MULTIVARIABLE CALCULUS

for MIT Interphase students

Samuel S. Watson
sswatson@mit.edu

Preface

This text was written to accompany the multivariable calculus course taught during the Office of Minority Education's six-week summer *Interphase* program at the Massachusetts Institute of Technology. Students participating in this program have the opportunity to take a full-semester version of multivariable calculus in the fall semester.

For this reason, the material presented here is substantially streamlined compared to a typical multivariable calculus course. Rather than thoroughly covering the first half of the standard curriculum and leaving students with an abrupt drop-off in content knowledge midway through the fall, I have carved out a selection of topics which approximately spans the standard curriculum but which nevertheless includes the most important ideas and is self-contained.

Another casualty of the abbreviated schedule is mathematical rigor. I emphasize concept visualization and physical intuition, and I include proofs only when they are sufficiently illuminating to be worth the effort. I fully embrace the physicist's perspective that one should derive formulas by reasoning about small positive quantities like Δx, trusting that one may replace sums with integrals and Δx with dx to obtain correct results. Turning these heuristics into proofs is left to later courses. While there are advantages to maintaining a higher standard of rigor, my top priority is for students to gain problem-solving facility with the methods covered in the course.

SW
2017

For the reader

Textbooks customarily include an abundance of material and leave you to figure out what to focus on. By contrast, I have attempted to pare the multivariable calculus story down to its essence. My goal is that you will be able to read each section carefully and work all the examples and exercises, thereby wasting no effort on figuring out what to skip. However, problem solving is the heart of mathematics, and you are likely to want additional exercises and more discussion or examples. I recommend the following sources:

1. The community calculus textbook, available at `communitycalculus.org`. This is an open source textbook available (for free) online, and it has *a lot* of exercises
2. Frederick Tsz-Ho Fong's multivariable calculus notes for Math 0200 at Brown University, available at `http://www.math.brown.edu/~sswatson/pdfs/frederick_multinotes.pdf`. These notes are very nicely written and have quite a few examples.
3. MathInsight (`mathinsight.org`) has webpages for many multivariable topics with nice exposition and some beautiful applets for exploring the ideas developed in this course
4. My Math 0200 webpage, with homework sets, practice tests, and solutions: `http://www.math.brown.edu/~sswatson/math0200.html`
5. A standard multivariable calculus textbook, such as Stewart or Edwards and Penney
6. 3blue1brown, a math video creator with an excellent series on linear algebra. Unfortunately he hasn't done multivariable calculus yet, so these videos will only be available for the vector topics

The following margin note icons in the text are clickable:

1. links to a CoCalc worksheet with a relevant calculation (see Appendix A.6 for more discussion)
2. links to a relevant 3blue1brown YouTube video.
3. mi links to a relevant page at mathinsight.org

Almost all of the **3D graphics in this PDF may be interactively manipulated**, but that feature requires that the PDF be viewed with Adobe's (free) Acrobat Reader (`https://get.adobe.com/reader/`). The references in this text are hyperlinked, which means for example that you can click on Theorem 6.3.1 to navigate directly to Green's theorem.

Please do not hesitate to contact me via email (sswatson@mit.edu) about any mistakes you find in these notes, no matter how minor. Enjoy!

Contents

1 Transformations of Euclidean space **6**
1.1 n-dimensional space . . . 6
1.2 Functions from $\mathbb{R}^n$ to $\mathbb{R}^n$ . . . 6
1.2.1 Visualizing functions . . . 6
1.2.2 Linear transformations . . . 7
1.3 The determinant . . . 9

2 Vectors **13**
2.1 Introduction to vectors . . . 13
2.2 The dot product . . . 15
2.3 The cross product . . . 16

3 Three-dimensional Geometry **18**
3.1 Lines and planes . . . 18
3.2 Motion in space . . . 21
3.3 Quadric surfaces . . . 22
3.4 Polar, cylindrical, and spherical coordinates . . . 24

4 Multivariable Differentiation **27**
4.1 Limits . . . 27
4.2 Partial derivatives . . . 31
4.3 Linear approximation . . . 33
4.4 Multivariable optimization . . . 36
4.5 Directional derivative and gradient . . . 38
4.6 The multivariable chain rule . . . 40
4.7 Optimization with Lagrange multipliers . . . 41

5 Multivariable Integration **45**
5.1 Double integration . . . 45
5.2 Triple integration . . . 48
5.3 Polar, cylindrical, and spherical integration . . . 52
5.4 Change of variables . . . 54

6 Vector Calculus **57**
6.1 Vector fields and line integrals 57
6.2 The fundamental theorem of vector calculus 59
6.3 Green's theorem 61
6.4 Surface integrals and flow 63
6.4.1 Surface integrals 63
6.4.2 Flow 64
6.5 Divergence and curl 65
6.5.1 Divergence 65
6.5.2 Curl 67
6.6 Divergence theorem 69
6.7 Stokes' theorem 71

A Appendix **74**
A.1 Sets and functions 74
A.2 Trig review 76
A.3 Visualizing functions 78
A.4 Summation notation 80
A.5 Polar, cylindrical, and spherical coordinate reference 81
A.6 SageMath 82

1 Transformations of Euclidean space

1.1 *n*-dimensional space

We visualize the set of **real numbers** (denoted by $\mathbb{R}$ or $\mathbb{R}^1$) as a line, called the *real number line* (Figure 1.1).

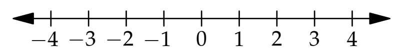


Figure 1.1 The real number line

The location of a number x on this line is the point whose *signed* distance from 0 is x. The word *signed* means that distances measured from 0 to a point which is left of 0 count as negative.

The set $\mathbb{R}^2$ of ordered pairs (x, y) where x and y are real numbers can be thought of as a *plane*, since a point in a plane can be specified by two real numbers: its signed distances from two lines which meet at a right angle. These two lines are called the x-axis and the y-axis.

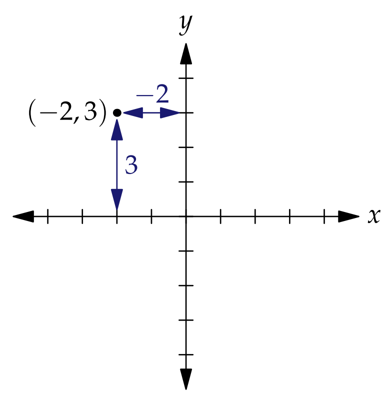


Figure 1.2 Coordinates in $\mathbb{R}^2$

The set $\mathbb{R}^3$ of ordered triples of real numbers (x, y, z) can be visualized as a point in space, since a point in space can be specified by three real numbers: its signed distances from three planes which meet one another at right angles. These planes are called the xy-plane, the yz-plane, and the xz-plane, and their lines of intersection are called the x-axis, the y-axis, and the z-axis.

Since we live in three dimensions, spatial visualization of $\mathbb{R}^n$ only works for $n \leq 3$. We will use other visualization tools or reason by analogy to build intuition in $\mathbb{R}^n$ for $n \geq 4$.

The set $\mathbb{R}^4$ is defined to be the set of ordered quadruples of real numbers, and similarly for $\mathbb{R}^5, \mathbb{R}^6, \ldots$. If n is a positive integer*, we refer to $\mathbb{R}^n$ as a **Euclidean space**. The superscript n is called the **dimension** of $\mathbb{R}^n$.

Exercise 1.1.1

Write down a formula for the distance on the number line between two real numbers x and y. Repeat for two points in the plane, and for two points in $\mathbb{R}^3$.

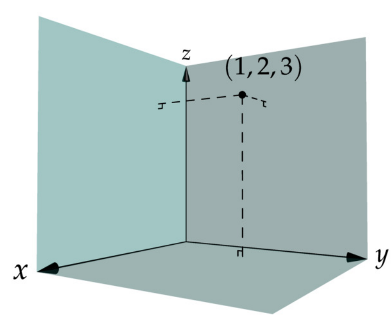


Figure 1.3 Coordinates in $\mathbb{R}^3$

1.2 Functions from $\mathbb{R}^n$ to $\mathbb{R}^n$

1.2.1 Visualizing functions

A **function** from $\mathbb{R}^1$ to $\mathbb{R}^1$ takes a real number x as input and returns another real number, denoted $f(x)$, as output. We can draw the *graph* of such a function in $1 + 1 = 2$ dimensions, by associating the horizontal axis with input values and the vertical axis with output values. For example, see Figure 1.4 for a graph of the squaring function.

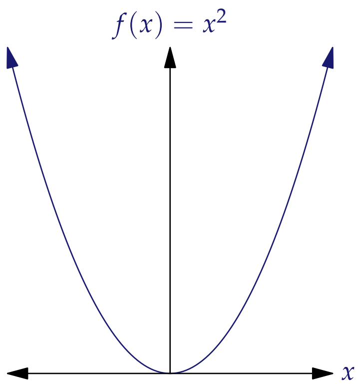


Figure 1.4 The graph of a function from $\mathbb{R}$ to $\mathbb{R}$

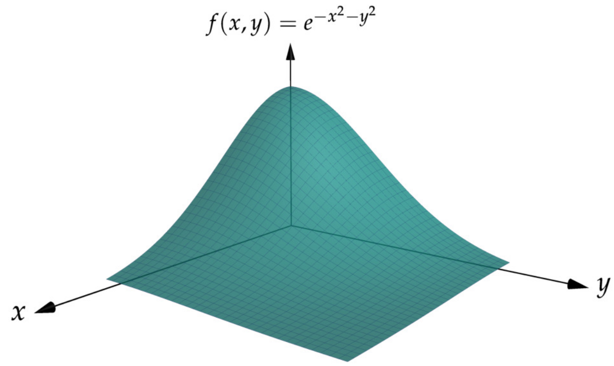


Figure 1.5 A graph of a function from $\mathbb{R}^2$ to $\mathbb{R}$

A function from $\mathbb{R}^2$ to $\mathbb{R}^1$ can be visualized in $2+1=3$ dimensions by using the xy-plane for the input values and the z-axis for the output value. In other words, we plot every triple of the form $(x, y, f(x,y))$ where x and y are real numbers. See Figure 1.5 for a graph of $f(x) = e^{-x^2-y^2}$ over a square-shaped region.

The graph of function from $\mathbb{R}^2$ to $\mathbb{R}^2$ would require $2+2=4$ dimensions to visualize, so we are out of luck there. However, we can visualize a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ drawing a picture of where all the grid lines go*—see Figure 1.6.

One way this method of visualization is different from graphing is that we separate the input values on the left side from the output values on the right side.

We often refer to a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ as a *transformation*, which is just a synonym of *function* but is meant to evoke this particular method of visualization. Functions from $\mathbb{R}^3$ to $\mathbb{R}^3$ are also called transformations and can be visualized in the same way, but for now we will focus on transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$.

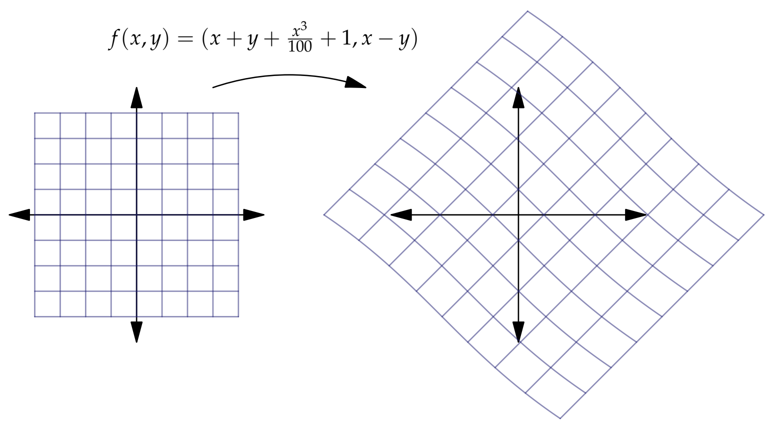


Figure 1.6 A transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$

on linear transformations

1.2.2 Linear transformations

One of the key ideas of differential calculus is to use *linear* functions to approximate curvy ones. Although linear functions are simple, they are very useful because all differentiable functions look increasingly linear as you zoom in. We will apply the same principle in higher dimensions: use linear transformations to approximate more complex transformations. Thus, as you did before you learned single-variable calculus, we will begin by learning about linear functions.

So what *is* a linear function from $\mathbb{R}^2$ to $\mathbb{R}^2$? Functions of the form $f(x) = mx + b$, where m and b are constants, are often called linear. However, we will take a slightly different view by requiring $b = 0$, so only functions of the form $f(x) = mx$ are considered linear. Our definition of linearity in higher dimensions will work similarly: only terms of the form "constant times variable" are allowed.

This notation means that a, b, c, d are in the set of real numbers.

Definition 1: Linearity

A function f from $\mathbb{R}^2$ to $\mathbb{R}^2$ is **linear** if there exist* $a, b, c, d \in \mathbb{R}$ so that

$$f(x, y) = (ax + by, cx + dy).$$

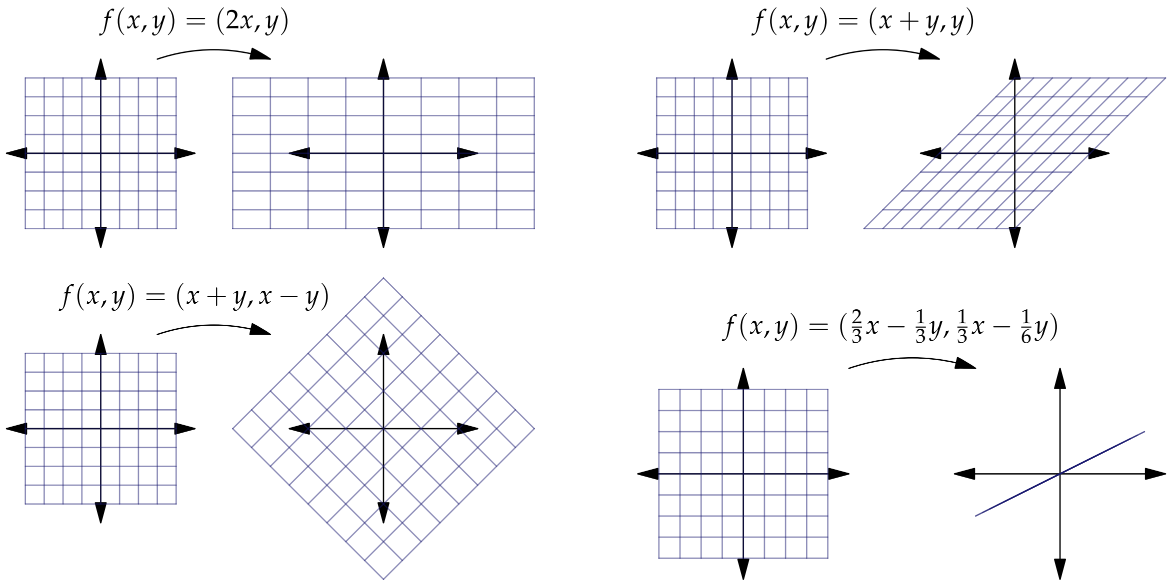


Figure 1.7 Four linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$

These transformations *scale*, *shear*, *rotate/scale*, and *project*, respectively.

Figure 1.7 shows four examples* of linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$. These pictures might lead us to conjecture that linear transformations map equally spaced lines to equally spaced lines, where coincident lines count as equally spaced (as in the last example). This is almost accurate: sometimes equally spaced lines can map to equally spaced *points* (Exercise 1.2.2).

Theorem 1.2.1

A function from $\mathbb{R}^2$ to $\mathbb{R}^2$ is linear if and only if it maps the origin to the origin and equally spaced lines to equally spaced lines or points.

Exercise 1.2.1

Use Theorem 1.2.1 to explain why if $f : \mathbb{R}^2 \to \mathbb{R}^2$ rotates every point counterclockwise about the origin by 30°, there necessarily exist exist $a, b, c, d \in \mathbb{R}$ such that $f(x, y) = (ax + by, cx + dy)$ for all $(x, y) \in \mathbb{R}^2$.

Exercise 1.2.2

Show that the linear function $f(x, y) = (2x, 0)$ maps any collection of equally spaced vertical lines to a collection of equally spaced points.

on the determinant

1.3 The determinant

The *slope* of a linear function from $\mathbb{R}^1$ to $\mathbb{R}^1$ measures how it distorts length. For example, the function $f(x) = 3x$ maps any interval $[a, b]$ to the interval $[3a, 3b]$ which is three times as long. The function $g(x) = -\frac{1}{2}x$ maps any interval to an interval which is half as long, and it also flips the interval around. We can say that the absolute value of the slope of a linear function is the *factor by which lengths are transformed*, and the sign of the slope tells us whether the function reverses the real number line.

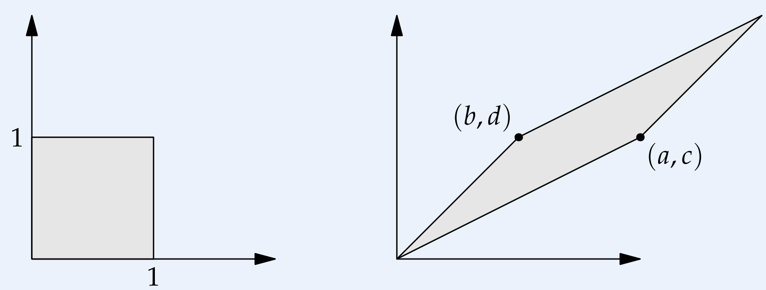
What is the corresponding idea for transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$? Can we look at a linear transformation and conveniently calculate the factor by which that transformation multiplies *areas*? Yes!

In each linear transformation picture in Section 1.2.2, the quadrilaterals on the image side of the picture are all congruent. This suggests that the linear transformation does indeed transform every area by the same factor. Taking this fact as given, it suffices for us to consider the image of a single square, which we will take to be $[0, 1]^2$, the set of points both of whose coordinates are between 0 and 1.

Example 1.3.1

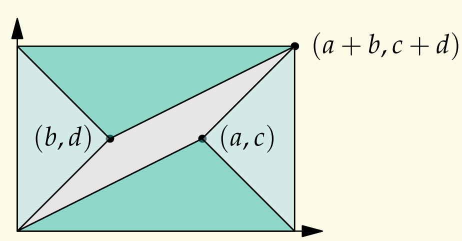
Find the area of the image of the unit square $[0, 1]^2$ under the transformation

$$f(x, y) = (ax + by, cx + dy).$$



Solution

The area of the unit square can be calculated by filling in some triangles to get a complete rectangle, as follows:



The area of the larger rectangle is $(a + b)(c + d)$, and the total area of the triangles we added is $2 \cdot \frac{1}{2}(a + b)(c) + 2 \cdot \frac{1}{2}(c + d)(b)$. Subtracting, we get that the area of the parallelogram is $ad - bc$.

Reversing the orientation of three points A, B, and C means that if these points are given in counterclockwise order around the triangle ABC, then their images are in clockwise order around the triangle they form.

We are not quite finished, however. Note that we assumed in our diagram that the line segment to from the origin to (a, c) is clockwise from the line segment from the origin to (b, d). If we switched these line segments around, the same reasoning would have given us the formula $bc - ad$. We can put this all together by saying that the factor by which areas are transformed is $\boxed{|ad - bc|}$.

We can interpret the result of Example 1.3.1: computing $ad - bc$ tells us how $f(x, y) = (ax + by, cx + dy)$ transforms areas (via its absolute value) and whether applying f to the three points $(0, 0)$, $(1, 0)$, and $(0, 1)$ reverses their orientation* (via its sign). This idea is important enough to deserve its own name. To simplify the definition, we refer to length as 1-dimensional volume and area as 2-dimensional volume.

!!!

Definition 2: Determinant

The **determinant** of a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ is the signed factor by it transforms n-dimensional volumes.

We have already figured out that the determinant of $f(x) = mx$ from $\mathbb{R}^1$ to $\mathbb{R}^1$ is the slope m, and the determinant of a function $f(x, y) = (ax + by, cx + dy)$ from $\mathbb{R}^2$ to $\mathbb{R}^2$ is given by the formula $ad - bc$.

For convenience, we sometimes represent a linear function by arranging its coefficients into grid of numbers called a *matrix*. By convention, rows correspond to coordinates of the output of the function, and columns correspond to the variables. So, for example, $f(x, y) = (ax + by, cx + dy)$ is represented by the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. So we have

$$\det[m] = m, \text{ and } \qquad \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

Exercise 1.3.1

Find the determinant of each of the following matrices, and draw the image of the unit square under the corresponding linear transformations to see that value of the determinant you computed makes sense.

(a) $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

(b) $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$

The absolute value of the determinant of a linear transformation

$$f(x, y, z) = (ax + by + cz,\ dx + ey + fz,\ hx + iy + jz)$$

is the volume of the three-dimensional shape, called a *parallelepiped* whose vertices are the images under f of the vertices of the unit cube $[0, 1]^3$:

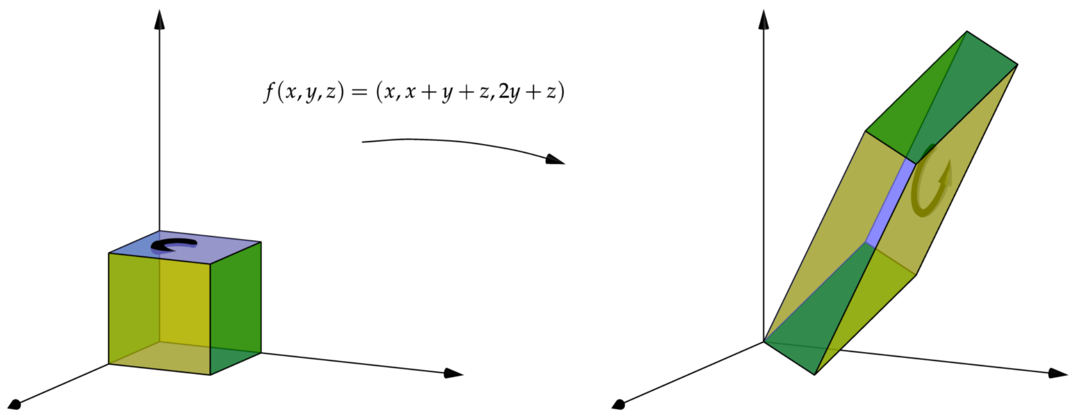


Figure 1.8 A linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$

The sign of the determinant depends on whether the orientation of a small loop drawn on a face is reversed (as shown above), from the point of view of a small person standing on the shape with their head pointing toward the outside.

The formula for the determinant of a 3×3 matrix may be derived analogously to the 2×2 case, by filling in the parallelepiped with polyhedra to get a rectangular prism with edges parallel to the coordinate axes. However, this derivation is quite messy, and you will develop a more principled approach in a linear algebra class. So let's skip straight to the formula:

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = aei - afh - bdi + bfg + cdh - ceg.$$

Unlike the formula $ad - bc$ for the 2×2 matrix, this formula is not easy to memorize. Let's abbreviate $\det[\ ]$ to $|\ |$ and write this formula as

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = +a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & i \\ f & g \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}.$$

This formula is *still* not easy to memorize, so let's break it down: each term on the right-hand side consists of a factor of $+1$ or -1 which alternates starting with $+1$, then an entry from the top row (going from left to right), then the determinant of the matrix you get when you remove the row and column of that entry from the original matrix. These smaller matrices are called *minors*, and this method of calculating the determinant is called **expansion by minors** along the first row. You can also expand by minors along any row or column (see Exercise 1.3.3 below), but if it's an even-numbered row or column, then the signs start with -1 instead of $+1$.

Exercise 1.3.2

Calculate each determinant.

(a) $\begin{vmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{vmatrix}$

(b) $\begin{vmatrix} -4 & 2 & 1 \\ 5 & 0 & 3 \\ -2 & 1 & 3 \end{vmatrix}$

Exercise 1.3.3

Expand by minors along the first *column* of this matrix, and show that you get the same result as when you expand by minors along the first row.

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

Exercise 1.3.4

Find the values of t for which the determinant of the following matrix is zero.

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

2 Vectors

on vectors

2.1 Introduction to vectors

The norm of a vector is sometimes called the **norm** or the **magnitude** of the vector

A **vector** in $\mathbb{R}^n$ is an arrow from one point in $\mathbb{R}^n$ (the *tail*) to another (the *head*). See Figure 2.1). The **length*** of a vector is the distance from the head to the tail. Two vectors are considered equivalent if the have the same length and the same direction.

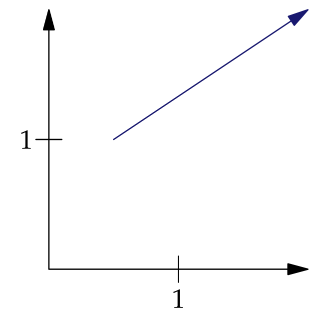


Figure 2.1 A vector in $\mathbb{R}^2$

The **components** of a vector are the coordinates of its head when it is translated so that its tail at the origin. In other words, to find the components of a vector, we subtract each coordinate of its tail from the correponding coordinate of the head. The components of the vector in Figure 2.1 are $\langle \frac{3}{2}, 1\rangle$—note that we use the pointy brackets to distinguish components of a vector from coordinates of a point. We can calculate the components by subtracting the coordinates of the tail from the coordinates of the head. Two vectors are equivalent if and only if their components are the same.*

Actually, vectors and points kind of *are* the same, since they are both specified by an ordered tuple of real numbers. Their distinction is in how we use them and visualize them.

The main things we will do with vectors are (i) add two of them together and (ii) multiply a vector by a real number (which is called a **scalar** in this context). These are defined as follows:

$$\langle u_1, u_2\rangle + \langle v_1, v_2\rangle = \langle u_1 + v_1, u_2 + v_2\rangle$$
$$c\langle u_1, u_2\rangle = \langle cu_1, cu_2\rangle.$$

These natural definitions of vector addition and scalar multiplication lead to natural geometric interpretations, as shown in Figures 2.2 and 2.3.

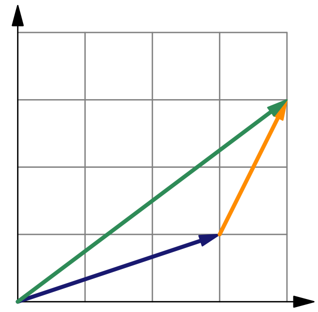


Figure 2.2 Vector addition: $\langle 3, 1\rangle + \langle 1, 2\rangle = \langle 4, 3\rangle$

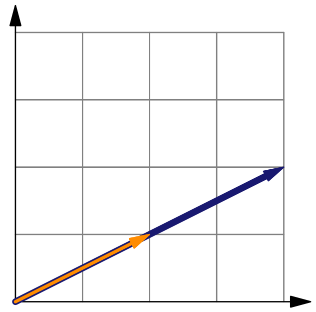


Figure 2.3 Scalar multiplication: $2\langle 2, 1\rangle = \langle 4, 2\rangle$

We typically assign names for vectors which are lowercase boldface letters, like **u** or **v**. Looking at Figure 2.3, we make the following observation

!!!

Observation 2.1.1: Parallel vectors

Two vectors **u** and **v** have the same direction if $\mathbf{u} = c\,\mathbf{v}$ for some scalar c.

That is, we use the same notation as we use for multiplication/addition of real numbers because these operations satisfy many of the same properties.

The following exercise shows that vector operations satisfy one of the properties suggested by the notation.* They satisfy several others, such as commutativity ($\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{w}$), associativity ($(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$), and so on. The basic strategy for proving all these property-verification exercises is the same: write out what each side of the equations in terms of components, and then simplify until you can see that both sides are equal.

Exercise 2.1.1

Show that scalar multiplication distributes across vector addition. In other words, show that for all $c \in \mathbb{R}$ and vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, we have

$$c(\mathbf{u}+\mathbf{v}) = c\mathbf{u}+c\mathbf{v}.$$

Exercise 2.1.2

Choose two vectors $\mathbf{u}$ and $\mathbf{v}$ with small integer coordinates and verify that $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u}-\mathbf{v}$ fit together to form a triangle.

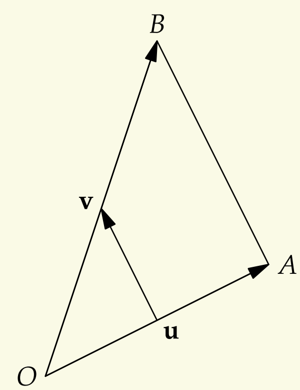
The following example shows how vectors can be applied to geometry problems.

Example 2.1.1

Use vectors to prove that the line joining the midpoints of two sides of a triangle is parallel to the third side and half its length.

Solution

Define $\mathbf{u}$ and $\mathbf{v}$ be two vectors with a common tail at one vertex O of the triangle and heads at the other two vertices A and B as shown. Then the vectors from O to the midpoints of OA and OB are $\frac{1}{2}\mathbf{u}$ and $\frac{1}{2}\mathbf{v}$, since the midpoint of a line segment is defined to be the point which is halfway between the endpoints.
Therefore, the vector $\mathbf{w}$ from one midpoint to another is $\frac{1}{2}\mathbf{v}-\frac{1}{2}\mathbf{u}$. By the distributive property, this is equal to $\frac{1}{2}(\mathbf{v}-\mathbf{u})$. The vector from A to B is $\mathbf{v}-\mathbf{u}$. Therefore, $\mathbf{w}$ has the same direction as the vector from A to B (by Observation 2.1.1) and is half as long.



Exercise 2.1.3

Use vectors to show that the diagonals of a parallelogram bisect one another.

on the dot product

2.2 The dot product

The fundamental vector operations of scalar multiplication and vector addition are not sufficient to capture information about a really important geometric concept: *angle*. So we introduce a new vector operation.

Definition 3: Dot product

The **dot product** of two three-dimensional vectors $\mathbf{u} = \langle u_1, u_2, u_3 \rangle$ and $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$ is defined by

$$\mathbf{u} \cdot \mathbf{v} = \langle u_1, u_2, u_3 \rangle \cdot \langle u_1, u_2, u_3 \rangle = u_1 v_1 + u_2 v_2 + u_3 v_3.$$

The dot product distributes across vector addition, and it is closely related to length, as shown in the following exercise. We denote by $|\mathbf{u}|$ the length of $\mathbf{u}$.

Exercise 2.2.1

Verify that $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ and that $\mathbf{u} \cdot \mathbf{u} = |\mathbf{u}|^2$.

Now we establish the relationship between the dot product and angle.

Example 2.2.1

Use the law of cosines to show that $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos \theta$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

Solution

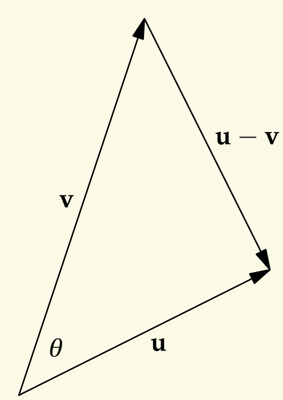
We apply the law of cosines to the triangle with sides $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} - \mathbf{v}$. We get

$$|\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2|\mathbf{u}|\mathbf{v}| \cos \theta$$

The left-hand side works out to

$$|\mathbf{u} - \mathbf{v}|^2 = (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2\mathbf{u} \cdot \mathbf{v}.$$

Subtracting these equations yields $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos \theta$.



Particularly noteworthy is the case where θ is a right angle: We say that two vectors are **perpendicular** or **orthogonal** or **normal** if they meet at a right angle.

!!!

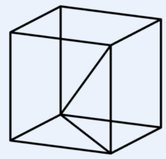
Observation 2.2.1: Perpendicular vectors

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are perpendicular if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

The following example shows how handy the relation $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos\theta$ can be.

Example 2.2.2

Find the angle between the diagonal of a cube and a diagonal of one of its faces.



Solution

The vector from the origin to the opposite corner of the cube is $\langle 1, 1, 1\rangle$. The vector from the origin to the opposite corner of the bottom face is $\langle 1, 1, 0\rangle$. Therefore, the angle is given by

$$\theta = \cos^{-1}\left(\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}\right) = \cos^{-1}\left(\frac{1+1+0}{\sqrt{1^2+1^2+1^2}\sqrt{1^2+1^2+0^2}}\right) = \boxed{\cos^{-1}\left(\frac{2}{\sqrt{6}}\right)}.$$

Exercise 2.2.2

Sketch the vectors $\mathbf{u} = \langle 4, 2\rangle$ and $\mathbf{v} = \langle -1, 2\rangle$ and show geometrically that they are perpendicular.

Then verify that the coordinate formula for dot product indeed gives $\mathbf{u} \cdot \mathbf{v} = 0$ for these two vectors.

We conclude this section by pointing out that one can begin with the geometric formula $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos\theta$ as the *definition* of the dot product, derive the distributive property of the dot product across vector addition, and then obtain the formula for the dot product in the following simple manner: define $\mathbf{i} = \langle 1, 0, 0\rangle$, $\mathbf{j} = \langle 0, 1, 0\rangle$, and $\mathbf{k} = \langle 0, 0, 1\rangle$. Then a vector $\mathbf{u} = \langle u_1, u_2, u_3\rangle$ can be written as

$$\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k},$$

and similarly for $\mathbf{v}$. Then

$$\begin{aligned}\mathbf{u} \cdot \mathbf{v} &= (u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k})(v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \\ &= u_1v_1\mathbf{i} \cdot \mathbf{i} + u_1v_2\mathbf{i} \cdot \mathbf{j} + u_1v_3\mathbf{i} \cdot \mathbf{k} + \\ &\quad u_2v_1\mathbf{j} \cdot \mathbf{i} + u_2v_2\mathbf{j} \cdot \mathbf{j} + u_2v_3\mathbf{j} \cdot \mathbf{k} + \\ &\quad u_3v_1\mathbf{k} \cdot \mathbf{i} + u_3v_2\mathbf{k} \cdot \mathbf{j} + u_3v_3\mathbf{k} \cdot \mathbf{k},\end{aligned}$$

by the distributive property. This looks like a mess, but since $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are perpendicular, six of these nine terms are zero. Furthermore, since $\mathbf{i} \cdot \mathbf{i} = 1$ and similarly for $\mathbf{j}$ and $\mathbf{k}$, we end up with $\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$, as desired.

on the cross product

2.3 The cross product

The last section introduced a vector product which reveals information about *angle*; in this section we'll see a new vector product which gives us information about *area*.

We put *determinant* in scare quotes because the matrix entries are not numbers.


A beautiful explanation of this formula

The **cross product** of $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ is defined by expanding the following 'determinant' by minors along the first row:*

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2 v_3 - u_3 v_2)\mathbf{i} - (u_1 v_3 - u_3 v_1)\mathbf{j} + (u_1 v_2 - u_2 v_1)\mathbf{k}.$$

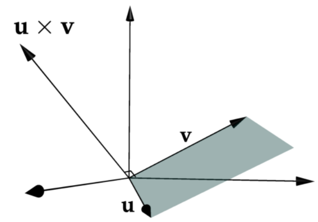


Figure 2.4 The cross product of **u** and **v**, whose length is equal to the area of the parallelogram shown

Note that the dot product of two vectors is a scalar, while the cross product of two vectors is another vector. It turns out that this vector is orthogonal to *both* of the first two.

Example 2.3.1

Confirm that $\mathbf{u} \times \mathbf{v}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$.

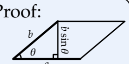
Solution

We compute

$$\langle u_2 v_3 - u_3 v_2, -(u_1 v_3 - u_3 v_1), u_1 v_2 - u_2 v_1 \rangle \cdot \langle u_1, u_2, u_3 \rangle = (u_2 v_3 - u_3 v_2) u_1 - (u_1 v_3 - u_3 v_1) u_2 + (u_1 v_2 - u_2 v_1) u_3 = 0.$$

This implies that $\mathbf{u} \times \mathbf{v}$ is orthogonal to $\mathbf{u}$. Swapping out $\mathbf{u}$ for $\mathbf{v}$, we see that $\mathbf{u} \times \mathbf{v}$ is orthogonal to $\mathbf{v}$ too.

Proof:



The following exercise provides the advertised connection to area. Recall from geometry that the area of a parallelogram with sides of length a and b meeting at an angle θ is equal to $ab \sin \theta$.*

Exercise 2.3.1

Verify that $|\mathbf{u} \times \mathbf{v}|^2 = |\mathbf{u}|^2 |\mathbf{v}|^2 - (\mathbf{u} \cdot \mathbf{v})^2$. Use this fact to show that

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

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

So to sum up: $\mathbf{u} \times \mathbf{v}$ is a vector which is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$ and whose length is equal to the area of the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$. Note that there are only two vectors satisfying both of these conditions. To determine which one is $\mathbf{u} \times \mathbf{v}$, we use the *right-hand rule*: imagine orienting your right hand so that you can curl your fingers from $\mathbf{u}$ towards $\mathbf{v}$. The direction of your thumb is the direction of $\mathbf{u} \times \mathbf{v}$.

Exercise 2.3.2

Find the volume of the parallelepiped spanned by $\langle 3, 4, 1 \rangle$, $\langle -2, 4, 0 \rangle$, and $\langle -5, 5, 2 \rangle$. (Hint: first find the area of the base, then figure out how to use a dot product to multiply by the height.)

3 Three-dimensional Geometry

3.1 Lines and planes

There are various ways to describe a line in 2D space using an equation, including point-slope form and y-intercept form. In this section we will learn the 3D analogue: equation descriptions of lines and planes in space. We begin with an example.

Example 3.1.1

Describe the line L in $\mathbb{R}^3$ passing through the points $A = (3, -4, 1)$ and $B = (2, -1, 4)$.

Solution

We can tell whether a given point (x, y, z) in $\mathbb{R}^3$ is on the line L using vectors: (x, y, z) is on L if and only if the vector from $(3, -4, 1)$ to (x, y, z) is a scalar multiple of the vector from $(3, -4, 1)$ to $(2, 1, -4)$ (see Figure 3.1). We can turn this into an equation: a point (x, y, z) is on L if and only if there exists $t \in \mathbb{R}$ such that

$$t\langle 2-3, 1-(-4), -4-1\rangle = \langle x-3, y-(-4), z-1\rangle.$$

Setting components equal, we find that (x, y, z) is on L if and only if there exists t so that

$$\begin{aligned} x &= 3 - t, \\ y &= -4 + 5t, \text{ and} \\ z &= 1 - 5t. \end{aligned} \tag{3.1.1}$$

Note that the solution above involves a new variable t; this is called a *parameter*, and the form we gave as an answer is called **parametric form**. You can imagine drawing the line by starting with $t = 0$, so that your pen begins at A, and then sweeping t through the values from 0 to 1, changing the location of your pen according to the parametric equations (3.1.1). This gives you the portion of the line between A and B. Then you can let t vary beyond 1 to get the rest of the ray past B, and you can let t vary over the negative numbers to get the part of the line on the other side of A.

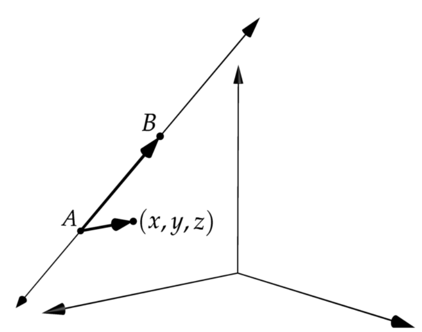


Figure 3.1 Checking whether (x, y, z) is on the line through A and B

If we didn't want to involve t, note that we could solve for t in one equation and substitute into the other two, thereby obtaining *two* equations involving x, y, and z. This makes sense: starting from the plane, imposing one equation on x and y cuts the dimension down by one and gives a line. However, starting from 3D space, we need to reduce the dimension by *two*. So we need two equations.

The procedure developed in Example 3.1.1 works in general: the line through $A = (a, b, c)$ and B has parametric form

$$\begin{aligned} x &= a + v_1 t, \\ y &= b + v_2 t \\ z &= c + v_3 t, \end{aligned}$$

Note that this representation is not unique, since we could've used B (or any other point on the line) in place of A.

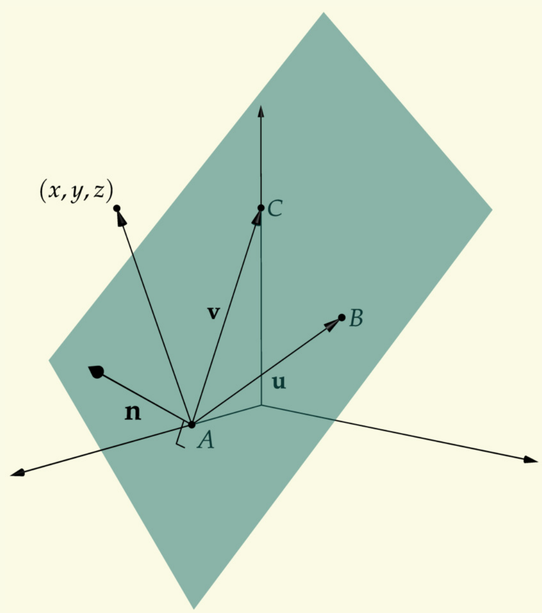
where $\langle v_1, v_2, v_3 \rangle = \overrightarrow{AB}$ is the vector from A to B.*

Example 3.1.2

Describe the plane P passing through the points $A = (1,0,0)$, $B = (0,1,1)$, and $C = (0,0,2)$.

Solution

We can tell whether (x,y,z) is on P using vectors. Define $\mathbf{u}$ and $\mathbf{v}$ to be the vectors from A to B and from A to C, respectively.



If we can find a vector $\mathbf{n}$ which is orthogonal to P, then we can say (x,y,z) is on P if and only if the vector from A to (x,y,z) is orthogonal to $\mathbf{n}$. But we can take $\mathbf{n} = \mathbf{u} \times \mathbf{v}$, since the cross product of two vectors is orthogonal to both of them. So

$$\mathbf{n} = \langle -1, 0, 2 \rangle \cdot \langle 0, -1, 1 \rangle = \langle 2, 1, 1 \rangle.$$

Now we can say that (x,y,z) is on P if and only if

$$\mathbf{n} \cdot \langle x - 1, y - 0, z - 0 \rangle = 0,$$

which simplifies to $\boxed{2x + y + z = 2}$.

!!!

Observation 3.1.1: Vector normal to a plane

A vector $\mathbf{n}$ normal to the plane $ax + by + cz = d$ can be read off from the coefficients:

$$\mathbf{n} = \langle a, b, c \rangle.$$

We can ask about the distance from a point to a point, a point to a line, a point to a plane, a line to a line, a line to a plane, or a plane to a plane.

One important type of 3D geometry problem that of finding distances between points, lines, and planes.* We define the distance between two sets to be the *minimum* distance between two points in those sets.

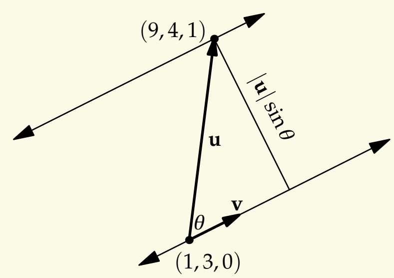
Example 3.1.3

Consider the line ℓ given by the parametric equation $(x, y, z) = (1 - 2t, 3, t)$. Find the distance from ℓ to the line m which is parallel to ℓ and which passes through the point $(9, 4, 1)$.

Solution

The parametric equations give us convenient access to a point on each of the two lines as well as a vector $\mathbf{v}$ which is parallel to both lines. So we make a figure with this information.
If we define $\mathbf{u}$ to be the vector from connecting the two given points, we can see by applying right-triangle trigonometry to the figure that the desired distance d is equal to $|\mathbf{u}| \sin\theta$. Therefore,

$$d = \frac{|\mathbf{u}||\mathbf{v}| \sin\theta}{|\mathbf{v}|} = \frac{|\mathbf{u} \times \mathbf{v}|}{|\mathbf{v}|} = \frac{\sqrt{105}}{\sqrt{5}} = \boxed{\sqrt{21}}.$$



Note the basic strategy: (i) draw a figure containing the information that the problem gives us (a schematic diagram suffices; there is no need to make it particularly precise), (ii) use right triangle trigonometry to express the desired distance terms of vectors we have, (iii) use vector formulas to calculate the desired quantity using a dot or cross product.

Exercise 3.1.1

Find the equation of the plane passing through the points $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. Find the distance from that plane to the origin.

Exercise 3.1.2

Find the distance between the planes $x + y - 2z = 3$ and $x + y - 2z = 0$.

Exercise 3.1.3

Find the distance between the lines $(x, y, z) = (2t, 1 - t, 4)$ and $(x, y, z) = (1 + t, -2t, -1 - t)$. Hint: these lines are *skew*, meaning that they are not parallel but do not intersect. Begin by using a cross product to find a vector which is perpendicular to both lines.

3.2 Motion in space

The derivative of a path $\mathbf{r}$ is the vector obtained by taking the derivative of each component.

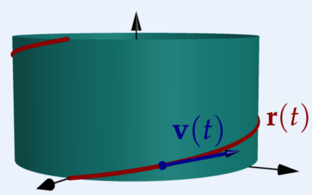
For a function from $\mathbb{R}^1$ to $\mathbb{R}^2$ or $\mathbb{R}^3$ to count as a path, we require that each of its components be continuous. For example, $\mathbf{r}(t) = (e^t, t^2 \sin t)$ is a path.

Consider a particle moving along the number line in such a way that its position at time t is given by $r(t)$. Then the velocity of the particle at time t is given by the first derivative $v(t) = r'(t)$. The velocity specifies the *speed* of the particle as well as its *direction* (left if negative, right if positive).

The same is true of a particle moving in 2D or 3D space: its location is specified by a function customarily denoted $\mathbf{r}(t)$ from $\mathbb{R}$ to either $\mathbb{R}^2$ or $\mathbb{R}^3$, and its derivative* $\mathbf{v}(t) = \mathbf{r}'(t)$ at time t tells us the speed of the particle at that time (via its length) as well as the direction. We call $\mathbf{r}$ a *path**.

Example 3.2.1

Consider a bug which crawls around a cylinder of radius 1 from $(1, 0, 0)$ to $(1, 0, 1)$ as shown:



Assuming the bug moves at constant speed and makes the whole journey in one second, find a formula for the position and velocity of the bug at time t.

Solution

We can see that the z-coordinate of the bug's position increases at a constant rate from 0 to 1 as t goes from 0 to 1, so the z-coordinate of $\mathbf{r}(t)$ is t.

For the x and y coordinates, we need a pair of functions $(x(t), y(t))$ that traces out the unit circle in one second. Recall that cosine and sine are defined to be the functions that trace out the unit circle according to angle, so we can scale them so they make it around in 1 second instead of 2π seconds:

$$(x(t), y(t)) = (\cos 2\pi t, \sin 2\pi t).$$

So all together we have

$$\mathbf{r}(t) = (\cos 2\pi t, \sin 2\pi t, t),$$

which means that

$$\mathbf{v}(t) = (-2\pi \sin 2\pi t, 2\pi \cos 2\pi t, 1).$$

Similarly, the acceleration $\mathbf{a}(t)$ of a particle whose position at time t is given by $\mathbf{r}(t)$ is defined to be $\mathbf{v}'(t) = \mathbf{r}''(t)$.

Exercise 3.2.1

An astronaut is using a rope to move in space in such a way that his position at time t is given by $\mathbf{r}(t) = (2+t)\mathbf{i} + (2+\ln t)\mathbf{j} + \left(7 - \frac{4}{t^2+1}\right)\mathbf{k}$. The coordinates of the space station doorway are $(5, 4, 9)$. When should the astronaut let go of the rope so as to drift into the doorway?

3.3 Quadric surfaces

The *graph* of an equation involving the variables x, y, z is the set of points (x, y, z) in $\mathbb{R}^3$ which satisfy the equation. For example, we have seen that the graph of $x + y + z = 1$ is a plane in $\mathbb{R}^3$. More generally, the graph of any linear equation in $\mathbb{R}^3$ is a plane. So let's step it up a notch and consider *quadratic* equations.* A graph of a quadratic equation in the variables x, y, z is called a *quadric surface*.

The 2D analogues are *conic sections*: parabolas, ellipses, and hyperbolas. These are graphs of various quadratic equations in two variables.

Perhaps the simplest quadratic equation to reason about is $x^2 + y^2 + z^2 = 1$. The left-hand side has a geometric interpretation as *the squared distance from* (x, y, z) *to the origin*. Therefore, a point (x, y, z) satisfies this equation if and only if its squared distance to the origin is 1. We have a name for the set of such points: the *sphere* of radius 1, centered at the origin.

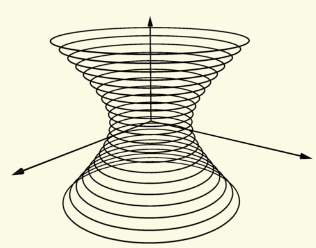
The situation is not always so simple. So here's a key idea for tackling 3D geometry problems: **slice it up**. Consider planes of the form $z =$ constant, $y =$ constant, or $z =$ constant and see what your graph looks like in these planes. Here's an archetypal example.

Example 3.3.1

Figure out what the graph of $x^2 + y^2 - z^2 = 1$ looks like.

Solution

We begin by finding all the points which satisfy this equation and the equation $z = 0$. If (x, y, z) satisfies this equation and $z = 0$, then that means that $x^2 + y^2 = 1$. Furthermore, if $x^2 + y^2 = 1$ and $z = 0$, then (x, y, z) satisfies the equation $x^2 + y^2 - z^2 = 1$. This means that the intersection of the desired graph and the line $z = 0$ is the circle of radius 1 centered at the origin.



Similarly, the intersection of the desired graph and the plane $z = 1$ is circle centered at $(0, 0, 1)$ and having radius $\sqrt{2}$. Drawing in several more of these traces*, we get a picture that looks like this: This is already a pretty clear picture of what the graph looks like: it's rotationally symmetric about the z-axis and "flares out" as you move away from the xy-plane. This graph is called a *one-sheeted hyperboloid*.

A *trace* of a figure is an intersection of that figure with a plane

Exercise 3.3.1

Sketch $\frac{z}{c} = \frac{x^2}{a^2} + \frac{y^2}{b^2}$, where $a = b = c = 1$. This is called an elliptic paraboloid.

Exercise 3.3.2

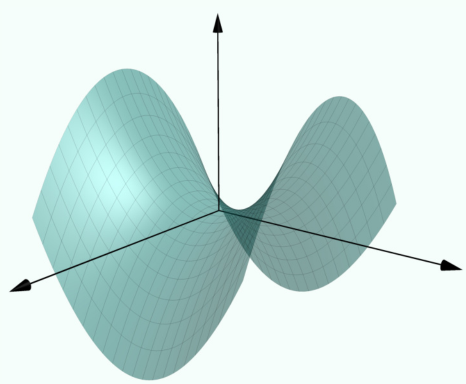
Sketch $\frac{z^2}{c^2} = \frac{x^2}{a^2} + \frac{y^2}{b^2}$, where $a = b = c = 1$. This is called an elliptic cone.

Exercise 3.3.3

Sketch the graph of $x^2 + y^2 - z^2 = -1$. This is called a two-sheeted hyperboloid.

Exercise 3.3.4

Show that the graph of $z = y^2 - x^2$ looks like the figure below. This is called a hyperbolic paraboloid.



3.4 Polar, cylindrical, and spherical coordinates

A coordinate system is a way of identifying locations using pairs or triples of real numbers. Rectangular coordinates—the ones commonly denoted (x, y) or (x, y, z)—have some nice properties, but some tasks are much more convenient in other coordinate systems.

For example, a captain at sea wishing to communicate the location of a nearby pirate ship would probably describe its location in terms of the distance r between the two ships and an angle θ (which might be given with reference to the ship's orientation or as a cardinal direction). The captain is using *polar coordinates*.

Given a point in the plane, we define r to be its distance from the origin and θ to be the angle formed between the positive horizontal axis and the vector from the origin to the point. The values r and θ are called the radial and angular polar coordinates of the point, respectively.*

Properly speaking, coordinates are *functions on the plane*. For example, $r = r(x, y) = \sqrt{x^2 + y^2}$

Exercise 3.4.1

Show that if the polar coordinates of a point (x, y) are r and θ, then we have

$$x = r\cos\theta, \quad \text{and} \quad y = r\sin\theta.$$

Correspondingly, we can coordinatize three-dimensional space by replacing either one or two spatial coordinates with an angular coordinate. The simplest way to do this is leave z the same and replace (x, y) with polar coordinates (r, θ). In other words, we define

$$\begin{aligned} r &= \text{distance to the } z\text{-axis} \\ \theta &= \text{angle of } (x, y) \text{ with respect to positive } x\text{-axis} \\ &= \text{angle of rotation about the } z\text{-axis necessary to hit the positive } yz\text{-plane} \\ z &= \text{signed distance to the } xy\text{-plane.} \end{aligned}$$

Then we can describe a point by its **cylindrical coordinates** (r, θ, z) rather than its rectangular coordinates, and therefore we can describe a solid in $\mathbb{R}^3$ by giving inequalities in the variables r, θ, and z which are all satisfied for a point if and only if that point's cylindrical coordinates satisfy them.

Example 3.4.1

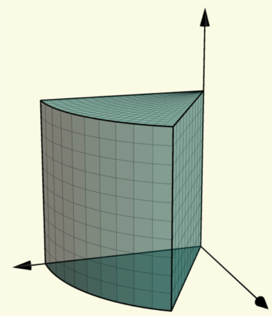
Graph* the system of cylindrical coordinate inequalities $r \le 4$, $0 \le \theta \le \pi/3$, $0 \le z \le 2$. Find the volume of the resulting region.

Familiarity with coordinate slices (Table A.2 in Appendix A.5) is helpful for graphing inequalities

Solution

The problem is asking us to find the points whose cylindrical coordinates satisfy all of the given inequalities. Such a point is less than or equal to 4 units from the z axis, has polar angle between 0 and $\pi/3$, and is between 0 and 2 units from the z-axis (and above it). The set of such points is shown to the right.

This region is one-sixth of a cylinder whose volume is $\pi r^2 h = 2\pi$, so its volume is $\boxed{\frac{\pi}{3}}$.



Exercise 3.4.2

Graph the system of inequalities $0 \leq r \leq z, \;\; \pi \leq \theta \leq 2\pi$.

Cylindrical coordinates have two distance coordinates and one angular coordinate. How can we specify a point in space using one distance coordinate and two angular coordinates? The most natural candidate for the distance coordinate is the distance from the origin. In other words, we define $\rho(x, y, z) = \sqrt{x^2 + y^2 + z^2}$. We call this coordinate ρ instead of r to distinguish it from the radial polar coordinate.

As for the angular coordinates, let's use the cylindrical coordinate θ for one of them. For the other, we measure the angle between the positive z-axis and the vector from the origin to (x, y, z). This pair of angular coordinates might be familiar: we use them to describe locations on the surface of the earth. In that context, the angle θ is called longitude and the angle ϕ is called latitude.

Note that θ varies from 0 to 2π as one loops around the z-axis. However, the angle ϕ varies only from 0 to π as one goes from the north pole to the south pole. Thus the angles θ and ϕ do not play symmetric roles*

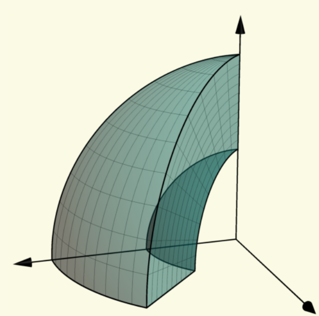
*This is because ϕ measures the angle required to rotate a vector freely so as to align with the positive z-axis, while θ measures the angle needed to rotate the vector *about the z-axis* to get to the positive half of the *xz-plane*.

Example 3.4.2

Graph the system of inequalities $\frac{1}{2} < \rho \leq 1, \;\; 0 \leq \theta \leq \pi/2, \;\; 0 \leq \phi \leq \pi/4$.

Solution

The set of points with $\rho \leq 1$ is the set of points on or inside of the sphere of radius 1 centered at the origin. Imposing the additional constraint $\rho > \frac{1}{2}$ removes the sphere of radius $\frac{1}{2}$ centered at the origin. Then the angular constraints carve out a portion of this hollowed out sphere, as shown.



Exercise 3.4.3

Find a system of inequalities in spherical coordinates to describe the portion of the unit ball above the plane $z = \frac{1}{2}$.

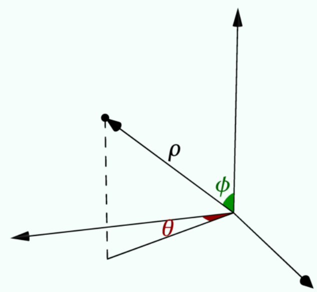
The unit ball is the set of points satisfying $x^2 + y^2 + z^2 \leq 1$.

Exercise 3.4.4

Use the given figure to show that

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

Hint: use right-triangle trigonometry to write $\sqrt{x^2 + y^2}$ and z in terms of ρ and ϕ, and then use a different right triangle to write $(x, y, 0)$ in terms of ρ, ϕ, and θ.



Exercise 3.4.5

Determine the graph of the spherical-coordinate equation $\rho = 2\cos\phi$. (Hint: multiply both sides by ρ and then switch to rectangular coordinates.)

Exercise 3.4.6

Determine the graph of $\rho = \sin\phi \sin\theta$.

Exercise 3.4.7

Sketch the set of points satisfying $1 < \rho < 2$ and $\phi < \pi/4$.

4 Multivariable Differentiation

In this chapter, we will be considering functions from $\mathbb{R}^n$ to $\mathbb{R}^1$, where $n \geq 2$. The main objectives will be to extend various important notions in single-variable calculus to the higher-dimensional setting.

4.1 Limits

Recall that $f(x)$ converges to L as $x \to a$ if $f(x)$ is as close to L as desired throughout a sufficiently small neighborhood of a. More precisely, $f(x)$ converges to L as $x \to a$ if and only if for every $\epsilon > 0$, there is $\delta > 0$ so that $f(x)$ within ϵ of L for all x satisfying $0 < |x - a| < \delta$ (see Figure 4.1).

We abbreviate "$f(x)$ converges to L as $x \to a$" to $\lim_{x\to a} f(x) = L$.

It can be helpful to think of this definition as a game against an adversary: the adversary chooses a positive real number ϵ which can be as small as they like. Then, after seeing the ϵ value, you get to choose a number $\delta > 0$, as small as you like. Finally, the adversary chooses an x value other than a in the interval $(a - \delta, a + \delta)$. If it turns out that $|f(x) - L| \geq \epsilon$, then the adversary wins. Otherwise, you win. We call this the **limit game**.

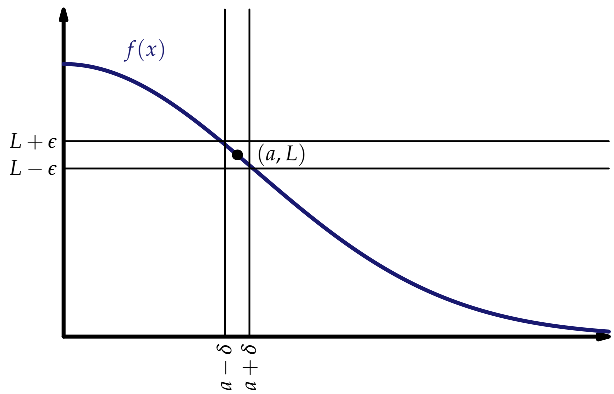


Figure 4.1 The ϵ-δ definition of a limit.

If you have a strategy for winning this game, then the limit of $f(x)$ as x approaches a exists and equals L. If the adversary has a strategy for winning, then it is not true that $\lim_{x\to a} f(x) = L$ (either because the limit does not exist, or because the limit exists and equals a number other than L).

Exercise 4.1.1

Show that $\lim_{x\to 3} \dfrac{x^2 - 9}{x - 3} = 6$ by explaining the winning strategy in the above-described game.

Exercise 4.1.2

Suppose that f is a function from $\mathbb{R}$ to $\mathbb{R}$ and that $a \in \mathbb{R}$. Show that if $f(x)$ converges to L as $x \to a$ and $f(x)$ converges to L' as $x \to a$, then $L = L'$. This fact is called *uniqueness of limits*.

We reviewed the definition of a limit for a single-variable function so we could think about how to generalize the definition for a function $f : \mathbb{R}^n \to \mathbb{R}$. For simplicity, let's take $n = 2$. What should it mean to say $\lim_{(x,y)\to(a,b)} f(x, y) = L$? The only aspect of the definition that requires revision is the part about x being within δ of a. But we can use standard Euclidean distance to compare (x, y) to (a, b). This leads to the following definition.

Definition 4: Limit of a function of two variables

We say $\lim_{(x,y)\to(a,b)} f(x,y) = L$ if and only if for every $\epsilon > 0$, there is $\delta > 0$ so that $|f(x) - L| < \epsilon$ for all (x,y) satisfying $0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta$.

Shadow here means the set of points $(x, y, f(x, y))$ where $(x, y) \in B((a, b), \delta)$.

One way to think about this definition is to consider the shadow* of the disk $B((a,b),\delta)$ on the graph (see Figure 4.2). The limit exists and equals L if for every ϵ, there exists δ small enough that this shadow lies entirely in the slab $L - \epsilon < z < L + \epsilon$.

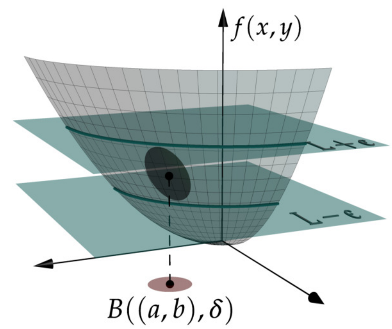


Figure 4.2 The definition of a limit for a two-variable function

A plane and a line are geometrically different in a way that has major implications for thinking about limits: there are only two directions from which to approach a point on a line, but there are infinitely many ways of approaching a point in the plane. The following example illustrates a sort of convergence failure which can occur in the higher dimensional case.

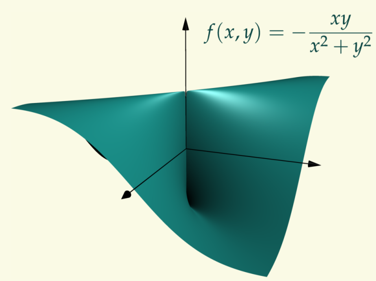
Example 4.1.1

Show that $\lim\limits_{(x,y)\to(0,0)} \dfrac{-xy}{x^2+y^2}$ does not exist.

Solution

Let's begin by graphing the function. It would appear, because of the sharp "crease" along the z-axis, that the shadow of any small disk around the origin (in the plane) includes some points which are well above the origin and some points which are well below. This suggests that the limit does not exist.

The graph suggests that the largest values occur along the line $y = -x$, while the smallest values occur along the line $y = x$. We can check this algebraically. If t is a nonzero number, no matter how small, then



$$f(t,-t) = \frac{-t(-t)}{t^2 + (-t)^2} = \frac{1}{2},$$

while

$$f(t,t) = \frac{-t(t)}{t^2+t^2} = -\frac{1}{2}.$$

Therefore, the adversary has a strategy for winning the limit game, no matter what L is. For example, if $L \geq 0$, then the adversary can choose $\epsilon = \frac{1}{4}$, and then no matter which δ you choose, the adversary can select $(x,y) = (\frac{\delta}{10}, \frac{\delta}{10})$. Then $f(x,y) = -\frac{1}{2}$, which is not within ϵ of L. Similarly, if $L < 0$, then the adversary can choose $\epsilon = \frac{1}{4}$ again and then $(x,y) = (-\frac{\delta}{10}, \frac{\delta}{10})$. Since the adversary has a winning strategy for any value of L, the limit does not exist.

Indeed, these values are simply equal to $-\frac{1}{2}$

In Example 4.1.1, there are two directions of approach along which f has different limits. Along the line $y = x$, the values of $f(x,y)$ approach* $-\frac{1}{2}$. Along the line $y = -x$, $f(x,y)$ converges to $\frac{1}{2}$. This is always an obstruction to the existence of a limit:

Exercise 4.1.3

Suppose that $\mathbf{r}_1$ and $\mathbf{r}_2$ are paths in the plane with the property that $\mathbf{r}_1(0) = (a, b)$ and $\mathbf{r}_2(0) = (a, b)$. If $\lim_{t\to 0} f(\mathbf{r}_1(t))$ and $\lim_{t\to 0} f(\mathbf{r}_1(t))$ exist and are unequal, then $\displaystyle\lim_{(x,y)\to(a,b)} f(x, y)$ does not exist.

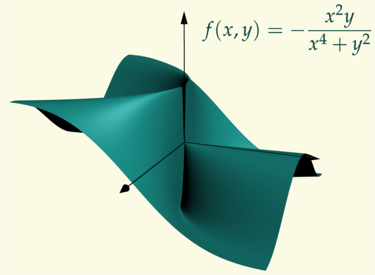
Now, suppose we know that the limits of $f(x, y)$ along every line passing through the origin exist, and that they are all equal to some common value L. Does this imply that $\lim_{(x,y)\to(0,0)} f(x, y) = L$? It seems like perhaps it should, since we've accounted for every possible angle of approach. Remarkably, this turns out not to be the case:

Example 4.1.2

Show that $\lim_{(x,y)\to(0,0)} \frac{-x^2y}{x^4+y^2}$ does not exist, even though the limits along every line through the origin exist and are equal.

Solution

We begin by checking the limit along the line $\mathbf{r}(t) = (t\cos\theta, t\sin\theta)$ (which is the line passing through the origin as well as the point on the unit circle whose angle with respect to the positive x-axis is θ). We find



$$\begin{aligned} f(t\cos\theta, t\sin\theta) &= \frac{-t^3\cos\theta\sin\theta}{t^4\cos^4\theta + t^2\sin^2\theta} \\ &= \frac{-t\cos\theta\sin\theta}{t^2\cos^4\theta + \sin^2\theta}. \end{aligned}$$

Since we're considering each value of θ individually, the $\cos\theta$ and $\sin\theta$ factors are constants. So we see that the numerator converges to 0 and the denominator converges to $\sin^2\theta$. Therefore, as long as $\sin\theta \neq 0$, we have $\lim_{t\to 0} f(t\cos\theta, t\sin\theta) = 0/\sin^2\theta = 0$. However, if $\sin\theta = 0$, then $f(t\cos\theta, t\sin\theta) = 0$ for all t, so $\lim_{t\to 0} f(t\cos\theta, t\sin\theta) = 0$ in that case too.

However, note that if we consider the limit along the parabolic path $\mathbf{r}(t) = (t, -t^2)$, we get

$$f(t, t^2) = \frac{t^2(-t^2)}{t^4 + (-t^2)^2} = \frac{1}{2}$$

Therefore, the limit along this path is equal to $-\frac{1}{2}$. Thus there are two paths (this one, as well as any straight-line path through the origin) along which f has different limits. Therefore, the limit of $f(x, y)$ as $(x, y) \to (0, 0)$ does not exist.

Note: this makes sense graphically, because this function also has a crease along the z-axis. But now we have to follow a parabolic path to travel along the top "ridge" and realize a limiting value other than zero.

With the notion of a multidimensional limit in hand, we can define continuity the same as in the one-dimensional case.

Definition 5

Suppose $n \geq 2$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is continuous at a point in $\mathbb{R}^n$ if and only if the limit of f exists at that point and equals the value of the function there.
A function $f : \mathbb{R}^n to \mathbb{R}$ is said to be continuous if it is continuous at every point in $\mathbb{R}^n$.

More generally, a function $f : D \to \mathbb{R}$—where $D \subset \mathbb{R}^n$—is said to be continuous if it is continuous at each point in its domain D. The following theorem gives us some tools for establishing continuity.

Theorem 4.1.1: Continuous functions

1. If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and $g : \mathbb{R} \to \mathbb{R}$ is continuous, then $g \circ f : \mathbb{R}^n \to \mathbb{R}$ is continuous.
2. A sum or product of continuous functions is continuous.
3. The "coordinate-extracting" functions $f(x, y, z) = x$, $f(x, y, z) = y$, etc., are continuous.

Example 4.1.3

Show that $\displaystyle\lim_{(x,y,z)\to(0,0,0)} \left(e^{\sin x} + \frac{xyz}{1 + x^2z^2} \right) = 1.$

Solution

We begin by showing that $e^{\sin x} + \frac{xyz}{1+x^2z^2}$ is continuous. Since $e^{\sin x}$ is a composition of continuous functions:

$$(x, y, z) \mapsto x \mapsto \sin x \mapsto e^{\sin x},$$

it's continuous by Theorem 4.1.1. Similarly, $\frac{xyz}{1+x^2z^2}$ is continuous wherever $1 + x^2z^2 \neq 0$, which is everywhere since $(xz)^2 \geq 0$. Finally, the sum of two continuous functions is continuous, so $e^{\sin x} + \frac{xyz}{1+x^2z^2}$ is continuous.

Since the function above is continuous, its limit at each point is equal to its value at that point. So we substitute $x = y = z = 0$ and find the that value of the function at the origin is $e^0 + \frac{0}{1+0} = 1$.

We conclude this section with one more tool for showing that a limit *does* exist: **polar coordinates**.

Example 4.1.4

Show that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^6y}{x^4 + y^4} = 0.$

Solution

For any point (x, y), let's define r and θ to be the polar coordinates of that point and calculate

$$\frac{x^6y}{x^4 + y^4} = \frac{(r^6 \cos^6 \theta)(r \sin \theta)}{r^4 \cos^4 \theta + r^4 \sin^4 \theta} = r^3 \left(\frac{\cos^6 \theta \sin \theta}{\cos^4 \theta + \sin^4 \theta} \right).$$

This means that the expression is never less than $-C$ or greater than C.

Actually, $C = 0.34$ works.

Now note that the expression $\frac{\cos^6\theta\sin\theta}{\cos^4\theta+\sin^4\theta}$ is continuous over $[0, 2\pi]$ and therefore bounded in absolute value* by some constant* C.

We can use this observation to describe a winning strategy in the limit game. Whatever ϵ is selected by the adversary, we choose δ to be $\sqrt[3]{\frac{\epsilon}{C}}$. Then, no matter which (x, y) pair the adversary selects, the fact that the polar coordinate r of (x, y) has to be less than $\sqrt[3]{\frac{\epsilon}{C}}$ implies that, regardless (x, y)'s θ value, we have

$$|f(x, y)| < \left(\sqrt[3]{\frac{\epsilon}{C}}\right)^3 C = \epsilon,$$

as desired.

Exercise 4.1.4

Show that $\lim_{(x,y)\to(0,0)} \frac{x^3 + y^3}{x^2 + y^2} = 0$.

Exercise 4.1.5

Consider the function f defined by $f(x, y) = \frac{x-y}{x^3-y}$ whenever $y \neq x^3$, and $f(x, y) = 1$ when $y = x^3$. Show that f is not continuous at $(1, 1)$. Evaluate the limits along along $x = 1$ and along $y = 1$.

on partial derivatives

4.2 Partial derivatives

Suppose f is a function from $\mathbb{R}$ to $\mathbb{R}$. The derivative f' of f is the answer to the question "how does $f(x)$ change when x changes just a little?" More precisely, if $a \in \mathbb{R}$, we define

$$f'(a) = \lim_{h\to 0} \frac{\overbrace{f(a+h) - f(a)}^{\text{how much } f \text{ changes}}}{\underbrace{h}_{\text{how much the input changes}}}$$

This means that if we know $f'(a)$, then we can estimate $f(a+h) - f(a)$ for h very small:

$$f(a+h) - f(a) \approx hf'(a).$$

So the derivative measures **how sensitive $f(x)$ is to small changes in x.**

What is the most natural corresponding idea for the derivative at some point (a, b) of a function f from $\mathbb{R}^2$ to $\mathbb{R}$? We were only able to adjust a value $x \in \mathbb{R}$ by increasing or decreasing it a little. A point in $\mathbb{R}^2$, by contrast, can be moved in any direction. Two directions are particularly easy to study: (i) move x a little while holding y fixed, and (ii) move y a little while holding x fixed. Accordingly, we define **partial**

derivatives

∂_x is read "partial x". Also, the role of x here is purely as a label that means "with respect to the first coordinate". It does not represent a number, as the symbol x usually does.

$$(\partial_x f)(a,b) = \lim_{h\to 0} \frac{f(a+h,b) - f(a)}{h}, \text{ and}$$
$$(\partial_y f)(a,b) = \lim_{h\to 0} \frac{f(a,b+h) - f(a)}{h}.$$

Calculating partial derivatives is easy because *you already know how to do it*. Since one of the two variables is being held constant, we are effectively taking a derivative with respect to a single-variable function.

Example 4.2.1

Find the partial derivatives of $f(x,y) = e^x \sin(xy)$ at $(x,y) = (1,0)$.

Solution

If you have difficulty getting used to holding a variable constant, consider replacing it with some number like 17; then substitute back at the end

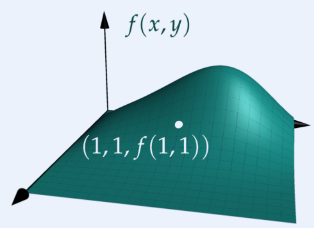
We can find the partial derivative with respect to x at *any* point (x,y) by treating y as constant and applying single-variable differentiation rules:*

$$(\partial_x f)(x,y) = e^x y \cos(xy) + e^x \sin(xy)$$
$$(\partial_y f)(x,y) = x \cos(xy)\, e^x$$

So the partial derivatives at $(1,0)$ with respect to x and y are 0 and e, respectively.

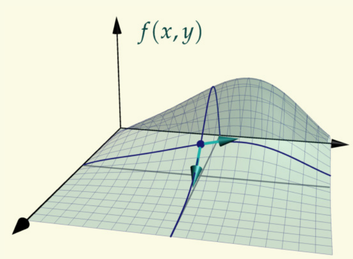
Example 4.2.2

Consider the function f whose graph is shown. Determine the sign of $(\partial_x f)(1,1)$ and the sign of $(\partial_y f)(1,1)$.



Solution

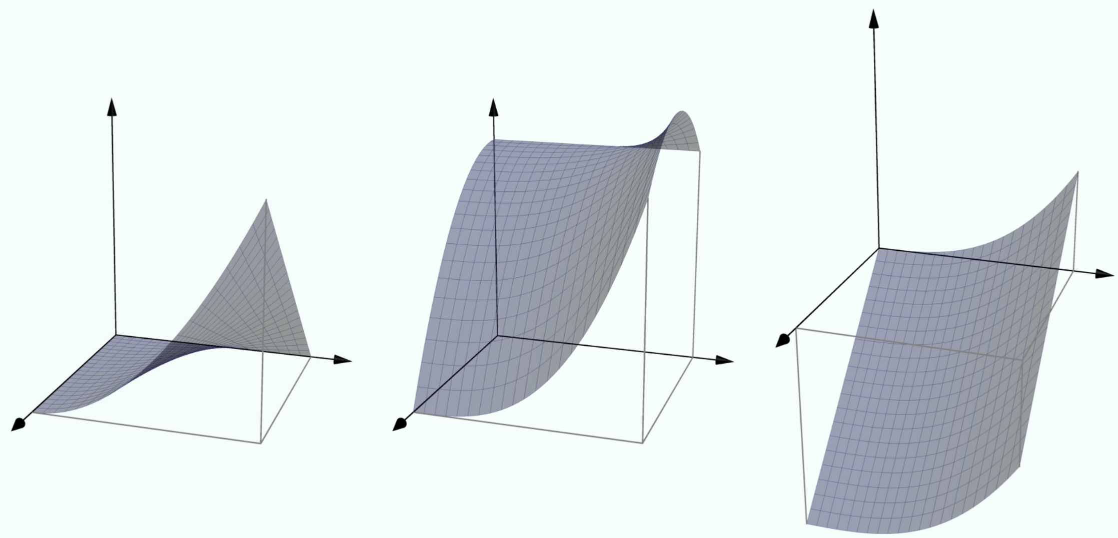
If we increase x a little while holding y constant, then f decreases. Therefore, $(\partial_x f)(1,1) < 0$. If we increase y a little while holding x constant, then f increases. Therefore, $(\partial_y f)(1,1) > 0$.
Graphically, the partial derivative with respect to x at a point is equal to the slope of the trace of the graph in the "$y =$ constant" plane passing through that point. Similarly, the partial derivative with respect to y at a point is equal to the slope of the trace of the graph in the "$x =$ constant" plane passing through that point.



Thus we can think of partial derivatives as an application of our "slice it up" strategy for understanding three dimensional objects through two dimension traces

Exercise 4.2.1

The following three graphs represent a function f and its two partial derivatives $\partial_x f$ and $\partial_y f$, in some order. Which is which?



The following theorem says that order doesn't matter when successively taking partial derivatives.

Theorem 4.2.1: Clairaut's theorem

Suppose $f : D \to \mathbb{R}$, where D is a disk in $\mathbb{R}^2$. If $\partial_x \partial_y f$ and $\partial_y \partial_x f$ exist and are continuous, then $\partial_x \partial_y f = \partial_y \partial_x f$ throughout D.

Exercise 4.2.2

Verify the conclusion of Clairaut's theorem for $f(x, y) = e^{xy} \sin y$.

on linear approximation

4.3 Linear approximation

The following example shows that partial derivatives don't tell the whole story when it comes to differentiating functions of multiple variables.

Example 4.3.1

Consider the function f for which $f(0,0) = 0$ and $f(x,y) = -\frac{xy}{x^2+y^2}$ for all $(x,y) \neq (0,0)$. Show that both partial derivatives of f at the origin are equal to zero.

Solution

If we move x a little from $x = 0$ while holding $y = 0$ fixed, the value of f doesn't change at all. Therefore,

$$(\partial_x f)(0,0) = \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h \to 0} \frac{0}{h} = \lim_{h \to 0} 0 = 0.$$

The same is true for the partial derivative with respect to y.

However, recall from Example 4.1.1 that the function in Example 4.3.1 isn't even continuous at the origin! We haven't said yet what it is required for a function of two variables to be considered differentiable, but whatever the definition, we surely cannot allow functions which aren't continuous.

This shouldn't be surprising: the partial derivatives only look at the behavior of the function along two lonely slices. A good definition of differentiability at (a, b) should account for how the function behaves in every direction around (a, b).

Another perspective on differentiability in the single-variable context is that *differentiable functions are the ones which are well-approximated by linear functions*:

Theorem 4.3.1

A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ if and only if there exists a linear function $L(x) = c_0 + c_1(x - a)$ such that

$$\lim_{x \to a} \frac{f(x) - L(x)}{|x - a|} = 0.$$

This perspective on differentiability turns out to generalize very nicely to functions of multiple variables. Let's make it a definition.

Definition 6: Differentiability for a function of two variables

A function $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at $(a, b) \in \mathbb{R}^2$ if and only if there exists a linear function $L(x, y) = c_0 + c_1(x - a) + c_2(y - b)$ such that

$$\lim_{(x,y) \to (a,b)} \frac{f(x, y) - L(x, y)}{\sqrt{(x - a)^2 + (y - b)^2}} = 0.$$

Lots of functions are differentiable. The following theorem establishes a handy way to check differentiability.

Theorem 4.3.2: Criterion for differentiability

If the partial derivatives $\partial_x f$ and $\partial_y f$ exist in some disk centered at (a, b) and are continuous at (a, b), then f is differentiable at (a, b).

The most common situation is that partial derivatives exist and are continuous everywhere, in which case Theorem 4.3.2 implies that f is differentiable everywhere.

Example 4.3.2

Show that $f(x,y) = e^{xy}\sin(x^2+y^2)$ is differentiable at every point in $\mathbb{R}^2$.

Solution

We can take partial derivatives of f with respect to both x and y and get functions which are built from x and y using addition/multiplication as well as the continuous functions $x \mapsto e^x$ and $x \mapsto \sin x$. Therefore, the partial derivatives exist and are continuous everywhere. Thus Theorem 4.3.2 implies that f is differentiable everywhere.

In the denominator we replaced $|x-a|$, whose geometric meaning is the distance from x to a on the number line, with the formula for the distance from (x,y) to (a,b) in the plane.

Graphically, Definition 6 says that a function is differentiable at (a,b) if we can draw a plane which is tangent* to the graph of f at the point $(a,b,f(a,b))$.

This is a little bit of a lie, because to be true strictly speaking it would require that we define what it means to be *tangent* in terms of Definition 6, which would be circular. So take this as intuition only.

In Theorem 4.3.1, the coefficients of the approximating function L are the value of the function f at a and the derivative of f at a. The coefficients in Definition 6 are also named quantities: as suggested by Figure 4.3, c_1 is the value of the function at (a,b) and c_1 and c_2 are the two partial derivatives at (a,b):

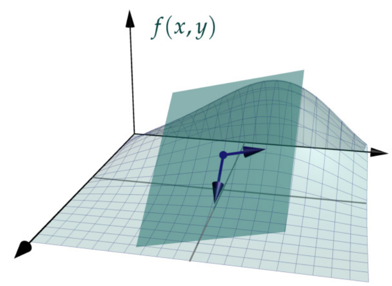


Figure 4.3 A plane tangent to the graph of a function f

Theorem 4.3.3: Linear Approximation

If $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at $(a,b) \in \mathbb{R}^2$, then

$$\lim_{(x,y)\to(a,b)} \frac{f(x,y) - \overbrace{\left[f(a,b) + (\partial_x f)(a,b)(x-a) + (\partial_y f)(a,b)(y-b)\right]}^{L(x,y)}}{\sqrt{(x-a)^2+(y-b)^2}} = 0.$$

Let's see how this theorem can be used numerically.

Example 4.3.3

Consider the function $f(x,y) = \frac{e^{xy}}{e(1+x^2)}$. Use a tangent plane to approximate $f(0.99, 0.98)$.

Solution

Noticing that $(0.99, 0.98)$ is very close to $(1,1)$, we differentiate $f(x,y)$ with respect and with respect

The bar notation means "substitute"

to y and find that*

$$(\partial_x f)(1,1) = \left(\frac{ye^{xy}}{e(x^2+1)} - \frac{2\,xe^{xy}}{e\,(x^2+1)^2} \right)\Bigg|_{(x,y)=(1,1)} = 0.$$

$$(\partial_y f)(1,1) = \frac{xe^{xy}}{e(x^2+1)}\Bigg|_{(x,y)=(1,1)} = \frac{1}{2}.$$

The actual value is 0.490197...

Therefore, $f(0.99, 0.98) \approx f(1,1) + 0(0.99-1) + \frac{1}{2}(0.98-1) = \frac{1}{2} + \frac{1}{2} \cdot (-\frac{1}{50}) = 0.49.$

on local extrema

4.4 Multivariable optimization

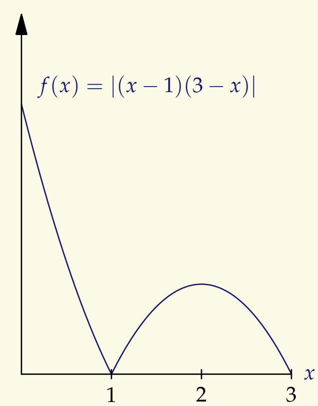
The following problem is a typical example of a single-variable optimization problem.

Example 4.4.1

Find the maximum and minimum of $f(x) = |(x-1)(3-x)|$ over the interval $[0,3]$.

Solution

Since f is continuous over the closed and bounded interval $[0,3]$, we know by the extreme value theorem that it has a maximum and a minimum value over $[0,3]$. Furthermore, these extrema must be realized at either a *critical point*, at which f is either not differentiable or has derivative zero, or else at an endpoint of the interval.
We check that f is not differentiable at 1 or 3 (see the graph). Also, we can solve $f'(x) = 0$ to find that f has a horizontal tangent line at $x = 2$.
Finally, we can check the values of f at the endpoints 0 and 3, as well as the critical points strictly between them, namely 1 and 2. We find that the maximum value is $f(0) = \boxed{3}$, and the minimum value is $\boxed{0}$, which occurs at $x = 1$ and at $x = 3$.



How does this story change when we consider a function of multiple variables? For concreteness, let's suppose $D = [0,1]^2$ and that $f : D \to \mathbb{R}$ is a continuous function. Consider the graph of the function

$$f(x,y) = -x^2 - y^2 + x + \frac{2}{3}y + \frac{23}{36},$$

shown in Example 4.4.2. We can see that any extremum must occur either (i) at a point somewhere on the boundary of the square, or (ii) a point inside the square where the tangent plane is parallel to the xy-plane (or where the function is not differentiable).

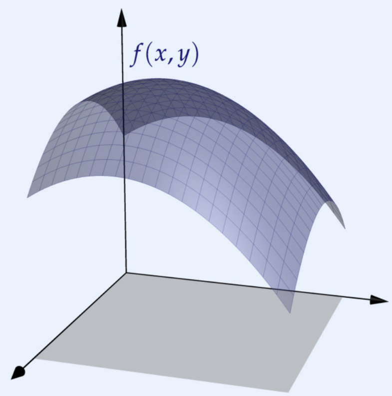
The notation ∂D means "the boundary of D", which is the set of points $p \in \mathbb{R}^2$ such that any small disk centered at p includes points in D and points not in D.

This observation gives us a strategy for finding the extrema of a function $f : D \to \mathbb{R}$, where $D \subset \mathbb{R}^2$: (i) set both partial derivatives of f equal to 0 and solve to find critical points inside D (also include any points where f is not differentiable), and (ii) find the extreme values of f on* ∂D.

We can see that this is considerably more complicated that the single-variable optimization: we have to solve a system of equations to find interior critical points, and we have to find any boundary critical points as well. If D is a rectangle, for example, then finding the extrema of f on the boundary of D boils down to doing four single-variable optimization problems (one for each side of the square). Let's see how this works out for the function shown above.

Example 4.4.2

Find the extreme values of the function $f(x,y) = -x^2 - y^2 + x + \frac{2}{3}y + \frac{23}{36}$ over the square $[0,1]^2$.



Solution

We begin by finding the critical points inside the square. We find

$$(\partial_x f)(x,y) = -2x + 1$$
$$(\partial_y f)(x,y) = -2y + \frac{2}{3}.$$

These quantities are both equal to zero only when $x = \frac{1}{2}$ and $y = \frac{2}{3}$. So $(1/2, 2/3)$ is the only critical point inside the square.

To optimize f along the $x = 0$ side, we look at

$$f(0,y) = -y^2 + \frac{2}{3}y + \frac{23}{36},$$

which has a critical point at $y = 1/3$. So $(0, 1/3)$ is a **boundary critical point**, and we should also check the two endpoints $(0,0)$ and $(0,1)$. Similarly, for the other three sides, we identify the points $(1,1/3)$, $(1/2,0)$, $(1/2,1)$, as boundary critical points as well as the other two corners $(1,1)$ and $(1,0)$. So, all together:

(x,y)	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$	$(0,1/3)$	$(1,1/3)$	$(1/2,0)$	$(1/2,1)$	$(1/2,1/3)$
$f(x,y)$	$23/36$	$23/36$	$11/36$	$11/36$	$3/4$	$3/4$	$8/9$	$5/9$	1
$f(x,y)$	0.62	0.62	0.31	0.31	0.75	0.75	0.88	0.56	1

So the maximum value is $\boxed{1}$ and the minimum value is $\boxed{\frac{11}{36}}$.

Exercise 4.4.1

Find the maximum value of $f(x,y) = 10x^2y - x$ over the closed triangle with vertices $(0,0)$, $(1,0)$, and $(0,1)$.

4.5 Directional derivative and gradient

on directional derivatives and the gradient

That is: up/down, left/right.

The two partial derivatives of a function $f : \mathbb{R}^2 \to \mathbb{R}$ tell us how f changes when (x,y) is wiggled a bit, but only in the four cardinal directions.* What about all the other directions? Suppose that $\mathbf{u}$ is a **unit vector** in $\mathbb{R}^2$, meaning that its length is 1. (see Figure 4.4).

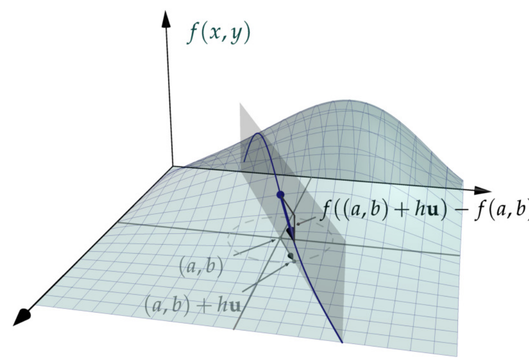


Figure 4.4 The derivative of f in the direction $\mathbf{u}$

Definition 7: Directional derivative

The **directional derivative** of a function $f : \mathbb{R}^2 \to \mathbb{R}$ in the direction $\mathbf{u} \in \mathbb{R}^2$ is defined by

$$D_{\mathbf{u}}(f)(a,b) = \lim_{h\to 0} \frac{f((a,b)+h\mathbf{u}) - f(a,b)}{h}.$$

In other words, move (x,y) a small distance h in the $\mathbf{u}$ direction, measure how much f changed, and then divide h.

If the partial derivatives of f exist and are continuous around (a,b), then we can work out the derivative in the $\mathbf{u} = \langle u_2, u_2\rangle$ direction in terms of the partial derivatives at (a,b) by breaking down a $\mathbf{u}$-step into a $\langle u_1, 0\rangle$ step and a $\langle 0, u_2\rangle$ step: the value of f changes by approximately $u_1 h(\partial_x f)(a,b)$ as we change the input value from (a,b) to $(a+u_1h, b)$, and then by $u_2h(\partial_y f)(a+u_1h,b)$ as we move from $(a+u_1h,b)$ to $(a+u_1h, b+u_2h)$. But $(\partial_y f)(a+u_1h,b)$ is approximately $(\partial_y f)(a,b)$ for small h, since $\partial_y f$ is continuous around (a,b). This leads to the following theorem.

Theorem 4.5.1: Directional derivative formula

If f is differentiable at (a,b) and $\mathbf{u} \in \mathbb{R}^2$, then

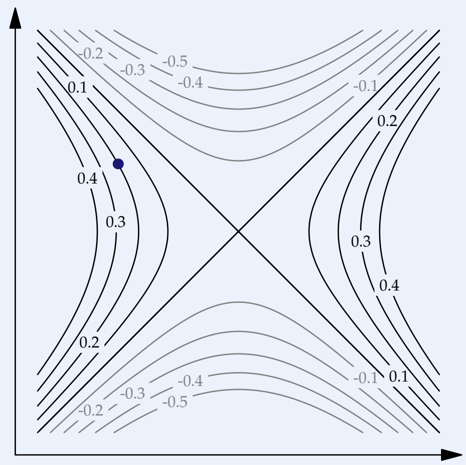
$$D_{\mathbf{u}} f(a,b) = (\partial_x f)(a,b)u_1 + (\partial_y f)(a,b)u_2 = (\nabla f)(a,b)\cdot \mathbf{u},$$

where $\nabla f = \langle \partial_x f, \partial_y f\rangle$.

The quantity ∇f introduced in Theorem 4.5.1—the vector partial derivatives of f—is called the **gradient** of f. Observe that the directional derivative of f in the $\mathbf{u}$ direction is equal to $\nabla f \cdot \mathbf{u} = |\nabla f|\cos\theta$, where θ is the angle between ∇f and $\mathbf{u}$. Since $\cos\theta$ is maximized when $\theta = 0$, we see that **the gradient of f at** (a,b) **is f's direction of maximum increase at** (a,b). Furthermore, the direction opposite to the gradient is the direction of maximum decrease, and f has zero derivative in any direction orthogonal to the gradient.

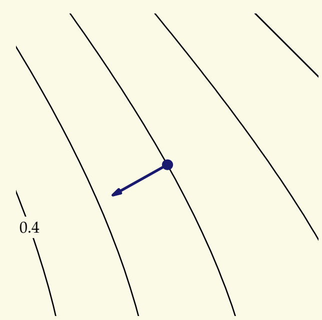
Example 4.5.1

Some of the level curves of a function $f(x,y)$ are shown. Sketch the direction of the gradient at the marked point.



Solution

The key idea here is that a function neither increases or decreases along its level curve. Therefore, f has zero directional derivative in the direction of any line tangent the level curve passing through a given point. This means that the **gradient of f is orthogonal to f's level curve** at any given point. So the gradient looks like the figure shown (zoomed in).



More precisely, the gradient at each point is the vector of partial derivatives of f at that point. Thus the gradient is actually a function from $\mathbb{R}^3$ to $\mathbb{R}^3$.

The gradient of a function from $\mathbb{R}^3$ to $\mathbb{R}$ is also defined to be the vector of partial derivatives*: $\nabla f = \langle \partial_x f, \partial_y f, \partial_z f \rangle$. The formula $D_{\mathbf{u}} f = \nabla f \cdot \mathbf{u}$ holds for differentiable $f : \mathbb{R}^3 \to \mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^3$.

Example 4.5.2

Find the equation of a plane tangent to the ellipsoid $x^2 + y^2 + 2z^2 = 4$ at the point $(1,1,1)$.

Solution

The ellipsoid is a level set of the function $f(x,y,z) = x^2 + y^2 + 2z^2$. Since the gradient of a function at a point is orthogonal to the level set of the function at that point, it follows that the vector

$$(\nabla f)(1,1,1) = \langle 2x, 2y, 4z \rangle|_{(x,y,z)=(1,1,1)} = \langle 2,2,4 \rangle$$

is orthogonal to the desired plane. So a point (x,y,z) is on the plane if and only if the vector from

$(1,1,1)$ to (x,y,z) is orthogonal to $\langle 2,2,4\rangle$. Thus the equation of the plane is

$$\langle 2,2,4\rangle \cdot \langle x-1, y-1, z-1\rangle = 0 \implies \boxed{x+y+2z=4}.$$

Exercise 4.5.1

Show that if f is a differentiable function with a local maximum at some point p inside its domain, then $(D_{\mathbf{u}}f)(p) = 0$ for any vector $\mathbf{u}$.

on the chain rule

4.6 The multivariable chain rule

The basic idea of the chain rule is that when considering how $f(g(t))$ changes when we increase t by some small amount h, we can note that $g(t)$ changes by approximately $hg'(t)$, and that change in the input to f induces a change of

$$\left(\overbrace{hg'(t)}^{\text{change in input to } f}\right)\left(\overbrace{f'(g(t))}^{\text{sensitivity of } f \text{ to change in input}}\right)$$

in the value of $f(g(t))$.

The simplest multivariable generalization of this idea is make a function from $\mathbb{R}$ to $\mathbb{R}$ by composing function from $\mathbb{R}$ to $\mathbb{R}^2$ with a function $f : \mathbb{R}^2 \to \mathbb{R}$. Let's look at an example.

Example 4.6.1

Suppose $f(x,y) = \sin xy \cos y$ and $\mathbf{r}(t) = (e^t, t^2)$. Find the derivative of $f \circ \mathbf{r}$.

Solution

We can calculate directly

$$(f \circ \mathbf{r})(t) = f(\mathbf{r}(t)) = \sin(t^2e^t)\cos t^2.$$

So the desired derivative is

$$\begin{aligned}\cos(t^2e^t)\left[t^2e^t + 2te^t\right]\cos t^2 - \sin(t^2e^t)2t\sin t^2 \\ = t^2e^t\cos(t^2e^t)\cos t^2 + 2te^t\cos(t^2e^t)\cos t^2 - 2t\sin(t^2e^t)\sin t^2.\end{aligned}$$

The multivariable chain rule gives as an alternative approach which takes advantage of partial derivatives. Let's write $\mathbf{r}(t) = \langle r_1(t), r_2(t)\rangle$. When we change t by h, the value of $f(\mathbf{r}(t))$ changes as follows:

$$f\left(\overbrace{r_1(t)}^{\text{changes by } hr_1'(t)}, \overbrace{r_1(t)}^{\text{changes by } hr_2'(t)}\right)$$

The change of $hr_1'(t)$ in the first argument induces a change of $hr_1'(t)(\partial_x f)(\mathbf{r}(t))$ in the value of f, while the change of $hr_2'(t)$ in the second argument induces a change of $hr_2'(t)(\partial_y f)(\mathbf{r}(t))$. Dividing by h and taking $h \to 0$ gives the following theorem.

Theorem 4.6.1: Multivariable chain rule

If $f : \mathbb{R}^2 \to \mathbb{R}$ and $\mathbf{r} = \langle r_1, r_2 \rangle : \mathbb{R} \to \mathbb{R}^2$, are differentiable, then

$$(f \circ \mathbf{r})'(t) = (\partial_x f)(\mathbf{r}(t))r_1'(t) + (\partial_y f)(\mathbf{r}(t))r_2'(t). \tag{4.6.1}$$

You can write (4.6.1) using the more suggestive notation

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

where x and y represent r_1 and r_2. Although this formula is more memorable, it does involve some abuse of notation: the symbols x and y are being used* as independent variables (in the partial derivative expressions) and as function names (in dx/dt and dy/dt). Also, on the left-hand side f looks like it's being treated as a function of a single variable; actually $f \circ \mathbf{r}$ is the single-variable function one has in mind here.

using the same symbol to mean different things (relying on context to distinguish) is called *overloading*

Exercise 4.6.1

Verify that applying the multivariable chain rule to Example 4.6.1 gives the same result we found by calculating that derivative directly.

Exercise 4.6.2

Find the derivative with respect to t of the function $g(t) = t^t$ by writing the function as $f(x(t), y(t))$ where $f(x, y) = x^y$ and $x(t) = t$ and $y(t) = t$.

4.7 Optimization with Lagrange multipliers

Consider the function

$$f(x, y) = -x^2 - y^2 + x + \frac{2}{3}y + \frac{23}{36},$$

which we optimized over the square $[0, 1]^2$ in Example 4.4.2. In that case, we identified possible extreme values on the boundary of the square by doing a single-variable optimization along each edge of the square. But suppose that we want to find the maximum and minimum values of f over a disk D (see Figure 4.5)? Let's consider the disk D of radius $\frac{1}{2}$ centered at $\left(\frac{1}{2}, \frac{1}{2}\right)$.

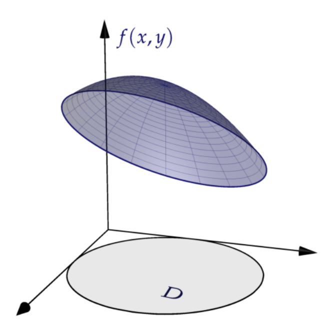


Figure 4.5 The graph of a function f defined on a disk D

We could actually take a similar approach to this problem. We can parametrize* the boundary of the disk as

To *parametrize* a curve means to find a path which traces it out

$$\mathbf{r}(t) = \left\langle \tfrac{1}{2} + \tfrac{1}{2}\cos t, \tfrac{1}{2} + \tfrac{1}{2}\sin t \right\rangle, \quad 0 \le t < 2\pi.$$

The minimum is $0.\overline{5}$

Then the single-variable function $t \mapsto f(\mathbf{r}(t))$ can be optimized over $[0, 2\pi]$ using the standard single-variable technique (as in Example 4.4.1).

However, this approach is limited because it requires a parametrization of the boundary of D, which is not always convenient. Suppose that ∂D is specified as a level set of some function $g : \mathbb{R}^2 \to \mathbb{R}$. For example, the circle in Figure 4.5 is a level set $\left\{(x, y) : g(x, y) = \frac{1}{2}\right\}$ of the function

$$g(x, y) = \left(x - \tfrac{1}{2}\right)^2 + \left(y - \tfrac{1}{2}\right)^2.$$

Let's derive an approach to finding the extreme values on the boundary which begins with the functions f (the *objective* function) and g (the *constraint* function).

Imagine a bug moving around on the edge of the graph in Figure 4.5. How can it tell that it is *not* at a maximum or minimum? One approach is to calculate the gradient of f at its location. If the gradient of f is not orthogonal to ∂D, then the value of the function can be increased by sliding a bit in one direction* and can be decreased by sliding a bit in the opposite direction. So, for example, in Figure 4.6, a bug at the point p could increase the value of f at its location by moving slightly clockwise and decrease the value of f by moving slightly counterclockwise around ∂D.

Specifically, the direction in which

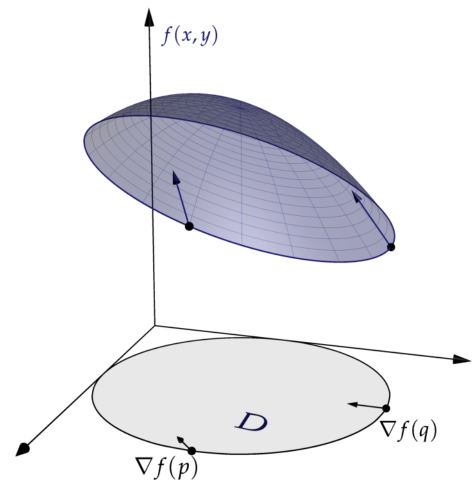


Figure 4.6 q is a boundary critical point and p is not

Therefore, if the gradient of f at a point is *not* orthogonal to ∂D, then f does not have an extreme value there. So to find points where f might have an extreme value on ∂D, we can restrict our attention to the points where ∂f is orthogonal to ∂D.

We can simplify this idea further: recall that the gradient of g at each point is orthogonal to the level set of g passing through that point. It follows that if ∂D is a level set of g and $p \in \partial D$ is a point where f has an extreme value, then ∇g and ∇f are both orthogonal to ∂D. This means that they are parallel! By Observation 2.1.1, this means that there exist a scalar λ such that $\nabla f = \lambda \nabla g$.

Theorem 4.7.1: Method of Lagrange multipliers

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are differentiable functions and $c \in \mathbb{R}$. If the restriction of f to the level set $\{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) = c\}$ has a local extremum at $\mathbf{x} \in \mathbb{R}^n$, then

$$\nabla f(\mathbf{x}) = \lambda \nabla g(\mathbf{x}).$$

So far we've been considering the restriction of f to the boundary of a region D, but the region D; any level set of a differential function g will do

Let's see how this works in practice.

Example 4.7.1

Find the maximum and minimum values of

$$f(x,y) = -x^2 - y^2 + x + \frac{2}{3}y + \frac{23}{36}$$

over the disk of radius $\frac{1}{2}$ centered at $\left(\frac{1}{2}, \frac{1}{2}\right)$.

Solution

The only interior critical point is $\left(\frac{1}{2}, \frac{1}{3}\right)$, as in Example 4.4.2. To find boundary critical points, we set up the Lagrange equations:

$$\begin{aligned} \partial_x f &= \lambda\, \partial_x g \\ \partial_y f &= \lambda\, \partial_y g \implies \\ -2x + 1 &= \lambda(2x - 1) && (4.7.1) \\ -2y + \tfrac{2}{3} &= \lambda(2y - 1). && (4.7.2) \end{aligned}$$

We're looking for pairs (x, y) which satisfy both of these equations **and** the equation

$$g(x,y) = \left(x - \tfrac{1}{2}\right)^2 + \left(y - \tfrac{1}{2}\right)^2 = \tfrac{1}{4} \tag{4.7.3}$$

since (x, y) must be on ∂D. So have three equations and three variables: x, y, and λ. The first equation implies that either $\lambda = -1$ or $x = \frac{1}{2}$.

In the case $x = \frac{1}{2}$, we can use the constraint equation to conclude that $y = 0$ or $y = 1$. In either case, we can substitute into (4.7.2) to get a value for λ so that all three equations are satisfied. So, we have $(x, y) = (1/2, 0)$ and $(1/2, 1)$ as boundary critical points.

If $x \neq \frac{1}{2}$, then $\lambda = -1$. Substituting into (4.7.2) gives a contradiction, which means that we've already found all the boundary critical points.

Finally, we evaluate f at the interior critical point and the two boundary critical points:

(x,y)	$\left(\frac{1}{2}, 0\right)$	$\left(\frac{1}{2}, 1\right)$	$\left(\frac{1}{2}, \frac{1}{3}\right)$
$f(x,y)$	$\frac{8}{9}$	$\frac{5}{9}$	1

So the maximum of f over D is 1, and the minimum is $\frac{5}{9}$.

The following example is a 3D application of Lagrange multipliers.

Example 4.7.2

Find the maximum possible volume of a box made with 80 square centimeters of cardboard and having sides and a bottom but no top.

Solution

Denote by x, y, and z the dimensions (in centimeters) of the cardboard. Then the amount of cardboard used is

$$g(x, y, z) = 2yz + 2xz + xy = 72,$$

while the objective function is the volume $f(x, y, z) = xyz$. Setting up the Lagrange equations, we get

$$\begin{aligned} yz &= \lambda(2z + y) \\ xz &= \lambda(2z + x) \\ xy &= \lambda(2x + 2y) \\ 2yz + 2xz + xy &= 72, \end{aligned}$$

where the last one is the constraint equation. Multiplying the first two equations by x and y, respectively, and the setting the resulting right-hand sides equal implies that either $\lambda = 0$ or $z = 0$ or $x = y$. Since $\lambda = 0$ or $z = 0$ clearly give zero volume (and thus not the maximum volume), it follows that $x = y$. Substituting y for x in the third equation gives*

Once again, we can divide by x because we know that $x = 0$ wouldn't make sense for the maximum volume.

$$x^2 = 4\lambda x \implies \lambda = \tfrac{x}{4}.$$

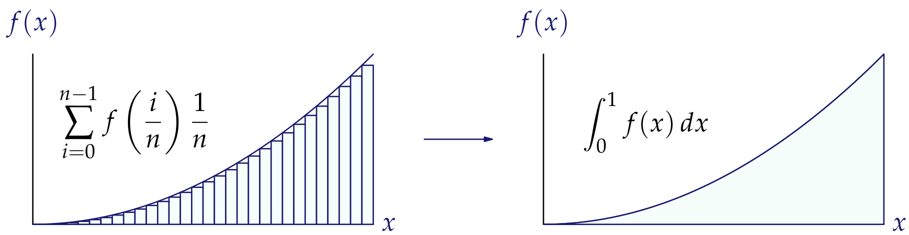
Substituting this into the second equation and simplifying, we get $x = 2z$. Finally, substituting into the constraint equation gives $z = \sqrt{6}$, which in turn implies $x = y = 2\sqrt{6}$.

Exercise 4.7.1

Find the points on the ellipse $\left(\frac{x-1}{2}\right)^2 + (y-2)^2 = 1$ which are nearest and farther from the origin. Hint: for the objective function, use *squared* distance rather than distance.

5 Multivariable Integration

To find the area under the graph of a continuous function f over the unit interval $[0,1]$, we first approximate the area by splitting $[0,1]$ into many short intervals and sum up the areas of rectangles approximating the area under the graph over each short interval:



This approximation converges to the actual area under the graph as $n \to \infty$.

In this section we will work out how to generalize this concept to integrate of functions of multiple variables over regions in $\mathbb{R}^2$ or $\mathbb{R}^3$.

on Double integrals

5.1 Double integration

Recall our convention that 1D volume is length and 2D volume is area.

We can state the definition of the integral, described above, more informally and generally: break the region of integration into small pieces, multiply the volume* of each piece by the value of the function at some point on that piece*, and add up the results. If we take the number of pieces to ∞ and the piece size to zero, then this sum should converge to a number, and if it does then we declare that number to be the value of the integral.

It doesn't ultimately matter where we evaluate the function, since the piece is very small

Stated at this level of generality, the definition of the integral applies to a function $f : \mathbb{R}^2 \to \mathbb{R}$ over a region $D \subset \mathbb{R}^2$. See Figure 5.1.

Definition 8: Integral over a 2D region

Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is a continuous function, and that D is a region in $\mathbb{R}^2$. Then the integral of f over D, denoted $\iint_D f\,dA$, is defined to be

$$\iint_D f\,dA = \lim_{n\to\infty} \sum_{(i,j):\left(\frac{i}{n},\frac{j}{n}\right)\in D} f\left(\frac{i}{n},\frac{j}{n}\right)\overbrace{\frac{1}{n^2}}^{\Delta A}, \tag{5.1.1}$$

The A in dA stands for area.

where the sum includes one term for each integer pair (i,j) such that $(i/n, j/n)$ is in the region D.*

As in the single-variable case, this Riemann-sum definition is not generally practical for precise evaluation

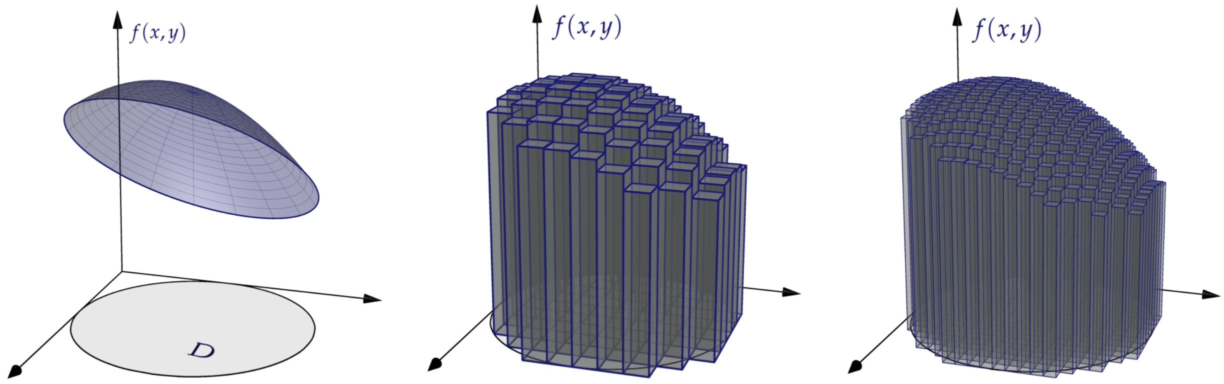


Figure 5.1 The integral of f over a disk D, defined as a limit of sums of volumes of narrow boxes

of integrals. The fundamental theorem of calculus* is the primary tool for evaluating integrals in single-variable calculus, and fortunately we can bootstrap our way up from 1D integrals to 2D integrals by applying our primary strategy for tackling higher dimensional problems: slice it up. Let's start by considering integrals over rectangular regions D.

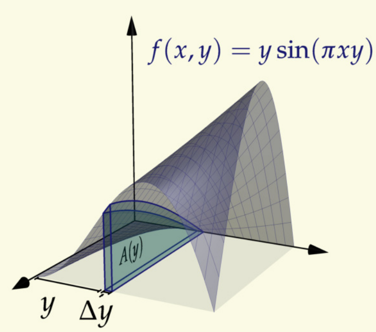
Example 5.1.1

Find the integral of $f(x, y) = y \sin(\pi xy)$ over the square $[0, 1]^2$.

Solution

Let's slice up the desired solid using many 'y = constant' cuts, producing many thin slices like the one shown. The volume of one of these slices, situated at a particular y-value, is given by* the thickness Δy times the area $A(y)$ under the graph of the single-variable function $x \mapsto f(x, y)$. So we can use the fundamental theorem of calculus to compute

$$A(y) = \int_0^1 y \sin(\pi xy)\, dx = -\frac{\cos(\pi xy)}{\pi}\bigg|_0^1$$
$$= \frac{1 - \cos \pi y}{\pi}.$$



Once we have each area $A(y)\Delta y$, we can add them all up and take $\Delta y \to 0$ (as the number of slices goes to ∞) to find that the desired volume is

$$\sum_{\text{all slices}} A(y)\Delta y \to \int_0^1 A(y)\, dy.$$

We can again evaluate this integral using the fundamental theorem to get a final answer* of $\boxed{\frac{1}{\pi}}$.

…ignoring an error, having to do with the top of the slice not being flat—this error tends to zero as the number of slices tends to infinity

For some confidence that our answer is reasonable, we can calculate a Riemann sum for this integrand.

We can express this process more succinctly as

$$\int_0^1 \int_0^1 y \sin(\pi x y)\, dx\, dy = \int_0^1 \frac{1-\cos \pi y}{\pi}\, dy = \frac{1}{\pi}. \tag{5.1.2}$$

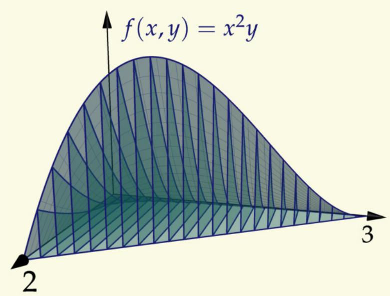
The first expression in (5.1.2) is called an **iterated integral**, since it expresses an integral over a 2D region in terms of two successive single-variable integrals.

Let's see how this works over a non-rectangular region.

Example 5.1.2

Find the integral over the triangle T with vertices $(0,0)$, $(2,0)$, and $(0,3)$ of the function $f(x,y) = x^2 y$, by first finding the area under each '$y =$ constant' slice.

Solution



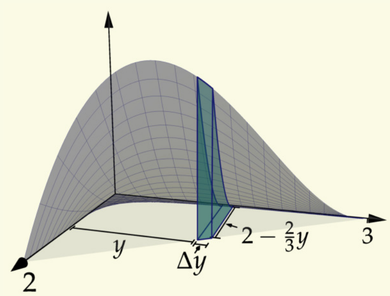
As in the previous example, we slice up the desired volume making many '$y =$ constant' cuts of thickness Δy, yielding thin slices such that each one has volume (very close to) $A(y)\Delta y$, where y is the slice's signed distance from the xz-plane and $A(y)$ is the area of the cross-section (see figure). Since this cross section is an area under a curve, we can find it by integrating $x \mapsto f(x,y)$ over the set of relevant x-values:*

We can find the formula $2 - \frac{2}{3}y$ by writing an equation for the line connecting $(2,0)$ to $(0,3)$ and solving for x

$$A(y) = \int_0^{2-\frac{2}{3}y} f(x,y)\, dx.$$

Thus $A(y) = \frac{1}{3}\left(2 - \frac{2}{3}y\right)^3 y$. Finally, adding up all these areas and taking $\Delta y \to 0$ gives the result

$$\int_0^3 A(y)\, dy = \int_0^3 \left(-\frac{8}{81}y^4 + \frac{8}{9}y^3 - \frac{8}{3}y^2 + \frac{8}{3}y\right) dy = \boxed{\frac{6}{5}}.$$



Let's summarize what we figured out in Example 5.1.2.

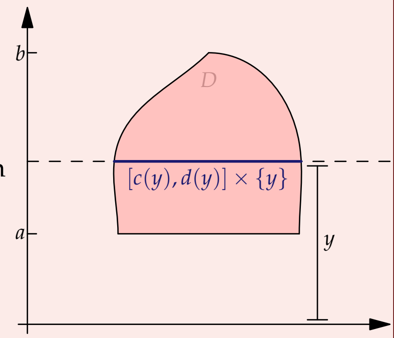
Theorem 5.1.1: Iterated integrals for two-variable functions

Suppose that

- D is a region in $\mathbb{R}^2$,
- $f : D \to \mathbb{R}$ is a continuous function, and
- for all $y \in \mathbb{R}$, the intersection of D and horizontal line through $(0, y)$ is a segment $[c(y), d(y)] \times \{y\}$.

Then

$$\iint_D f\, dA = \int_a^b \int_{c(y)}^{d(y)} f(x,y)\, dx\, dy.$$



Think of dA merely as a reminder that the positive quantity ΔA involved in the corresponding Riemann sums represents an area

In light of Theorem 5.1.1, we sometimes write the area differential* as $dA = dx\, dy$. We can describe the procedure in Theorem 5.1.1 more casually:

Observation 5.1.1: Limits of integration over a 2D region

To set up an iterated integral to evaluate $\iint_D f\, dA$ (where f is continuous and D is a region such that the intersection of every horizontal line with D is a segment):

1. Find the least and greatest y values for any point in D. These are your **outer limits** of integration.
2. For each fixed horizontal line which intersects D, identify the least and greatest values of x for any point which is in D *and on that line*, expressed in terms of the vertical position y of the line. These are the **inner limits** of integration, and they may depend on y.

The role of x and y in Observation 5.1.1 can be reversed (in which case we have vertical rather than horizontal lines in Step 2). The following exercise shows how this can be useful.

Exercise 5.1.1

Find

$$\int_0^{1/2} \int_{2y}^{1} 4e^{x^2}\, dx\, dy$$

by first rewriting it as an integral over a 2D region and then reversing the order of integration.

on triple integrals

5.2 Triple integration

We interpret the integral of a single-variable function as an area and the integral of a two-variable function as a volume. So how should we interpret the integral of a function of *three* variables over a region D in $\mathbb{R}^3$? *Four-dimensional volume* is a reasonable answer, but of course this is unsatisfactory from a visualization point of view, since we don't have access to four spatial dimensions with which to visualize.

Therefore, let's consider a physics interpretation of integration which permits a visualization *not* involving the graph of the function being integrated.

Example 5.2.1

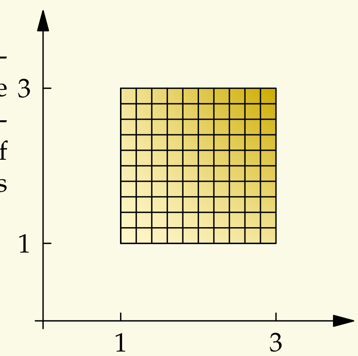
See the figure in the solution, where darker color indicates a denser portion of the plate

Consider a square plate occupying the square $[1,3]^2$ whose density at each point is $\sigma(x,y) = xy$ kilograms per square meter.* Find the mass of the plate.

Solution

Let's imagine physically cutting the plate into small squares, computing the mass of each one, and adding up the resulting masses. The mass of a small plate of area $\Delta x \Delta y$ containing the point (x,y) is approximately the area density times the area: $\sigma(x,y)\,\Delta x\,\Delta y$. The sum of these masses is a Riemann sum (see Definition 8) which converges as the number of small squares goes to ∞ to the integral

$$\int_1^3 \int_1^3 xy\,dx\,dy = \boxed{8} \text{ kilograms.}$$



Let's do a three-dimensional example.

Example 5.2.2

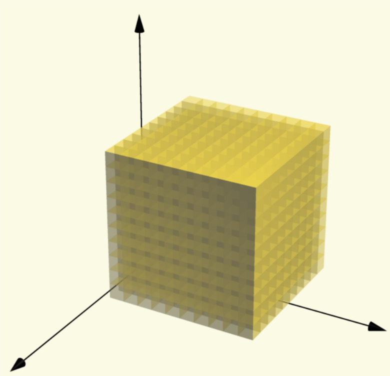
Consider a cubical block occupying $D = [1,2]^3$ whose density at each point is $\rho(x,y) = x^2 + y^2 + z^2$ kilograms per cubic meter.* Find the mass of the block.

Solution

Any corner, or indeed any point in the cube, would give the same result—we choose the bottom, back corner (the point nearest the origin) for concreteness

We cut the cube into n^3 small cubes, where n is a large integer. The mass one of these cubes with bottom, back* corner (x,y,z) is approximately equal to the product of its volume $\frac{1}{n^3}$ and the approximate density $\rho(x,y,z)$ throughout the small cube. So the approximate volume is

$$\sum_{\text{all cubes}} \rho\,(x,y,z)\,\frac{1}{n^3}.$$



Intuitively, this sum should converge to a limit as $n \to \infty$, and if so, then we should define the limiting value to be the integral of ρ over D. Let's state this idea for any continuous function $f : \mathbb{R}^3 \to \mathbb{R}$: we define the integral of f over D by

$$\iiint_D f(x,y,z)\,dV = \lim_{n\to\infty} \sum_{(i,j,k):\left(\frac{i}{n},\frac{j}{n},\frac{k}{n}\right)\in D} f\left(\frac{i}{n},\frac{j}{n},\frac{k}{n}\right)\frac{1}{n^3}.$$

We can calculate the integral by slicing up the region of integration into thin slabs along 'z = constant' slices, and then performing double integrals to find the area of each slab. This works the same as double iterated integration, but with one extra step. Rather than writing Δz and then

This is a general theme: we contract the following two steps into a single step (by writing dz instead of Δz from the outset): (i) reason about sums involving a small but positive quantity Δz, and (ii) replace the sum with an integral over the relevant z values and replace Δz with dz

taking a limit to turn Δz into dz, we'll skip to the limit and work directly with dz*

$$\begin{aligned}
\text{mass} &= \int_1^2 \overbrace{\int_1^2 \int_1^2 (x^2+y^2+z^2)\,dx\,dy\,dz}^{\text{mass of slice from } z \text{ to } z+dz} \\
&= \int_1^2 \int_1^2 \left(y^2+z^2+\frac{7}{3}\right) dy\,dz \\
&= \int_1^2 \left(z^2+\frac{14}{3}\right) dz \\
&= \boxed{7} \text{ kilograms.}
\end{aligned}$$

The following theorem summarizes the idea of integrating in 3D by breaking down the 3D region of integration into 2D slices.

Theorem 5.2.1: Iterated integrals for three-variable functions

D_z is the region obtained by intersecting D with the plane which is z units from the xy-plane and then dropping off the third coordinate.

Suppose f is a continuous function over a region D which is bounded between the planes $z = a$ and $z = b$. For each $z \in (a, b)$, define $D_z \subset \mathbb{R}^2$ to be the region*

$$D_z = \{(x, y) \in \mathbb{R}^2 \,:\, (x, y, z) \in D\}.$$

Then

$$\iiint_D f\,dV = \int_a^b \left[\iint_{D_z} f(x, y, z)\,dx\,dy\right] dz,$$

Let's break this theorem down into a simple algorithm (the following observation is the 3D analogue of Observation 5.1.1):

Observation 5.2.1: Limits of integration over a 3D region

To set up an iterated integral to evaluate $\iiint_D f\,dV$:

1. Find the least and greatest z values for any point in D. These are your **outer limits** of integration.
2. For each fixed 'z = constant' plane which intersects D, identify the least and greatest values of y for any point which is in D *and on that plane*, expressed in terms of the vertical position z of the plane. These are the **middle limits** of integration, and they may depend on z.
3. For each line of the form 'z = constant and y = constant', find the least and greatest values of x for any point which is in D *and on that line*. These are your **inner limits** of integration, and they may depend on both z and y.

Example 5.2.3

Find the volume of the tetrahedron with vertices $(0, 0, 0)$, $(2, 0, 0)$, $(0, 3, 0)$, and $(0, 0, 4)$ using a triple integral.

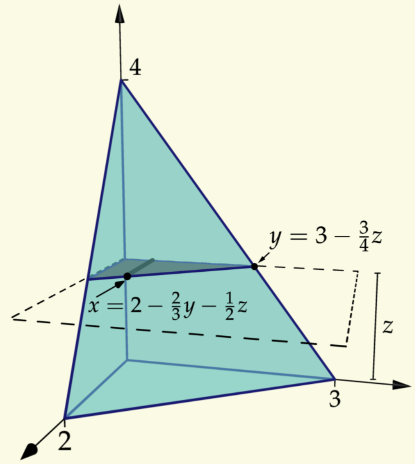
Solution

The volume of a region is equal to the integral of the constant function 1 over that region:

$$\text{volume}(D) = \iiint_D 1\,dV,$$

because $\iiint_D 1\,dV$ is equal to the mass of a solid occupying the region D and having density 1 at every point. But if a solid has a constant mass density of 1, then its mass is equal to its volume.*

Alternatively, note that calculating the Riemann sums approximating $\iiint_D 1\,dV$ amounts to splitting D into small pieces and summing their volumes.



So we set up our iterated integral: the least and greatest values of z are 0 and 4, so those are our outer limits. For a fixed value of z, the least and greatest values of y for a point in D are 0 and $3 - \frac{3}{4}z$, respectively. Finally, for fixed y and z, the least and greatest values of x for a point in D are 0 and the point on the plane $6x + 4y + 3z = 12$ with the given values of y and z (see figure).

So we get

$$\begin{aligned}
\text{volume}(D) &= \int_0^4 \int_0^{3-\frac{3}{4}z} \int_0^{2-\frac{2}{3}y-\frac{1}{2}z} 1\,dx\,dy\,dz \\
&= \int_0^4 \int_0^{3-\frac{3}{4}z} \left(2 - \frac{2}{3}y - \frac{1}{2}z\right) dy\,dz \\
&= \int_0^4 \frac{3}{16}(z-4)^2\,dz \\
&= \boxed{4}.
\end{aligned}$$

There is nothing special about the order $dx\,dy\,dz$—any way of slicing up the region gives the same result. The following exercise

Exercise 5.2.1

Write the iterated integral

$$\int_0^1 \int_{\sqrt{x}}^1 \int_0^{1-y} f(x,y,z)\,dz\,dy\,dx.$$

as an integral over a 3D region. Then sketch that region and use your figure to rewrite the integral in five other ways, using the five other permutations of x-y-z.

5.3 Polar, cylindrical, and spherical integration

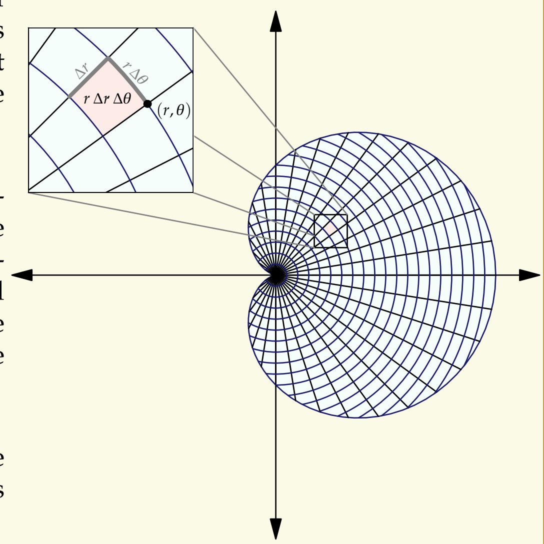
Some regions in $\mathbb{R}^2$ are more conveniently described in polar coordinates than rectangular coordinates. If we are integrating a function $f : \mathbb{R}^2 \to \mathbb{R}$ over such a region, it is helpful to work directly in polar coordinates. Let's do an example.

Example 5.3.1

(i) Find the area of the region D enclosed by solution set of the polar coordinate equation $r = 1 + \cos\theta$. (ii) Integrate the function $f(x,y) = x + y$ over D.

Solution

(i) Let's slice D into small pieces using equally spaced cuts along rays and circles of the form $r =$ constant and $\theta =$ constant, as shown. This decomposes D into a set of **coordinate patches**. The figure suggests that these pieces farther away from the origin are larger than the ones that are close to the origin, which leads us to investigate the area of each patch.



To find the area of the set of points with radial polar coordinate between r and $r + \Delta r$ and angular polar coordinate between θ and $\theta + \Delta\theta$, we note that this region is approximately a rectangle. The straight side length is Δr, and the curvy side length is $r\Delta\theta$, because the perimeter of the circle of radius r is $2\pi r$, and the angle represents $\frac{\theta}{2\pi}$ of the whole circle. So the area is approximately $r\Delta r\Delta\theta$.

Now, for fixed θ, we can add up all the coordinate patches between θ and $\theta + \Delta\theta$, and this sum of areas is approximately equal to the product of $\Delta\theta$ the integral

$$\int_0^{1+\cos\theta} r\,dr.$$

Adding up these areas over all θ from 0 to 2π, we get

$$\int_0^{2\pi}\int_0^{1+\cos\theta} r\,dr\,d\theta = \int_0^{2\pi} \frac{1}{2}(1+\cos\theta)^2\,d\theta\,d\theta = \boxed{\frac{3\pi}{2}}.$$

(ii) We can find this integral using the same procedure as above, except that at the step where we calculate the area of a patch, we also need to multiply it by the value of the function at some point in the patch. Since our function is defined in terms of x and y, we need to substitute $x = r\cos\theta$ and $x = r\sin\theta$ to discover the value of f at the point whose polar coordinates are (r,θ). So we get*

This one is tedious if done by hand, so we use computer algebra assistance.

$$\int_0^{2\pi}\int_0^{1+\cos\theta} (r\cos\theta + r\sin\theta)r\,dr\,d\theta = \boxed{\frac{5\pi}{4}}.$$

We can see from this example that the ideas for setting up an iterated polar integral are similar to those for rectangular integration:

Observation 5.3.1: Iterated polar integration

1. Find the least and greatest values of θ for any point in the region of integration, and
2. For each fixed value of θ, find the least and greatest values of r for any point on the ray of angle θ and in the region of integration
3. Include the area differential* $dA = r\,dr\,d\theta$
4. Substitute $x = r\cos\theta$ and $y = r\sin\theta$ into f, so that your integrand varies appropriately as r and θ vary

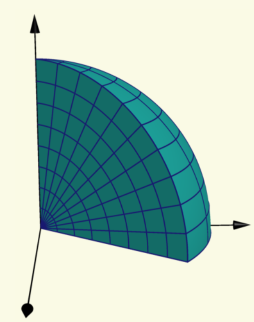
Don't forget the extra factor of r in the polar area differential!

This same basic idea can be carried out in cylindrical and spherical coordinates. The ingredients we need are (i) the volume differential dV expressed in terms of cylindrical and spherical coordinates, and (ii) the formulas for x, y, z in terms of r, θ, z and in terms of ρ, θ, ϕ. This information is listed in Appendix A.5. The only surprising entry in the tables of that appendix is the spherical coordinate volume differential $dV = \rho^2 \sin\phi\, d\rho\, d\phi\, d\theta$.

Example 5.3.2: Spherical coordinate volume differential

Explain why the volume differential in spherical coordinates is $dV = \rho^2 \sin\phi\, d\rho\, d\phi\, d\theta$.

Solution

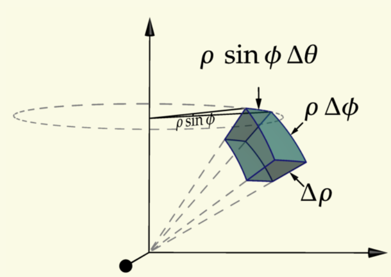


The volume differential arises from slicing up the region of integration into small coordinate "patches", each of which consists of those points whose three spherical coordinates lie in the intervals* $[\rho, \rho + \Delta\rho]$, $[\theta, \theta + \Delta\theta]$, and $[\phi, \phi + \Delta\phi]$, respectively (see the top figure, where a wedge has been decomposed into spherical coordinate patches). Thus we must calculate the approximate volume of one such patch.

Here (ρ, θ, ϕ) are the spherical coordinates of one of the corners of the patch.

When $\Delta\rho$, $\Delta\theta$, and $\Delta\phi$ are all very small, the coordinate patch is approximately a rectangular prism. The dimensions of this rectangular prism, as marked in the lower figure, are $\Delta\rho$, $\rho\,\Delta\phi$, and $\rho\sin\phi\,\Delta\theta$.

To see why the top edge length is $\rho\sin\phi\Delta\theta$, note that the dashed circle in the figure has radius $\rho\sin\phi$, since the *cylindrical* radial coordinate r satisfies the equation $r = \rho\sin\phi$. Thus the volume of the patch is approximately $\rho^2 \sin\phi\Delta\rho\,\Delta\phi\,\Delta\theta$.

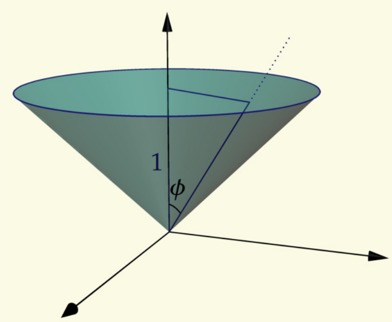


Example 5.3.3

Consider a solid whose density at each point (x,y,z) is $\rho(x,y,z) = \frac{1}{x^2+y^2+z^2}$ and which occupies the region enclosed by the cone $z = \sqrt{x^2+y^2}$ and the plane $z = 1$. Find the mass of the solid.

Solution

Let's set up the iterated integral with the order $d\rho\, d\phi\, d\theta$. The solid has points with θ values as small as 0 and as large as 2π, so the outer limits will be 0 and 2π.
For any given value of θ, there are points with that θ value whose ϕ value is as small as zero (for the points on the positive z-axis) and as large as $\frac{\pi}{4}$ (for the points on the cone $z = \sqrt{x^2+y^2}$). So the middle limits are 0 and $\frac{\pi}{4}$.
Finally, for any given ϕ and θ, the solid contains points with z as small as 0 and as large as $\frac{1}{\cos\phi}$ (by right-triangle trigonometry; see figure).



Help with calculating this integral

For the integrand, we should substitute the spherical coordinate formulas for x, y, and z. However, we know that it will simplify to $\frac{1}{\rho^2}$, since $\rho^2 = x^2+y^2+z^2$. So we get*

$$\int_0^{2\pi}\int_0^{\pi/4}\int_0^{\sec\phi} \rho^{-2}\rho^2 \sin\phi\, d\rho\, d\phi\, d\theta = \int_0^{2\pi}\int_0^{\pi/4} \sec\phi \sin\phi\, d\rho\, d\phi\, d\theta = (2\pi)(\tfrac{1}{2}\ln 2) = \boxed{\pi \ln 2}.$$

Exercise 5.3.1

What proportion of the volume of the unit sphere lies above the plane $z = \frac{1}{2}$?

Exercise 5.3.2

Find

$$\int_{-2}^{2}\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}}\int_{\sqrt{x^2+y^2}}^{2} (x^2+y^2)\, dz\, dy\, dx$$

by writing it as integral over a 3D region and then rewriting that integral using cylindrical coordinates.

on change of variables

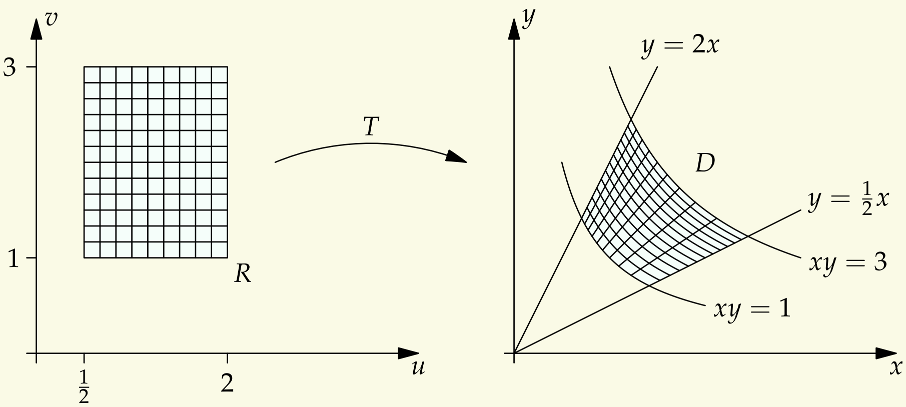
5.4 Change of variables

Sometimes we want to integrate a function over a region which is not conveniently described using any of the standard coordinate systems. In this section we will develop a general program for integrating with respect to a custom-designed coordinate system.

Example 5.4.1

Find $\iint_D y^2\,dA$, where D is the region in the first quadrant bounded by the lines through the origin of slope $\frac{1}{2}$ and 2, as well as the hyperbolas $xy = 1$ and $xy = 3$.

Solution



Let's cut the region along lines of the form $y = ux$ where u ranges over equally spaced values between $\frac{1}{2}$ and 2, and along hyperbolas of the form $xy = v$ where v ranges over $[1, 3]$, as shown in the figure on the right above.

Note that each point (x, y) in the first quadrant can be identified by its u and v values.* So u and v provide a coordinate system for the first quadrant. We can visualize the relationship between (u, v) and (x, y) as a transformation T that maps each (u, v) pair to its corresponding (x, y) pair, as shown in the figure. If we want to find a formula for this map, we can solve the system $y = ux$ and $xy = v$ for x and y to find that $y = \sqrt{uv}$ and $x = \sqrt{v/u}$.

These specify, respectively, which line through the origin and which hyperbola the point is on

To integrate $f(x, y) = x$ over D, we must find the area of each of these small patches, multiply each area by the value of f somewhere on the patch, and sum the resulting products. Each patch is the image under T of a rectangle of the form $[u, u + \Delta u] \times [v, v + \Delta v]$. The area of this rectangle is $\Delta u \Delta v$, and the transformation distorts its area by an amount that we can approximate by treating the transformation as linear around (u, v) and using the fact that area distortion is measured by the determinant.

Writing $T(u, v) = (\sqrt{v/u}, \sqrt{uv}) = (g(u, v), h(u, v))$, we see that T maps the point $(u + \Delta u, v)$ to

$$T(u, v) + ((\partial_u g)(u, v)\Delta u, (\partial_u h)(u, v)\Delta u), \tag{5.4.1}$$

and $(v, v + \Delta v)$ to

$$T(u, v) + ((\partial_v g)(u, v)\Delta v, (\partial_v h)(u, v)\Delta v). \tag{5.4.2}$$

Thus the area of the image of $[u, u + \Delta u] \times [v, v + \Delta v]$ under T is approximately*

The four entries of the matrix below come from the coefficients of Δu and Δv in (5.4.2) and (5.4.1)

$$\left|\det \begin{bmatrix} \partial_u g & \partial_v g \\ \partial_u h & \partial_v h \end{bmatrix}\right| \Delta u \Delta v, = \left|\det \begin{bmatrix} -\frac{1}{2}\sqrt{v}u^{-3/2} & \frac{1}{2\sqrt{uv}} \\ \sqrt{v}(2\sqrt{u}) & \sqrt{u}(2\sqrt{v}) \end{bmatrix}\right| \Delta u \Delta v = \left(\frac{v}{u} + 1\right) \Delta u \Delta v.$$

for help with this integral

As usual, we may take $\Delta u, \Delta v \to 0$ to get* $\displaystyle\int_1^3 \int_{1/2}^2 uv \left(\frac{v}{u} + 1\right) = du\,dv = \boxed{\frac{41}{2}}$.

Note that in this new notation, the symbols g and h are replaced by x and y. Thus we are abusing notation by regarding x and y as functions of u and v, even though they also represent independent variables

The matrix $\begin{bmatrix} \partial_u g & \partial_v g \\ \partial_u h & \partial_v h \end{bmatrix}$ is called the *Jacobian matrix*, and its determinant is called the *Jacobian determinant*. Often we just say "Jacobian", relying on context to distinguish. It can be written as $\left|\frac{\partial(x,y)}{\partial(u,v)}\right|$ for short.*

The following theorem summarizes the technique we developed in Example 5.4.1.

Theorem 5.4.1: Multivariable change of variables

Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a differentiable transformation that maps a region R one-to-one onto a region D. Then for any continuous function f, we have

$$\iint_D f(x,y)\,dx\,dy = \iint_R f(T^{-1}(x,y)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv.$$

Exercise 5.4.1

Use the change of variables $x = u^2 - v^2$, $y = 2uv$ to evaluate the integral $\iint_R y\,dA$, where R is the region above by the x-axis bounded by the parabolas $y^2 = 4 - 4x$ and $y^2 = 4 + 4x$.

Exercise 5.4.2

Find the integral of $\frac{(x-y)^2}{x+y+2}$ over the square whose vertices are the four points of intersection between the axes and the unit circle.

6 Vector Calculus

6.1 Vector fields and line integrals

on vector fields

So far we have been considering functions from $\mathbb{R}^n$ to $\mathbb{R}^1$ (where n is 2 or 3). In this chapter we work with functions from $\mathbb{R}^n$ to $\mathbb{R}^2$ or $\mathbb{R}^3$. How should such functions be visualized? Let's begin by considering how they arise in applications.

Many other physical phenomena, such as electromagnetic forces, heat flow, fluid flow are modeled as vector fields.

To describe the gravitational force* felt by a particle, we would use a function with a three-dimensional input (to specify the particle's location) as well as a three-dimensional output, to specify the direction and magnitude of the force. It is natural to represent this function by drawing a small arrow indicating the output vector at several points in space, because this makes it easy to imagine how the force changes as the particle moves around (see Figure 6.1).

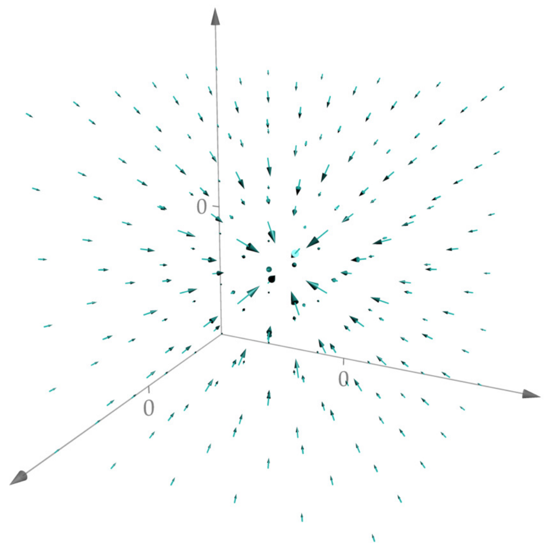


Figure 6.1 A gravitational vector field

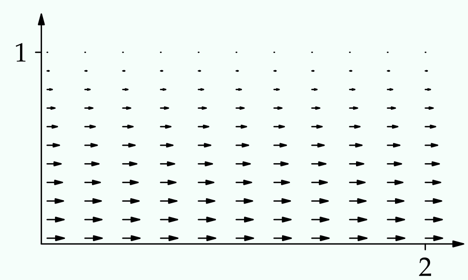
This picture suggests the term **vector field** for a function from $\mathbb{R}^n$ to $\mathbb{R}^n$, where $n > 1$. The gravitational vector field plotted in Figure 6.1 is

$$\mathbf{F}(x, y, z) = -\frac{GMm}{(x^2 + y^2 + z^2)^{3/2}} \langle x, y, z \rangle . \tag{6.1.1}$$

where G is the gravitational constant and Mm is the product of the masses of particle and the attracting body.

Exercise 6.1.1

The vector plot shown represents the velocity of water on the surface of a river. The water is flowing due east, and it is flowing faster near the south end of the river than the north. Come up with a vector field $\mathbf{F}$ whose vector plot looks approximately like the one shown.



on line integrals

A vector field in which the vectors represent a physical force

Now suppose that rather than remaining stationary, our particle moves along a path in the presence of a force field (see Figure 6.2).* Sometimes the particle is moving with the force field and getting a boost from it, whereas other times it's working against the force field. How much net work does it take to move along the path?

If the force field were constant and the path straight, then we get the answer from physics: the work is equal to the product of the magnitude F of the force, the distance d traveled, and the cosine of the angle θ between the force and the path. Alternatively, we may interpret the force and distance as vectors $\mathbf{F}$ and $\mathbf{d}$

and write the work as a dot product:

$$W = Fd\cos\theta = \mathbf{F}\cdot\mathbf{d}.$$

So how do we bootstrap our way from constant force and a straight path to varying force and a curvy path? We can cut up the path into small pieces, handle each small piece by treating the force as approximately constant and the path as approximately straight, and then adding up the amount of work for each small piece. We will assume that our path $\mathbf{r}(t)$ is differentiable.*

The curve could alternatively by *piecewise* differentiable, meaning that the curve is non-differentiable at only finitely many points

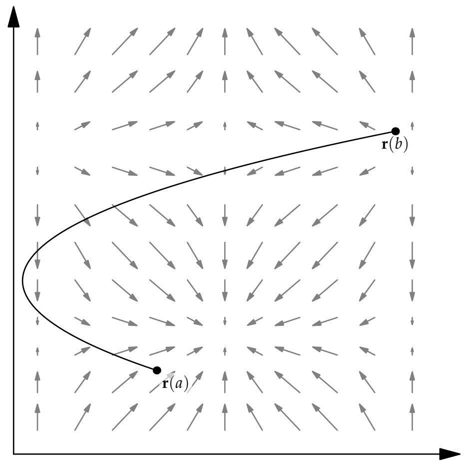


Figure 6.2 The path of a particle moving through a vector field

Suppose we have a path C parametrized as $\mathbf{r}(t)$ where t ranges from a to b. Over the time interval $[t, t+\Delta t]$, the particle is displaced by* the vector $\mathbf{r}'(t)\,\Delta t$, and the force it feels over that time is* to $\mathbf{F}(\mathbf{r}(t))$. Therefore, the contribution from the time period $[t, t+\Delta t]$ is equal to

Approximately, with an error that vanishes as $\Delta t \to 0$

$$\mathbf{F}(\mathbf{r}(t))\cdot\mathbf{r}'(t)\,\Delta t.$$

Summing all these contributions and taking $\Delta t \to 0$, we arrive at the formula

$$W = \int_a^b \mathbf{F}(\mathbf{r}(t))\cdot\mathbf{r}'(t)\,dt = \int_C \mathbf{F}\cdot d\mathbf{r},$$

where the last expression is an abbreviation for the middle expression. We call an integral of the form $\int_C \mathbf{F}\cdot d\mathbf{r}$ a **line integral**.

Example 6.1.1

Find the line integral of $\mathbf{F}(x,y) = \langle xy, y\rangle$ along the parabola $y = x^2$ from $(0,0)$ to $(2,4)$

Solution

Let's parametrize the parabola using the x coordinate as the parameter:

$$\mathbf{r}(t) = \langle t, t^2\rangle.$$

Note that the point $(2,4)$ is visited at time $t = 2$, while the origin is visited at time $t = 0$. Therefore,

$$W = \int_0^2 \langle t(t^2), t^2\rangle\cdot\langle 1, 2t\rangle\,dt = \int_0^2 (t^3 + 2t^3)\,dt = \boxed{12}.$$

The following theorem states that the choice of parametrization of a curve doesn't matter when computing a line integral. This makes sense physically, since the formula $W = Fd$ does not involve time, and our derivation of the line integral formula was based on $W = Fd$. The role of the parametrization was merely to provide a convenient way to split up the path into short pieces. Exercise 6.1.2 below gives an example.

Theorem 6.1.1: Independence of parametrization

If C is a curve parametrized by $\mathbf{r}_1$ over $[a,b]$ and also by $\mathbf{r}_2$ over $[c,d]$, then

$$\int_a^b \mathbf{F}(\mathbf{r}_1(t)) \cdot \mathbf{r}_1'(t)\,dt = \int_c^d \mathbf{F}(\mathbf{r}_2(t)) \cdot \mathbf{r}_2'(t)\,dt.$$

In other words, $\int_C \mathbf{F} \cdot d\mathbf{r}$ depends only on the curve C, not the choice of parametrization.

Exercise 6.1.2

(i) Compute the line integral of $\mathbf{F} = \langle x^2, -xy \rangle$ over the portion of the unit circle in the first quadrant, using the parametrization $\mathbf{r}(t) = \langle \sin t, \cos t \rangle$.

(ii) Perform the same line integral using the parametrization $\mathbf{r}(t) = \langle t, \sqrt{1-t^2} \rangle$.

Exercise 6.1.3

Consider the vector field $\mathbf{F}$ and path C shown in Figure 6.2. Is $\int_C \mathbf{F} \cdot d\mathbf{r}$ positive or negative?

6.2 The fundamental theorem of vector calculus

m:i on the gradient theorem for line integrals

In general, the line integral of $\mathbf{F}$ over a path between two points depends on the path, not just the starting and ending points. For example, in Figure 6.3, the line integral along the blue (top) path is positive, while the line integral along the orange (bottom) path is negative.

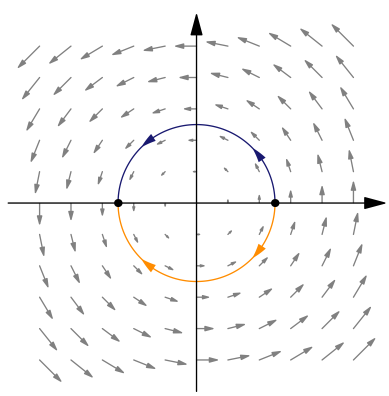


Figure 6.3 The vector field $\mathbf{F}(x,y) = \langle -y, x \rangle$ and two semicircular paths

However, there is an important class of vector fields which are path-independent, meaning that the value of $\int_C \mathbf{F} \cdot d\mathbf{r}$ depends only on the starting and ending points of C. These are the vector fields which can be written as the gradient of a function from $\mathbb{R}^n$ to $\mathbb{R}^1$. For example, $\mathbf{F}(x,y,z) = \langle -2x, -2y, z \rangle$ is the gradient of the function

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

Such vector fields are called **conservative**.

If we calculate the line integral of ∇f along a curve C parametrized by $\mathbf{r}(t) = \langle r_1(t), r_2(t), r_3(t) \rangle$, then the contribution from the portion of the curve from $\mathbf{r}(t)$ to $\mathbf{r}(t+\Delta t)$ is*

approximately, with an error that vanishes a $\Delta t \to 0$

$$\langle \partial_x f, \partial_y f, \partial_z f \rangle \cdot \langle r_1'(t), r_2'(t), r_3'(t) \rangle\,\Delta t,$$

which by the chain rule is* the change in $f(\mathbf{r})(t)$ over that interval. Therefore, the line integral of ∇f along a path is equal to the change in f from the beginning to the end of the path.

Theorem 6.2.1: Fundamental theorem for line integrals

If C is a path from $\mathbf{a}$ to $\mathbf{b}$ and f is a differentiable function, then

$$\int_C \nabla f \, d\mathbf{r} = f(\mathbf{b}) - f(\mathbf{a}).$$

Example 6.2.1

Suppose $\mathbf{F}(x,y,z) = \langle 2xy^3z, 3x^2y^2z + y, x^2y^3 \rangle$ and that C is circular arc from the origin to the point $(1,1,1)$ and passing through the point $(1/2, 1/2, 1)$. Find $\int_C \mathbf{F} \cdot d\mathbf{r}$.

Solution

Finding a parametrization for C seems computationally messy. However, if $\mathbf{F}$ is conservative, then we can use Theorem 6.2.1. Integrating $2xy^3z$ with respect to x, we see that if $\mathbf{F} = \nabla f$ for some function f, then we would have

$$f(x,y,z) = x^2y^3z + C_1(y,z),$$

where $C_1(y,z)$ denotes a function not depending on x. Similarly, we can integrate the second and third components with respect to y and z to find that

$$\begin{aligned} f(x,y,z) &= x^2y^3z + \frac{1}{2}y^2 + C_2(x,z) \\ f(x,y,z) &= x^2y^3z + C_3(x,y). \end{aligned}$$

We see that these three conditions are simultaneously satisfied by the function $f(x,y,z) = x^2y^3z + \frac{1}{2}y^2$. So the desired line integral is equal to

$$f(1,1,1) - f(0,0,0) = \frac{3}{2} - 0 = \boxed{\frac{3}{2}}.$$

The following theorem provides a convenient way to check whether a two-dimensional vector field is conservative.

Theorem 6.2.2

A vector field $\mathbf{F}(x,y) = \langle M(x,y), N(x,y) \rangle$ which is differentiable on $\mathbb{R}^2$ is conservative if and only if

$$\partial_x N = \partial_y M \tag{6.2.1}$$

To see where (6.2.1) comes from, note that this equation follows directly from Clairaut's theorem for conservative fields $\mathbf{F}$. So the more interesting aspect of Theorem 6.2.2 is the converse direction: merely checking $\partial_y M = \partial N_x$ establishes existence or nonexistence of a gradient function.

Exercise 6.2.1

Show that the gravitational force in (6.1.1) is conservative.

Exercise 6.2.2

(i) Try to apply Theorem 6.2.2 to the vector field

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

(ii) Show by plotting this vector field that it is not conservative. How does this square with Theorem 6.2.2.

mi on Green's theorem

6.3 Green's theorem

Invented in 1854

Is it possible to engineer a simple mechanical device that displays the area bounded by a curve traced out on paper? This seems surprising, since computing the area would seem to require some inspection of the region inside the curve. However, the *planimeter** can calculate area of a region based on the motion of its wheels as its tip traverses the boundary of the region. The design of the planimeter takes advantage of the following beautiful relationship between line integrals along the boundary of a curve and and area integrals over the region it encloses.

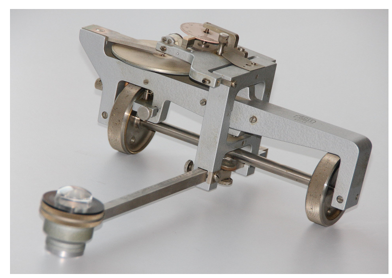


Figure 6.4 A planimeter

Theorem 6.3.1: Green's theorem

If $\mathbf{F} = \langle M, N \rangle$ is a vector field* on $\mathbb{R}^2$ with continuous partial derivatives and if D is a region bounded by a simple, counterclockwise oriented, piecewise smooth curve C, then

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \iint_D \left(\partial_x N - \partial_y M\right) \, dA.$$

$\mathbf{F} = \mathbf{F}(x, y)$ and similarly for M and N; we will begin using these abbreviations for notational convenience; bear in mind that $\mathbf{F}$ always represents a vector field which implicitly depends on (x, y) or (x, y, z)

Example 6.3.1

Verify Green's theorem in the case where D is the unit disk and $\mathbf{F}(x, y) = \langle 0, x \rangle$.

Solution

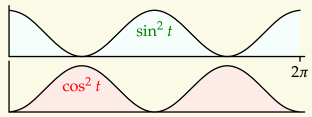
We parametrize the unit disk trigonometrically as $(\cos t, \sin t)$, and we calculate the line integral

$$\int_0^{2\pi} \langle 0, \cos t \rangle \cdot \langle -\sin t, \cos t \rangle \, dt = \int_0^{2\pi} \cos^2 t \, dt = \pi.$$

This last integral can be done with a trick: note that

$$\int_0^{2\pi} (\cos^2 t + \sin^2 t)\, dt = \int_0^{2\pi} 1\, dt = 2\pi.$$

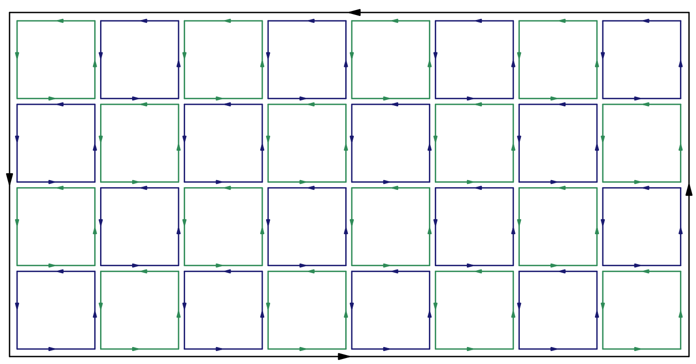
However, the contributions of $\int_0^{2\pi} \cos^2 t\, dt$ and $\int_0^{2\pi} \sin^2 t\, dt$ are equal, since their graphs over the region of integration are the same up to a shift. So each is equal to π.



The integrand for the double integral is $\partial_x N - \partial_y M = 1 - 0 = 1$, so the value of the double integral is the area of the unit disk, which is equal to π. Thus the conclusion of Green's theorem is satisfied.

Proving Green's theorem

The idea of the proof of Green's theorem is to cut D into small rectangles along grid lines (shown with small gaps for visual clarity). Green's theorem holds approximately on each small rectangle $R = [x, x + \Delta x] \times [y, y + \Delta y]$, because the left and right sides of R contribute to $\int\int_R \mathbf{F} \cdot d\mathbf{r}$ approximately



$$\overbrace{N(x + \Delta x, y)\Delta y}^{\text{integral over right side}} - \overbrace{N(x, y)\Delta y}^{\text{integral over left side}} \approx (\partial_x N)(x, y)\, \Delta x\, \Delta y.$$

Similarly, the contribution of the top and bottom sides is $-(\partial_y M)(x, y)\, \Delta x\, \Delta y$. So all together, **circulation** of $\mathbf{F}$ around R is $(\partial_x N - \partial_y M)\, \Delta x\, \Delta y$. This is also approximately equal to the integral of $(\partial_x N - \partial_y M)$ over R, since we may regard $\partial_x N - \partial_y M$ as constant over R.

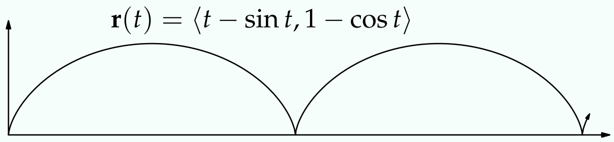
The line integrals of $\mathbf{F}$ over the small rectangles sum to the line integral of $\mathbf{F}$ around the boundary of D, because each interior segment is integrated along twice (once for each adjoining rectangle) and in opposite directions. These contributions sum to zero, leaving only the integrals along the outer edges. Since these outer edges fit together to form ∂D, the line integrals along them sum to the line integral along ∂D.

Exercise 6.3.1

Use Green's theorem to find the line integral of $\mathbf{F} = \langle \sqrt{x^2 + 1}, \arctan x \rangle$ along a counterclockwise traversal of the triangle with vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$.

Exercise 6.3.2

Use Green's theorem to find the area under each arch of the cycloid shown below.



6.4 Surface integrals and flow

mi on surface integrals

6.4.1 Surface integrals

What is the average temperature on the surface of the earth? Let's overlook the scientific challenges of this problem and imagine that the earth is a sphere S and that we have a reading of the temperature T at every point on its surface at a particular point in time. What do we do with this information to find the average temperature?

To interpret the result as an average, we should be using an area unit for which the total area of the earth's surface is 1

We use the same approach we use throughout calculus when handling continuously varying quantities: split S into tiny patches over which T may be treated as constant, multiply the area* of each patch by the value of T somewhere on that patch, and sum the resulting products. As the size of the patches tends to zero, we expect this sum to converge to some limiting value, and we can declare that limit to be the value of the **surface integral*** of f over S, denoted $\iint_S f\, dA$.

Or scalar surface integral, to distinguish from vector surface integral introduced in the next subsection

Example 6.4.1

Find the surface integral of $f(x, y, z) = 2x^2 + 2y^2 + 2z^2$ over the unit sphere S.

Solution

This function is equal to 2 everywhere on the unit sphere, so the integral is

$$\iint_S 2\, dA = 2 \iint_S 1\, dA = 2 \times \text{surface area}(S) = 8\pi,$$

since the surface area formula for a sphere of radius r is $4\pi r^2$.

Example 6.4.1 was special because the function happened to be constant over the surface. A general method for computing surface integrals can be developed in a manner analogous to change-of-variables technique in Section 5.4 (see Exercise 6.4.2). However, surface integrals can be calculated in the same manner as double integrals if the surface is contained in a plane (or if it consists of several pieces each of which is planar—see Exercise 6.4.1).

Exercise 6.4.1

Find the surface integral of $f(x, y, z) = xyz$ over the rectangular prism $[0, 1] \times [0, 2] \times [0, 3]$.

on surface parametrization

Exercise 6.4.2: $\star$ (Surface integral formula)

If D is a planar domain and $\mathbf{r} : D \to \mathbb{R}^3$ is a function mapping D bijectively onto a surface S in $\mathbb{R}^3$ (in other words, $\mathbf{r}$ is a *parametrization** of S), then

$$\iint_S f\, dA = \iint_D f(\mathbf{r}(u, v))|\partial_u \mathbf{r} \times \partial_v \mathbf{r}|\, dA. \tag{6.4.1}$$

Explain why $|\partial_u \mathbf{r} \times \partial_v \mathbf{r}|$ is the appropriate Jacobian, and use (6.4.1) to calculate the average value of z over the top half of the unit sphere.

It might help to note that

$$\langle \cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi \rangle,$$

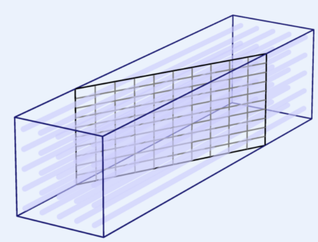
where (θ, ϕ) ranges over $[0, 2\pi] \times [0, \frac{\pi}{2}]$, is a parametrization of the upper unit hemisphere.

on surface integrals of a vector field

6.4.2 Flow

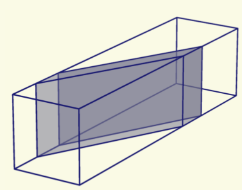
Example 6.4.2

Consider a constant-velocity river flowing through a net as shown. Find the volume of water flowing through the net per unit time, in terms of the area A of the net, the velocity v, and the angle θ between the direction of the river's flow and a vector normal to the plane of the net.



Solution

Imagine letting the water flow for one time unit and then taking a snapshot. The locations of all the water molecules which have flowed through the net during this period occupy a parallelepiped, as shown. The base area of this parallelepiped is A, while its height is equal to* $v \cos\theta$. Therefore, the volume of water passing through the net per unit of time is $\boxed{Av \cos\theta}$.



This is a right-triangle trigonometry exercise

Let's define the vector $\mathbf{A}$ whose length is equal to A and whose direction is orthogonal to the net. Then the **flow** $Av \cos\theta$ can be written as

$$\text{flow} = \mathbf{A} \cdot \mathbf{v},$$

where $\mathbf{v}$ is the river's velocity vector.

Example 6.4.3

Suppose that the velocity field of a body of water is given by $\mathbf{F} = \langle -2y, 4z, x\rangle$ (in meters per second) and that a rectangular net is positioned in the water with corners at $(1,1,3)$, $(1,4,3)$, $(1,4,5)$, and $(1,1,5)$. Find the volume of water flowing through the frame of the net per unit time.

Solution

Since the velocity field isn't constant, we divide the net into small patches and treat the velocity as constant on each one. The flow through a small patch of area ΔA located at (x,y,z) is approximately equal to

$$\mathbf{A}\cdot\mathbf{v} = \big\langle \Delta A, 0, 0\big\rangle \cdot \big\langle -2y, 4z, x\big\rangle = -2y\,\Delta A.$$

If we sum the flow through each patch across the whole rectangular region occupied by the net, we get a Riemann sum that converges as $\Delta A \to 0$ to

$$\int_1^4\int_3^5 (-2y)\,dx\,dy = \boxed{-30}.$$

The ideas in Example 6.4.3 yield the following definition, which develops the notion of vector field integration over surfaces in terms of the scalar surface integral.

Definition 9

The **flow** of a vector field $\mathbf{F}$ through a surface S is defined by

$$\iint_S \mathbf{F}\cdot d\mathbf{A} = \iint_S \mathbf{F}\cdot\mathbf{n}\,dA,$$

where $\mathbf{n} = \mathbf{n}(x,y,z)$ is a vector which is orthogonal to S at each point (x,y,z).

Exercise 6.4.3

Find the flow of the vector field $\mathbf{F} = \langle x^2, y^2, z^2\rangle$ through the unit sphere.

6.5 Divergence and curl

One of the classic and famous vector calculus books is called *div, grad, curl, and all that* by H.M. Schey. We have already discussed the gradient, and in this section we will develop the two other fundamental vector calculates derivative operators: divergence and curl. We will emphasize the underlying physical intuition.

mi on divergence

6.5.1 Divergence

Definition 10: Divergence

The **divergence** of a vector field $\mathbf{F} = \langle M, N, P\rangle$ is a scalar function defined by

$$\nabla \cdot \mathbf{F} = \partial_x M + \partial_y N + \partial_z P.$$

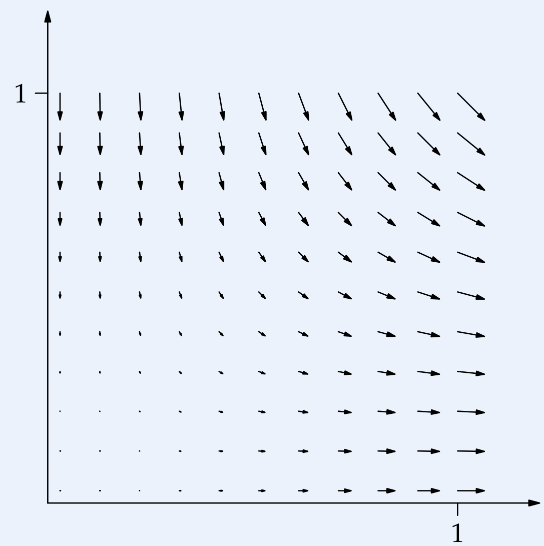
We dot the normal vector $\langle 0,0,1\rangle$ with $\langle M,N,P\rangle$ to account for the cosine factor

For example, the divergence of $\langle x^2, xy, z, \rangle$ is $2x + x + 1 = 3x + 1$. The divergence of $\mathbf{F}$ can be interpreted physically as the **net flow** of $\mathbf{F}$ out of a small region located at (x, y, z) per volume of that region.

To see this, note that if we put a small box of dimensions $\Delta x, \Delta y, \Delta z$ around (x, y, z), then the net flow through the top of the box is approximately* $P(x, y, z + \Delta z)\Delta x \Delta y$, and the net flow through the bottom of the box is approximately $P(x, y, z)\, \Delta x\, \Delta y$. Thus the difference is approximately $(\partial_z P)(x, y, z)\, \Delta x\, \Delta y\, \Delta z$, and he difference per unit volume is just $\partial_z P$. Similarly, the front/back and left/right sides contribute $\partial_x M$ and $\partial_y N$ to the net flow density through the small box.

Example 6.5.1

Figure out where $\nabla \cdot \mathbf{F}$ is positive for the vector field $\mathbf{F}$ shown.



Solution

We can see that in the top left of the diagram that there is more flow into each small region than out of it, since the vectors are downward-pointing and longer than the vectors below them. Therefore, the divergence is negative in the top left.

Similarly, the divergence is positive in the bottom right part of the figure. The dividing line between regions of positive and negative divergence is $y = x$, since points along that line have vectors of equal length pointing towards and away from them.

Exercise 6.5.1

Find the divergence of the gravitational vector field in (6.1.1).

for drawing a vector plot

Exercise 6.5.2

Look at a vector plot* to figure out where $\nabla \times \mathbf{F} > 0$, using the approach of Example 6.5.1, for the vector field $\mathbf{F} = \langle xy, y^2 \rangle$. Then evaluate $\nabla \cdot \mathbf{F}$ and find where $\nabla \cdot \mathbf{F} > 0$ algebraically.

on curl

6.5.2 Curl

Definition 11: Curl

The **curl** of a vector field $\mathbf{F} = \langle M, N, P \rangle$ is is a vector field on $\mathbb{R}^3$ defined by

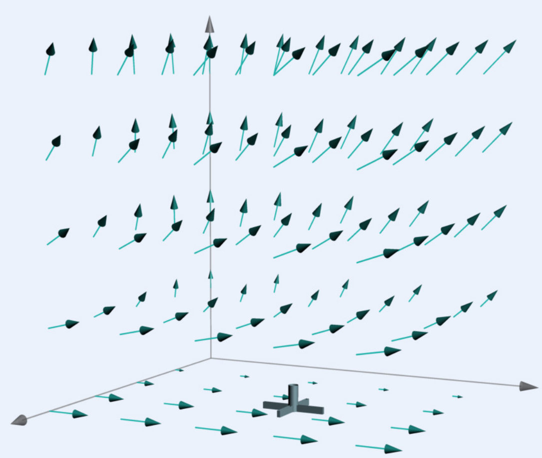
$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ M & N & P \end{vmatrix} = \langle \partial_y P - \partial_z N,\ -\partial_x P + \partial_z M,\ \partial_x N - \partial_y M \rangle.$$

We've already seen the third component, $\partial_x N - \partial_y M$, in Green's theorem. In the context, $\partial_x N - \partial_y M$ measures the line integral (per unit of enclosed area) around a small loop perpendicular to the z-axis. This quantity is called **circulation density**. The other two components of $\nabla \times \mathbf{F}$ supplement this information by providing the circulation density with respect to the x and y directions.

We can visualize this idea physically by interpreting the vector field as a flow velocity and imagining placing a small paddle wheel (with an axis of rotation in the x, y, or z direction) at a particular point in this field. The corresponding component of the curl measures how rapidly and in which direction this paddle wheel turns.

Example 6.5.2

Consider the vector field $\mathbf{F}$ shown. Find the sign of the z-component of the curl of $\mathbf{F}$ at any point in the xy-plane.



Solution

We can see that if we place a small paddlewheel at a point of interest (as shown in the figure above) that it will rotate in the counterclockwise direction, because the vectors on the right (meaning the side where x is larger) push harder than the vectors on the left. Therefore, the z-component of the curl is $\boxed{\text{positive}}$.

When we studied gradients, we learned that directional derivatives of a function in the coordinate directions determine its directional derivatives in all directions: the derivative in the $\langle u_1, u_2 \rangle$ direction is equal a linear combination with weights u_1 and u_2 of the derivatives in the coordinate directions. The same idea holds for the curl:

Theorem 6.5.1

!!!

The circulation density of a vector field $\mathbf{F}$ with respect to a unit vector $\mathbf{u}$ is equal to $(\nabla \times \mathbf{F}) \cdot \mathbf{u}$.

Example 6.5.3

Find the orientation for which a paddle wheel at the point $(1, 1, 1)$ in the velocity field $\langle xyz, x^2 - y, z \rangle$ which will maximize how fast it spins.

Solution

We calculate the curl: $\nabla \times \mathbf{F} = \langle 0, xy, -xz + 2x \rangle$, which at the point $(1, 1, 1)$ is equal to $\langle 0, 1, 1 \rangle$. Since the dot product of a fixed vector $\mathbf{v}$ with a unit vector is maximized when the unit vector is aligned with $\mathbf{v}$, we see that the paddle wheel should be oriented so that its axis is in the direction $\boxed{\left\langle 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle}$.

Exercise 6.5.3

Calculate $\nabla \times \mathbf{F}$, where $\mathbf{F} = \langle e^{\sin \log x} + y^2, -2z, y^3 + \cos z \rangle$.

Exercise 6.5.4

Show that the curl of a conservative vector field is zero.

on the divergence theorem

6.6 Divergence theorem

Suppose that $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field. As discussed in the previous section, the divergence of $\mathbf{F}$ at each point measures the net flow of $\mathbf{F}$ out of a small region around that point, per unit volume. Note that to integrate $\nabla \cdot \mathbf{F}$ over a region D in $\mathbb{R}^3$, we divide D into many small pieces, multiply the volume of each piece by the value of $\nabla \cdot \mathbf{F}$ there, and sum the results. The contribution of each piece is the net flow out of that piece*, so when we add the contributions of all the pieces we get the net flow out of D. This idea is called the *divergence theorem*.

The contribution of each piece is the divergence somewhere in the piece times its volume, and the divergence equals flow per unit volume, so the product is the net flow out of the piece

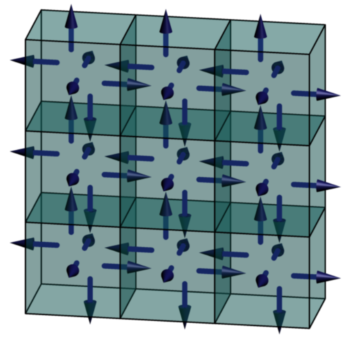


Figure 6.5 The sum of the flows of **F** (not shown) out of each cell is equal to the flow out of the whole box

Theorem 6.6.1: Divergence theorem

If $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field with continuous partial derivatives and D is a region in $\mathbb{R}^3$ bounded by a piecewise smooth surface $S = \partial D$, then

$$\overbrace{\iiint_D \nabla \cdot \mathbf{F}\, dV}^{\text{pointwise net flow density integrated over } D} = \overbrace{\iint_{\partial D} \mathbf{F} \cdot d\mathbf{S}}^{\text{total flow through } \partial D}.$$

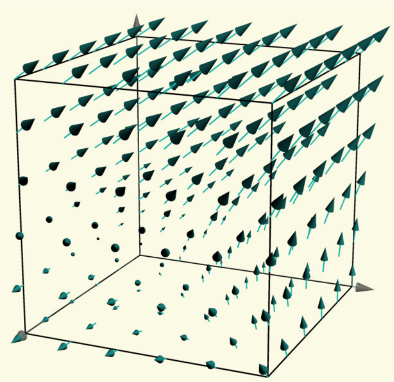
Example 6.6.1

Verify that the divergence theorem holds in the case where $\mathbf{F} = \langle x^2, 3z^2, 2z^2 + y^2 \rangle$ and $D = [0,1]^3$.

Solution

The divergence of $\mathbf{F}$ is $2x + 4z$, so the divergence theorem asserts that

$$\iiint_D (2x + 4z)\, dV = \iint_{\partial D} \langle x^2, 3z^2, 2z^2 + y^2 \rangle \cdot d\mathbf{A}.$$



The left-hand side equals

$$\int_0^1 \int_0^1 \int_0^1 (2x + 4z)\, dx\, dy\, dz = 3.$$

To evaluate the right-hand side directly, we split the boundary of the cube into its six square faces. For the top face, we get

$$\iint_{\text{top face}} \mathbf{F} \cdot d\mathbf{A} = \iint_{\text{top face}} \langle x^2, 3z^2, 2z^2 + y^2 \rangle \cdot \langle 0, 0, 1 \rangle dA = \int_0^1 \int_0^1 (2 + y^2)\, dx\, dy = \frac{7}{3},$$

where we've substituted 1 for z since $z = 1$ for every point in the top face. Likewise, the integral over the top face is

$$-\int_0^1 \int_0^1 (0 + y^2)\, dx\, dy = -\frac{1}{3},$$

where the negative sign comes from the fact that the outward-pointing normal on the bottom face is $\langle 0, 0, -1 \rangle$.

Similarly, the integral over the $x = 1$ face is

$$\int_0^1 \int_0^1 1\, dy\, dz = 1,$$

while the $x = 0$ face contributes 0. The $y = 1$ face yields

$$\int_0^1 \int_0^1 3z^2\, dx\, dz = 1,$$

and the $y = 0$ face gives $-\int_0^1 \int_0^1 3z^2\, dx\, dz = -1$. Indeed,

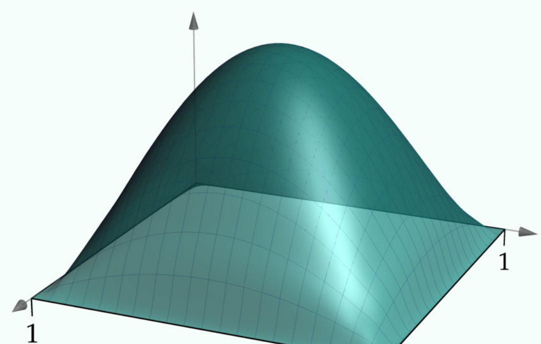
$$3 \overset{\checkmark}{=} \frac{7}{3} + \left(-\frac{1}{3}\right) + 1 + 0 + 1 + (-1).$$

Exercise 6.6.1

Verify the divergence theorem in the case where $\mathbf{F} = \langle x^2, y^2, z^2 \rangle$ and S is the unit sphere.

Exercise 6.6.2

Consider the vector field $\mathbf{F} = \langle x^3, xz, 1 - 3zx^2 \rangle$. Verify that the divergence of $\mathbf{F}$ is everywhere zero. Then use the divergence theorem to calculate the flow of $\mathbf{F}$ through the surface S shown. Note that this is not a closed surface: it excludes the square $[0, 1]^2 \times \{0\}$ on the bottom.



on Stokes' theorem

6.7 Stokes' theorem

The planar domain D in Green's theorem can be thought of as a surface S, in which case the conclusion of Green's theorem can be written as

$$\iint_S \nabla \times \mathbf{F} \cdot d\mathbf{A} = \int_{\partial S} \mathbf{F} \cdot d\mathbf{r}.$$

The argument for Green's theorem now applies even if S doesn't lie in a plane, because Theorem 6.5.1 tells us that $\nabla \times \mathbf{F} \cdot d\mathbf{A}$ measures the circulation around a small patch dA of the surface S. As we discussed for Green's theorem, if you divide a surface into many small patches and sum the circulations around all of them, you get the circulation around the boundary of the surface. This generalization of Green's theorem is known as *Stokes' theorem*.*

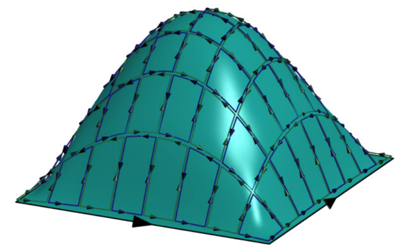


Figure 6.6 The circulation of **F** (not shown) around each patch sums to the circulation around the boundary of the surface

If you take a differential geometry course, you will learn a far more general result of the same name which implies Green's theorem and the divergence theorem as special cases

Theorem 6.7.1: Stokes' theorem

If $\mathbf{F} : \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field with continuous partial derivatives and S is a surface in $\mathbb{R}^3$, then

$$\overbrace{\iint_S \nabla \times \mathbf{F} \cdot d\mathbf{A}}^{\text{pointwise circulation density integrated over } S} = \overbrace{\int_{\partial S} \mathbf{F} \cdot d\mathbf{r}}^{\text{circulation around } \partial S}.$$

Example 6.7.1

Let $\mathbf{F} = \langle x\sin(\pi y), e^x, -\cos(\pi z)\rangle$. Find the flow of $\nabla \times \mathbf{F}$ through the surface shown in Exercise 6.6.2.

Solution

Stokes' theorem tells us that the flow of $\nabla \times \mathbf{F}$ through S is equal to the line integral of $\mathbf{F}$ around ∂S. We can see from the figure that ∂S consists four line segments, so we calculate $\int \mathbf{F}\, d\mathbf{r}$ along each one and sum the results. Integrating from $(0,0,0)$ to $(1,0,0)$, we get

$$\int_0^1 \mathbf{F}(x,0,0) \cdot \langle 1,0,0\rangle\, dx = 0.$$

From $(1,0,0)$ to $(1,1,0)$, we get

$$\int_0^1 \mathbf{F}(1,y,0) \cdot \langle 0,1,0\rangle\, dy = e.$$

From $(1,1,0)$ to $(0,1,0)$ we get

$$\int_0^1 \mathbf{F}(x,1,0) \cdot \langle -1,0,0\rangle\, dx = 0.$$

And finally from $(0,1,0)$ back to the origin we get

$$\int_0^1 \mathbf{F}(0,y,0) \cdot \langle 0,-1,0\rangle\, dy = -1.$$

So altogether the circulation of **F** around the boundary of S is $\boxed{e-1}$.

Example 6.7.1 shows that a surface can be deformed without changing the the flow of $\nabla \times \mathbf{F}$ through it, as long as it is deformed in such a way that its boundary is preserved:

Observation 6.7.1

In the context of Stokes' theorem, if S_1 and S_2 are surfaces whose boundaries are the same, then

$$\iint_{S_1} (\nabla \times \mathbf{F}) \cdot d\mathbf{A} = \iint_{S_2} (\nabla \times \mathbf{F}) \cdot d\mathbf{A}.$$

Exercise 6.7.1

Suppose that S is the surface consisting of the points on the sphere $x^2 + y^2 + z^2 = 1$ which are not inside the sphere $x^2 + y^2 + (z+1)^2 = 1$. Find $\iint_S \nabla \times \mathbf{F} \cdot d\mathbf{A}$, where $\mathbf{F} = \langle yz, x, e^{xyz} \rangle$.

Exercise 6.7.2

Suppose $\mathbf{F} = \langle xy, y, xz \rangle$. Find $\iint_S \nabla \times \mathbf{F} \cdot d\mathbf{A}$ where S is the portion of the unit sphere $x^2 + y^2 + z^2 = 1$ in the first octant.

Colophon

This text was typeset with pdfTEX, using `tcolorbox` and a version of the `mathpazo` package's Palatino fonts which was modified to borrow Greek symbols from Utopia and blackboard bold symbols from Computer Modern. The cover art is rendered using Ti*k*Z, from code on this StackExchange thread:

`http://tex.stackexchange.com/questions/85904/showcase-of-beautiful-title-page-done-in-tex`

The figures are all produced in Asymptote and are included using the `asymptote` LaTeX package. All the files necessary to produce this document are available at github.com/sswatson.

A Appendix

A.1 Sets and functions

A **set** is a collection of elements. These elements can be numbers, points, shapes, vectors, other sets, whatever. For example,

$$A = \{1, 4, 9\}$$

is the set consisting of the positive, single-digit perfect squares. The main thing you can do with a set is check whether a particular element is in it. For example, we say that $1 \in A$ (read "1 is an element of A"), while $2 \notin A$ ("2 is not an element of A").

Some sets with standard and specially typeset names include

- $\mathbb{R}$, the set of real numbers,
- $\mathbb{Q}$, the set of rational numbers,
- $\mathbb{Z}$, the set of integers, and
- $\mathbb{N}$, the set of natural numbers.

Subsets and set equality

We say that $A \subset B$ (read "A is a **subset** of B") if every element of A is an element of B. For example,

$$\{1, 4, 9\} \subset \{1, 2, 3, 4, 9, 10\}.$$

We say that two sets A and B are **equal** if $A \subset B$ and $B \subset A$. Note that

$$\{1, 1, 2\} = \{1, 2\} = \{2, 1\}.$$

since each element of each set is in the others. Thus we can see that sets "don't care" about repeated elements or order. All that matters is what is in and what is not. It is customary to write sets with repeats omitted, for clarity.

Intersections and unions

We write $A \cap B$, the **intersection** of A and B, for the set of all the elements that are in both A and B. So, for example,

$$\{1, 4, 9\} \cap \{x \in \mathbb{R} : x^2 > 15\} = \{4, 9\}.$$

That second set on the left-hand side, which is written in *set-builder* notation, is read as "the set of all real numbers x such that the square of x is greater than 15".

We write $A \cup B$, the **union** of A and B, for the set of all the elements that are in either A or B. So, for example,

$$\{1, 4, 9\} \cup \{1, 9, 25\} = \{1, 4, 9, 25\}.$$

Functions

If A and B are sets, then a function $f : A \to B$ is a rule that assigns a single element of B to each element of A. The set A is called the **domain** of f and B is called the **codomain** of f. Given a subset A' of A, we define the **image** $f(A')$ to be

$$f(A') = \{b \in B \, : \, \text{there exists } a \in A' \text{ so that } f(a) = b\}. \tag{A.1.1}$$

This is the set of all elements of B that get mapped to from some element of A'. The **range** of f is defined to be the set $f(A)$, which contains all the elements of B that get mapped to at least once.

Similarly, if $B' \subset B$, then the **preimage** $f^{-1}(B')$ of B' is defined by

$$f^{-1}(B') = \{a \in A \, : f(a) \in B'\}.$$

This is the subset of A consisting of every element of A that maps to some element of B'.

A function f is **injective** if no two elements in the domain map to the same element in the codomain; in other words if $f(a) = f(a')$ implies $a = a'$.

A function f is **surjective** if the range of f is equal to the codomain of f; in other words, if $b \in B$ implies that there exists $a \in A$ with $f(a) = b$.

A function f is **bijective** if it is both injective and surjective. This means that for every $b \in B$, there is exactly one $a \in A$ such that $f(a) \in b$. If f is bijective, then the function from B to A that maps $b \in B$ to the element $a \in A$ that satisfies $f(a) = b$ is called the **inverse** of f.

If $f : A \to B$ and $A' \subset A$, then the **restriction** of f to A is the function $f|_{A'} : A' \to B$ defined by $f|_{A'}(x) = f(x)$ for all $x \in A'$.

If the rule defining a function is sufficiently simple, we can describe the function using **anonymous function notation**. For example, $x \in \mathbb{R} \mapsto x^2 \in \mathbb{R}$, or $x \mapsto x^2$ for short, is the squaring function from $\mathbb{R}$ to $\mathbb{R}$. Note that bar on the left edge of the arrow, which distinguishes the arrow in anonymous function notation from the arrow between the domain and codomain of a named function.

Cartesian product

The **Cartesian product** of two sets A and B, denoted $A \times B$, is the set of all pairs (a, b) where $a \in A$ and $b \in B$. For example, $[0, 3] \times [0, 2]$ is a rectangle in the plane. We sometimes use exponents for a Cartesian product of a set with itself. Thus $[0, 1]^2$ is a unit square in $\mathbb{R}^2$, and $[0, 1]^3$ is a unit cube in $\mathbb{R}^3$.

A.2 Trig review

This appendix provides a streamlined presentation of trig which is intended to provide enough starting off points to recover everything else you need.

Trig Review

1. **Cosine and sine**. The basic trig functions are $\cos\theta$ and $\sin\theta$. The most important definition of these functions is the following: the cosine of an angle θ is equal to the **x-coordinate** of the point obtained by rotating $(1, 0)$ an angle of θ about the origin. Sine is the same, but with the y-coordinate instead of x.

   !!! This idea bears repeating: the point on the unit circle obtained by rotating $(1, 0)$ an angle θ about the origin is equal to $(\cos\theta, \sin\theta)$, **by definition of cosine and sine**.

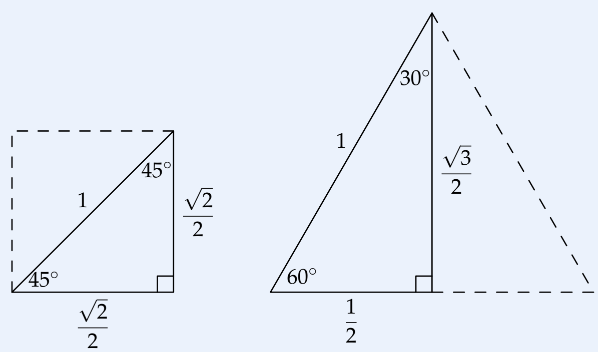
2. **The other ones**. The other four trig functions are simply abbreviations for various combinations of sine and cosine:

$$\sin\theta = \sin\theta \qquad \sec\theta = \tfrac{1}{\cos\theta}$$

$$\cos\theta = \cos\theta \qquad \csc\theta = \tfrac{1}{\sin\theta}$$

$$\tan\theta = \tfrac{\sin\theta}{\cos\theta} \qquad \cot\theta = \tfrac{\cos\theta}{\sin\theta}$$

3. **Special right triangles**. The following two triangles, each half of a regular polygon, can be hand for evaluating trig functions at special angles.



4. **Pythagorean identities** The famous identity $\sin^2\theta + \cos^2\theta = 1$ follows from the definition of sine and cosine combined with the Pythagorean theorem. Dividing both sides of this equation by $\sin^2\theta$ or $\cos^2\theta$, we get
$$\tan^2\theta + 1 = \sec^2\theta \qquad \text{and} \qquad 1 + \cot^2\theta = \csc^2\theta.$$

5. **Sum-angle formulas**. The sine sum-angle formula is worth memorizing: for all α and β,
$$\boxed{\sin(\alpha+\beta) = \sin\alpha\cos\beta + \sin\beta\cos\alpha}$$
The cosine sum-angle formula is worth memorizing too, although it can be derived fairly easily from the sine formula by substituting $\frac{\pi}{2} - \alpha$ for α and $-\beta$ for β. We get
$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta.$$

From these identities, we can derive many others. For example, setting $\alpha = \beta$ in the cosine sum-angle formula, we get

$$\cos 2\alpha = \cos^2 \alpha - \sin^2 \alpha.$$

Substituting $\cos^2 \alpha = 1 - \sin^2 \alpha$, we find that

$$\cos 2\alpha = 1 - 2 \sin^2 \alpha.$$

which can be solved to express $\sin^2 \alpha$ in terms of $\cos 2\alpha$.

A.3 Visualizing functions

Graphical visualization is an important conceptual tool for reasoning about the behavior of functions. There are a variety of different methods for visualizing functions (see the table on the next page for pictures):

Function Visualization Methods

1. **Graphs**. The graph of a function $f : \mathbb{R}^1 \to \mathbb{R}^1$ is the set of points of the form $(x, f(x))$, where x is in the domain of f. The graph of a function $f : \mathbb{R}^2 \to \mathbb{R}^1$ is the set of points of the form $(x, y, f(x, y))$, where (x, y) is in the domain of f. The graph uses one or two dimensions for the input and one dimension for the output, so it only works (as a visualization tool) for $f : \mathbb{R}^n \to \mathbb{R}^m$ if $m + n \leq 3$.

   The graph involves no loss of information; in principle, you can read off anything you want to know about a function from its graph. It depicts the domain and the codomain in the same picture.

2. **Level sets**. A level set of a function $f : \mathbb{R}^2 \to \mathbb{R}$ is the solution set of an equation of the form $f(x, y) = c$, where c is some constant. For example, the $c = 1$ level set of the function $x^2 + y^2$ is the unit circle. The level set of a function $f : \mathbb{R}^3 \to \mathbb{R}$ is typically a *surface*, For example, the level sets of $x^2 + y^2 + z^2$ are spheres.

   We can visualize a function by drawing its level sets. However, there are a couple drawbacks: we have to choose a discrete number of level sets to draw, and the picture doesn't tell us which c value corresponds to each level set, unless we draw that information in with colors or labels. When we visualize a function in this way, we are looking only at the *domain* of the function.

3. **Grid lines**. For a function T from $\mathbb{R}^2 to \mathbb{R}^2$, we can understand T as a transformation which moves points in the plane to other points in the plane, and we can visualize this geometric action by drawing the images of various grid lines under T. This visualization is drawn on the codomain side and loses some information about which grid lines match to which images.

4. **Traces**. We can visualize a function $\mathbf{r}$ from $\mathbb{R}^1$ to $\mathbb{R}^2$ or $\mathbb{R}^3$ by highlighting every point in $\mathbb{R}^2$ or $\mathbb{R}^3$ which is equal to $\mathbf{r}(t)$ for some $t \in \mathbb{R}$. This set of points is called the *trace* of $\mathbf{r}$.

   The trace is drawn entirely on the codomain side, which means that this visualization lacks information about which t value or values mapped to each highlighted point.

5. **Vector Fields**. For functions from $\mathbb{R}^2$ to $\mathbb{R}^2$ or $\mathbb{R}^3$ to $\mathbb{R}^3$, we can visualize them by interpreting the output value as a vector and depicting a discrete set of these vectors as small arrows drawn in place at the corresponding input values. Doing this requires scaling the vectors down so the picture doesn't get chaotic. This visualization incorporates inputs and outputs in the same picture, and information is lost about the absolute size of each vector and about what happens between the discrete set of input values shown.

Table A.1 below shows examples of each type of visualization, with the input (domain) dimension varying by row and the output (codomain) dimension by column. An example of a common type of visualization is shown for input-output pair of dimensions.

In a couple cases, the method shown isn't the only one in common use: a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ can also be drawn as a vector field (particularly if one is thinking of the outputs as vectors rather than points in $\mathbb{R}^2$), and the level set method shown for a function from $\mathbb{R}^3$ to $\mathbb{R}^1$ can also be applied to a function from $\mathbb{R}^2$ to $\mathbb{R}^1$.

Conventional names are used for each function; note that these vary by input and output dimension. Functions from $\mathbb{R}^3$ to $\mathbb{R}^2$ don't have a dedicated visualization method, although one could visualize each component of such a function separately, or identify the codomain $\mathbb{R}^2$ with the xy-plane in $\mathbb{R}^3$ to make a vector field representation.

Output Dimension

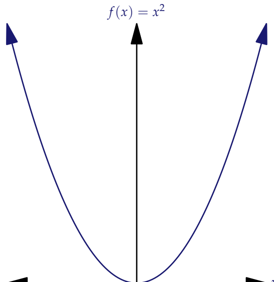
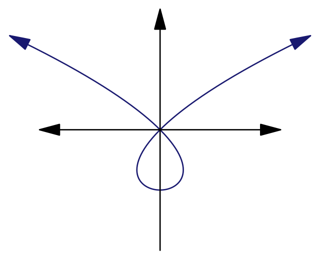
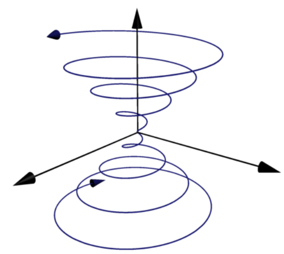
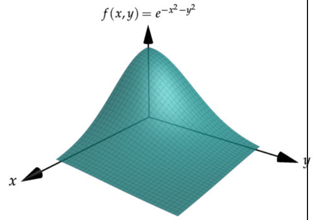
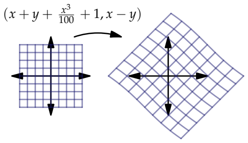
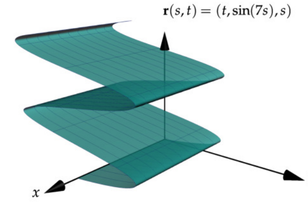
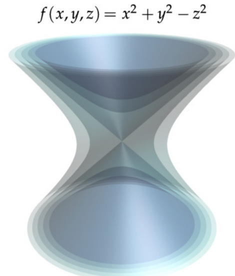
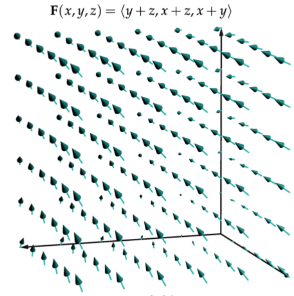
Input Dimension	1	2	3
1	$f(x) = x^2$  graph	$\mathbf{r}(t) = (t^3 - t, t^2 - 1)$  trace	$\mathbf{r}(t) = (t\cos(20t), t\sin(20t), t)$  trace
2	$f(x,y) = e^{-x^2-y^2}$  graph	$T(x,y) = (x + y + \frac{x^3}{100} + 1, x - y)$  grid lines	$\mathbf{r}(s,t) = (t, \sin(7s), s)$  trace
3	$f(x,y,z) = x^2 + y^2 - z^2$  level surfaces	(none)	$\mathbf{F}(x,y,z) = \langle y+z, x+z, x+y \rangle$  vector field

Table A.1 Different methods of visualizing functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, arranged by (n, m) pairs

A.4 Summation notation

Some shorthand is essential for writing sums with many terms. Perhaps the most common approach is to use ellipses:

$$1+2+3+\cdots+99+100=5050.$$

However, this approach is not ideal because the reader is left to infer the pattern.

When more precision is required, we would like to specify a formula for the kth term as well as a starting and ending value. For example, the sum

$$1+4+9+16+\cdots+100$$

can be written as "the sum of k^2 as k ranges from 1 to 10". The math notation that has been adopted for abbreviating this English phrase is the following:

$$\sum_{k=1}^{10} k^2$$

The variable k is called a *dummy variable*, since it is only there as a way to specify a formula for generating the terms. We could change each k to a different symbol without changing the essential meaning, which is "sum the first 10 positive perfect squares".

Exercise A.4.1

Find $\displaystyle\sum_{k=1}^{5} \frac{1}{k(k+1)}$.

Exercise A.4.2

Find $\displaystyle\sum_{k=1}^{5} \sum_{j=1}^{k} j$.

A.5 Polar, cylindrical, and spherical coordinate reference

Polar to Cartesian

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

Cylindrical to Cartesian

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

Spherical to Cartesian

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

Area/Volume differentials

$$\begin{cases} dA = r\,dr\,d\theta \\ dV = r\,dr\,d\theta\,dz \\ dV = \rho^2\sin\phi\,d\rho\,d\phi\,d\theta \end{cases}$$

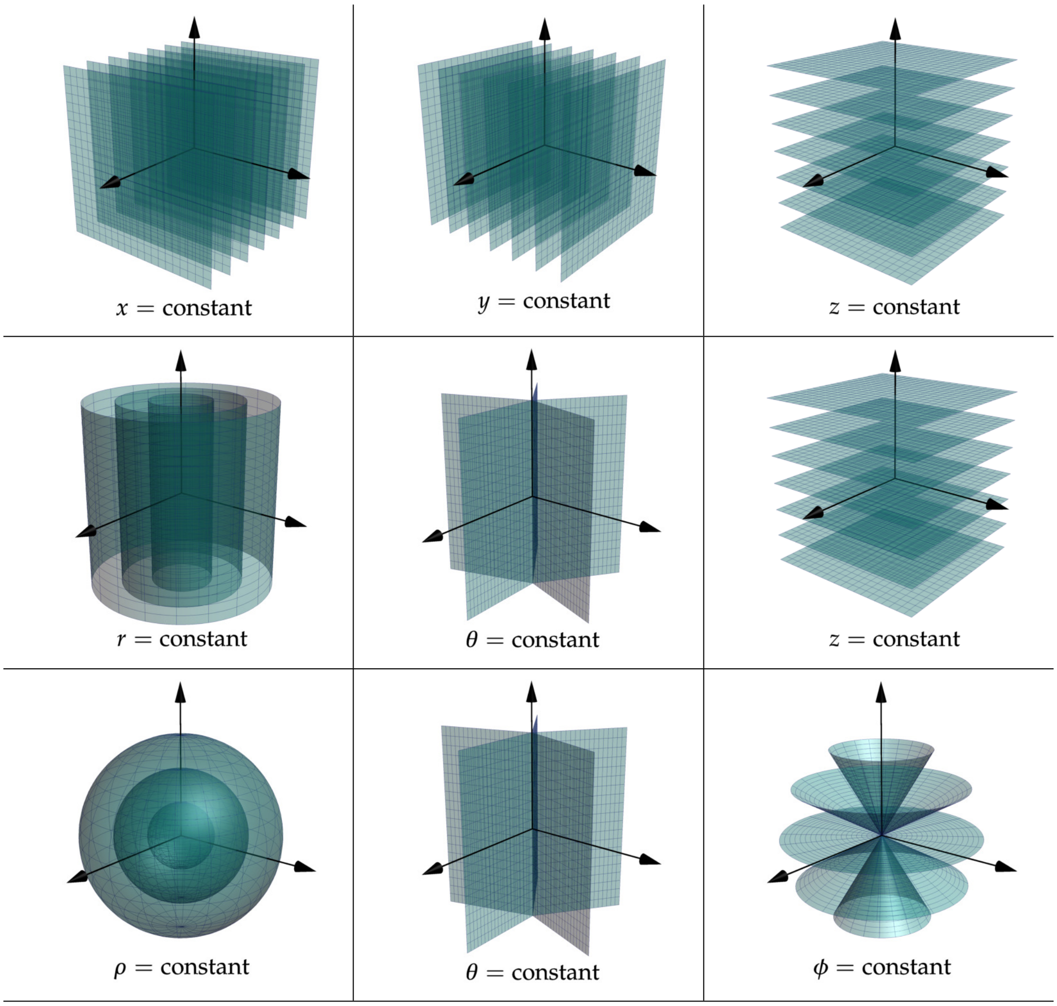


Table A.2 "Coordinate = constant" surfaces for each coordinate in the rectangular, cylindrical, and spherical systems.

A.6 SageMath

Math is more fun when you learn how to take advantage of computational resources to assist your learning. Some problem solving tasks are done much better by computers than people, and while in some cases it is important to gain facility with performing such calculations by hand, at some point you want to delegate tedious tasks to the computer and spend your time and attention on the more creative aspects of problem solving.

The open source project which has arguably made the most concerted and successful effort to be broadly useful to math students with minimal fuss is SageMath (or just Sage). You can use Sage in your browser without having to install anything (at sagecell.sagemath.org, or at cocalc.com for more extensive resources). You can use it essentially as a calculator (i.e., with minimal programming) for many things. However, if you do want to do something a little more involved, this is quite convenient, since Sage uses famously beginner-friendly Python as its language. This is where Sage really shines in comparison to a proprietary tool like Wolfram|Alpha, which is great for one-liners but not so much if you want to do a several-step computation.

Here are some examples of calculations you can do with Sage. You have to tell it that you're going to use `x` as a symbol using the `var` function; you can declare several symbols in this way using spaces. The text following the hashtag is a comment and is ignored by Sage.

```
> var('x y') # tell sage that you want to use x and y as symbols
> integrate(sqrt(x^2+1),x) + integrate(e^y * sin(y), y)
-1/2*(cos(y) - sin(y))*e^y + 1/2*sqrt(x^2 + 1)*x + 1/2*arcsinh(x)

> limit(sin(x^2)/x^2,x=0) # find a limit
1

> cos(3*x).trig_expand() # work with trig functions
cos(x)^3 - 3*cos(x)*sin(x)^2

> diff(x^x,x) # differentiate the function x^x
x^x*(log(x) + 1)

> find_local_maximum(sin(x) + cos(x), 0, 2*pi) # find the maximum of sin+cos over [0,2pi]
(1.414213562373095, 0.78539814681742492)

> plot(sin(x) + cos(x), 0, 2*pi) # plot a function

> [factor(x^n - 1) for n in [1..5]] # factor the first five polynomials of the form x^n - 1
[x - 1,
 (x + 1)*(x - 1),
 (x^2 + x + 1)*(x - 1),
 (x^2 + 1)*(x + 1)*(x - 1),
 (x^4 + x^3 + x^2 + x + 1)*(x - 1)]
```

Actually, some of these linked code snippets are in Julia, which is a newer language more suited to numerical work

Throughout the text, some computations are performed using Sage*. They are indicated with the CoCalc icon which can be clicked to open a page at cocalc.com showing the result as well as the code used to generate it.

If you want to learn more about Sage, I recommend the (freely available) book *Sage for Undergraduates* by Gregory Bard.