation indicated by the arrows in the fine picture—meditate on the covering by approximating rectangles and you will see why the orientation of the "inside" curves is clockwise. The line integral on the left side is simply the sum of the integrals over the pieces of the boundary curve.

Example

Let's evaluate the line integral $\int_C [5y\boldsymbol{i} + 3(x+1)\boldsymbol{j}] \cdot d\boldsymbol{r}$, where C is the circle of radius 2 centered at the origin, oriented counterclockwise. First, note that

$$\frac{\partial q}{\partial \mathrm{x}} = 3, \text{ and } \frac{\partial p}{\partial y} = 5 \text{ .}$$

Thus,

$$\int_C [5y\boldsymbol{i} + 3(x+1)\boldsymbol{j}] \cdot d\boldsymbol{r} = \iint_R \left[\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right] dA$$

$$= -2\iint_R dA = -8\pi$$

Exercises

5. Evaluate $\int_C [(\sin x + 3y^2)\boldsymbol{i} + (2x - e^{-y^2})\boldsymbol{j}] \cdot d\boldsymbol{r}$, where C is the boundary of the half-disc $x^2 + y^2 \leq 9,\ y \geq 0$ oriented counterclockwise.

6. Evaluate $\int_C \left[(\tan^{-1} \frac{y}{x}) \boldsymbol{i} + \log(x^2 + y^2) \boldsymbol{j} \right] \cdot d\boldsymbol{r}$, where C is the boundary of the region $1 \le r \le 2,\ 0 \le q \le p$, oriented clockwise. (These are the usual polar coordinates.)

7. Evaluate the line integral $\int_C [ye^{x^2} \boldsymbol{i} + x^3 e^y \boldsymbol{j}] \cdot d\boldsymbol{r}$, where C is the curve given by $\boldsymbol{r}(t) = \sin t\, \boldsymbol{i} + \sin 2t\, \boldsymbol{j},\ 0 \le t \le 2p$ by using Green's Theorem.

17.3 A Pleasing Application

Here we shall use Green's Theorem to find the area of a region R bound by a polygon P with vertices $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. How do we do this? We simply apply Greens' Theorem to the function

$$\mathbf{F}(x, y) = p(x, y)\mathbf{i} + q(x, y)\mathbf{j} = x\mathbf{j}.$$

Then Green's Theorem tells us that

$$\iint_R \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial x} \right) dA = \int_P \mathbf{F}(x, y) \cdot d\mathbf{r} ,$$

which becomes

$$\iint_R dA = \int_P x\mathbf{j} \cdot d\mathbf{r} .$$

We thus find the area by evaluating the line integral on the right side. This is easy. We simply integrate over each line segment of the polygon and add up the integrals.

Let's integrate along the line segment L_k from (x_k, y_k) to (x_{k+1}, y_{k+1}). A vector description of this segment is

$$\mathbf{r}(t) = (1-t)(x_k \mathbf{i} + y_k \mathbf{j}) + t(x_{k+1}\mathbf{i} + y_{k+1}\mathbf{j}),\ 0 \le t \le 1.$$

Thus $\mathbf{r}'(t) = (x_{x+1} - x_k)\mathbf{i} + (y_{k+1} - y_k)\mathbf{j}$, and we have

$$\int_{L_k} x\mathbf{j}\cdot d\mathbf{r} = \int_0^1 [(1-t)x_k + tx_{k+1}]\mathbf{j}\cdot\mathbf{r}'(t)dt$$

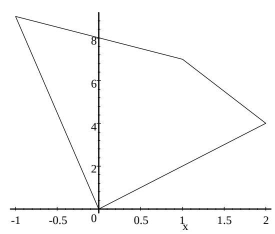
$$= (y_{k+1} - y_k)\int_0^1 [(1-t)x_k + tx_{k+1}]dt$$

$$= \frac{(y_{k+1} - y_k)(x_{k+1} + x_k)}{2}$$

Thus, Area = $\iint_R dA = \sum_{k=1}^{n-1} \frac{(y_{k+1} - y_k)(x_{k+1} + x_k)}{2} + \frac{(y_1 - y_n)(x_1 + x_n)}{2}$.

Meditate on this result. It is really a very simple formula for the area enclosed by a polygon.

Example. We shall find the area of the quadrilateral with vertices (0, 0), (2, 4), (1, 7), and (-1, 9):



Area = $\frac{1}{2}\left[(4-0)(2+0)+(7-4)(2+1)+(9-7)(-1+1)+(9-0)(-1+0)\right] = 4\right]$

Exercises

8. Find the area enclosed by the octagon with vertices (0, 0), (1, 0), (2, 3), (0, 5), (-2, 2), (-1, -1), (-2, -2), (-1, -3).

9. By means of a clever choice of the function $\mathbf{F}(x, y)$, use Green's Theorem and derive a recipe for the integral $\iint_R x dA$, where R is the region enclosed by the polygon with vertices $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

10. By means of a clever choice of the function $\mathbf{F}(x, y)$, use Green's Theorem and derive a recipe for the integral $\iint_R y\,dA$, where R is the region enclosed by the polygon with vertices $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

11. Find the centroid of the region enclosed by the triangle with vertices (1, 1), (2, 8), and (5, 5).

Chapter Eighteen

Stokes

18.1 Stokes's Theorem

Let $\boldsymbol{F}: D \to \boldsymbol{R}^3$ be a nice vector function. If

$$\boldsymbol{F}(x,y,z) = p(x,y,z)\boldsymbol{i} + q(x,y,z)\boldsymbol{j} + r(x,y,z)\boldsymbol{k}\,,$$

the ***curl*** of ***F*** is defined by

$$curl\boldsymbol{F} = \left(\frac{\partial r}{\partial y} - \frac{\partial q}{\partial z}\right)\boldsymbol{i} + \left(\frac{\partial p}{\partial z} - \frac{\partial r}{\partial x}\right)\boldsymbol{j} + \left(\frac{\partial q}{\partial x} - \frac{\partial p}{\partial y}\right)\boldsymbol{k}\,.$$

Here also the so-called del operator $\nabla = \boldsymbol{i}\frac{\partial}{\partial x} + \boldsymbol{j}\frac{\partial}{\partial y} + \boldsymbol{k}\frac{\partial}{\partial z}$ provides a nice memory device:

$$curl\boldsymbol{F} = \nabla \times \boldsymbol{F} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ p & q & r \end{vmatrix}.$$

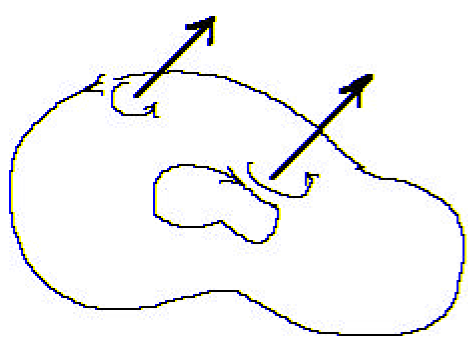
This definition allows us to look at Green's Theorem from a new perspective by observing that in case $\boldsymbol{F}(x,y) = p(x,y)\boldsymbol{i} + q(x,y)\boldsymbol{j}$, Green's Theorem becomes

$$(\heartsuit) \qquad \int_C \boldsymbol{F}\cdot d\boldsymbol{r} = \iint_R curl\boldsymbol{F}\cdot d\boldsymbol{S}\,,$$

where we are thinking of the region R as an oriented surface with its orientation pointing in the direction of ***k***.

We want to look at this formula in case the region R is not necessarily in the ***i-j*** plane, in which case, the word "clockwise" doesn't help in deciding on the orientation of the boundary C. Once again, we orient things according to our familiar "right-hand" rule.

Here's the way it goes. Suppose now S is any surface bounded by a finite number of disjoint curves $C_1, C_2, \ldots C_n$. We say simply that $C = C_1 \cup C_2 \cup \ldots \cup C_n$ is the boundary of S. Now choose an orientation for the surface S. Look at one of these normal vectors "close" to a curve C_j and imagine a little circle around the base of the normal oriented so that the normal vector points in the right-hand direction with respect to the direction of the circle. Then the orientation, or direction, of C_j that is ***consistent*** with the given orientation of the surface S is the one that "lines up" with the direction on this little circle. Look at this picture:



The surface and its boundary in this case are said the be ***consistently oriented***.

Now we do what we have done so many times in the past. Look at a surface S in three space bounded by C. (Here neither S nor C are assumed to lie in a plane.) Approximate the surface by a bunch of plane regions tangent to S, apply the equation (♥) to each of these approximating plane regions, and then sum these equations. The sum of the surface integrals is just the surface integral over the union of the approximating pieces, and the sum of the line integrals is just the line integral around the boundary of the union of the pieces—as in the plane case, the line integrals over the boundaries of adjacent regions cancel. Then, of course, we think of looking at the limit as we take more and more approximating regions, *etc*., and we obtain the equation

$$\int_C \boldsymbol{F} \cdot d\boldsymbol{r} = \iint_S curl \boldsymbol{F} \cdot d\boldsymbol{S} \,,$$

where S and C are oriented consistently. This result is the celebrated ***Stokes's Theorem***.

Example

Let's use Stokes's Theorem to evaluate the line integral

$$\int_C [-y^3\boldsymbol{i} + x^3\boldsymbol{j} - z^3\boldsymbol{k}] \cdot d\boldsymbol{r},$$

where C is the intersection of the cylinder $x^2 + y^2 = 1$ and the plane $x + y + z = 1$ oriented in the clockwise direction when viewed from above (*i.e.*, looking in the direction of **-*k***.). The curve C bounds the part of the plane $x + y + z = 1$ that lies above the set of (x, y) such that $x^2 + y^2 \le 1$. A vector description is thus given by

$$\boldsymbol{r}(s,t) = s\cos t\boldsymbol{i} + s\sin t\boldsymbol{j} + (1 - s\cos t - s\sin t)\boldsymbol{j},\ 0 \le s \le 1,\ 0 \le t \le 2\pi .$$

Hence,

$$\frac{\partial \boldsymbol{r}}{\partial s} \times \frac{\partial \boldsymbol{r}}{\partial t} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \cos t & \sin t & -(\cos t + \sin t) \\ -s\sin t & s\cos t & s(\sin t - \cos t) \end{vmatrix}$$

$$= s\boldsymbol{i} + s\boldsymbol{j} + s\boldsymbol{k}$$

I hope this result is no surprise. Notice that this is the opposite of the orientation consistent with that specified for the curve C, and so we must use

$$\frac{\partial \boldsymbol{r}}{\partial t} \times \frac{\partial \boldsymbol{r}}{\partial s} = -s(\boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k})$$

in our surface integral. The surface integral $\iint_S curl\boldsymbol{F} \cdot d\boldsymbol{S}$ looks like

$$\iint_S curl\boldsymbol{F}\cdot d\boldsymbol{S} = \int_0^{2\pi}\int_0^1 curl\boldsymbol{F}\cdot\left(\frac{\partial \boldsymbol{r}}{\partial t}\times\frac{\partial \boldsymbol{r}}{\partial s}\right)dsdt\ .$$

We must find *curl* ***F***:

$$curl\boldsymbol{F} = \nabla\times\boldsymbol{F} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ -y^3 & x^3 & -z^3 \end{vmatrix} = 3(x^2+y^2)\boldsymbol{k}\ .$$

Hence,

$$\iint_S curl\boldsymbol{F}\cdot d\boldsymbol{S} = \int_0^{2\pi}\int_0^1 curl\boldsymbol{F}\cdot\left(\frac{\partial \boldsymbol{r}}{\partial t}\times\frac{\partial \boldsymbol{r}}{\partial s}\right)dsdt$$

$$= \int_0^{2\pi}\int_0^1 3(s^2)(-s)dsdt = -2\pi\frac{3}{4} = -\frac{3}{2}\pi$$

Exercises

1. Let S be the surface $S = S_1 \cup S_2$, where $S_1 = \{(x,y,z): x^2+y^2=1,\ 0\le z\le 1\}$, and $S_2 = \{(x,y,z): x^2+y^2+(z-1)^2=1,\ z\ge 1\}$. Let the function **F** be given by

$$\mathbf{F}(x,y,z) = (x^2z^3+y)\mathbf{i} + (xy+z)\mathbf{j} + (5x\sqrt{z}+y^4)\mathbf{k}\ .$$

Compute the flux integral

$$\iint_S \nabla\times\mathbf{F}\cdot d\mathbf{S}\ ,$$

where S has the orientation pointing away from the z- axis.

2. Let S be the hemisphere $x^2 + y^2 + z^2 = 1,\ z \le 0$ with the orientation pointing toward the origin.

a)Describe the boundary of S and its orientation that is consistent with the orientation of S.

b)Evaluate the flux $\iint_S \nabla \times \mathbf{F} \cdot d\mathbf{S}$, where $\mathbf{F}(x, y, z) = 2y\mathbf{i} + x\mathbf{j} + z\mathbf{k}$

3. Let S_1 and S_2 be two surfaces with a common boundary C. Draw a picture indicating the orientations these surfaces must have to insure that

$$\iint_{S_1} \nabla \times \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{S_{21}} \nabla \times \boldsymbol{F} \cdot d\boldsymbol{S} \ .$$

4. Let S be a surface with boundary C . Suppose they are consistently oriented. Suppose $\boldsymbol{a}$ is a constant vector. Prove that

$$\int_C (\boldsymbol{a} \times \boldsymbol{r}) \cdot d\boldsymbol{r} = 2\iint_S \boldsymbol{a} \cdot d\boldsymbol{S} \ .$$

[Remember, $\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$.]

5. Suppose S is a surface with boundary C and **F** is a vector function such that $\nabla \times \mathbf{F}$ is tangent to S at each point of S. Prove that $\int_C \mathbf{F} \cdot d\mathbf{r} = 0$.

6. Let $\mathbf{r} : B \to \mathbf{R}^3$ be a vector description of the surface S with boundary C. Let **F** be a vector function such that

$$\nabla \times \mathbf{F} = \frac{1}{\left|\frac{\partial \mathbf{r}}{\partial s} \times \frac{\partial \mathbf{r}}{\partial t}\right|}\left(\frac{\partial \mathbf{r}}{\partial s} \times \frac{\partial \mathbf{r}}{\partial t}\right).$$

Show that $\left|\int_C \mathbf{F} \cdot d\mathbf{c}\right| =$ area of S.

7. Suppose the vector function **F** on a domain *D* is conservative. Prove that $\nabla \times \mathbf{F} = 0$ everywhere in *D*.

8. Let $\mathbf{F}(x, y, z) = \frac{-y}{x^2 + y^2}\mathbf{i} + \frac{x}{x^2 + y^2}\mathbf{j}, \quad x^2 + y^2 \neq 0$.

a)Compute $\nabla \times \mathbf{F}$.

b)Prove that **F** is not conservative. [Hint: Evaluate the line integral $\int_C \mathbf{F} \cdot d\mathbf{r}$, where *C* is the circle $x^2 + y^2 = 1$, $z = 0$, with the usual counterclockwise orientation.]

18.2 Path Independence Revisited

Problem **7** at the end of the previous section perhaps raised our hopes that an easy test for a function **F** to be conservative in a domain *D* is simply to see if $\nabla \times \mathbf{F}$ =0. If so, these hopes were quickly dashed by Problem **8**. In this section, we shall see just what we can do along this line. The concept introduced next provides the key to understanding and enlightenment.

An open subset *D* of $\mathbf{R}^3$ is called ***simply connected*** if every simple closed curve in *D* is the boundary of some surface contained entirely in *D*. Thus for instance the region

$$D = \{(x, y, z) : x^2 + y^2 + z^2 < 1\}$$

is simply connected, while the region

$$R = \{(x, y, z) : x^2 + y^2 > 1\}$$

is not.

Now it easy to see that if **F** has as domain a simply connected region D, then $\nabla\times\mathbf{F}=0$ everywhere in D implies that **F** is indeed conservative. We show that **F** is conservative by showing that the integral of **F** around any closed curve is 0. This is easy to do. Let C be any closed curve in D. Then D is simply connected, so there is a surface S the boundary of which is C. Now unleash Stokes's Theorem:

$$\int_C \mathbf{F}\cdot d\mathbf{r} = \iint_S \nabla\times\mathbf{F}\cdot d\mathbf{S} = 0.$$

How about that!

Exercises

9. Explain how you know that $\mathbf{F}(x,y,z)=\frac{-y}{x^2+y^2}\mathbf{i}+\frac{x}{x^2+y^2}\mathbf{j}$, $x>0$. is conservative.

10. Find a potential function for the vector function **F** given in Problem **9**.

Chapter Nineteen

Some Physics

19.1 Fluid Mechanics

Suppose $\mathbf{v}(x, y, z, t)$ is the velocity at $\mathbf{r} = (x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ of a fluid flowing smoothly through a region in space, and suppose $r(x, y, z, t)$ is the density at $\mathbf{r}$ at time t. If S is an oriented surface, it is not hard to convince yourself that the flux integral

$$\iint_S r\mathbf{v} \cdot d\mathbf{r}$$

is the rate at which mass flows through the surface S. Now, if S is a closed surface, then the mass in the region B bounded by S is, of course

$$\iiint_B r\,dV \; .$$

The rate at which this mass is changing is simply

$$\frac{\partial}{\partial t}\iiint_B r\,dV = \iiint_B \frac{\partial r}{\partial t}\,dV \, .$$

This is the same as the rate at which mass is flowing across S into B: $-\iint_S r\mathbf{v} \cdot d\mathbf{r}$, where S is given the outward pointing orientation. Thus,

$$\iiint_B \frac{\partial r}{\partial t}\,dV = -\iint_S r\mathbf{v} \cdot d\mathbf{r} \, .$$

We now apply Gauss's Theorem and get

$$\iiint_B \frac{\partial r}{\partial t}\,dV = -\iint_S r\mathbf{v} \cdot d\mathbf{r} = \iiint_B -\nabla \cdot (r\mathbf{v})\,dV .$$

Thus,

$$\iiint_B \left(\frac{\partial r}{\partial t} + \nabla \cdot (r\mathbf{v}) \right) dV \, .$$

Meditate on this result. The region B is *any* region, and so it must be true that the integrand itself is everywhere 0:

$$\frac{\partial r}{\partial t} + \nabla \cdot (r\mathbf{v}) = 0 .$$

This is one of the fundamental equations of fluid dynamics. It is called the ***equation of continuity***.

In case the fluid is incompressible, the continuity equation becomes quite simple. Incompressible means simply that the density r is constant. Thus $\frac{\partial r}{\partial t} = 0$ and so we have

$$\frac{\partial r}{\partial t} + \nabla \cdot (r\mathbf{v}) = \nabla \cdot (r\mathbf{v}) = r\nabla \cdot \mathbf{v} = 0, \text{ or}$$

$$\nabla \cdot \mathbf{v} = 0 .$$

Exercise

1. Consider a one dimensional flow in which the velocity of the fluid is given by $\mathbf{v} = f(x)$, where $f(x) > 0$. Suppose further that the density r of the fluid does not vary with time *t*. Show that

$$r(x) = \frac{k}{f(x)},$$

where *k* is a constant.

19.2 Electrostatics

Suppose there is a point charge *q* fixed at the point ***s***. Then the electric field $\mathbf{E}_q(\mathbf{r})$ due to *q* is given by

$$\mathbf{E}_q(\mathbf{r}) = kq\frac{\mathbf{r} - \mathbf{s}}{|\mathbf{r} - \mathbf{s}|^3} .$$

It is easy to verify, as we have done in a previous chapter, that this field, or function, is conservative, with a potential function

$$P_q(\mathbf{r}) = \frac{-kq}{|\mathbf{r} - \mathbf{s}|};$$

so that $\mathbf{E}_q = \nabla P_q$. Physicists do not like to be bothered with the minus sign in P_q, so they define the electric potential V_q to be $-P_q$. Thus,

$$V_q(\mathbf{r}) = \frac{kq}{|\mathbf{r}-\mathbf{s}|},$$

and

$$\mathbf{E}_q(\mathbf{r}) = -\nabla V_q(\mathbf{r}).$$

We have already seen that the flux out of a closed surface S is

$$\iint_S \mathbf{E}_q \cdot d\mathbf{S} = \begin{cases} 0 & \text{if } S \text{ does not enclose the origin} \\ 4pkq & \text{if } S \text{ does enclose the origin} \end{cases}$$

Some meditation will convince you there is nothing special here about the origin; that is, if the point charge is at **s**, then

$$\iint_S \mathbf{E}_q \cdot d\mathbf{S} = \begin{cases} 0 & \text{if } S \text{ does not enclose } \mathbf{s} \\ 4pkq & \text{if } S \text{ does enclose } \mathbf{s} \end{cases}$$

Next, suppose there are a finite number of point charges q_1 at $\mathbf{s}_1$, q_2 at $\mathbf{s}_2, \ldots,$ and q_n at $\mathbf{s}_n$. Suppose $\mathbf{E}_j$ is the electric intensity due to q_j. Then it should be clear that the electric field due to these charges is simply the sum

$$\mathbf{E}(\mathbf{r}) = \sum_{j=1}^{n} \mathbf{E}_j = k\sum_{j=1}^{n} q_j \frac{\mathbf{r}-\mathbf{s}_j}{|\mathbf{r}-\mathbf{s}_j|^3}.$$

Also,

$$V(\mathbf{r}) = k\sum_{j=1}^{n} \frac{q_j}{|\mathbf{r}-\mathbf{s}_j|}; \text{ and}$$

$$\mathbf{E}(\mathbf{r}) = -\nabla V(\mathbf{r}).$$

Finally,

$$\iint_S \mathbf{E} \cdot d\mathbf{S} = 4pk\sum q_j$$

where the sum is over those charges enclosed by S.

Things become more exciting if instead of point charges, we have a charge distribution in space with charge density r. To find the electric field $\mathbf{E}(\mathbf{r})$ produced by this distribution of charge in space, we need to integrate:

$$\mathbf{E}(\mathbf{r}) = \iiint_U k r(\mathbf{s}) \frac{(\mathbf{r}-\mathbf{s})}{|\mathbf{r}-\mathbf{s}|^3} dV_s .$$

But this appears to be a serious breach of decorum. We are integrating over everything, and at $\mathbf{s} = \mathbf{r}$ we have the dreaded 0 in the denominator. Thus what we see above is an ***improper*** integral—that is, it is actually a limit of integrals. Specifically, we integrate not over everything but over everything outside a spherical solid region of radius a centered at **r**. We then look at the limit as $a \to 0$ of this integral. With the integral for the electric field, this limit exists, and so there is no problem with 0 on the bottom of the integrand. In the same way, we are safe in writing for the potential

$$V(\mathbf{r}) = k\iiint_U \frac{r(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} dV_s .$$

Everything works nicely so that we also have

$$\mathbf{E}(\mathbf{r}) = -\nabla V(\mathbf{r}).$$

If R is a solid region bounded by a closed surface S, then we can also integrate to get

$$(*) \quad \iint_S \mathbf{E} \cdot d\mathbf{S} = 4pk \iiint_R r(\mathbf{s}) dV.$$

The divergence of **E** is the troublesome item in extending matters to distributed charge. If we simply try to calculate the divergence by $div \iiint_U \text{stuff}\, dV = \iiint_U div(\text{stuff})\, dV$, then things go wrong because the improper integral of the divergence does not exist. Gauss saves the day. Let R be any region and let S be the closed surface bounding R. Then

$$\iint_S \mathbf{E} \cdot d\mathbf{S} = \iiint_R \nabla \cdot \mathbf{E}\, dV .$$

But from equation (*) we have

$$\iint_S \mathbf{E} \cdot d\mathbf{S} = 4pk \iiint_R r(\mathbf{s}) dV = \iiint_R \nabla \cdot \mathbf{E}\, dV .$$

This gives us

$$\iiint_R 4pkr\,dV = \iiint_R \nabla \cdot \mathbf{E}\,dV \text{ , or}$$

$$\iiint_R (\nabla \cdot \mathbf{E} - 4pkr)\,dV \text{ .}$$

But R is *any* region, and so it must be true that

$$\nabla \cdot \mathbf{E} = 4pkr$$

for all **r**.

Finally, remembering that $\mathbf{E} = -\nabla V$, we get

$$\nabla \cdot \mathbf{E} = -\nabla \cdot (\nabla V) = 4pkr \text{ ;}$$

$$\nabla^2 V = -4pkr \text{ , or}$$

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -4pkr \text{ .}$$

This is the celebrated ***Poisson's Equation***, a justly famous partial differential equation, the study of which is beyond the scope of this course.

.

Taylor's Theorem

1. **Introduction**. Suppose f is a one-variable function that has $n+1$ derivatives on an interval about the point $x = a$. Then recall from Ms. Turner's class the single variable version of Taylor's Theorem tells us that there is exactly one polynomial p of degree $\leq n$ such that $p(a) = f(a)$, $p'(a) = f'(a)$, $p''(a) = f''(a), \ldots, p^{(n)}(a) = f^{(n)}(a)$. This polynomial is given by

$$p(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \ldots + \frac{f^{(n)}}{n!}(x-a)^n$$

We also know the difference between $f(x)$ and $p(x)$:

$$f(x) - p(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1},$$

where ξ is somewhere between a and x.

The polynomial p is called the **Taylor Polynomial** of degree $\leq n$ for f at a.

Before we worry about what the Taylor polynomial might be in higher dimensions, we need to be sure we understand what is a polynomial in more than one dimension. In two dimensions, a polynomial $p(x,y)$ of degree $\leq n$ is a function of the form

$$p(x,y) = \sum_{i,j=0}^{i+j=n} a_{ij}x^i y^j.$$

Thus a polynomial of degree ≤ 2 (perhaps more commonly known as a quadratic) looks like

$$p(x,y) = a_{00} + a_{10}x + a_{01}y + +a_{11}xy + a_{20}x^2 + a_{02}y^2.$$

I hope it easy to guess what one means by a polynomial in three variables, (x,y,z), or indeed, in any number of variables.

Now, how might we extend the idea of the Taylor polynomial of degree $\leq n$ for a function f at a point $\mathbf{a}$? Simple enough. It's a polynomial $p(\mathbf{x})$ of degree $\leq n$ so that

$$\frac{\partial^{i_1+\ldots+i_q} f(a)}{\partial x_1^{i_1} \partial x_2^{i_2} \ldots \partial x_q^{i_q}} = \frac{\partial^{i_1+\ldots+i_q} p(a)}{\partial x_1^{i_1} \partial x_2^{i_2} \ldots \partial x_q^{i_q}},$$

for all $i_1, i_2, \ldots, i_q$ such that $i_1 + i_2 + \ldots + i_q \leq n$.

This looks pretty ferocious in general, so let's see what it says for just two variables. In this case, we have $\mathbf{a} = (a,b)$ and the Taylor polynomial $p(x,y)$ at $\mathbf{a}$ becomes the polynomial such that

$$\frac{\partial^{i+j} f(\mathbf{a})}{\partial^i x \partial^j y} = \frac{\partial^{i+j} p(\mathbf{a})}{\partial^i x \partial^j y},$$

for all $i + j \leq n$.

Example

Let $f(x,y) = \cos(x+y)$, and let $p(x,y) = 1 - \frac{x^2}{2} - xy - \frac{y^2}{2}$. Let's verify that p is the Taylor polynomial of degree ≤ 2 for f at $(0,0)$. He we go.

$$f(0,0) = 1, \text{ and } p(0,0) = 1;$$
$$\frac{\partial f}{\partial x} = -\sin(x+y), \text{ and } \frac{\partial p}{\partial x} = -x - y;$$
$$\frac{\partial f}{\partial y} = -\sin(x+y), \text{ and } \frac{\partial p}{\partial y} = -x - y;$$
$$\frac{\partial^2 f}{\partial x^2} = -\cos(x+y), \text{ and } \frac{\partial^2 p}{\partial x^2} = -1,$$
$$\frac{\partial^2 f}{\partial y^2} = -\cos(x+y), \text{ and } \frac{\partial^2 p}{\partial y^2} = -1,$$
$$\frac{\partial^2 f}{\partial x \partial y} = -\cos(x+y), \text{ and } \frac{\partial^2 p}{\partial x \partial y} = -1.$$

Now it's easy to see that

$$f(0,0) = 0 = p(0,0);$$
$$\frac{\partial f}{\partial x}(0,0) = 0 = \frac{\partial p}{\partial x}(0,0);$$
$$\frac{\partial f}{\partial y}(0,0) = 0 = \frac{\partial p}{\partial y}(0,0);$$
$$\frac{\partial^2 f}{\partial x^2}(0,0) = -1 = \frac{\partial^2 p}{\partial x^2}(0,0);$$
$$\frac{\partial^2 f}{\partial y^2}(0,0) = -1 = \frac{\partial^2 p}{\partial y^2}(0,0); \text{ and}$$
$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = -1 = \frac{\partial^2 p}{\partial x \partial y}(0,0).$$

Exercises

1. Verify that the polynomial in the Example is also the Taylor polynomial for f at (0,0) of degree ≤ 3.

2. Let $f(x,y) = \sin(x+y)$.Which Which of the following is the Taylor polynomial of degree ≤ 2 for f at (0,0)? Explain.

a) $p(x,y) = 1 + x^2 + y^2$ b) $p(x,y) = xy$

c) $p(x,y) = x^2 + xy + 2y$ d) $p(x,y) = x + y$

2. **Derivatives**. Prior to finding a general recipe for the Taylor polynomial, we need look at finding higher order derivatives of certain composite functions. Let f be a real-valued function defined on a subset of $\mathbf{R}^q$. Suppose that in a neighborhood of the point $\mathbf{x}$, the function f has a lot of continuous partial derivatives. Define the function g by

$$g(t) = f(\mathbf{a} + t\mathbf{h}),$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_q)$ and $\mathbf{h} = (h_1, h_2, \ldots, h_q)$. We know from the chain rule that $g'(t)$ is given by

$$\begin{aligned} g'(t) &= \nabla f(\mathbf{a} + t\mathbf{h}) \bullet \mathbf{h} \\ &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_q} \right) \bullet (h_1, h_2, \ldots, h_q) \\ &= \left(h_1 \frac{\partial}{\partial x_1} + h_2 \frac{\partial}{\partial x_2} + \ldots + h_q \frac{\partial}{\partial x_q} \right) f \Big|_{(\mathbf{a}+t\mathbf{h})} \end{aligned}$$

In keeping with our general practice of restricting ourselves to dimensions one, two, or three, let's look first at the case $q = 2$. As usual, we'll write $\mathbf{x} = (x, y)$ and $\mathbf{h} = (h, k)$. The expression for $g'(t)$ now looks like:

$$g'(t) = \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right) f \Big|_{(\mathbf{x}+t\mathbf{h})}$$

We are now in business, for we have a nice recipe for higher order derivatives of g :

$$g^{(m)}(t) = \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right)^m f \Big|_{(\mathbf{x}+t\mathbf{h})}$$

For example,

$$\begin{aligned} g''(t) &= \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y} \right)^2 f \\ &= \left(h^2 \frac{\partial^2}{\partial x^2} + 2hk\frac{\partial^2}{\partial x \partial y} + k^2 \frac{\partial^2}{\partial y^2} \right) f \\ &= h^2 \frac{\partial^2 f}{\partial x^2} + 2hk\frac{\partial^2 f}{\partial x \partial y} + k^2 \frac{\partial^2 f}{\partial y^2} \end{aligned}$$

Example

Suppose $f(x,y) = x^2 y^3 + y^2$. Let's find the second derivative of the function

$$g(t) = f(1 + 3t, -2 + t)$$

First,

$$g''(t) = \left(3\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^2 f$$
$$= 9\frac{\partial^2 f}{\partial x^2} + 6\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2}$$

Now, $\frac{\partial f}{\partial x} = 2xy^3$, and $\frac{\partial f}{\partial y} = 3x^2y^2 + 2y$, and so $\frac{\partial^2 f}{\partial x^2} = 2y^3$, $\frac{\partial^2 f}{\partial y \partial x} = 6y^2$, and $\frac{\partial^2 f}{\partial y^2} = 6x^2y + 2$. Thus,

$$g''(t) = 18(-2+t)^3 + 36(-2+t)^2 + 6(1+3t)^2(-2+t) + 2$$

Exercises

3. Let $f(x,y) = xe^y$. Find the derivative of $g(t) = f(1+t, 3-4t)$.

4. Find the second derivative of the function g defined in **Problem 3**.

5. Let $F(u,v) = u^3v + v^2$. Find the second derivative of $R(z) = F(z, 3z)$.

6. Find $g'''(t)$, where g is the function defined in the Example.

3. **The Taylor polynomial**. To find the Taylor polynomial for a function f of several variables at a point $\mathbf{a}$, we shall simply apply the one-dimensional results to the function

$$g(t) = f(\mathbf{a} + t\mathbf{h}).$$

Thus,

$$g(t) = \sum_{m=0}^{n} \frac{g^{(m)}(0)}{m!} t^m + \frac{g^{(n+1)}(\xi)}{(n+1)!} t^{n+1},$$

where ξ is a number between 0 and t. Next, substitute $t = 1$ into the above:

$$g(1) = f(\mathbf{a}) = \sum_{m=0}^{n} \frac{g^{(m)}(0)}{m!} + \frac{g^{(n+1)}(\xi)}{(n+1)!}$$

We know the value of $g^{(k)}$ from **Section 2**:

$$f(\mathbf{a}+\mathbf{h}) = \sum_{m=0}^{n} \frac{1}{m!}\left(h_1\frac{\partial}{\partial x_1} + h_2\frac{\partial}{\partial x_2} + \ldots + h_q\frac{\partial}{\partial x_q}\right)^m f(\mathbf{a})$$

$$+\frac{1}{(n+1!}\left(h_1\frac{\partial}{\partial x_1}+h_2\frac{\partial}{\partial x_2}+\ldots+h_q\frac{\partial}{\partial x_q}\right)^{n+1}f(\mathbf{c})$$

The point **c** lies somewhere on the line segment joining **a** and **a** + **h**.
The polynomial

$$p(\mathbf{h}) = p(h_1,h_2,\ldots,h_q) = \sum_{m=0}^{n}\frac{1}{m!}\left(h_1\frac{\partial}{\partial x_1}+h_2\frac{\partial}{\partial x_2}+\ldots+h_q\frac{\partial}{\partial x_q}\right)^{m}f(\mathbf{a})$$

is the Taylor polynomial of degree $\leq n$ for f at **a**; the last term is traditionally called the **error term** or sometimes, the **remainder term**. Actually, if we let $\mathbf{h} = \mathbf{x}-\mathbf{a}$, then $q(\mathbf{x}) = p(\mathbf{x}-\mathbf{a})$ is the thing we called the Taylor polynomial in the first section.

This is pretty fierce looking. Let's look at the two variable case:

$$f(a_1+h,a_2+k) = \sum_{m=0}^{n}\frac{1}{m!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^{m}f(a_1,a_2)$$
$$+\frac{1}{(n+1!)}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^{n+1}f(c_1,c_2)$$

where (c_1,c_2) is on the line joining (a_1,a_2) and (a_1+h,a_2+k).

Example

Let $f(x,y) = \sin x\sin y$. For $n = 2$ and $\mathbf{a} = (0,0)$, Taylor's polynomial becomes

$$p(h,k) = f(0,0)+h\frac{\partial f}{\partial x}(0,0)+k\frac{\partial f}{\partial y}(0,0)+\frac{h^2}{2}\frac{\partial^2 f}{\partial x^2}(0,0)+hk\frac{\partial^2 f}{\partial x\partial y}(0,0)+\frac{k^2}{2}\frac{\partial^2 f}{\partial y^2}(0,0)$$

We have
$\frac{\partial f}{\partial x} = \cos x\sin y$; $\frac{\partial f}{\partial y} = \sin x\cos y$; $\frac{\partial^2 f}{\partial x^2} = -\sin x\sin y$; $\frac{\partial^2 f}{\partial x\partial y} = \cos x\cos y$; $\frac{\partial^2 f}{\partial y^2} = -\sin x\sin y$.
Thus,

$$p(h,k) = hk.$$

Let's get an estimate for how well this approximates $\sin x\sin y$ near $(0,0)$. We know that

$$|\sin x\sin y - xy| = \left|\frac{1}{3!}\left(x\frac{\partial}{\partial x}+y\frac{\partial}{\partial y}\right)^{3}f(\xi,\mu)\right|$$

where (ξ,μ) is one the segment joining (x,y) and the origin. Now,

$$\left(x\frac{\partial}{\partial x}+y\frac{\partial}{\partial y}\right)^3 f = x^3\frac{\partial^3 f}{\partial x^3}+3x^2y\frac{\partial^3 f}{\partial x^2\partial y}+3xy^2\frac{\partial^3 f}{\partial x\partial y^2}+y^3\frac{\partial^3 f}{\partial x^3}.$$

Next, let's suppose that $|x| \le c$ and $|y| \le c$ for some constant c. Noting that all the partial derivatives in the above expression are simply products of sine and cosines, we can estimate

$$\left|\left(x\frac{\partial}{\partial x}+y\frac{\partial}{\partial y}\right)^3 f\right| \le 8c^3,$$

and so, at last,

$$|\sin x \sin y - xy| \le \frac{8c^3}{6} = \frac{4}{3}c^3$$

Exercises

7. Find the Taylor polynomial of degree ≤ 1 for $f(x,y) = e^{xy}$ at $(0,0)$.

8. Find the Taylor polynomial of degree ≤ 2 for $f(x,y) = e^{xy}$ at $(0,0)$.

9. Find the Taylor polynomial of degree ≤ 3 for $f(x,y) = e^{xy}$ at $(0,0)$.

10. Find the Taylor polynomial of degree ≤ 1 for $f(x,y) = e^x \cos y$ at $(0,0)$.

11. Use Taylor's Theorem to find a quadratic approximation of $e^x \cos y$ at the origin.

12. Estimate the error in the approximation found in Problem **11** if $|x| \le 0.1$ and $|y| \le 0.1$.