Calculus in 3D

Geometry, Vectors, and Multivariate Calculus

Zbigniew H. Nitecki
Tufts University

August 14, 2009

Preface

The present volume is a sequel to my earlier book, *Calculus Deconstructed: A Second Course in First-Year Calculus*, published by the Mathematical Association in 2009. It is designed, however, to be able to stand alone as a text in multivariate calculus. The current version is still very much a work in progress, and is subject to copyright.

The treatment here continues the basic stance of its predecessor, combining hands-on drill in techniques of calculation with rigorous mathematical arguments. However, there are some differences in emphasis. On one hand, the present text assumes a higher level of mathematical sophistication on the part of the reader: there is no explicit guidance in the rhetorical practices of mathematicians, and the theorem-proof format is followed a little more brusquely than before. On the other hand, there is much less familiar material being developed here, so more effort is expended on motivating various approaches and procedures. Where possible, I have followed my own predilection for geometric arguments over formal ones, although the two perspectives are naturally intertwined. At times, this feels more like an analysis text, but I have studiously avoided the temptation to give the general, n-dimensional versions of arguments and results that would seem natural to a mature mathematician: the book is, after all, aimed at the mathematical novice, and I have taken seriously the limitation implied by the "3D" in my title. This has the advantage, however, that many ideas can be motivated by natural geometric arguments. I hope that this approach lays a good intuitive foundation for further generalization that the reader will see in later courses.

Perhaps the fundamental subtext of my treatment is the way that the theory developed for functions of one variable interacts with geometry to handle higher-dimension situations. The progression here, after an initial chapter developing the tools of vector algebra in the plane and in space (including dot products and cross products), is first to view vector-valued functions of a single real variable in terms of parametrized curves—here, much of the theory translates very simply in a coordinate-wise way—then to consider real-valued functions of several variables both as functions with a vector input and in terms of surfaces in space (and level curves in the plane), and finally to vector fields as vector-valued functions of vector variables. This progression is not followed perfectly, as Chapter 4 intrudes between the differential and the integral calculus of real-valued functions of several variables to establish the change-of-variables formula for multiple integrals.

Idiosyncracies

There are a number of ways, some apparent, some perhaps more subtle, in which this treatment differs from the standard ones:

Parametrization: I have stressed the parametric representation of curves and surfaces far more, and beginning somewhat earlier, than many multivariate texts. This approach is essential for applying calculus to geometric objects, and it is also a beautiful and satisfying interplay between the geometric and analytic points of view. While Chapter 2 begins with a treatment of the conic sections from a classical point of view, this is followed by a catalogue of parametrizations of these curves, and in § 2.4 a consideration of what should constitute a curve in general. This leads naturally to the formulation of path integrals in § 2.5. Similarly, quadric surfaces are introduced in § 3.4 as level sets of quadratic polynomials in three variables, and the (three-dimensional) Implicit Function Theorem is introduced to show that any such curve is locally the graph of a function of two variables. The notion of parametrization of a surface is then introduced and exploited in § 3.5 to obtain the tangent planes of surfaces. When we get to surface integrals in § 5.4, this gives a natural way to define and calculate surface area and surface integrals of functions. This approach comes to full fruition in Chapter 6 in the formulationof the integral theorems of vector calculus.

Determinants and Cross-Products: There seem to be two approaches to determinants prevalent in the literature: one is formal and dogmatic, simply giving a recipe for calculation and proceeding from there with little motivation for it, the other is even more formal but elaborate, usually involving the theory of permutations. I believe I have come up with an approach to 2×2 and 3×3 determinants which is both motivated and rigorous, in § 1.6. Starting with the problem of calculating the area of a planar triangle from the coordinates of its vertices, we deduce a formula which is naturally written as the absolute value of a 2×2 determinant; investigation of the determinant itself leads to the notion of signed (*i.e.*, oriented) area (which has its own charm and prophesies the introduction of 2-forms in Chapter 6). Going to the analogous problem in space, we have the notion of an oriented area, represented by a vector (which we ultimately take as the definition of the cross-product, an approach taken for example by David Bressoud). We note that oriented areas project nicely, and from the projections of an oriented area vector onto the coordinate planes we come up with the formula for a cross-product as the expansion by minors along the first row of a 3×3 determinant. In the present treatment, various algebraic properties of determinants are developed as needed, and the relation to linear independence is argued geometrically.

I have found in my classes that the majority of students have already encountered (3×3) matrices and determinants in high school. I have therefore put some of the basic material about determinants in a separate appendix (Appendix F).

"Baby" Linear Algebra: I have tried to interweave into my narrative some of the basic ideas of linear algebra. As with determinants, I have found that the majority of my students (but not all) have already encountered vectors and matrices in their high school courses, so the basic material on matrix algebra and row reduction is covered quickly in the text but in more leisurely fashion in Appendix E. Linear independence and spanning for vectors in 3-space are introduced from a primarily geometric point of view, and the matrix representative of a linear function (*resp.* mapping) are introduced in § 3.2 (*resp.* § 4.1). The most sophisticated topics from linear algebra are eigenvectors and eigenfunctions, introduced in connection with the Principal Axis Theorem in § 3.9. The 2×2 case is treated separately in § 3.6, without the use of these tools, and the more complicated 3×3 case can be treated as optional. I have chosen to include this theorem, however, both because it leads to a nice understanding of quadratic forms (useful in understanding the second derivative test for critical points) and because its proof is a wonderful illustration of the synergy between calculus (Lagrange multipliers) and algebra.

Implicit and Inverse Function Theorems: I believe these theorems are among the most neglected important results in multivariate calculus. They take some time to absorb, and so I think it a good idea to introduce them at various stages in a student's mathematical education. In this treatment, I prove the Implicit Function Theorem for real-valued functions of two and three variables in § 3.4, and then formulate the Implicit Mapping Theorem for mappings $\mathbb{R}^3 \to \mathbb{R}^2$, as well as the Inverse Mapping Theorem for mappings $\mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbb{R}^3 \to \mathbb{R}^3$ in § 4.4. I use the geometric argument attributed to Goursat by [32] rather than the more sophisticated one using the contraction mapping theorem. Again, this is a more "hands on" approach than the latter.

Vector Fields vs. Differential Forms: A number of relatively recent treatments of vector calculus have been based exclusively on the theory of differential forms, rather than the traditional formulation using vector fields. I have tried this approach in the past, and find that it confuses the students at this level, so that they end up simply dealing with the theory on a purely formal basis. By contrast, I find it easier to motivate the operators and results of vector calculus by treating a vector field as the velocity of a moving fluid, and so have used this as my primary approach. However, the formalism of differential forms is very slick as a calculational device, and so I have also introduced this interwoven with the vector field approach. The main strength of the differential forms approach, of course, is that it generalizes to dimensions higher than 3; while I hint at this, it is one place where my self-imposed limitation to "3D" pays off.

Format

In general, I have continued the format of my previous book in this one.

As before, **exercises** come in four flavors:

Practice Problems serve as drill in calculation.

Theory Problems involve more ideas, either filling in gaps in the argument in the text or extending arguments to other cases. Some of these are a bit more sophisticated, giving details of results that are not sufficiently central to the exposition to deserve explicit proof in the text.

Challenge Problems require more insight or persistence than the standard theory problems. In my class, they are entirely optional, extra-credit assignments.

Historical Notes explore arguments from original sources. So far, there are many fewer of these then in the previous volume; I hope to remedy this as I study the history of the subject further.

There are more **appendices** in this volume than the previous one. To some extent, these reflect topics that seemed to add too much to the flow of the central exposition, but which I am loath to delete from the book. Very likely, some will be dropped from the final version. To summarize their contents:

Appendix A and Appendix B give the details of the classical arguments in Apollonius' treatment of conic sections and Pappus' proof of the focus-directrix property of conics. The results themselves are presented in § 2.1 of the text.

Appendix C gives a vector-based version of Newton's observations that Kepler's law of areas is equivalent to a central force field (*Principia*, Prop. I.1 and I.2) and the derivation of the inverse-square law from the fact that motion is along conic sections (*Principia*, Prop. I.11-13; we only do the first case, of an ellipse). An exercise at the end gives Newton's geometric proof of his Prop. I.1.

Appendix D develops the Frenet-Serret formulas for curves in space.

Appendix E gives a more leisurely and motivated treatment than is in the text of matrix algebra, row reduction, and rank of matrices.

Appendix F explains why 2×2 and 3×3 determinants can be calculated via expansion by minors along any row or column, that each is a multilinear function of its rows, and the relation between determinants and singularity of matrices.

Appendix G presents H. Schwartz's example showing that the definition of arclength as the supremum of lengths of piecewise linear approximations cannot be generalized to surface area. This helps justify the resort to differential formalism in defining surface area in § 5.4.

What's Missing?

The narrative so far includes far less historical material than the previous book. While before I was able to draw extensively on Edwards' history of (single-variable) calculus, among many other treatments, the history of multivariate calculus is far less well documented in the literature. I hope to draw out more information in the near future, but this requires digging a bit deeper than I needed to in the previous account.

I have also not peppered this volume with epigraphs. These were fun, and I might try to dig out some appropriate quotes for the present volume if time and energy permit. The jury is still out on this.

My emphasis on geometric arguments in this volume should result in more figures. I have been learning to use the packages `pst-3d` and `pst-solides3D`, which can create lovely 3D figures, and hope to expand the selection of pictures supplementing the text.

Acknowledgements

As with the previous book, I want to thank Jason Richards who as my grader in this course over several years contributed many corrections and useful comments about the text. I have also benefited greatly from much help with TeX packages especially from the e-forum on pstricks run by Herbert Voss. My colleague Loring Tu helped me better understand the role of orientation in the integration of differential forms. On the history side, Sandro Cappirini helped introduced me to the early history of vectors, and Lenore Feigenbaum and especially Michael N. Fried helped me with some vexing questions concerning Apollonius' classification of the conic sections.

Contents

1 Coordinates and Vectors

1.1 Locating Points in Space

Rectangular Coordinates

The geometry of the number line $\mathbb{R}$ is quite straightforward: the location of a real number x relative to other numbers is determined—and specified—by the inequalities between it and other numbers x': if $x < x'$ then x is to the *left* of x', and if $x > x'$ then x is to the *right* of x'. Furthermore, the **distance** between x and x' is just the difference $\triangle x = x' - x$ (*resp.* $x - x'$) in the first (*resp.* second) case, a situation summarized as the **absolute value**

$$|\triangle x| = \left|x - x'\right|.$$

When it comes to points in the plane, more subtle considerations are needed. The most familiar system for locating points in the plane is a **rectangular** or **Cartesian coordinate system**. We pick a distinguished point called the **origin** and denoted $\boldsymbol{\mathcal{O}}$.

Now we draw two **axes** through the origin: the first is called the **$\boldsymbol{x}$-axis** and is by convention *horizontal*, while the second, or **$\boldsymbol{y}$-axis**, is *vertical*. We regard each axis as a copy of the real line, with the origin corresponding to zero. Now, given a point P in the plane, we draw a rectangle with $\mathcal{O}$ and P as opposite vertices, and the two edges emanating from $\mathcal{O}$ lying along our axes (see Figure 1.1): thus, one of the vertices between $\mathcal{O}$ and P is a point on

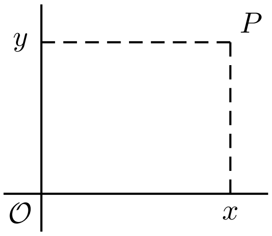


Figure 1.1: Rectangular Coordinates

the x-axis, corresponding to a number x called the **abcissa** of P; the other lies on the y-axis, and corresponds to the **ordinate** y of P. We then say that the (rectangular or Cartesian) **coordinates** of P are the two numbers (x, y). Note that the ordinate (*resp.* abcissa) of a point on the x-axis (*resp.* y-axis) is zero, so the point on the x-axis (*resp.* y-axis) corresponding to the number $x \in \mathbb{R}$ (*resp.* $y \in \mathbb{R}$) has coordinates $(x, 0)$ (*resp.* $(0, y)$).

The correspondence between points of the plane and pairs of real numbers, as their coordinates, is **one-to-one** (distinct points correspond to distinct pairs of numbers, and vice-versa), and **onto** (every point P in the plane corresponds to some pair of numbers (x, y), and conversely every pair of numbers (x, y) represents the coordinates of some point P in the plane). It will prove convenient to ignore the distinction between pairs of numbers and points in the plane: we adopt the notation $\mathbb{R}^{\mathbf{2}}$ for the collection of all pairs of real numbers, and we identify $\mathbb{R}^2$ with the collection of all points in the plane. We shall refer to "the point $P(x, y)$" when we mean "the point P in the plane whose (rectangular) coordinates are (x, y)".

The preceding description of our coordinate system did not specify which direction along each of the axes is regarded as positive (or increasing). We adopt the convention that (using geographic terminology) the x-axis goes "west-to-east", with "eastward" the increasing direction, and the y-axis goes "south-to-north", with "northward" increasing. Thus, points to the "west" of the origin (and of the y-axis) have negative abcissas, and points "south" of the origin (and of the x-axis) have negative ordinates (Figure 1.2).

The idea of using a pair of numbers in this way to locate a point in the plane was pioneered in the early seventeenth cenury by Pierre de Fermat (1601-1665) and René Descartes (1596-1650). By means of such a scheme, a plane curve can be identified with the **locus** of points whose coordinates satisfy some equation; the study of curves by analysis of the correspond-

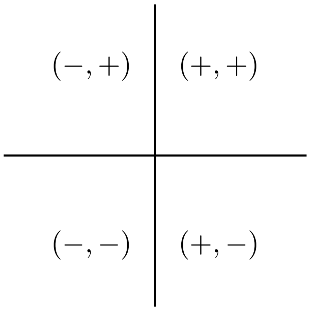


Figure 1.2: Direction Conventions

ing equations, called **analytic geometry**, was initiated in the research of these two men. Actually, it is a bit of an anachronism to refer to rectangular coordinates as "Cartesian", since both Fermat and Descartes often used **oblique coordinates**, in which the axes make an angle other than a right one.[1] Furthermore, Descartes in particular didn't really consider the meaning of negative values for the abcissa or ordinate.

One particular advantage of a rectangular coordinate system over an oblique one is the calculation of distances. If P and Q are points with respective rectangular coordinates (x_1, y_1) and (x_2, y_2), then we can introduce the point R which shares its last coordinate with P and its first with Q—that is, R has coordinates (x_2, y_1) (see Figure 1.3); then the triangle with vertices P, Q, and R has a right angle at R. Thus, the line segment PQ is

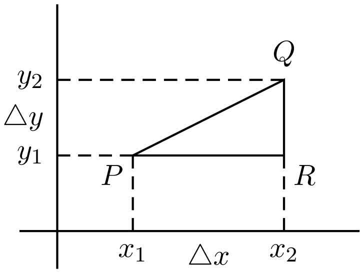


Figure 1.3: Distance in the Plane

the hypotenuse, whose length $|PQ|$ is related to the lengths of the "legs" by

[1] We shall explore some of the differences between rectangular and oblique coordinates in Exercise 14.

Pythagoras' Theorem

$$|PQ|^2 = |PR|^2 + |RQ|^2 .$$

But the legs are parallel to the axes, so it is easy to see that

$$\begin{aligned} |PR| &= |\triangle x| = |x_2 - x_1| \\ |RQ| &= |\triangle y| = |y_2 - y_1| \end{aligned}$$

and the distance from P to Q is related to their coordinates by

$$|PQ| = \sqrt{\triangle x^2 + \triangle y^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \tag{1.1}$$

In an oblique system, the formula becomes more complicated (Exercise 14).

The rectangular coordinate scheme extends naturally to locating points in space. We again distinguish one point as the **origin** $\mathcal{O}$, and draw a horizontal plane through $\mathcal{O}$, on which we construct a rectangular coordinate system. We continue to call the coordinates in this plane x and y, and refer to the horizontal plane through the origin as the ***xy*-plane**. Now we draw a new ***z*-axis** vertically through $\mathcal{O}$. A point P is located by first finding the point P_{xy} in the xy-plane that lies on the vertical line through P, then finding the signed "height" z of P above this point (z is negative if P lies below the xy-plane): the rectangular coordinates of P are the three real numbers (x, y, z), where (x, y) are the coordinates of P_{xy} in the rectangular system on the xy-plane. Equivalently, we can define z as the number corresponding to the intersection of the z-axis with the horizontal plane through P, which we regard as obtained by moving the xy-plane "straight up" (or down). Note the standing convention that, when we draw pictures of space, we regard the x-axis as pointing toward us (or slightly to our left) out of the page, the y-axis as pointing to the right in the page, and the z-axis as pointing up in the page (Figure 1.4).

This leads to the identification of the set $\mathbb{R}^3$ of triples (x, y, z) of real numbers with the points of space, which we sometimes refer to as **three dimensional space** (or **3-space**).

As in the plane, the distance between two points $P(x_1, y_1, z_1)$ and $Q(x_2, y_2, z_2)$ in $\mathbb{R}^3$ can be calculated by applying Pythagoras' Theorem to the right triangle PQR, where $R(x_2, y_2, z_1)$ shares its last coordinate with P and its other coordinates with Q. Details are left to you (Exercise 12); the resulting formula is

$$|PQ| = \sqrt{\triangle x^2 + \triangle y^2 + \triangle z^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \tag{1.2}$$

In what follows, we will denote the distance between P and Q by $\mathbf{dist}(\boldsymbol{P}, \boldsymbol{Q})$.

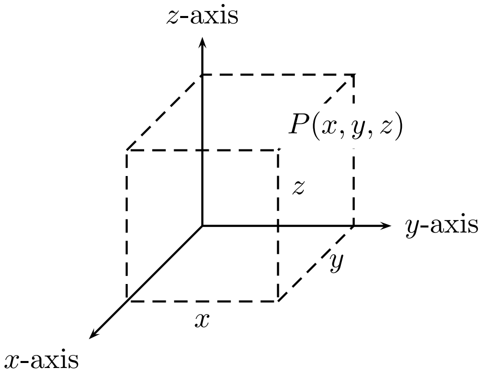


Figure 1.4: Pictures of Space

Polar and Cylindrical Coordinates

Rectangular coordinates are the most familiar system for locating points, but in problems involving rotations, it is sometimes convenient to use a system based on the direction and distance of a point from the origin.

For points in the plane, this leads to **polar coordinates**. Given a point P in the plane, we can locate it relative to the origin $\mathcal{O}$ as follows: think of the line ℓ through P and $\mathcal{O}$ as a copy of the real line, obtained by rotating the x-axis θ radians counterclockwise; then P corresponds to the real number r on ℓ. The relation of the *polar* coordinates (r, θ) of P to its *rectangular* coordinates (x, y) is illustrated in Figure 1.5, from which we see that

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta. \end{aligned} \tag{1.3}$$

The derivation of Equation (1.3) from Figure 1.5 requires a pinch of salt: we have drawn θ as an acute angle and x, y, and r as positive. In fact, when y is negative, our triangle has a *clockwise* angle, which can be interpreted as *negative* θ. However, as long as r is *positive*, relation (1.3) amounts to Euler's definition of the trigonometric functions (*Calculus Deconstructed*, p. 86). To interpret Figure 1.5 when r is *negative*, we move $|r|$ units in the *opposite* direction along ℓ. Notice that a reversal in the direction of ℓ amounts to a (further) rotation by π radians, so the point with polar coordinates (r, θ) also has polar coordinates $(-r, \theta + \pi)$.

In fact, while a given geometric point P has only one pair of *rectangular* coordinates (x, y), it has many pairs of *polar* coordinates. Given (x, y), r

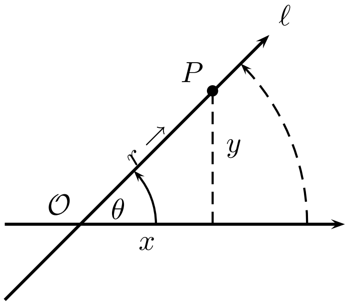


Figure 1.5: Polar Coordinates

can be either solution (positive or negative) of the equation

$$r^2 = x^2 + y^2 \tag{1.4}$$

which follows from a standard trigonometric identity. The angle by which the x-axis has been rotated to obtain ℓ determines θ only up to adding an even multiple of π: we will tend to measure the angle by a value of θ between 0 and 2π or between $-\pi$ and π, but any appropriate real value is allowed. Up to this ambiguity, though, we can try to find θ from the relation

$$\tan\theta = \frac{y}{x}.$$

Unfortunately, this determines only the "tilt" of ℓ, not its direction: to *really* determine the geometric angle of rotation (given r) we need both equations

$$\begin{array}{ll} \cos\theta & = \frac{x}{r} \\ \sin\theta & = \frac{y}{r}. \end{array} \tag{1.5}$$

Of course, either of these alone determines the angle up to a rotation by π radians (a "flip"), and only the *sign* in the other equation is needed to decide between one position of ℓ and its "flip".

Thus we see that the polar coordinates (r, θ) of a point P are subject to the ambiguity that, if (r, θ) is one pair of polar coordinates for P then so are $(r, \theta + 2n\pi)$ and $(-r, (2n+1)\pi)$ for any integer n (positive or negative).

Finally, we see that $r = 0$ precisely when P is the origin, so then the line ℓ is indeterminate: $r = 0$ together with *any* value of θ satisfies Equation (1.3), and gives the origin.

For example, to find the polar coordinates of the point P with rectangular coordinates $(-2\sqrt{3}, 2)$, we first note that

$$r^2 = (-2\sqrt{3})^2 + (2)^2 = 16.$$

Using the positive solution of this

$$r = 4$$

we have

$$\begin{aligned}\cos\theta &= -\frac{2\sqrt{3}}{4} = -\frac{\sqrt{3}}{2}\\ \sin\theta &= -\frac{2}{4} = \frac{1}{2}.\end{aligned}$$

The first equation says that θ is, up to adding multiples of 2π, one of $\theta = 5\pi/6$ or $\theta = 7\pi/6$, while the fact that $\sin\theta$ is positive picks out the first value. So one set of polar coordinates for P is

$$\begin{aligned}r &= 4\\ \theta &= \frac{5\pi}{6} + 2n\pi\end{aligned}$$

where n is any integer, while another set is

$$\begin{aligned}r &= -4\\ \theta &= \left(\frac{5\pi}{6} + \pi\right) + 2n\pi\\ &= \frac{11\pi}{6} + 2n\pi.\end{aligned}$$

It may be more natural to write this last expression as

$$\theta = -\frac{\pi}{6} + 2n\pi.$$

For problems in *space* involving rotations (or rotational symmetry) about a single axis, a convenient coordinate system locates a point P relative to the origin as follows (Figure 1.6): if P is not on the z-axis, then this axis together with the line $\mathcal{O}P$ determine a (vertical) plane, which can be regarded as the xz-plane rotated so that the x-axis moves θ radians counterclockwise (in the horizontal plane); we take as our coordinates the angle θ together with the abcissa and ordinate of P in *this* plane. The angle θ can be identified with the polar coordinate of the projection P_{xy} of P on the horizontal plane; the abcissa of P in the rotated plane is its distance from the z-axis, which is the

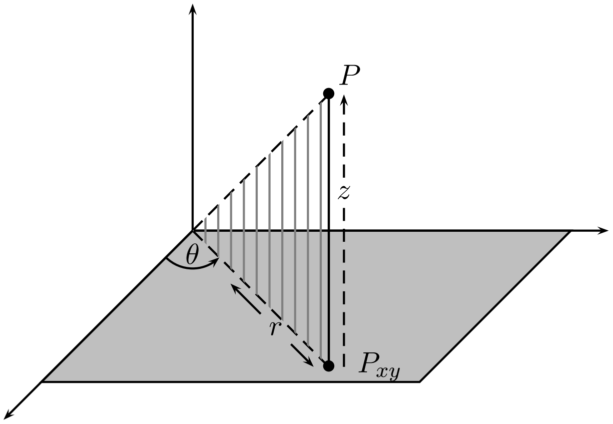


Figure 1.6: Cylindrical Coordinates

same as the polar coordinate r of P_{xy}; and its ordinate in this plane is the same as its vertical rectangular coordinate z.

We can think of this as a hybrid: combine the *polar* coordinates (r, θ) of the projection P_{xy} with the vertical *rectangular* coordinate z of P to obtain the **cylindrical coordinates** (r, θ, z) of P. Even though in principle r could be taken as negative, in this system it is customary to confine ourselves to $r \geq 0$. The relation between the cylindrical coordinates (r, θ, z) and the rectangular coordinates (x, y, z) of a point P is essentially given by Equation (1.3):

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= z. \end{aligned} \tag{1.6}$$

We have included the last relation to stress the fact that this coordinate is the same in both systems. The inverse relations are given by (1.4), (1.5) and the trivial relation $z = z$.

The name "cylindrical coordinates" comes from the geometric fact that the locus of the equation $r = c$ (which in polar coordinates gives a circle of radius c about the origin) gives a vertical cylinder whose axis of symmetry is the z-axis with radius c.

Cylindrical coordinates carry the ambiguities of polar coordinates: a point on the z-axis has $r = 0$ and θ arbitrary, while a point off the z-axis has θ determined up to adding *even* multiples of π (since r is taken to be positive).

For example, the point P with rectangular coordinates $(-2\sqrt{3}, 2, 4)$ has cylindrical coordinates

$$\begin{aligned} r &= 4 \\ \theta &= \frac{5\pi}{6} + 2n\pi \\ z &= 4. \end{aligned}$$

Spherical Coordinates

Another coordinate system in space, which is particularly useful in problems involving rotations around various axes through the origin (for example, astronomical observations, where the origin is at the center of the earth) is the system of **spherical coordinates**. Here, a point P is located relative to the origin $\mathcal{O}$ by measuring the distance of P from the origin

$$\rho = |\mathcal{O}P|$$

together with two angles: the angle θ between the xz-plane and the plane containing the z-axis and the line $\mathcal{O}P$, and the angle ϕ between the (positive) z-axis and the line $\mathcal{O}P$ (Figure 1.7). Of course, the *spherical* coordinate θ of P is identical to the *cylindrical* coordinate θ, and we use the same letter to indicate this identity. While θ is sometimes allowed to take on all real

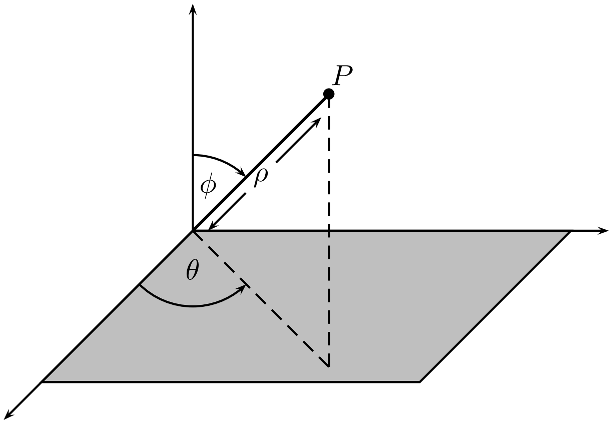


Figure 1.7: Spherical Coordinates

values, it is customary in spherical coordinates to restrict ϕ to $0 \leq \phi \leq \pi$. The relation between the cylindrical coordinates (r, θ, z) and the spherical coordinates (ρ, θ, ϕ) of a point P is illustrated in Figure 1.8 (which is drawn in the vertical plane determined by θ): [2]

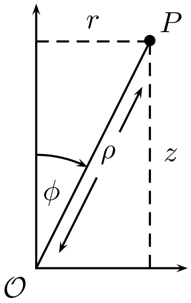


Figure 1.8: Spherical *vs.* Cylindrical Coordinates

$$\begin{aligned} r &= \rho \sin\phi \\ \theta &= \theta \\ z &= \rho\cos\phi. \end{aligned} \tag{1.7}$$

To invert these relations, we note that, since $\rho \geq 0$ and $0 \leq \phi \leq \pi$ by convention, z and r completely determine ρ and ϕ:

$$\begin{aligned} \rho &= \sqrt{r^2 + z^2} \\ \theta &= \theta \\ \phi &= \arccos\tfrac{z}{\rho}. \end{aligned} \tag{1.8}$$

The ambiguities in spherical coordinates are the same as those for cylindrical coordinates: the origin has $\rho = 0$ and both θ and ϕ arbitrary; any other point on the z-axis ($\phi = 0$ or $\phi = \pi$) has arbitrary θ, and for points off the z-axis, θ can (in principle) be augmented by arbitrary even multiples of π.

Thus, the point P with cylindrical coordinates

$$\begin{aligned} r &= 4 \\ \theta &= \frac{5\pi}{6} \\ z &= 4 \end{aligned}$$

[2] Be warned that in some of the engineering and physics literature the names of the two spherical angles are reversed, leading to potential confusion when converting between spherical and cylindrical coordinates.

has spherical coordinates

$$\begin{aligned}\rho &= 4\sqrt{2}\\ \theta &= \frac{5\pi}{6}\\ \phi &= \frac{\pi}{4}.\end{aligned}$$

Combining Equations (1.6) and (1.7), we can write the relation between the *spherical* coordinates (ρ, θ, ϕ) of a point P and its *rectangular* coordinates (x, y, z) as

$$\begin{aligned}x &= \rho\sin\phi\cos\theta\\ y &= \rho\sin\phi\sin\theta\\ z &= \rho\cos\phi.\end{aligned} \tag{1.9}$$

The inverse relations are a bit more complicated, but clearly, given x, y and z,

$$\rho = \sqrt{x^2 + y^2 + z^2} \tag{1.10}$$

and ϕ is completely determined (if $\rho \neq 0$) by the last equation in (1.9), while θ is determined by (1.4) and (1.6).

In spherical coordinates, the equation

$$\rho = R$$

describes the sphere of radius R centered at the origin, while

$$\phi = \alpha$$

describes a cone with vertex at the origin, making an angle α (*resp.* $\pi - \alpha$) with its axis, which is the positive (*resp.* negative) z-axis if $0 < \phi < \pi/2$ (*resp.* $\pi/2 < \phi < \pi$).

Exercises for § 1.1

Practice problems:

1. Find the distance between each pair of points (the given coordinates are rectangular):

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

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

   (c) $(-1, 2)$, $(2, 5)$

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

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

(f) $(3,5,7)$, $(1,7,5)$

2. What conditions on the components signify that $P(x,y,z)$ (rectangular coordinates) belongs to

   (a) the x-axis?

   (b) the y-axis?

   (c) the z-axis?

   (d) the xy-plane?

   (e) the xz-plane?

   (f) the yz-plane?

3. For each point with the given rectangular coordinates, find (i) its cylindrical coordinates, and (ii) its spherical coordinates:

   (a) $x = 0$, $y = 1$,, $z = -1$

   (b) $x = 1$, $y = 1$, $z = 1$

   (c) $x = 1$, $y = \sqrt{3}$, $z = 2$

   (d) $x = 1$, $y = \sqrt{3}$, $z = -2$

   (e) $x = -\sqrt{3}$, $y = 1$, $z = 1$

4. Given the spherical coordinates of the point, find its rectangular coordinates:

   (a) $\rho = 2, \quad \theta = \frac{\pi}{3}, \quad \phi = \frac{\pi}{2}$

   (b) $\rho = 1, \quad \theta = \frac{\pi}{4}, \quad \phi = \frac{2\pi}{3}$

   (c) $\rho = 2, \quad \theta = \frac{2\pi}{3}, \quad \phi = \frac{\pi}{4}$

   (d) $\rho = 1, \quad \theta = \frac{4\pi}{3}, \quad \phi = \frac{\pi}{3}$

5. What is the geometric meaning of each transformation (described in cylindrical coordinates) below?

   (a) $(r,\theta,z) \to (r,\theta,-z)$

   (b) $(r,\theta,z) \to (r,\theta+\pi,z)$

(c) $(r,\theta,z) \to (-r,\theta-\frac{\pi}{4},z)$

6. Describe the locus of each equation (in cylindrical coordinates) below:

   (a) $r = 1$

   (b) $\theta = \frac{\pi}{3}$

   (c) $z = 1$

7. What is the geometric meaning of each transformation (described in spherical coordinates) below?

   (a) $(\rho,\theta,\phi) \to (\rho,\theta+\pi,\phi)$

   (b) $(\rho,\theta,\phi) \to (\rho,\theta,\pi-\phi)$

   (c) $(\rho,\theta,\phi) \to (2\rho,\theta+\frac{\pi}{2},\phi)$

8. Describe the locus of each equation (in spherical coordinates) below:

   (a) $\rho = 1$

   (b) $\theta = \frac{\pi}{3}$

   (c) $\phi = \frac{\pi}{3}$

9. Express the plane $z = x$ in terms of (a) cylindrical and (b) spherical coordinates.

10. What conditions on the spherical coordinates of a point signify that it lies on

    (a) the x-axis?

    (b) the y-axis?

    (c) the z-axis?

    (d) the xy-plane?

    (e) the xz-plane?

    (f) the yz-plane?

11. A disc in space lies over the region $x^2+y^2 \leq a^2$, and the highest point on the disc has $z = b$. If $P(x,y,z)$ is a point of the disc, show that it has cylindrical coordinates satisfying

$$\begin{aligned} &0 \leq r \leq a \\ &0 \leq \theta \leq 2\pi \\ &z \leq b. \end{aligned}$$

Theory problems:

12. Prove the distance formula for $\mathbb{R}^3$ (Equation (1.2))

$$|PQ| = \sqrt{\triangle x^2 + \triangle y^2 + \triangle z^2} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

as follows (see Figure 1.9). Given $P(x_1, y_1, z_1)$ and $Q(x_2, y_2, z_2)$, let R be the point which shares its last coordinate with P and its first two coordinates with Q. Use the distance formula in $\mathbb{R}^2$ (Equation (1.1)) to show that

$$\operatorname{dist}(P, R) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

and then consider the triangle $\triangle PRQ$. Show that the angle at R is a right angle, and hence by Pythagoras' Theorem again,

$$\begin{aligned} |PQ| &= \sqrt{|PR|^2 + |RQ|^2} \\ &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}. \end{aligned}$$

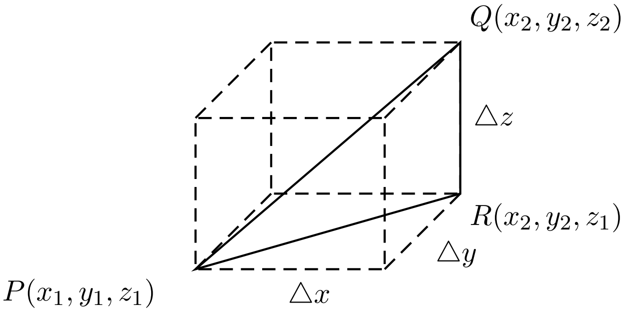


Figure 1.9: Distance in 3-Space

Challenge problem:

13. Use Pythagoras' Theorem and the angle-summation formulas to prove the **Law of Cosines**: If ABC is any triangle with sides

$$\begin{aligned} a &= |AC| \\ b &= |BC| \\ c &= |AB| \end{aligned}$$

and the angle at C is $\angle ACB = \theta$, then

$$c^2 = a^2 + b^2 - 2ab\cos\theta. \tag{1.11}$$

Here is one way to proceed (see Figure 1.10) Drop a perpendicular

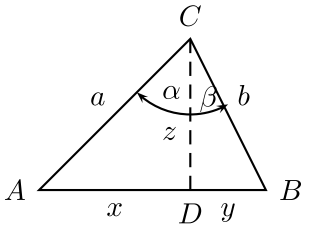


Figure 1.10: Law of Cosines

from C to AB, meeting AB at D. This divides the angle at C into two angles, satisfying

$$\alpha + \beta = \theta$$

and divides AB into two intervals, with respective lengths

$$|AD| = x$$
$$|DB| = y$$

so

$$x + y = c.$$

Finally, set

$$|CD| = z.$$

Now show the following:

$$\begin{aligned} x &= a\sin\alpha \\ y &= b\sin\beta \\ z &= a\cos\alpha = b\cos\beta \end{aligned}$$

and use this, together with Pythagoras' Theorem, to conclude that

$$\begin{aligned} a^2 + b^2 &= x^2 + y^2 + 2z^2 \\ c^2 &= x^2 + y^2 + 2xy \end{aligned}$$

and hence

$$c^2 = a^2 + b^2 - 2ab\cos(\alpha + \beta).$$

See Exercise 16 for the version of this which appears in Euclid.

14. **Oblique Coordinates:** Consider an **oblique coordinate system** on $\mathbb{R}^2$, in which the vertical axis is replaced by an axis making an angle of α radians with the horizontal one; denote the corresponding coordinates by (u, v) (see Figure 1.11).

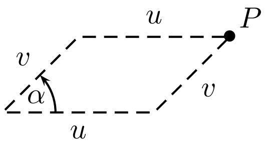


Figure 1.11: Oblique Coordinates

(a) Show that the oblique coordinates (u, v) and rectangular coordinates (x, y) of a point are related by

$$\begin{aligned} x &= u + v\cos\alpha \\ y &= v\sin\alpha. \end{aligned}$$

(b) Show that the distance of a point P with oblique coordinates (u, v) from the origin is given by

$$\operatorname{dist}(P, \mathcal{O}) = \sqrt{u^2 + v^2 + 2\,|uv|\cos\alpha}.$$

(c) Show that the distance between points P (with oblique coordinates (u_1, v_1)) and Q (with oblique coordinates (u_2, v_2)) is given by

$$\operatorname{dist}(P, Q) = \sqrt{\triangle u^2 + \triangle v^2 + 2\triangle u \triangle v\cos\alpha}$$

where

$$\begin{aligned}\triangle u &:= u_2 - u_1\\ \triangle v &:= v_2 - v_1.\end{aligned}$$

(*Hint:* There are two ways to do this. One is to substitute the expressions for the rectangular coordinates in terms of the oblique coordinates into the standard distance formula, the other is to use the law of cosines. Try them both.)

History note:

15. Given a right triangle with "legs" of respective lengths a and b and hypotenuse of length c (Figure 1.12) **Pythagoras' Theorem** says

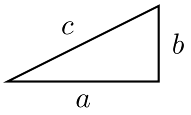


Figure 1.12: Right-angle triangle

that

$$c^2 = a^2 + b^2.$$

In this problem, we outline two quite different proofs of this fact.

First Proof: Consider the pair of figures in Figure 1.13.

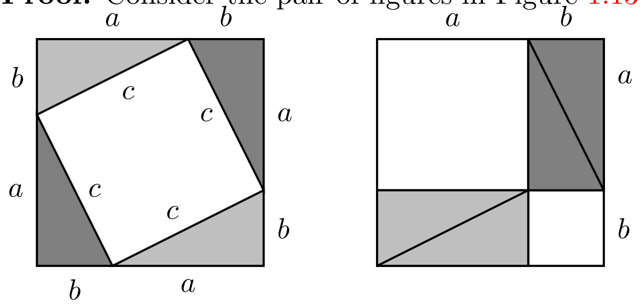


Figure 1.13: Pythagoras' Theorem by Dissection

(a) Show that the white quadrilateral on the left is a square (that is, show that the angles at the corners are right angles).

(b) Explain how the two figures prove Pythagoras' theorem.

A variant of Figure 1.13 was used by the twelfth-century Indian writer Bhāskara (b. 1114) to prove Pythagoras' Theorem. His proof consisted of a figure related to Figure 1.13 (without the shading) together with the single word "Behold!".

According to Eves [13, p. 158] and Maor [35, p. 63], reasoning based on Figure 1.13 appears in one of the oldest Chinese mathematical manuscripts, the *Caho Pei Suang Chin*, thought to date from the Han dynasty in the third century B.C.

The Pythagorean Theorem appears as Proposition 47, Book I of Euclid's *Elements* with a different proof (see below). In his translation of the *Elements*, Heath has an extensive commentary on this theorem and its various proofs [27, vol. I, pp. 350-368]. In particular, he (as well as Eves) notes that the proof above has been suggested as possibly the kind of proof that Pythagoras himself might have produced. Eves concurs with this judgement, but Heath does not.

Second Proof: The proof above represents one tradition in proofs of the Pythagorean Theorem, which Maor [35] calls "dissection proofs." A second approach is via the theory of proportions. Here is an example: again, suppose $\triangle ABC$ has a right angle at C; label the sides with lower-case versions of the labels of the opposite vertices (Figure 1.14) and draw a perpendicular CD from the right angle to the hypotenuse. This cuts the hypotenuse into two pieces of respective lengths c_1 and c_2, so

$$c = c_1 + c_2. \tag{1.12}$$

Denote the length of CD by x.

(a) Show that the two triangles $\triangle ACD$ and $\triangle CBD$ are both similar to $\triangle ABC$.

(b) Using the similarity of $\triangle CBD$ with $\triangle ABC$, show that

$$\frac{a}{c} = \frac{c_1}{a}$$

or

$$a^2 = cc_1.$$

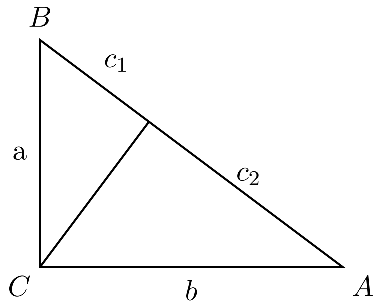


Figure 1.14: Pythagoras' Theorem by Proportions

(c) Using the similarity of $\triangle ACD$ with $\triangle ABC$, show that

$$\frac{c}{b} = \frac{b}{c_2}$$

or

$$b^2 = cc_2.$$

(d) Now combine these equations with Equation (1.12) to prove Pythagoras' Theorem.

The basic proportions here are those that appear in Euclid's proof of Proposition 47, Book I of the *Elements* , although he arrives at these via different reasoning. However, in Book VI, Proposition 31 , Euclid presents a generalization of this theorem: draw any polygon using the hypotenuse as one side; then draw similar polygons using the legs of the triangle; Proposition 31 asserts that the sum of the areas of the two polygons on the legs equals that of the polygon on the hypotenuse. Euclid's proof of this proposition is essentially the argument given above.

16. The Law of Cosines for an *acute* angle is essentially given by Proposition 13 in Book II of Euclid's *Elements*[27, vol. 1, p. 406] :

> *In acute-angled triangles the square on the side subtending the acute angle is less than the squares on the sides containing the acute angle by twice the rectangle contained by one of the sides about the acute angle, namely that on which the*

> *perpendicular falls, and the straight line cut off within by the perpendicular towards the acute angle.*

Translated into algebraic language (see Figure 1.15, where the acute angle is $\angle ABC$) this says

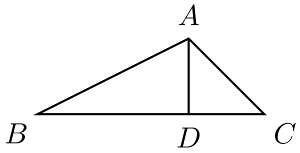


Figure 1.15: Euclid Book II, Proposition 13

$$|AC|^2 = |CB|^2 + |BA|^2 - |CB|\,|BD|\,.$$

Explain why this is the same as the Law of Cosines.

1.2 Vectors and Their Arithmetic

Many quantities occurring in physics have a magnitude and a direction—for example, forces, velocities, and accelerations. As a prototype, we will consider **displacements**.

Suppose a rigid body is pushed (without being turned) so that a distinguished spot on it is moved from position P to position Q (Figure 1.16). We represent this motion by a directed line segment, or arrow, going from P to Q and denoted $\overrightarrow{PQ}$. Note that this arrow encodes all the information about the motion of the *whole* body: that is, if we had distinguished a different spot on the body, initially located at P', then *its* motion would be described by an arrow $\overrightarrow{P'Q'}$ parallel to $\overrightarrow{PQ}$ and of the same length: in other words, the important characteristics of the displacement are its *direction* and *magnitude*, but *not* the location in space of its *initial* or *terminal points* (*i.e.*, its **tail** or **head**).

A second important property of displacement is the way different displacements combine. If we first perform a displacement moving our distinguished spot from P to Q (represented by the arrow $\overrightarrow{PQ}$) and then perform a second displacement moving our spot from Q to R (represented by the arrow $\overrightarrow{QR}$), the net effect is the same as if we had pushed directly from P to R. The arrow $\overrightarrow{PR}$ representing this net displacement is formed by putting

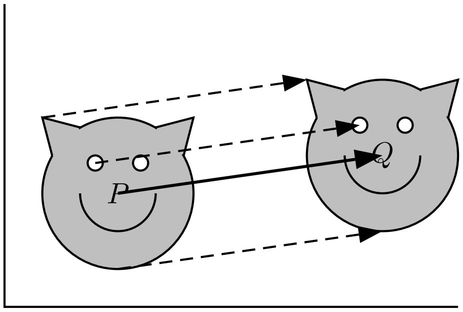


Figure 1.16: Displacement

arrow $\overrightarrow{QR}$ with its tail at the head of $\overrightarrow{PQ}$ and drawing the arrow from the tail of $\overrightarrow{PQ}$ to the head of $\overrightarrow{QR}$ (Figure 1.17). More generally, the net effect of several successive displacements can be found by forming a broken path of arrows placed tail-to-head, and forming a new arrow from the tail of the first arrow to the head of the last.

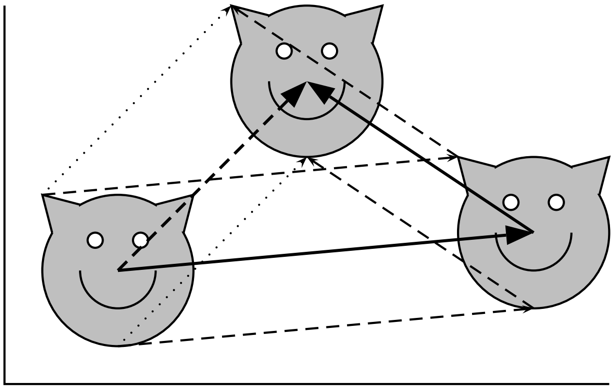


Figure 1.17: Combining Displacements

A representation of a physical (or geometric) quantity with these characteristics is sometimes called a **vectorial representation**. With respect to velocities, the "parallelogram of velocities" appears in the *Mechanica*, a work incorrectly attributed to, but contemporary with, Aristotle (384-322 BC) [24, vol. I, p. 344], and is discussed at some length in the *Me-*

chanics by Heron of Alexandria (*ca.* 75 AD) [24, vol. II, p. 348]. The vectorial nature of some physical quantities, such as velocity, acceleration and force, was well understood and used by Isaac Newton (1642-1727) in the *Principia* [39, Corollary 1, Book 1 (p. 417)]. In the late eighteenth and early nineteenth century, Paolo Frisi (1728-1784), Leonard Euler (1707-1783), Joseph Louis Lagrange (1736-1813), and others realized that other physical quantities, associated with rotation of a rigid body (torque, angular velocity, moment of a force), could also be usefully given vectorial representations; this was developed further by Louis Poinsot (1777-1859), Siméon Denis Poisson (1781-1840), and Jacques Binet (1786-1856). At about the same time, various geometric quantities (*e.g.*, areas of surfaces in space) were given vectorial representations by Gaetano Giorgini (1795-1874), Simon Lhuilier (1750-1840), Jean Hachette (1769-1834), Lazare Carnot (1753-1823)), Michel Chasles (1793-1880) and later by Hermann Grassmann (1809-1877) and Giuseppe Peano (1858-1932). In the early nineteenth century, vectorial representations of complex numbers (and their extension, quaternions) were formulated by several researchers; the term *vector* was coined by William Rowan Hamilton (1805-1865) in 1853. Finally, extensive use of vectorial properties of electromagnetic forces was made by James Clerk Maxwell (1831-1879) and Oliver Heaviside (1850-1925) in the late nineteenth century. However, a general theory of vectors was only formulated in the very late nineteenth century; the first elementary exposition was given by Edwin Bidwell Wilson (1879-1964) in 1901 [54], based on lectures by the American mathematical physicist Josiah Willard Gibbs (1839-1903)[3] [17].

By a **geometric vector** in $\mathbb{R}^3$ (or $\mathbb{R}^2$) we will mean an "arrow" which can be moved to any position, provided its direction and length are maintained.[4] We will denote vectors with a letter surmounted by an arrow, like this: $\overrightarrow{v}$. We define two operations on vectors. The **sum** of two vectors is formed by moving $\overrightarrow{w}$ so that its "tail" coincides in position with the "head" of $\overrightarrow{v}$, then forming the vector $\overrightarrow{v} + \overrightarrow{w}$ whose tail coincides with that of $\overrightarrow{v}$ and whose head coincides with that of $\overrightarrow{w}$ (Figure 1.18). If instead we place $\overrightarrow{w}$ with its tail at the position previously occupied by the tail of $\overrightarrow{v}$ and then move $\overrightarrow{v}$ so that its tail coincides with the head of $\overrightarrow{w}$, we form $\overrightarrow{w} + \overrightarrow{v}$, and it is clear that these two configurations form a parallelogram with diagonal

$$\overrightarrow{v} + \overrightarrow{w} = \overrightarrow{w} + \overrightarrow{v}$$

(Figure 5.18). This is the **commutative property** of vector addition.

[3] I learned much of this from Sandro Caparrini [6, 7, 8]. This narrative differs from the standard one, given by Michael Crowe [10]

[4] This mobility is sometimes expressed by saying it is a **free vector**.

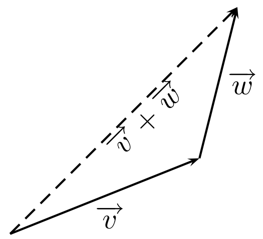


Figure 1.18: Sum of two vectors

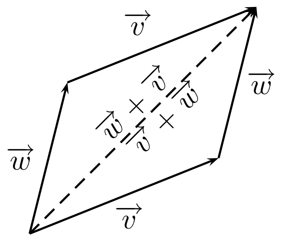

$\overrightarrow{v}$
$\overrightarrow{w} + \overrightarrow{v}$
$\overrightarrow{v} + \overrightarrow{w}$
$\overrightarrow{w}$
$\overrightarrow{w}$
$\overrightarrow{v}$

Figure 1.19: Parallelogram Rule (Commutativity of Vector Sums)

A second operation is **scaling** or **multiplication of a vector by a number**. We naturally define

$$\begin{aligned}
1\overrightarrow{v} &= \overrightarrow{v} \\
2\overrightarrow{v} &= \overrightarrow{v} + \overrightarrow{v} \\
3\overrightarrow{v} &= \overrightarrow{v} + \overrightarrow{v} + \overrightarrow{v} = 2\overrightarrow{v} + \overrightarrow{v}
\end{aligned}$$

and so on, and then define *rational* multiples by

$$\overrightarrow{v} = \frac{m}{n}\overrightarrow{w} \Leftrightarrow n\overrightarrow{v} = m\overrightarrow{w};$$

finally, suppose

$$\frac{m_i}{n_i} \to \ell$$

is a convergent sequence of rationals. For any fixed vector $\overrightarrow{v}$, if we draw arrows representing the vectors $(m_i/n_i)\overrightarrow{v}$ with all their tails at a fixed position, then the heads will form a convergent sequence of points along a line, whose limit is the position for the head of $\ell\overrightarrow{v}$. Alternatively, if we pick a unit of length, then for any vector $\overrightarrow{v}$ and any positive real number r, the vector $r\overrightarrow{v}$ has the same direction as $\overrightarrow{v}$, and its length is that of

$\overrightarrow{v}$ multiplied by r. For this reason, we refer to real numbers (in a vector context) as **scalars**.

If

$$\overrightarrow{u} = \overrightarrow{v} + \overrightarrow{w}$$

then it is natural to write

$$\overrightarrow{v} = \overrightarrow{u} - \overrightarrow{w}$$

and from this (Figure 1.20) it is natural to define the negative $-\overrightarrow{w}$ of a vector $\overrightarrow{w}$ as the vector obtained by interchanging the head and tail of $\overrightarrow{w}$. This allows us to also define multiplication of a vector $\overrightarrow{v}$ by any *negative* real number $r = -|r|$ as

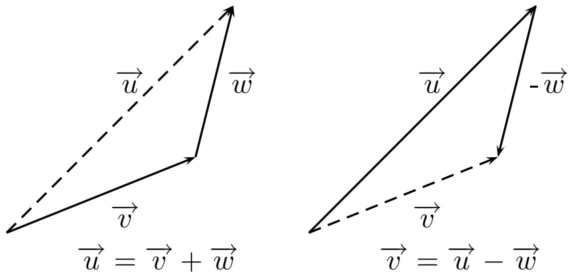

$\overrightarrow{u}$
$\overrightarrow{w}$
$\overrightarrow{v}$
$\overrightarrow{u} = \overrightarrow{v} + \overrightarrow{w}$
$\overrightarrow{u}$
-$\overrightarrow{w}$
$\overrightarrow{v}$
$\overrightarrow{v} = \overrightarrow{u} - \overrightarrow{w}$

Figure 1.20: Difference of vectors

$$(-|r|)\overrightarrow{v} := |r|(-\overrightarrow{v})$$

—that is, we reverse the direction of $\overrightarrow{v}$ and "scale" by $|r|$.

Addition of vectors (and of scalars) and multiplication of vectors by scalars have many formal similarities with addition and multiplication of numbers. We list the major ones (the first of which has already been noted above):

- Addition of vectors is

  commutative: $\overrightarrow{v} + \overrightarrow{w} = \overrightarrow{w} + \overrightarrow{v}$, and

  associative: $\overrightarrow{u} + (\overrightarrow{v} + \overrightarrow{w}) = (\overrightarrow{u} + \overrightarrow{v}) + \overrightarrow{w}$.

- Multiplication of vectors by scalars

  distributes over vector sums: $r(\overrightarrow{v} + \overrightarrow{w}) = r\overrightarrow{w} + r\overrightarrow{v}$, and

  distributes over scalar sums: $(r+s)\overrightarrow{v} = r\overrightarrow{v} + s\overrightarrow{v}$.

We will explore some of these properties further in Exercise 3.

The interpretation of displacements as vectors gives us an alternative way to represent vectors. We will say that an arrow representing the vector $\overrightarrow{v}$ is in **standard position** if its tail is at the origin. Note that in this case the vector is completely determined by the position of its head, giving us a natural correspondence between *vectors* $\overrightarrow{v}$ in $\mathbb{R}^3$ (or $\mathbb{R}^2$) and *points* $P \in \mathbb{R}^3$ (*resp.* $\mathbb{R}^2$). $\overrightarrow{v}$ corresponds to P if the arrow $\overrightarrow{\mathcal{O}P}$ from the origin to P is a representation of $\overrightarrow{v}$: that is, $\overrightarrow{v}$ is the vector representing that displacement of $\mathbb{R}^3$ which moves the origin to P; we refer to $\overrightarrow{v}$ as the **position vector** of P. We shall make extensive use of the correspondence between vectors and points, often denoting a point by its position vector $\overrightarrow{p} \in \mathbb{R}^3$.

Furthermore, using rectangular coordinates we can formulate a numerical specification of vectors in which addition and multiplication by scalars is very easy to calculate: if $\overrightarrow{v} = \overrightarrow{\mathcal{O}P}$ and P has rectangular coordinates (x, y, z), we identify the vector $\overrightarrow{v}$ with the triple of numbers (x, y, z) and write $\overrightarrow{v} = (x, y, z)$. We refer to x, y and z as the **components** or **entries** of $\overrightarrow{v}$. Then if $\overrightarrow{w} = \overrightarrow{\mathcal{O}Q}$ where $Q = (\triangle x, \triangle y, \triangle z)$ (that is, $\overrightarrow{w} = (\triangle x, \triangle y, \triangle z)$), we see from Figure 1.21 that

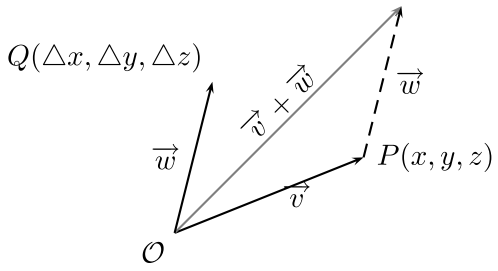


Figure 1.21: Componentwise addition of vectors

$$\overrightarrow{v} + \overrightarrow{w} = (x + \triangle x, y + \triangle y, z + \triangle z);$$

that is, *we add vectors componentwise.*

Similarly, if r is any scalar and $\overrightarrow{v} = (x, y, z)$, then

$$r\overrightarrow{v} = (rx, ry, rz):$$

a scalar multiplies all entries of the vector.

This representation points out the presence of an exceptional vector—the **zero vector**

$$\overrightarrow{0} := (0,0,0)$$

which is the result of either multiplying an arbitrary vector by the scalar zero

$$0\overrightarrow{v} = \overrightarrow{0}$$

or of subtracting an arbitrary vector from itself

$$\overrightarrow{v} - \overrightarrow{v} = \overrightarrow{0}.$$

As a *point*, $\overrightarrow{0}$ corresponds to the origin $\mathcal{O}$ itself. As an "*arrow*", its tail and head are at the same position. As a *displacement*, it corresponds to not moving at all. Note in particular that *the zero vector does not have a well-defined direction*—a feature which will be important to remember in the future. From a formal, algebraic point of view, the zero *vector* plays the role for *vector* addition that is played by the *number* zero for addition of *numbers*: it is an **additive identity element**, which means that adding it to any vector gives back that vector:

$$\overrightarrow{v} + \overrightarrow{0} = \overrightarrow{v} = \overrightarrow{0} + \overrightarrow{v}.$$

A final feature that is brought out by thinking of vectors in $\mathbb{R}^3$ as triples of numbers is that we can recover the entries of a vector geometrically. Note that if $\overrightarrow{v} = (x,y,z)$ then we can write

$$\begin{aligned}\overrightarrow{v} &= (x,0,0) + (0,y,0) + (0,0,z)\\ &= x(1,0,0) + y(0,1,0) + z(0,0,1).\end{aligned}$$

This means that any vector in $\mathbb{R}^3$ can be expressed as a sum of scalar multiples (or **linear combination**) of three specific vectors, known as the **standard basis** for $\mathbb{R}^3$, and denoted

$$\begin{aligned}\overrightarrow{\imath} &= (1,0,0)\\ \overrightarrow{\jmath} &= (0,1,0)\\ \overrightarrow{k} &= (0,0,1).\end{aligned}$$

It is easy to see that these are the vectors of length 1 pointing along the three (positive) coordinate axes (see Figure 1.22). Thus, every vector $\overrightarrow{v} \in \mathbb{R}^3$ can be expressed as

$$\overrightarrow{v} = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}.$$

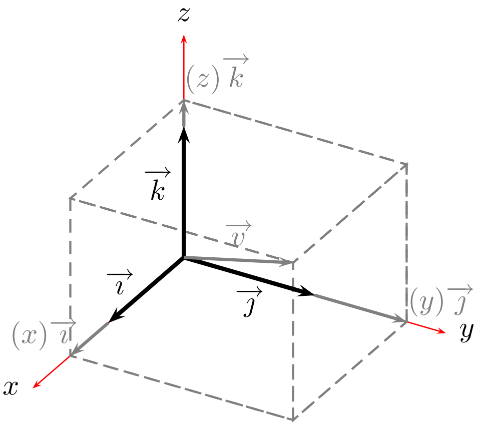

z
$(z)\overrightarrow{k}$
$\overrightarrow{k}$
$\overrightarrow{v}$
$\overrightarrow{\imath}$
$\overrightarrow{\jmath}$
$(y)\overrightarrow{\jmath}$
y
$(x)\overrightarrow{\imath}$
x

Figure 1.22: The Standard Basis for $\mathbb{R}^3$

We shall find it convenient to move freely between the coordinate notation $\overrightarrow{v} = (x, y, z)$ and the "arrow" notation $\overrightarrow{v} = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$; generally, we adopt coordinate notation when $\overrightarrow{v}$ is regarded as a position vector, and "arrow" notation when we want to picture it as an arrow in space.

We began by thinking of a vector $\overrightarrow{v}$ in $\mathbb{R}^3$ as determined by its magnitude and its direction, and have ended up thinking of it as a triple of numbers. To come full circle, we recall that the vector $\overrightarrow{v} = (x, y, z)$ has as its standard representation the arrow $\overrightarrow{\mathcal{O}P}$ from the origin $\mathcal{O}$ to the point P with coordinates (x, y, z); thus its magnitude (or **length**, denoted $\left|\overrightarrow{\boldsymbol{v}}\right|$) is given by the distance formula

$$|\overrightarrow{v}| = \sqrt{x^2 + y^2 + z^2}.$$

When we want to specify the **direction** of $\overrightarrow{v}$, we "*point*", using as our standard representation the **unit vector**—that is, the vector of length 1—in the direction of $\overrightarrow{v}$. From the scaling property of multiplication by real numbers, we see that the unit vector in the direction of a vector $\overrightarrow{v}$ ($\overrightarrow{v} \neq \overrightarrow{0}$) is

$$\overrightarrow{u} = \frac{1}{|\overrightarrow{v}|}\overrightarrow{v}.$$

In particular, the standard basis vectors $\overrightarrow{\imath}$, $\overrightarrow{\jmath}$, and $\overrightarrow{k}$ are unit vectors along the (positive) coordinate axes.

This formula for unit vectors gives us an easy criterion for deciding whether two vectors point in parallel directions. Given (nonzero[5]) vectors $\overrightarrow{v}$ and $\overrightarrow{w}$, the respective unit vectors in the same direction are

$$\begin{aligned}\overrightarrow{u}_{\overrightarrow{v}} &= \frac{1}{|\overrightarrow{v}|}\overrightarrow{v}\\ \overrightarrow{u}_{\overrightarrow{w}} &= \frac{1}{|\overrightarrow{w}|}\overrightarrow{w}.\end{aligned}$$

The two vectors $\overrightarrow{v}$ and $\overrightarrow{w}$ point in the *same* direction precisely if the two unit vectors are equal

$$\overrightarrow{u}_{\overrightarrow{v}} = \overrightarrow{u}_{\overrightarrow{w}} = \overrightarrow{u}$$

or

$$\begin{aligned}\overrightarrow{v} &= |\overrightarrow{v}|\,\overrightarrow{u}\\ \overrightarrow{w} &= |\overrightarrow{w}|\,\overrightarrow{u}.\end{aligned}$$

This can also be expressed as

$$\begin{aligned}\overrightarrow{v} &= \lambda\overrightarrow{w}\\ \overrightarrow{w} &= \frac{1}{\lambda}\overrightarrow{v}\end{aligned}$$

where the (positive) scalar λ is

$$\lambda = \frac{|\overrightarrow{v}|}{|\overrightarrow{w}|}.$$

Similarly, the two vectors point in *opposite* directions if the two unit vectors are *negatives* of each other, or

$$\begin{aligned}\overrightarrow{v} &= \lambda\overrightarrow{w}\\ \overrightarrow{w} &= \frac{1}{\lambda}\overrightarrow{v}\end{aligned}$$

where the *negative* scalar λ is

$$\lambda = -\frac{|\overrightarrow{v}|}{|\overrightarrow{w}|}.$$

So we have shown

[5] A vector is **nonzero** if it is not equal to the zero vector. Thus, *some* of its entries can be zero, but *not all* of them.

Remark 1.2.1. *For two nonzero vectors* $\overrightarrow{v} = (x_1, y_1, z_1)$ *and* $\overrightarrow{w} = (x_2, y_2, z_2)$, *the following are equivalent:*

- $\overrightarrow{v}$ *and* $\overrightarrow{w}$ *point in parallel or opposite directions;*
- $\overrightarrow{v} = \lambda\overrightarrow{w}$ *for some nonzero scalar* λ*;*
- $\overrightarrow{w} = \lambda'\overrightarrow{v}$ *for some nonzero scalar* λ'*;*
- $\frac{x_1}{x_2} = \frac{y_1}{y_2} = \frac{z_1}{z_2} = \lambda$ *for some nonzero scalar* λ *(where if one of the entries of* $\overrightarrow{w}$ *is zero, so is the corresponding entry of* $\overrightarrow{v}$*, and the corresponding ratio is omitted from these equalities.)*
- $\frac{x_2}{x_1} = \frac{y_2}{y_1} = \frac{z_2}{z_1} = \lambda'$ *for some nonzero scalar* λ' *(where if one of the entries of* $\overrightarrow{w}$ *is zero, so is the corresponding entry of* $\overrightarrow{v}$*, and the corresponding ratio is omitted from these equalities.)*

The values of λ *(resp.* λ'*) are the same wherever they appear above, and* λ' *is the reciprocal of* λ*.*

λ *(hence also* λ'*) is* positive *precisely if* $\overrightarrow{v}$ *and* $\overrightarrow{w}$ *point in the* same *direction, and* negative *precisely if they point in* opposite *directions.*

Two (nonzero) vectors are **linearly dependent** if they point in *either* the same or opposite directions—that is, if we picture them as arrows from a common initial point, then the two heads and the common tail fall on a line (this terminology will be extended in Exercise 6—but for more than two vectors, the condition is more complicated). Vectors which are not linearly dependent are **linearly independent**.

Exercises for § 1.2

Practice problems:

1. In each part, you are given two vectors, $\overrightarrow{v}$ and $\overrightarrow{w}$. Find (i) $\overrightarrow{v} + \overrightarrow{w}$; (ii) $\overrightarrow{v} - \overrightarrow{w}$; (iii) $2\overrightarrow{v}$; (iv) $3\overrightarrow{v} - 2\overrightarrow{w}$; (v) the length of $\overrightarrow{v}$, $\|\overrightarrow{v}\|$; (vi) the unit vector $\overrightarrow{u}$ in the direction of $\overrightarrow{v}$:

   (a) $\overrightarrow{v} = (3, 4)$, $\overrightarrow{w} = (-1, 2)$

   (b) $\overrightarrow{v} = (1, 2, -2)$, $\overrightarrow{w} = (2, -1, 3)$

   (c) $\overrightarrow{v} = 2\overrightarrow{\imath} - 2\overrightarrow{\jmath} - \overrightarrow{k}$, $\overrightarrow{w} = 3\overrightarrow{\imath} + \overrightarrow{\jmath} - 2\overrightarrow{k}$

2. In each case below, decide whether the given vectors are linearly dependent or linearly independent.

   (a) $(1,2)$, $(2,4)$

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

   (c) $(-1,2)$, $(3,-6)$

   (d) $(-1,2)$, $(2,1)$

   (e) $(2,-2,6)$, $(-3,3,9)$

   (f) $(-1,1,3)$, $(3,-3,-9)$

   (g) $\overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}$, $2\overrightarrow{\imath} - 2\overrightarrow{\jmath} + 2\overrightarrow{k}$

   (h) $2\overrightarrow{\imath} - 4\overrightarrow{\jmath} + 2\overrightarrow{k}$, $-\overrightarrow{\imath} + 2\overrightarrow{\jmath} - \overrightarrow{k}$

Theory problems:

3. (a) We have seen that the commutative property of vector addition can be interpreted via the "parallelogram rule" (Figure 5.18). Give a similar pictorial interpretation of the associative property.

   (b) Give geometric arguments for the two distributive properties of vector arithmetic.

   (c) Show that if

   $$a\overrightarrow{v} = \overrightarrow{0}$$

   then either

   $$a = 0$$

   or

   $$\overrightarrow{v} = \overrightarrow{0}.$$

   (*Hint:* What do you know about the relation between lengths for $\overrightarrow{v}$ and $a\overrightarrow{v}$?)

   (d) Show that if a vector $\overrightarrow{v}$ satisfies

   $$a\overrightarrow{v} = b\overrightarrow{v}$$

   where $a \neq b$ are two specific, distinct scalars, then $\overrightarrow{v} = \overrightarrow{0}$.

(e) Show that vector subtraction is *not* associative.

4. (a) Show that if $\overrightarrow{v}$ and $\overrightarrow{w}$ are two linearly independent vectors in the plane, then every vector in the plane can be expressed as a linear combination of $\overrightarrow{v}$ and $\overrightarrow{w}$. (*Hint:* The independence assumption means they point along non-parallel lines. Given a point P in the plane, consider the parallelogram with the origin and P as opposite vertices, and with edges parallel to $\overrightarrow{v}$ and $\overrightarrow{w}$. Use this to construct the linear combination.)

   (b) Now suppose $\overrightarrow{u}$, $\overrightarrow{v}$ and $\overrightarrow{w}$ are *three* nonzero vectors in $\mathbb{R}^3$. If $\overrightarrow{v}$ and $\overrightarrow{w}$ are linearly independent, show that every vector lying in the plane that contains the two lines through the origin parallel to $\overrightarrow{v}$ and $\overrightarrow{w}$ can be expressed as a linear combination of $\overrightarrow{v}$ and $\overrightarrow{w}$. Now show that if $\overrightarrow{u}$ does not lie in this plane, then every vector in $\mathbb{R}^3$ can be expressed as a linear combination of $\overrightarrow{u}$, $\overrightarrow{v}$ and $\overrightarrow{w}$.

   The two statements above are summarized by saying that $\overrightarrow{v}$ and $\overrightarrow{w}$ (*resp.* $\overrightarrow{u}$, $\overrightarrow{v}$ and $\overrightarrow{w}$) **span** $\mathbb{R}^2$ (*resp.* $\mathbb{R}^3$).

Challenge problem:

5. Show (using vector methods) that the line segment joining the midpoints of two sides of a triangle is parallel to and has half the length of the third side.

6. Given a collection $\{\overrightarrow{v}_1, \overrightarrow{v}_2, \dots, \overrightarrow{v}_k\}$ of vectors, consider the equation (in the unknown coefficients $c_1,\dots,c_k$)

$$c_1\overrightarrow{v}_1 + c_2\overrightarrow{v}_2 + \cdots + c_k\overrightarrow{v}_k = \overrightarrow{0}; \tag{1.13}$$

   that is, an expression for the zero vector as a linear combination of the given vectors. Of course, regardless of the vectors $\overrightarrow{v}_i$, one solution of this is

$$c_1 = c_2 = \cdots = 0;$$

   the combination coming from this solution is called the **trivial combination** of the given vectors. The collection $\{\overrightarrow{v}_1, \overrightarrow{v}_2, \dots, \overrightarrow{v}_k\}$ is **linearly dependent** if there exists some **nontrivial** combination of these vectors—that is, a solution of Equation (1.13) with *at least one* nonzero coefficient. It is **linearly independent** if it is not linearly dependent—that is, if the *only* solution of Equation (1.13) is the trivial one.

(a) Show that any collection of vectors which includes the zero vector is linearly dependent.

(b) Show that a collection of *two* nonzero vectors $\{\overrightarrow{v}_1, \overrightarrow{v}_2\}$ in $\mathbb{R}^3$ is linearly independent precisely if (in standard position) they point along non-parallel lines.

(c) Show that a collection of *three* position vectors in $\mathbb{R}^3$ is linearly dependent precisely if at least one of them can be expressed as a linear combination of the other two.

(d) Show that a collection of three position vectors in $\mathbb{R}^3$ is linearly *independent* precisely if the corresponding points determine a plane in space that does *not* pass through the origin.

(e) Show that any collection of *four or more* vectors in $\mathbb{R}^3$ is linearly *dependent*. (*Hint:* Use either part (a) of this problem or part (b) of Exercise 4.)

1.3 Lines in Space

Parametrization of Lines

A line in the plane is the locus of a "linear" equation in the rectangular coordinates x and y

$$Ax + By = C$$

where A, B and C are real constants with at least one of A and B nonzero. A geometrically informative version of this is the **slope-intercept formula** for a non-vertical line

$$y = mx + b \tag{1.14}$$

where the **slope** m is the tangent of the angle the line makes with the horizontal and the ***y*-intercept** b is the ordinate (signed height) of its intersection with the y-axis.

Unfortunately, neither of these schemes extends verbatim to a three-dimensional context. In particular, the locus of a "linear" equation in the three rectangular coordinates x, y and z

$$Ax + By + Cz = D$$

is a *plane*, not a line. Fortunately, though, we can use vectors to implement the geometric thinking behind the point-slope formula (1.14). This formula separates two pieces of geometric data which together determine a line: the *slope* reflects the *direction* (or tilt) of the line, and then the *y-intercept* distinguishes between the various parallel lines with a given slope by specifying a point which must lie on the line. A direction in 3-space cannot be determined by a single number, but it is naturally specified by a nonzero vector, so the three-dimensional analogue of the slope of a line is a **direction vector**

$$\overrightarrow{v} = a\overrightarrow{\imath} + b\overrightarrow{\jmath} + c\overrightarrow{k}$$

to which it is parallel.[6] Then, to pick out one among all the lines parallel to $\overrightarrow{v}$, we specify a **basepoint** $P_0(x_0, y_0, z_0)$ through which the line is required to pass.

The points lying on the line specified by a particular direction vector $\overrightarrow{v}$ and basepoint P_0 are best described in terms of their position vectors. Denote the position vector of the basepoint by

$$\overrightarrow{p}_0 = x_0\overrightarrow{\imath} + y_0\overrightarrow{\jmath} + z_0\overrightarrow{k};$$

then to reach any other point $P(x, y, z)$ on the line, we travel parallel to $\overrightarrow{v}$ from P_0, which is to say the displacement $\overrightarrow{P_0P}$ from P_0 is a scalar multiple of the direction vector:

$$\overrightarrow{P_0P} = t\overrightarrow{v}.$$

The position vector $\overrightarrow{p}(t)$ of the point corresponding to this scalar multiple of $\overrightarrow{v}$ is

$$\overrightarrow{p}(t) = \overrightarrow{p}_0 + t\overrightarrow{v}$$

which defines a **vector-valued function** of the real variable t. In terms of coordinates, this reads

$$\begin{aligned} x &= x_0 + at \\ y &= y_0 + bt \\ z &= z_0 + ct. \end{aligned}$$

We refer to the vector-valued function $\overrightarrow{p}(t)$ as a **parametrization**; the coordinate equations are **parametric equations** for the line.

The vector-valued function $\overrightarrow{p}(t)$ can be interpreted kinematically: it gives the position vector at time t of the moving point whose position at time $t = 0$ is the basepoint P_0, and which is travelling at the constant velocity $\overrightarrow{v}$. Note that to obtain the full line, we need to consider *negative*

[6] In general, we do not need to restrict ourselves to *unit* vectors; any nonzero vector will do.

as well as *positive* values of t—that is, the domain of the function $\overrightarrow{p}(t)$ is the whole real line, $-\infty < t < \infty$.

It is useful to keep in mind the distinction between the *parametrization* $\overrightarrow{p}(t)$, which represents a moving point, and the *line* ℓ being parametrized, which is the *path* of this moving point. A given line ℓ has many different parametrizations: we can take any point on ℓ as P_0, and any nonzero vector pointing parallel to ℓ as the direction vector $\overrightarrow{v}$. This ambiguity means we need to be careful when making geometric comparisons between lines given parametrically. Nonetheless, this way of presenting lines exhibits geometric information in a very accessible form.

For example, let us consider two lines in 3-space. The first, ℓ_1, is given by the parametrization

$$\overrightarrow{p}_1(t) = (1,-2,3) + t(-3,-2,1)$$

or, in coordinate form,

$$\begin{array}{rcrr} x & = & 1 & -3t \\ y & = & -2 & -2t \\ z & = & 3 & +t \end{array}$$

while the second, ℓ_2, is given in coordinate form as

$$\begin{array}{rcrr} x & = & 1 & +6t \\ y & = & & 4t \\ z & = & 1 & -2t. \end{array}$$

We can easily read off from this that ℓ_2 has parametrization

$$\overrightarrow{p}_2(t) = (1,0,1) + t(6,4,-2).$$

Comparing the two direction vectors

$$\begin{aligned} \overrightarrow{v}_1 &= -3\overrightarrow{\imath} - 2\overrightarrow{\jmath} + \overrightarrow{k} \\ \overrightarrow{v}_2 &= 6\overrightarrow{\imath} + 4\overrightarrow{\jmath} - 2\overrightarrow{k} \end{aligned}$$

we see that

$$\overrightarrow{v}_2 = -2\overrightarrow{v}_1$$

so the two lines have the same direction—either they are parallel, or they coincide. To decide which is the case, it suffices to decide whether the basepoint of one of the lines lies on the other line. Let us see whether the basepoint of ℓ_2

$$\overrightarrow{p}_2(0) = (1, 0, 1)$$

lies on ℓ_1. This means we need to see if for some value of t we have $\overrightarrow{p}_2(0) = \overrightarrow{p}_1(t)$, or

$$\begin{array}{rcrr} 1 & = & 1 & -3t \\ 0 & = & -2 & -2t \\ 1 & = & 3 & +t. \end{array}$$

It is easy to see that the first equation requires $t = 0$, the second requires $t = -1$, and the third requires $t = -2$; there is no way we can solve all three simultaneously. It follows that ℓ_1 and ℓ_2 are distinct, but parallel, lines.

Now, consider a third line, ℓ_3, given by

$$\begin{array}{rcrr} x & = & 1 & +3t \\ y & = & 2 & +t \\ z & = & -3 & +t. \end{array}$$

We read off its direction vector as

$$\overrightarrow{v}_3 = 3\overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}$$

which is clearly *not* a scalar multiple of the other two. This tells us immediately that ℓ_3 is different from both ℓ_1 and ℓ_2 (it has a different direction). Let us ask whether ℓ_2 intersects ℓ_3. It might be tempting to try to answer this by looking for a solution of the vector equation

$$\overrightarrow{p}_2(t) = \overrightarrow{p}_3(t)$$

but *this would be a mistake.* Remember that these two parametrizations describe the positions of two points—one moving along ℓ_2 and the other moving along ℓ_3—at time t. The equation above requires the two points to be at the same place *at the same time*—in other words, it describes a *collision.* But all we ask is that the two *paths* cross: it would suffice to locate a *place* occupied by both moving points, but possibly *at different times.* This means we need to distinguish the parameters appearing in the two functions $\overrightarrow{p}_2(t)$ and $\overrightarrow{p}_3(t)$, by renaming one of them (say the first) as (say) s: the vector equation we need to solve is

$$\overrightarrow{p}_2(s) = \overrightarrow{p}_3(t)\,,$$

which amounts to the three equations in two unknowns

$$\begin{array}{rrcrr} 1 & +6s & = & 1 & +3t \\ & 4s & = & 2 & +t \\ 1 & -2s & = & -3 & +t. \end{array}$$

The first equation can be reduced to the condition

$$2s = t$$

and substituting this into the other two equations

$$\begin{array}{rrcrr} & 2t & = & 2 & +t \\ 1 & -t & = & -3 & +t \end{array}$$

we end up with the solution

$$\begin{aligned} t &= 2 \\ s &= 1 \end{aligned}$$

—that is, the lines ℓ_2 and ℓ_3 intersect at the point

$$\overrightarrow{p}_2(1) = (7, 4, -1) = \overrightarrow{p}_3(2)\,.$$

Now let us apply the same process to see whether ℓ_1 intersects ℓ_3. The vector equation

$$\overrightarrow{p}_1(s) = \overrightarrow{p}_3(t)$$

yields the three coordinate equations

$$\begin{array}{rrcrr} 1 & -3s & = & 1 & +3t \\ -2 & -2s & = & 2 & +t \\ 3 & +s & = & -3 & +t. \end{array}$$

You can check that these imply, respectively

$$\begin{array}{rcrr} s & = & & -t \\ -2s & = & 4 & +t \\ s & = & -6 & +t. \end{array}$$

Substituting the first equality into the other two yields, respectively

$$\begin{array}{rcrr} 2t & = & 4 & +t \\ -t & = & -6 & +t \end{array}$$

and the only value of t for which the first (*resp.* second) holds is, respectively,

$$\begin{aligned} t &= 4 \\ t &= 3. \end{aligned}$$

Thus our three coordinate equations cannot be satisfied simultaneously; it follows that ℓ_1 and ℓ_3 *do not intersect*, even though they are *not parallel*. Such lines are sometimes called **skew lines** .

Geometric Applications

A basic geometric fact is that *any pair of distinct points determines a line*. Given two points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$, how do we find a parametrization of the line ℓ they determine?

Suppose the position vectors of the two points are

$$\begin{aligned} \overrightarrow{p}_1 &= x_1\overrightarrow{\imath} + y_1\overrightarrow{\jmath} + z_1\overrightarrow{k} \\ \overrightarrow{p}_2 &= x_2\overrightarrow{\imath} + y_2\overrightarrow{\jmath} + z_2\overrightarrow{k}. \end{aligned}$$

The vector $\overrightarrow{P_1P_2}$ joining them lies along ℓ, so we can use it as a direction vector:

$$\overrightarrow{v} = \overrightarrow{p}_2 - \overrightarrow{p}_1 = \triangle x\overrightarrow{\imath} + \triangle y\overrightarrow{\jmath} + \triangle z\overrightarrow{k}$$

where $\triangle x = x_2 - x_1$, $\triangle y = y_2 - y_1$, and $\triangle z = z_2 - z_1$. Using P_1 as basepoint, this leads to the parametrization

$$\begin{aligned} \overrightarrow{p}(t) &= \overrightarrow{p}_1 + t\overrightarrow{v} \\ &= \overrightarrow{p}_1 + t(\overrightarrow{p}_2 - \overrightarrow{p}_1) \\ &= (1-t)\overrightarrow{p}_1 + t\overrightarrow{p}_2. \end{aligned}$$

Note that we have set this up so that

$$\begin{aligned} \overrightarrow{p}(0) &= \overrightarrow{p}_1 \\ \overrightarrow{p}(1) &= \overrightarrow{p}_2. \end{aligned}$$

The full line ℓ through these points corresponds to allowing the parameter to take on all real values. However, if we restrict t to the interval $0 \leq t \leq 1$, the corresponding points fill out the **line segment** P_1P_2. Since the point $\overrightarrow{p}(t)$ is travelling with constant velocity, we have the following observations:

Remark 1.3.1 (Two-Point Formula)**.** *Suppose $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ are distinct points. The line through P_1 and P_2 is given by the parametrization*[7]

$$\vec{p}(t) = (1-t)\vec{p}_1 + t\vec{p}_2$$

with coordinates

$$\begin{aligned} x &= (1-t)x_1 + tx_2 \\ y &= (1-t)y_1 + ty_2 \\ z &= (1-t)z_1 + tz_2. \end{aligned}$$

The ***line segment*** *P_1P_2 consists of the points for which $0 \leq t \leq 1$. The value of t gives the portion of P_1P_2 represented by the segment $P_1\vec{p}(t)$; in particular, the* ***midpoint*** *of P_1P_2 has position vector*

$$\frac{1}{2}(\vec{p}_1 + \vec{p}_2) = \left(\frac{1}{2}(x_1+x_2), \frac{1}{2}(y_1+y_2), \frac{1}{2}(z_1+z_2)\right).$$

We will use these ideas to prove the following.

Theorem 1.3.2. *In any triangle, the three lines joining a vertex to the midpoint of the opposite side meet at a single point.*

Proof. Label the vertices of the triangle A, B and C, and their position vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$, respectively. Label the midpoint of each side with the name of the opposite vertex, primed; thus the midpoint of BC (the side opposite vertex A) is A' (see Figure 1.23). From Remark 1.3.1 we see that the position vectors of the midpoints of the sides are

$$\begin{aligned} \overrightarrow{\mathcal{O}A'} &= \frac{1}{2}(\vec{b} + \vec{c}) \\ \overrightarrow{\mathcal{O}B'} &= \frac{1}{2}(\vec{c} + \vec{a}) \\ \overrightarrow{\mathcal{O}C'} &= \frac{1}{2}(\vec{a} + \vec{b}), \end{aligned}$$

and so the line ℓ_A through A and A' can be parametrized (using r as the parameter) by

$$\vec{p}_A(r) = (1-r)\vec{a} + \frac{r}{2}(\vec{b} + \vec{c}) = (1-r)\vec{a} + \frac{r}{2}\vec{b} + \frac{r}{2}\vec{c}.$$

[7] This is the parametrized form of the **two-point formula** for a line in the plane determined by a pair of points.

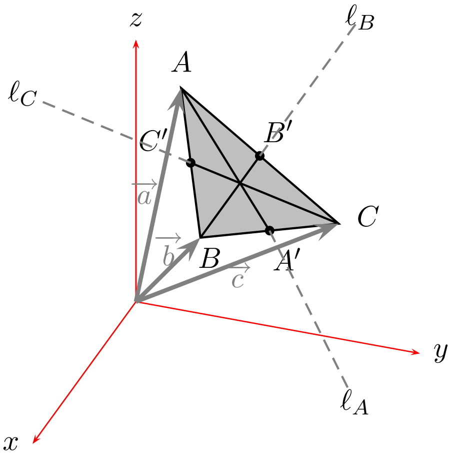


Figure 1.23: Theorem 1.3.2

Similarly, the other two lines ℓ_B and ℓ_C are parametrized by

$$\overrightarrow{p}_B(s) = \frac{s}{2}\overrightarrow{a} + (1-s)\overrightarrow{b} + \frac{s}{2}\overrightarrow{c}$$
$$\overrightarrow{p}_C(t) = \frac{t}{2}\overrightarrow{a} + \frac{t}{2}\overrightarrow{b} + (1-t)\overrightarrow{c}.$$

To find the intersection of ℓ_A with ℓ_B, we need to solve the vector equation

$$\overrightarrow{p}_A(r) = \overrightarrow{p}_B(s)$$

which, written in terms of $\overrightarrow{a}$, $\overrightarrow{b}$ and $\overrightarrow{c}$ is

$$(1-r)\overrightarrow{a} + \frac{r}{2}\overrightarrow{b} + \frac{r}{2}\overrightarrow{c} = \frac{s}{2}\overrightarrow{a} + (1-s)\overrightarrow{b} + \frac{s}{2}\overrightarrow{c}.$$

Assuming the triangle $\triangle ABC$ is nondegenerate—which is to say, none of $\overrightarrow{a}$, $\overrightarrow{b}$ or $\overrightarrow{c}$ can be expressed as a linear combination of the others—this equality can only hold if corresponding coefficients on the two sides are equal. This leads to three equations in two unknowns

$$\begin{aligned}(1-r) &= \frac{s}{2}\\ \frac{r}{2} &= (1-s)\\ \frac{r}{2} &= \frac{s}{2}.\end{aligned}$$

The last equation requires

$$r = s$$

turning the other two equations into

$$1 - r = \frac{r}{2}$$

and leading to the solution

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

In a similar way, the intersection of ℓ_B with ℓ_C is given by

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

and so we see that the three lines ℓ_A, ℓ_B and ℓ_C all intersect at the point given by

$$\vec{p}_A\left(\frac{2}{3}\right) = \vec{p}_B\left(\frac{2}{3}\right) = \vec{p}_C\left(\frac{2}{3}\right) = \frac{1}{3}\vec{a} + \frac{1}{3}\vec{b} + \frac{1}{3}\vec{c}.$$

□

The point given in the last equation is sometimes called the **barycenter** of the triangle $\triangle ABC$. Physically, it represents the **center of mass** of equal masses placed at the three vertices of the triangle. Note that it can also be regarded as the (vector) arithmetic **average** of the three position vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$. In Exercise 9, we shall explore this point of view further.

Exercises for § 1.3

Practice problems:

1. For each line in the plane described below, give (i) an equation in the form $Ax+By+C=0$, (ii) parametric equations, and (iii) a parametric vector equation:

   (a) The line with slope -1 through the origin.

   (b) The line with slope -2 and y-intercept 1.

(c) The line with slope 1 and y-intercept -2.

(d) The line with slope 3 going through the point $(-1, 2)$.

(e) The line with slope -2 going through the point $(-1, 2)$.

2. Find the slope and y-intercept for each line given below:

(a) $2x + y - 3 = 0$

(b) $x - 2y + 4 = 0$

(c) $3x + 2y + 1 = 0$

(d) $y = 0$

(e) $x = 1$

3. For each line in $\mathbb{R}^3$ described below, give (i) parametric equations, and (ii) a parametric vector equation:

(a) The line through the point $(2, -1, 3)$ with direction vector $\overrightarrow{v} = -\overrightarrow{\imath} + 2\overrightarrow{\jmath} + \overrightarrow{k}$.

(b) The line through the points $(-1, 2, -3)$ and $(3, -2, 1)$.

(c) The line through the points $(2, 1, 1)$ and $(2, 2, 2)$.

(d) The line through the point $(1, 3, -2)$ parallel to the line

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

4. For each pair of lines in the plane given below, decide whether they are parallel or if not, find their point of intersection.

(a) $x + y = 3$ and $3x - 3y = 3$

(b) $2x - 2y = 2$ and $2y - 2x = 2$

(c)

$$\begin{aligned} x &= 1 + 2t \\ y &= -1 + t \end{aligned} \quad \text{and} \quad \begin{aligned} x &= 2 - t \\ y &= -4 + 2t \end{aligned}$$

(d)

$$\begin{aligned} x &= 2 - 4t \\ y &= -1 - 2t \end{aligned} \quad \text{and} \quad \begin{aligned} x &= 1 + 2t \\ y &= -4 + t \end{aligned}$$

5. Find the points at which the line with parametrization

$$\overrightarrow{p}(t) = (3 + 2t, 7 + 8t, -2 + t)$$

that is,

$$\begin{aligned} x &= 3 + 2t \\ y &= 7 + 8t \\ z &= -2 + t \end{aligned}$$

intersects each of the coordinate planes.

6. Determine whether the given lines intersect:

(a)

$$\begin{aligned} x &= 3t + 2 \\ y &= t - 1 \\ z &= 6t + 1 \end{aligned}$$

and

$$\begin{aligned} x &= 3t - 1 \\ y &= t - 2 \\ z &= t; \end{aligned}$$

(b)

$$\begin{aligned} x &= t + 4 \\ y &= 4t + 5 \\ z &= t - 2 \end{aligned}$$

and

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

Theory problems:

7. Show that if $\overrightarrow{u}$ and $\overrightarrow{v}$ are both *unit* vectors, placed in standard position, then the line through the origin parallel to $\overrightarrow{u} + \overrightarrow{v}$ bisects the angle between them.

8. The following is implicit in the proof of Book V, Proposition 4 of Euclid's *Elements* [27, pp. 85-88] . Here, we work through a proof using vectors; we work through the proof of the same fact following Euclid in Exercise 11

Theorem 1.3.3 (Angle Bisectors). *In any triangle, the lines which bisect the three interior angles meet in a common point.*

Note that this is *different* from Theorem 1.3.2 in the text.

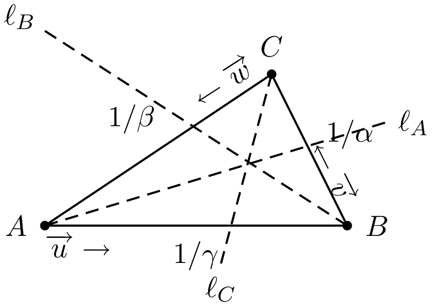


Figure 1.24: Theorem 1.3.3

Suppose the position vectors of the vertices A, B, and C are $\overrightarrow{a}$, $\overrightarrow{b}$ and $\overrightarrow{c}$ respectively.

(a) Show that the *unit* vectors pointing *counterclockwise* along the edges of the triangle (see Figure 1.24) are as follows:

$$\overrightarrow{u} = \gamma\overrightarrow{b} - \gamma\overrightarrow{a}$$
$$\overrightarrow{v} = \alpha\overrightarrow{c} - \alpha\overrightarrow{b}$$
$$\overrightarrow{w} = \beta\overrightarrow{a} - \beta\overrightarrow{c}$$

where

$$\alpha = \frac{1}{|BC|}$$
$$\beta = \frac{1}{|AC|}$$
$$\gamma = \frac{1}{|AB|}$$

are the reciprocals of the lengths of the sides (each length is labelled by the Greek analogue of the name of the opposite vertex).

(b) Show that the line ℓ_A bisecting the angle $\angle A$ can be given as

$$\vec{p}_A(r) = (1 - r\beta - r\gamma)\vec{a} + r\gamma\vec{b} + r\beta\vec{c}$$

and the corresponding bisectors of $\angle B$ and $\angle C$ are

$$\vec{p}_B(s) = s\gamma\vec{a} + (1 - s\alpha - s\gamma)\vec{b} + s\alpha\vec{c}$$
$$\vec{p}_C(t) = t\beta\vec{a} + t\alpha\vec{b} + (1 - t\alpha - t\beta)\vec{c}.$$

(c) Show that the intersection of ℓ_A and ℓ_B is given by

$$r = \frac{\alpha}{\alpha\beta + \beta\gamma + \gamma\alpha}$$
$$s = \frac{\beta}{\alpha\beta + \beta\gamma + \gamma\alpha}.$$

(d) Show that the intersection of ℓ_B and ℓ_C is given by

$$s = \frac{\beta}{\alpha\beta + \beta\gamma + \gamma\alpha}$$
$$t = \frac{\gamma}{\alpha\beta + \beta\gamma + \gamma\alpha}.$$

(e) Conclude that all three lines meet at the point given by

$$\begin{aligned} &\vec{p}_A\left(\frac{\alpha}{\alpha\beta + \beta\gamma + \gamma\alpha}\right) \\ &\quad = \vec{p}_B\left(\frac{\beta}{\alpha\beta + \beta\gamma + \gamma\alpha}\right) = \vec{p}_C\left(\frac{\gamma}{\alpha\beta + \beta\gamma + \gamma\alpha}\right) \\ &\qquad = \frac{1}{\alpha\beta + \beta\gamma + \gamma\alpha}\left(\beta\gamma\vec{a} + \gamma\alpha\vec{b} + \alpha\beta\vec{c}\right). \end{aligned}$$

Challenge problem:

9. **Barycentric Coordinates:** Show that if $\overrightarrow{a}$, $\overrightarrow{b}$ and $\overrightarrow{c}$ are the position vectors of the vertices of a triangle $\triangle ABC$ in $\mathbb{R}^3$, then the position vector v' of every point P in that triangle (lying in the plane determined by the vertices) can be expressed as a linear combination of $\overrightarrow{a}$, $\overrightarrow{b}$ and $\overrightarrow{c}$

$$v' = \lambda_1 \overrightarrow{a} + \lambda_2 \overrightarrow{b} + \lambda_3 \overrightarrow{c}$$

with

$$0 \leq \lambda_i \leq 1 \text{ for } i = 1, 2, 3$$

and

$$\lambda_1 + \lambda_2 + \lambda_3 = 1.$$

(*Hint:* (see Figure 1.25) Draw a line from vertex A through P, and observe where it meets the opposite side; call this point D. Use Remark 1.3.1 to show that the position vector $\overrightarrow{d}$ of D is a linear combination of $\overrightarrow{b}$ and $\overrightarrow{c}$, with coefficients between zero and one and summing to 1. Then use Remark 1.3.1 again to show that v' is a linear combination of $\overrightarrow{d}$ and $\overrightarrow{a}$.)

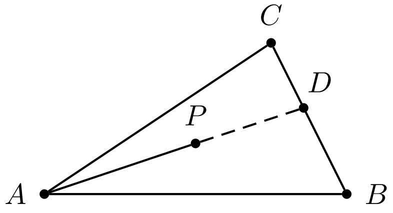


Figure 1.25: Barycentric Coordinates

The numbers λ_i are called the **barycentric coordinates** of P with respect to A, B and C. Show that P lies on an edge of the triangle precisely if one of its barycentric coordinates is zero.

Barycentric coordinates were introduced (in a slightly different form) by August Möbius (1790-1860)) in his book *Barycentrische Calcul*

(1827). His name is more commonly associated with “Möbius transformations” in complex analysis and with the “Möbius band” (the one-sided surface that results from joining the ends of a band after making a half-twist) in topology.[8]

10. Find a line that lies entirely in the set defined by the equation $x^2 + y^2 - z^2 = 1$.

History note:

11. Heath [27, pp. 85-88] points out that the proof of Proposition 4, Book IV of the *Elements* contains the following implicit proof of Theorem 1.3.3 (see Figure 1.26). This was proved by vector methods in Exercise 8.

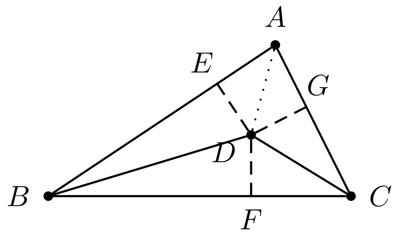


Figure 1.26: Euclid’s proof of Theorem 1.3.3

(a) The lines bisecting $\angle B$ and $\angle C$ intersect at a point D above BC, because of Book I, Postulate 5 (known as the **Parallel Postulate**):

> That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

Why do the interior angles between BC and the two angle bisectors add up to less than a right angle? (*Hint:* What do you know about the angles of a triangle?)

[8] The “Möbius band” was independently formulated by Johann Listing (1808-1882) at about the same time—in 1858, when Möbius was 68 years old. These two are often credited with beginning the study of topology. [31, p. 1165]

(b) Drop perpendiculars from D to each edge of the triangle, meeting the edges at E, F and G.

(c) Show that the triangles $\triangle BFD$ and $\triangle BED$ are congruent. (*Hint:* ASA—angle, side, angle!)

(d) Similarly, show that the triangles $\triangle CFD$ and $\triangle CGD$ are congruent.

(e) Use this to show that

$$|DE| = |DF| = |DG| \,.$$

(f) Now draw the line DA. Show that the triangles $\triangle AGD$ and $\triangle AED$ are congruent. (*Hint:* Both are right angles; compare one pair of legs and the hypotenuse.)

(g) Conclude that $\angle EAD = \angle ADG$—which means that DA bisects $\angle A$. Thus D lies on all three angle bisectors.

1.4 Projection of Vectors; Dot Products

Suppose a weight is set on a ramp which is inclined θ radians from the horizontal (Figure 1.27). The gravitational force $\overrightarrow{g}$ on the weight is directed downward, and some of this is countered by the structure holding up the ramp. The effective force on the weight can be found by expressing $\overrightarrow{g}$ as a sum of two (vector) forces, $\overrightarrow{g}_\perp$ *perpendicular* to the ramp, and $\overrightarrow{g}_\parallel$ *parallel* to the ramp. Then $\overrightarrow{g}_\perp$ is cancelled by the structural forces in the ramp, and the net unopposed force is $\overrightarrow{g}_\parallel$, the *projection* of $\overrightarrow{g}$ in the direction of the ramp.

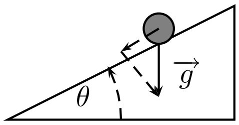


Figure 1.27: A weight on a ramp

To abstract this situation, recall that a direction is specified by a unit vector. The (vector) **projection** of an arbitrary vector $\overrightarrow{v}$ in the direction specified by the unit vector $\overrightarrow{u}$ is the vector

$$\text{proj}_{\overrightarrow{u}}\,\overrightarrow{v} := (|\overrightarrow{v}|\cos\theta)\overrightarrow{u}$$

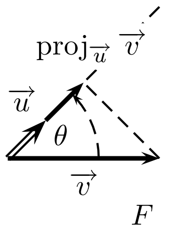


Figure 1.28: Projection of a Vector

where θ is the angle between $\overrightarrow{u}$ and $\overrightarrow{v}$ (Figure 1.28). Note that replacing $\overrightarrow{u}$ with its negative replaces θ with $\pi - \theta$, and the projection is unchanged:

$$\begin{aligned}\text{proj}_{-\overrightarrow{u}}\,\overrightarrow{v} &= (|\overrightarrow{v}|\cos(\pi - \theta))(-\overrightarrow{u})\\ &= (-|\overrightarrow{v}|\cos(\theta))(-\overrightarrow{u})\\ &= (|\overrightarrow{v}|\cos(\theta))(\overrightarrow{u})\\ &= \text{proj}_{\overrightarrow{u}}\,\overrightarrow{v}.\end{aligned}$$

This means that the projection of a vector in the direction specified by $\overrightarrow{u}$ depends only on the line parallel to $\overrightarrow{u}$ (not the direction along that line).

If $\overrightarrow{w}$ is any nonzero vector, we define the **projection** of $\overrightarrow{v}$ onto (the direction of) $\overrightarrow{w}$ as its projection onto the unit vector $\overrightarrow{u} = \overrightarrow{w}/|\overrightarrow{w}|$ in the direction of $\overrightarrow{w}$:

$$\text{proj}_{\overrightarrow{w}}\,\overrightarrow{v} = \text{proj}_{\overrightarrow{u}}\,\overrightarrow{v} = \left(\frac{|\overrightarrow{v}|}{|\overrightarrow{w}|}\cos\theta\right)\overrightarrow{w}. \tag{1.15}$$

How do we calculate this projection from the entries of the two vectors? To this end, we perform a theoretical detour.[9]

Suppose $\overrightarrow{v} = (x_1, y_1, z_1)$ and $\overrightarrow{w} = (x_2, y_2, z_2)$; how do we determine the angle θ between them? If we put them in standard position, representing $\overrightarrow{v}$ by $\overrightarrow{\mathcal{O}P}$ and $\overrightarrow{w}$ by $\overrightarrow{\mathcal{O}Q}$ (Figure 1.29), then we have a triangle $\triangle\mathcal{O}PQ$ with angle θ at the origin, and two sides given by

$$\begin{aligned}a &= |\overrightarrow{v}| = \sqrt{x_1^2 + y_1^2 + z_1^2}\\ b &= |\overrightarrow{w}| = \sqrt{x_2^2 + y_2^2 + z_2^2}.\end{aligned}$$

The distance formula lets us determine the length of the third side:

$$c = \text{dist}(P, Q) = \sqrt{\triangle x^2 + \triangle y^2 + \triangle z^2}.$$

[9] Thanks to my student Benjamin Brooks, whose questions helped me formulate the approach here.

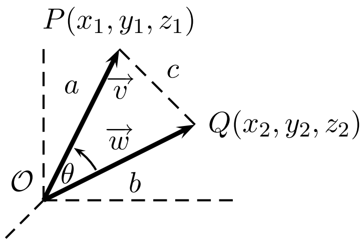


Figure 1.29: Determining the Angle θ

But we also have the **Law of Cosines** (Exercise 13):

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

or

$$2ab\cos\theta = a^2 + b^2 - c^2. \tag{1.16}$$

We can compute the right-hand side of this equation by substituting the expressions for a, b and c in terms of the entries of $\overrightarrow{v}$ and $\overrightarrow{w}$:

$$a^2 + b^2 - c^2 = (x_1^2 + y_1^2 + z_1^2) + (x_2^2 + y_2^2 + z_2^2) - (\triangle x^2 + \triangle y^2 + \triangle z^2).$$

Consider the terms involving x:

$$\begin{aligned} x_1^2 + x_2^2 - \triangle x^2 &= x_1^2 + x_2^2 - (x_1 - x_2)^2 \\ &= x_1^2 + x_2^2 - (x_1^2 - 2x_1x_2 + x_2^2) \\ &= 2x_1x_2. \end{aligned}$$

Similar calculations for the y- and z-coordinates allow us to conclude that

$$a^2 + b^2 - c^2 = 2(x_1x_2 + y_1y_2 + z_1z_2)$$

and hence, substituting into Equation (1.16), factoring out 2, and recalling that $a = |\overrightarrow{v}|$ and $b = |\overrightarrow{w}|$, we have

$$|\overrightarrow{v}|\,|\overrightarrow{w}|\cos\theta = x_1x_2 + y_1y_2 + z_1z_2. \tag{1.17}$$

This quantity, which is easily calculated from the entries of $\overrightarrow{v}$ and $\overrightarrow{w}$ (on the right) but has a useful geometric interpretation (on the left), is called the *dot product*[10] of $\overrightarrow{v}$ and $\overrightarrow{w}$. Equation (1.17) appears already (with somewhat different notation) in Lagrange's 1788 *Méchanique Analitique* [34, N.11], and

[10] Also the *scalar product*, *direct product*, or *inner product*

also as part of Hamilton's definition (1847) of the product of quaternions [22], although the scalar product of vectors was apparently not formally identified until Wilson's 1901 textbook [54], or more accurately Gibbs' earlier (1881) notes on the subject [17, p. 20].

Definition 1.4.1. *Given any two vectors $\overrightarrow{v} = (x_1, y_1, z_1)$ and $\overrightarrow{w} = (x_2, y_2, z_2)$ in $\mathbb{R}^3$, their **dot product** is the scalar*

$$\overrightarrow{v} \cdot \overrightarrow{w} = x_1x_2 + y_1y_2 + z_1z_2.$$

The definition of the dot product exhibits a number of algebraic properties, which we leave to you to verify (Exercise 3):

Proposition 1.4.2. *The dot product has the following algebraic properties:*

1. *It is **commutative**:*
$$\overrightarrow{v} \cdot \overrightarrow{w} = \overrightarrow{w} \cdot \overrightarrow{v}$$

2. *It **distributes over vector sums**[11]:*
$$\overrightarrow{u} \cdot (\overrightarrow{v} + \overrightarrow{w}) = \overrightarrow{u} \cdot \overrightarrow{v} + \overrightarrow{u} \cdot \overrightarrow{w}$$

3. *it **respects scalar multiples**:*
$$(r\overrightarrow{v}) \cdot \overrightarrow{w} = r(\overrightarrow{v} \cdot \overrightarrow{w}) = \overrightarrow{v} \cdot (r\overrightarrow{w}).$$

Also, the geometric interpretation of the dot product given by Equation (1.17) yields a number of geometric properties:

Proposition 1.4.3. *The dot product has the following geometric properties:*

1. $\overrightarrow{v} \cdot \overrightarrow{w} = |\overrightarrow{v}|\,|\overrightarrow{w}|\cos\theta$*, where θ is the angle between the "arrows" representing $\overrightarrow{v}$ and $\overrightarrow{w}$.*

2. $\overrightarrow{v} \cdot \overrightarrow{w} = 0$ *precisely if the arrows representing $\overrightarrow{v}$ and $\overrightarrow{w}$ are perpendicular to each other, or if one of the vectors is the zero vector.*

3. $\overrightarrow{v} \cdot \overrightarrow{v} = |\overrightarrow{v}|^2$

4. $\operatorname{proj}_{\overrightarrow{w}} \overrightarrow{v} = \left(\dfrac{\overrightarrow{v} \cdot \overrightarrow{w}}{\overrightarrow{w} \cdot \overrightarrow{w}}\right)\overrightarrow{w}$ *(provided $\overrightarrow{w} \neq \overrightarrow{0}$).*

[11] In this formula, $\overrightarrow{u}$ is an arbitrary vector, not necessarily of unit length.

We note the curiosity in the second item: the dot product of the zero vector with *any* vector is zero. While the zero vector has no well-defined direction, we will find it a convenient fiction to say that *the zero vector is perpendicular to every vector, including itself.*

Proof. 1. This is just Equation (1.17).

2. This is an (almost) immediate consequence: if $|\overrightarrow{v}|$ and $|\overrightarrow{w}|$ are both nonzero (*i.e.*, $\overrightarrow{v} \neq \overrightarrow{0} \neq \overrightarrow{w}$) then $\overrightarrow{v} \cdot \overrightarrow{w} = 0$ precisely when $\cos\theta = 0$, and this is the same as saying that $\overrightarrow{v}$ is perpendicular to $\overrightarrow{w}$ (denoted $\boldsymbol{\overrightarrow{v}} \perp \boldsymbol{\overrightarrow{w}}$). But if either $\overrightarrow{v}$ or $\overrightarrow{w}$ *is* $\overrightarrow{0}$, then clearly $\overrightarrow{v} \cdot \overrightarrow{w} = 0$ by either side of Equation (1.17).

3. This is just (1) when $\overrightarrow{v} = \overrightarrow{w}$, which in particular means $\theta = 0$, or $\cos\theta = 1$.

4. This follows from Equation (1.15) by substitution:

$$\begin{aligned}
\operatorname{proj}_{\overrightarrow{w}} \overrightarrow{v} &= \left(\frac{|\overrightarrow{v}|}{|\overrightarrow{w}|}\cos\theta\right)\overrightarrow{w} \\
&= \left(\frac{|\overrightarrow{v}|\,|\overrightarrow{w}|}{|\overrightarrow{w}|^2}\cos\theta\right)\overrightarrow{w} \\
&= \left(\frac{\overrightarrow{v}\cdot\overrightarrow{w}}{\overrightarrow{w}\cdot\overrightarrow{w}}\right)\overrightarrow{w}.
\end{aligned}$$

□

These interpretations of the dot product make it a powerful tool for attacking certain kinds of geometric and mechanical problems. We consider two examples below, and others in the exercises.

Distance from a point to a line: Given a point Q with coordinate vector $\overrightarrow{q}$ and a line ℓ parametrized via

$$\overrightarrow{p}(t) = \overrightarrow{p}_0 + t\overrightarrow{v}$$

let us calculate the distance from Q to ℓ. We will use the fact that this distance is achieved by a line segment from Q to a point R on the line such that QR is *perpendicular* to ℓ (Figure 1.30). We have

$$\overrightarrow{P_0Q} = \overrightarrow{q} - \overrightarrow{p}_0.$$

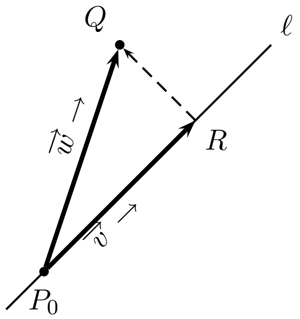


Figure 1.30: Distance from Point to Line

We will denote this, for clarity, by

$$\begin{aligned}\overrightarrow{w} &:= \overrightarrow{q} - \overrightarrow{p}_0 \\ \overrightarrow{P_0R} &= \operatorname{proj}_{\overrightarrow{v}} \overrightarrow{P_0Q} \\ &= \operatorname{proj}_{\overrightarrow{v}} \overrightarrow{w}\end{aligned}$$

so

$$|P_0R| = \frac{\overrightarrow{w} \cdot \overrightarrow{v}}{|\overrightarrow{v}|}$$

and thus by Pythagoras' Theorem

$$\begin{aligned}|QR|^2 &= |P_0Q|^2 - |P_0R|^2 \\ &= \overrightarrow{w} \cdot \overrightarrow{w} - \left(\frac{\overrightarrow{w} \cdot \overrightarrow{v}}{|\overrightarrow{v}|}\right)^2 \\ &= \frac{(\overrightarrow{w} \cdot \overrightarrow{w})(\overrightarrow{v} \cdot \overrightarrow{v}) - (\overrightarrow{w} \cdot \overrightarrow{v})^2}{\overrightarrow{v} \cdot \overrightarrow{v}}.\end{aligned}$$

Another approach is outlined in Exercise 7.

Angle cosines: A natural way to try to specify the direction of a line through the origin is to find the angles it makes with the three coordinate axes; these are sometimes referred to as the **Euler angles** of the line. In the plane, it is clear that the angle α between a line and the horizontal is complementary to the angle β between the line and the vertical. In space, the relation between the angles α, β and γ which a line makes with the positive x, y, and z-axes respectively is less obvious on purely geometric grounds. The relation

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1 \tag{1.18}$$

was implicit in the work of the eighteenth-century mathematicians Joseph Louis Lagrange (1736-1813) and Gaspard Monge (1746-1818), and explicitly stated by Leonard Euler (1707-1783) [4, pp. 206-7]. Using vector ideas, it is almost obvious.

Proof of Equation (1.18). Let $\overrightarrow{u}$ be a unit vector in the direction of the line. Then the angles between $\overrightarrow{u}$ and the unit vectors along the three axes are

$$\begin{aligned} \overrightarrow{u} \cdot \overrightarrow{\imath} &= \cos\alpha \\ \overrightarrow{u} \cdot \overrightarrow{\jmath} &= \cos\beta \\ \overrightarrow{u} \cdot \overrightarrow{k} &= \cos\gamma \end{aligned}$$

from which it follows that

$$\overrightarrow{u} = \cos\alpha\,\overrightarrow{\imath} + \cos\beta\,\overrightarrow{\jmath} + \cos\gamma\,\overrightarrow{k}$$

or in other words

$$\overrightarrow{u} = (\cos\alpha, \cos\beta, \cos\gamma).$$

But then the distance formula says that

$$1 = |\overrightarrow{u}| = \sqrt{\cos^2\alpha + \cos^2\beta + \cos^2\gamma}$$

and squaring both sides yields Equation (1.18).

□

Scalar Projection: The projection $\mathrm{proj}_{\overrightarrow{w}}\overrightarrow{v}$ of the vector $\overrightarrow{v}$ in the direction of the vector $\overrightarrow{w}$ is itself a vector; a related quantity is the **scalar projection** of $\overrightarrow{v}$ in the direction of $\overrightarrow{w}$, also called the **component** of $\overrightarrow{v}$ in the direction of $\overrightarrow{w}$. This is defined as

$$\mathrm{comp}_{\overrightarrow{w}}\overrightarrow{v} = \|\overrightarrow{v}\|\cos\theta$$

where θ is the angle between $\overrightarrow{v}$ and $\overrightarrow{w}$; clearly, this can also be expressed as $\overrightarrow{v} \cdot \overrightarrow{u}$, where

$$\overrightarrow{u} := \frac{\overrightarrow{w}}{\|\overrightarrow{w}\|}$$

is the unit vector parallel to $\overrightarrow{w}$. Thus we can also write

$$\mathrm{comp}_{\overrightarrow{w}}\, \overrightarrow{v} = \frac{\overrightarrow{v} \cdot \overrightarrow{w}}{\|\overrightarrow{w}\|}. \tag{1.19}$$

This is a scalar, whose absolute value is the length of the vector projection, which is positive if $\mathrm{proj}_{\overrightarrow{w}}\, \overrightarrow{v}$ is parallel to $\overrightarrow{w}$ and negative if it points in the opposite direction.

Exercises for § 1.4

Practice problems:

1. For each pair of vectors $\overrightarrow{v}$ and $\overrightarrow{w}$ below, find their dot product, their lengths, the cosine of the angle between them, and the (vector) projection of each onto the direction of the other:

   (a) $\overrightarrow{v} = (2,3)$, $\overrightarrow{w} = (3,2)$

   (b) $\overrightarrow{v} = (2,3)$, $\overrightarrow{w} = (3,-2)$

   (c) $\overrightarrow{v} = (1,0)$, $\overrightarrow{w} = (3,2)$

   (d) $\overrightarrow{v} = (1,0)$, $\overrightarrow{w} = (3,4)$

   (e) $\overrightarrow{v} = (1,2,3)$, $\overrightarrow{w} = (3,2,1)$

   (f) $\overrightarrow{v} = (1,2,3)$, $\overrightarrow{w} = (3,-2,0)$

   (g) $\overrightarrow{v} = (1,2,3)$, $\overrightarrow{w} = (3,0,-1)$

   (h) $\overrightarrow{v} = (1,2,3)$, $\overrightarrow{w} = (1,1,-1)$

2. A point travelling at the constant velocity $\overrightarrow{v} = \overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}$ goes through the position $(2,-1,3)$; what is its closest distance to $(3,1,2)$ over the whole of its path?

Theory problems:

3. Prove Proposition 1.4.2

4. The following theorem (see Figure 1.31) can be proved in two ways:

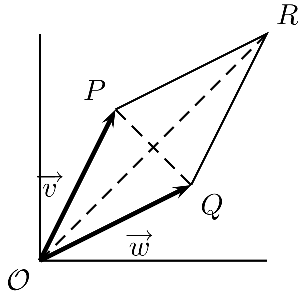


Figure 1.31: Theorem 1.4.4

Theorem 1.4.4. *In any parallelogram, the sum of the squares of the diagonals equals the sum of the squares of the (four) sides.*

(a) Prove Theorem 1.4.4 using the Law of Cosines (§ 1.2, Exercise 13).

(b) Prove Theorem 1.4.4 using vectors, as follows:

Place the parallelogram with one vertex at the origin: suppose the two adjacent vertices are P and Q and the opposite vertex is R (Figure 1.31). Represent the sides by the vectors

$$\begin{aligned}\overrightarrow{v} &= \overrightarrow{\mathcal{O}P} = \overrightarrow{QR}\\ \overrightarrow{w} &= \overrightarrow{\mathcal{O}Q} = \overrightarrow{PR}.\end{aligned}$$

i. Show that the diagonals are represented by

$$\begin{aligned}\overrightarrow{\mathcal{O}R} &= \overrightarrow{v} + \overrightarrow{w}\\ \overrightarrow{PQ} &= \overrightarrow{v} - \overrightarrow{w}.\end{aligned}$$

ii. Show that the squares of the diagonals are

$$\begin{aligned}|\mathcal{O}R|^2 &= |\overrightarrow{v} + \overrightarrow{w}|^2\\ &= |\overrightarrow{v}|^2 + 2\overrightarrow{v} \cdot \overrightarrow{w} + |\overrightarrow{w}|^2\end{aligned}$$

and

$$\begin{aligned}|PQ|^2 &= |\overrightarrow{v} - \overrightarrow{w}|^2\\ &= |\overrightarrow{v}|^2 - 2\overrightarrow{v} \cdot \overrightarrow{w} + |\overrightarrow{w}|^2 .\end{aligned}$$

iii. Show that

$$|\mathcal{O}R|^2 + |PQ|^2 = 2\left|\overrightarrow{v}\right|^2 + 2\left|\overrightarrow{w}\right|^2;$$

but of course

$$\begin{aligned} |\mathcal{O}P|^2 + |PR|^2 + |RQ|^2 + |Q\mathcal{O}|^2 &= \left|\overrightarrow{v}\right|^2 + \left|\overrightarrow{w}\right|^2 + \left|\overrightarrow{v}\right|^2 + \left|\overrightarrow{w}\right|^2 \\ &= 2\left|\overrightarrow{v}\right|^2 + 2\left|\overrightarrow{w}\right|^2. \end{aligned}$$

5. Show that if

$$\overrightarrow{v} = x\overrightarrow{\imath} + y\overrightarrow{\jmath}$$

is any nonzero vector in the plane, then the vector

$$\overrightarrow{w} = y\overrightarrow{\imath} - x\overrightarrow{\jmath}$$

is perpendicular to $\overrightarrow{v}$.

6. Consider the line ℓ in the plane defined as the locus of the linear equation in the two rectangular coordinates x and y

$$Ax + By = C.$$

Define the vector

$$\overrightarrow{N} = A\overrightarrow{\imath} + B\overrightarrow{\jmath}.$$

(a) Show that ℓ is the set of points P whose position vector $\overrightarrow{p}$ satisfies

$$\overrightarrow{N} \cdot \overrightarrow{p} = C.$$

(b) Show that if $\overrightarrow{p}_0$ is the position vector of a specific point on the line, then ℓ is the set of points P whose position vector $\overrightarrow{p}$ satisfies

$$\overrightarrow{N} \cdot (\overrightarrow{p} - \overrightarrow{p}_0) = 0.$$

(c) Show that $\overrightarrow{N}$ is perpendicular to ℓ.

7. Show that if ℓ is a line given by

$$Ax + By = C$$

then the distance from a point $Q(x, y)$ to ℓ is given by the formula

$$\operatorname{dist}(Q, \ell) = \frac{|Ax + By - C|}{\sqrt{A^2 + B^2}}. \tag{1.20}$$

(*Hint:* Let $\overrightarrow{N}$ be the vector given in Exercise 6, and $\overrightarrow{p}_0$ the position vector of any point P_0 on ℓ. Show that $\operatorname{dist}(Q, \ell) = \left|\operatorname{proj}_{\overrightarrow{N}} \overrightarrow{P_0Q}\right| = \left|\operatorname{proj}_{\overrightarrow{N}}(\overrightarrow{q} - \overrightarrow{p}_0)\right|$, and interpret this in terms of A, B, C, x and y.)

1.5 Planes

Equations of Planes

We noted earlier that the locus of a "linear" equation in the three rectangular coordinates x, y and z

$$Ax + By + Cz = D \tag{1.21}$$

is a plane in space. Using the dot product, we can extract a good deal of geometric information about this plane from Equation (1.21).

Let us form a vector from the coefficients on the left of (1.21):

$$\overrightarrow{N} = A\overrightarrow{\imath} + B\overrightarrow{\jmath} + C\overrightarrow{k}.$$

Using

$$\overrightarrow{p} = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$$

as the position vector of $P(x, y, z)$, we see that (1.21) can be expressed as the vector equation

$$\overrightarrow{N} \cdot \overrightarrow{p} = D. \tag{1.22}$$

In the special case that $D = 0$ this is the condition that $\overrightarrow{N}$ is perpendicular to $\overrightarrow{p}$. In general, for any two points P_0 and P_1 satisfying (1.21), the vector $\overrightarrow{P_0P_1}$ from P_0 to P_1, which is the difference of their position vectors

$$\begin{aligned}\overrightarrow{P_0P_1} &= \overrightarrow{\triangle p}\\ &= \overrightarrow{p}_1 - \overrightarrow{p}_0\end{aligned}$$

lies in the plane, and hence satisfies

$$\begin{aligned}\overrightarrow{N}\cdot\overrightarrow{\triangle p} &= \overrightarrow{N}\cdot(\overrightarrow{p}_1-\overrightarrow{p}_0)\\ &= \overrightarrow{N}\cdot\overrightarrow{p}_1-\overrightarrow{N}\cdot\overrightarrow{p}_0\\ &= D-D=0.\end{aligned}$$

Thus, letting the second point P_1 be an arbitrary point $P(x,y,z)$ in the plane, we have

Remark 1.5.1. *If $P_0(x_0,y_0,z_0)$ is any point whose coordinates satisfy* (1.21)

$$Ax_0+By_0+Cz_0=D$$

then the locus of Equation (1.21) *is the plane through P_0 perpendicular to the **normal vector***

$$\overrightarrow{N}:=A\overrightarrow{\imath}+B\overrightarrow{\jmath}+C\overrightarrow{k}.$$

This geometric characterization of a plane from an equation is similar to the geometric characterization of a line from its parametrization: the normal vector $\overrightarrow{N}$ formed from the left side of Equation (1.21) (by analogy with the direction vector $\overrightarrow{v}$ of a parametrized line) determines the "tilt" of the plane, and then the right-hand side D picks out from among the planes perpendicular to $\overrightarrow{N}$ (which are, of course, all parallel to one another) a particular one by, in effect, determining a point that must lie in this plane.

For example, the plane $\mathcal{P}$ determined by the equation

$$2x-3y+z=5$$

is perpendicular to the normal vector

$$\overrightarrow{N}=2\overrightarrow{\imath}-3\overrightarrow{\jmath}+\overrightarrow{k}.$$

To find an explicit point P_0 in $\mathcal{P}$, we can use one of many tricks. One such trick is to fix two of the values x y and z and then substitute to see what the third one must be. If we set

$$x=0=y,$$

then substitution into the equation yields

$$z=5,$$

so we can use as our basepoint

$$P_0(0,0,5)$$

(which is the intersection of $\mathcal{P}$ with the z-axis). We could find the intersections of $\mathcal{P}$ with the other two axes in a similar way. Alternatively, we could notice that

$$\begin{aligned} x &= 1 \\ y &= -1 \end{aligned}$$

means that

$$2x - 3y = 5,$$

so

$$z = 0$$

and we could equally well use as our basepoint

$$P_0'(1,-1,0).$$

Conversely, given a nonzero vector $\overrightarrow{N}$ and a basepoint $P_0(x_0, y_0, z_0)$, we can write an equation for the plane through P_0 perpendicular to $\overrightarrow{N}$ in vector form as

$$\overrightarrow{N} \cdot \overrightarrow{p} = \overrightarrow{N} \cdot \overrightarrow{p}_0$$

or equivalently

$$\overrightarrow{N} \cdot (\overrightarrow{p} - \overrightarrow{p}_0) = 0.$$

For example an equation for the plane through the point $P_0(3,-1,-5)$ perpendicular to $\overrightarrow{N} = 4\overrightarrow{\imath} - 6\overrightarrow{\jmath} + 2\overrightarrow{k}$ is

$$4(x-3) - 6(y+1) + 2(z+5) = 0$$

or

$$4x - 6y + 2z = 8.$$

Note that the point $P_0'(2,1,3)$ also lies in this plane. If we used this as our basepoint (and kept $\overrightarrow{N} = 4\overrightarrow{\imath} - 6\overrightarrow{\jmath} + 2\overrightarrow{k}$) the equation $\overrightarrow{N} \cdot (\overrightarrow{p} - \overrightarrow{p}_0) = 0$ would take the form

$$4(x-2) - 6(y-1) + 2(z-2) = 0$$

which, you should check, is equivalent to the previous equation.

An immediate corollary of Remark 1.5.1 is

Corollary 1.5.2. *The planes given by two linear equations*

$$\begin{aligned} A_1x + B_1y + C_1z &= D_1 \\ A_2x + B_2y + C_2z &= D_2 \end{aligned}$$

are parallel (or coincide) precisely if the two normal vectors

$$\begin{aligned} \overrightarrow{N}_1 &= A_1\overrightarrow{\imath} + B_1\overrightarrow{\jmath} + C_1\overrightarrow{k} \\ \overrightarrow{N}_2 &= A_2\overrightarrow{\imath} + B_2\overrightarrow{\jmath} + C_2\overrightarrow{k} \end{aligned}$$

are (nonzero) scalar multiples of each other; when the normal vectors are equal (i.e., the two left-hand sides of the two equations are the same) then the planes coincide if $D_1 = D_2$, and otherwise they are parallel and non-intersecting.

For example the plane given by the equation

$$-6x + 9y - 3z = 12$$

has normal vector

$$\begin{aligned} \overrightarrow{N} &= -6\overrightarrow{\imath} + 9\overrightarrow{\jmath} - 3\overrightarrow{k} \\ &= -\frac{3}{2}(4\overrightarrow{\imath} - 6\overrightarrow{\jmath} + 2\overrightarrow{k}) \end{aligned}$$

so multiplying the equation by $-2/3$, we get an equivalent equation

$$4x - 6y + 2z = -8$$

which shows that this plane is parallel to (and does not intersect) the plane specified earlier by

$$4x - 6y + 2z = 8$$

(since $8 \neq -8$).

We can also use vector ideas to calculate the **distance from a point** $Q(x,y,z)$ **to the plane** $\mathcal{P}$ given by an equation

$$Ax + By + Cz = D.$$

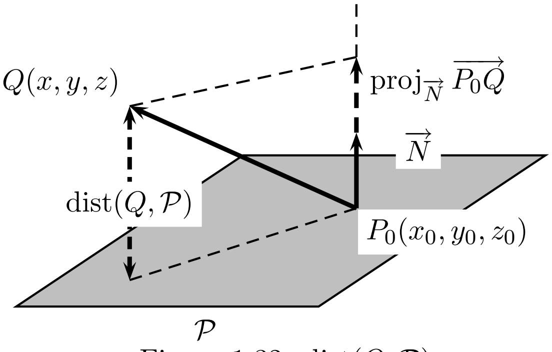

$Q(x,y,z)$
$\text{proj}_{\overrightarrow{N}}\,\overrightarrow{P_0Q}$
$\overrightarrow{N}$
$\text{dist}(Q,\mathcal{P})$
$P_0(x_0,y_0,z_0)$
$\mathcal{P}$

Figure 1.32: $\text{dist}(Q,\mathcal{P})$

If $P_0(x_0,y_0,z_0)$ is any point on $\mathcal{P}$ (see Figure 1.32) then the (perpendicular) distance from Q to $\mathcal{P}$ is the (length of the) projection of $\overrightarrow{P_0Q} = \triangle x\,\overrightarrow{\imath} + \triangle y\,\overrightarrow{\jmath} + \triangle z\,\overrightarrow{k}$ in the direction of $\overrightarrow{N} = A\,\overrightarrow{\imath} + B\,\overrightarrow{\jmath} + C\,\overrightarrow{k}$

$$\text{dist}(Q,\mathcal{P}) = \left|\text{proj}_{\overrightarrow{N}}\,\overrightarrow{P_0Q}\right|$$

which we can calculate as

$$\begin{aligned} &= \frac{\left|\overrightarrow{N}\cdot(\overrightarrow{q} - \overrightarrow{p}_0)\right|}{\left\|\overrightarrow{N}\right\|} \\ &= \frac{\left|\overrightarrow{N}\cdot\overrightarrow{q} - \overrightarrow{N}\cdot\overrightarrow{p}_0\right|}{\sqrt{\overrightarrow{N}\cdot\overrightarrow{N}}} \\ &= \frac{|(Ax+By+Cz)-D|}{\sqrt{A^2+B^2+C^2}}. \end{aligned}$$

For example, the distance from $Q(1,1,2)$ to the plane $\mathcal{P}$ given by

$$2x - 3y + z = 5$$

is

$$\begin{aligned}\operatorname{dist}(Q,\mathcal{P}) &= \frac{|(2)(1)-3(1)+1(2)-(5)|}{\sqrt{2^2+(-3)^2+1^2}}\\ &= \frac{4}{\sqrt{15}}.\end{aligned}$$

The **distance between two parallel planes** is the distance from any *point* Q on *one* of the planes to the *other plane*. Thus, the distance between the parallel planes discussed earlier

$$\begin{array}{rrrcr} 4x & -6y & +2z & = & 8\\ -6x & +9y & -3z & = & 12\end{array}$$

is the same as the distance from $Q(3,-1,-5)$, which lies on the first plane, to the second plane, or

$$\begin{aligned}\operatorname{dist}(\mathcal{P}_1,\mathcal{P}_2) &= \operatorname{dist}(Q,\mathcal{P}_2)\\ &= \frac{|(-6)(3)+(9)(-1)+(-3)(-5)-(12)|}{\sqrt{(-6)^2+(9)^2+(-3)^2}}\\ &= \frac{24}{3\sqrt{14}}.\end{aligned}$$

Finally, the **angle θ between two planes** $\mathcal{P}_1$ and $\mathcal{P}_2$ can be defined as follows (Figure 1.33): if they are parallel, the angle is zero. Otherwise, they intersect along a line ℓ_0: pick a point P_0 on ℓ_0, and consider the line ℓ_i in $\mathcal{P}_i$ $(i=1,2)$ through P_0 and perpendicular to ℓ_0. Then θ is by definition the angle between ℓ_1 and ℓ_2.

To relate this to the equations of $\mathcal{P}_1$ and $\mathcal{P}_2$, consider the plane $\mathcal{P}_0$ (through P_0) containing the lines ℓ_1 and ℓ_2. $\mathcal{P}_0$ is perpendicular to ℓ_0 and hence contains the arrows with tails at P_0 representing the normals $\overrightarrow{N}_1 = A_1\overrightarrow{\imath}+B_1\overrightarrow{\jmath}+C_1\overrightarrow{k}$ (*resp.* $\overrightarrow{N}_2 = A_2\overrightarrow{\imath}+B_2\overrightarrow{\jmath}+C_2\overrightarrow{k}$) to $\mathcal{P}_1$ (*resp.* $\mathcal{P}_2$). But since $\overrightarrow{N}_i$ is perpendicular to ℓ_i for $i=1,2$, the angle between the vectors $\overrightarrow{N}_1$ and $\overrightarrow{N}_2$ is the same as the angle between ℓ_1 and ℓ_2 (Figure 1.34), hence

$$\cos\theta = \frac{\overrightarrow{N}_1\cdot\overrightarrow{N}_2}{\left|\overrightarrow{N}_1\right|\left|\overrightarrow{N}_2\right|} \tag{1.23}$$

For example, the planes determined by the two equations

$$\begin{array}{rrrcr} x & +y & +z & = & 3\\ x & +\sqrt{6}y & -z & = & 2\end{array}$$

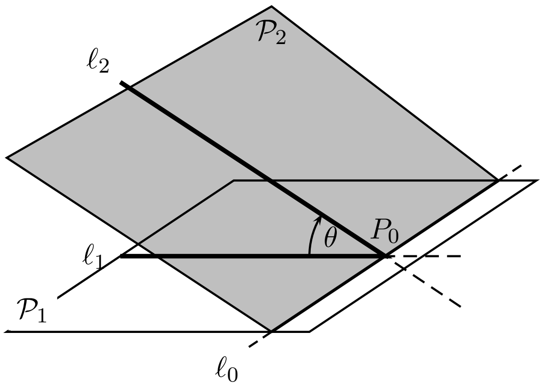


Figure 1.33: Angle between two planes

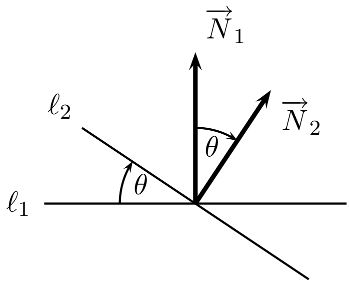


Figure 1.34: Angle between planes (cont'd)

meet at angle θ, where

$$\begin{aligned}\cos\theta &= \frac{\left|(1,1,1)\cdot(1,\sqrt{6},-1)\right|}{\sqrt{1^2+1^2+1^2}\sqrt{1^2+\sqrt{6}^2+(-1)^2}}\\ &= \frac{\left|1+\sqrt{6}-1\right|}{\sqrt{3}\sqrt{8}}\\ &= \frac{\sqrt{6}}{2\sqrt{6}}\\ &= \frac{1}{2}\end{aligned}$$

so θ equals $\pi/6$ radians.

Parametrization of Planes

So far, we have dealt with planes given as loci of linear equations. This is an *implicit* specification. However, there is another way to specify a plane, which is more *explicit* and in closer analogy to the parametrizations we have used to specify lines in space.

Suppose

$$\begin{aligned}\overrightarrow{v} &= v_1\overrightarrow{\imath} + v_2\overrightarrow{\jmath} + v_3\overrightarrow{k}\\ \overrightarrow{w} &= w_1\overrightarrow{\imath} + w_2\overrightarrow{\jmath} + w_3\overrightarrow{k}\end{aligned}$$

are two linearly independent vectors in $\mathbb{R}^3$. If we represent them via arrows in standard position, they determine a plane $\mathcal{P}_0$ through the origin. Note that any linear combination of $\overrightarrow{v}$ and $\overrightarrow{w}$

$$\overrightarrow{p}(s,t) = s\overrightarrow{v} + t\overrightarrow{w}$$

is the position vector of some point in this plane: when s and t are both positive, we draw the parallelogram with one vertex at the origin, one pair of sides parallel to $\overrightarrow{v}$, of length $s\left|\overrightarrow{v}\right|$, and the other pair of sides parallel to $\overrightarrow{w}$, with length $t\left|\overrightarrow{w}\right|$ (Figure 1.35). Then $\overrightarrow{p}(s,t)$ is the vertex opposite the origin in this parallelogram. Conversely, the position vector of any point P in $\mathcal{P}_0$ can be expressed uniquely as a linear combination of $\overrightarrow{v}$ and $\overrightarrow{w}$. We leave it to you to complete the details (see Exercise 4 in § 1.2).

Remark 1.5.3. *If $\overrightarrow{v}$ and $\overrightarrow{w}$ are linearly independent vectors in $\mathbb{R}^3$, then the set of all linear combinations of $\overrightarrow{v}$ and $\overrightarrow{w}$*

$$\mathcal{P}_0\left(\overrightarrow{v},\overrightarrow{w}\right) := \left\{s\overrightarrow{v} + t\overrightarrow{w} \mid s,t\in\mathbb{R}\right\}$$

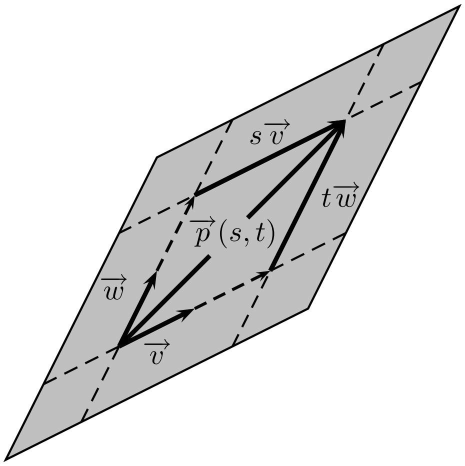


Figure 1.35: Linear Combination

is the set of position vectors for points in the plane (through the origin) determined by $\overrightarrow{v}$ and $\overrightarrow{w}$, called the ***span*** *of $\overrightarrow{v}$ and $\overrightarrow{w}$.*

Suppose now we want to describe a plane $\mathcal{P}$ parallel to $\mathcal{P}_0(\overrightarrow{v},\overrightarrow{w})$, but going through an arbitrarily given basepoint $P_0(x_0,y_0,z_0)$. If we let

$$\overrightarrow{p}_0 = x_0\overrightarrow{\imath} + y_0\overrightarrow{\jmath} + z_0\overrightarrow{k}$$

be the position vector of P_0, then displacement by $\overrightarrow{p}_0$ moves the origin $\mathcal{O}$ to P_0 and the plane $\mathcal{P}_0(\overrightarrow{v},\overrightarrow{w})$ to the plane $\mathcal{P}$ through P_0 parallel to $\mathcal{P}_0(\overrightarrow{v},\overrightarrow{w})$. It is clear from Remark 1.5.3 that the position vector

$$\overrightarrow{p} = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$$

of every point in $\mathcal{P}$ can be expressed as $\overrightarrow{p}_0$ plus some linear combination of $\overrightarrow{v}$ and $\overrightarrow{w}$

$$\overrightarrow{p}(s,t) = \overrightarrow{p}_0 + s\overrightarrow{v} + t\overrightarrow{w}$$

or

$$\begin{aligned}x &= x_0 + sv_1 + tw_1\\ y &= y_0 + sv_2 + tw_2\\ z &= z_0 + sv_3 + tw_3\end{aligned}$$

for a unique pair of scalars $s, t \in \mathbb{R}$. These scalars form an oblique coordinate system for points in the plane $\mathcal{P}$. Equivalently, we can regard these equations as defining a **vector-valued function** $\overrightarrow{p}(s,t)$ which assigns to each point (s,t) in the "st-plane" a point $\overrightarrow{p}(s,t)$ of the plane $\mathcal{P}$ in $\mathbb{R}^3$. This is a **parametrization** of the plane $\mathcal{P}$; by contrast with the parametrization of a line, which uses *one* parameter t, this uses *two* parameters, s and t.

We can use this to parametrize the plane determined by any three noncollinear points. Suppose $\triangle PQR$ is a nondegenerate triangle in $\mathbb{R}^3$. Set

$$\overrightarrow{p}_0 = \overrightarrow{\mathcal{O}P},$$

the position vector of the vertex P, and let

$$\overrightarrow{v} = \overrightarrow{PQ}$$

and

$$\overrightarrow{w} = \overrightarrow{PR}$$

be two vectors representing the sides of the triangle at this vertex. Then the parametrization

$$\begin{aligned}\overrightarrow{p}(s,t) &= \overrightarrow{p}_0 + s\overrightarrow{v} + t\overrightarrow{w}\\ &= \overrightarrow{\mathcal{O}P} + s\overrightarrow{PQ} + t\overrightarrow{PR}\end{aligned}$$

describes the plane containing our triangle; the vertices have position vectors

$$\begin{aligned}\overrightarrow{\mathcal{O}P} &= \overrightarrow{p}_0 = \overrightarrow{p}(0,0)\\ \overrightarrow{\mathcal{O}Q} &= \overrightarrow{p}_0 + \overrightarrow{v} = \overrightarrow{p}(1,0)\\ \overrightarrow{\mathcal{O}R} &= \overrightarrow{p}_0\overrightarrow{w} = \overrightarrow{p}(0,1)\,.\end{aligned}$$

For example, the three points one unit out along the three (positive) coordinate axes

$$\begin{array}{ll} P(1,0,0) & (\overrightarrow{\mathcal{O}P} = \overrightarrow{\imath}) \\ Q(0,1,0) & (\overrightarrow{\mathcal{O}Q} = \overrightarrow{\jmath}) \\ R(0,0,1) & (\overrightarrow{\mathcal{O}R} = \overrightarrow{k}) \end{array}$$

determine the plane with parametrization

$$\overrightarrow{p}(s,t) = \overrightarrow{\imath} + s(\overrightarrow{\jmath} - \overrightarrow{\imath}) + t(\overrightarrow{k} - \overrightarrow{\imath})$$

or

$$\begin{array}{rcllll} x & = & 1 & -s & -t \\ y & = & & s & \\ z & = & & & t. \end{array}$$

To see whether the point $P(3,1,-3)$ lies in this plane, we can try to solve

$$\begin{array}{rcllll} 3 & = & 1 & -s & -t \\ 1 & = & & s & \\ -3 & = & & & t; \end{array}$$

it is clear that the values of s and t given by the second and third equations also satisfy the first, so P does indeed lie in the plane through $\overrightarrow{\imath}$, $\overrightarrow{\jmath}$ and $\overrightarrow{k}$:

$$\overrightarrow{\mathcal{O}P} = \overrightarrow{p}(1,-3)\,.$$

Given a linear equation, we can parametrize its locus by finding three noncollinear points on the locus and using the procedure above. For example, to parametrize the plane given by

$$3x - 2y + 4z = 12$$

we need to find three noncollinear points in this plane. If we set

$$y = z = 0,$$

we have

$$x = 4,$$

and so we can take our basepoint P to be $(4,0,0)$, or

$$\overrightarrow{p}_0 = 4\overrightarrow{\imath}.$$

To find two other points, we could note that if

$$x = 4$$

then

$$-2y + 4z = 0,$$

so any choice with $y = 2z$ will work, for example $Q(4,2,1)$, or

$$\overrightarrow{v} = \overrightarrow{PQ} = 2\overrightarrow{\jmath} + \overrightarrow{k}$$

gives one such point. Unfortunately, any *third* point given by this scheme will produce $\overrightarrow{w}$ a scalar multiple of $\overrightarrow{v}$, so won't work. However, if we set

$$x = 0$$

we have

$$-2y + 4z = 12,$$

and one solution of this is

$$\begin{aligned} y &= -4, \\ z &= 1, \end{aligned}$$

so $R(0,-4,1)$ works, with

$$\overrightarrow{w} = \overrightarrow{PR} = -4\overrightarrow{\imath} - 4\overrightarrow{\jmath} + \overrightarrow{k}.$$

This leads to the parametrization

$$\overrightarrow{p}(s,t) = 4\overrightarrow{\imath} + s(2\overrightarrow{\jmath} + \overrightarrow{k}) + t(-4\overrightarrow{\imath} - 4\overrightarrow{\jmath} + \overrightarrow{k})$$

or

$$\begin{array}{rclcc} x & = & 4 & & -4t \\ y & = & & 2s & -4t \\ z & = & & s & +t. \end{array}$$

A different parametrization results from setting coordinates equal to zero in pairs: this yields the same basepoint, $P(4,0,0)$, but two new points, $Q(0,-6,0)$, $R(0,0,3)$. Then

$$\vec{p}(s,t) = 4\vec{\imath} + s(-4\vec{\imath} - 6\vec{\jmath}) + t(-4\vec{\imath} + 3\vec{k})$$

or

$$\begin{array}{ccccc} x & = & 4 & -4s & -4t \\ y & = & & -6s & \\ z & = & & & 3t. \end{array}$$

The converse problem—given a parametrization of a plane, to find an equation describing it—can sometimes be solved easily: for example, the plane through $\vec{\imath}$, $\vec{\jmath}$ and $\vec{k}$ easily leads to the relation $x + y + z = 1$. However, in general, it will be easier to handle this problem using cross products (§ 1.6).

Exercises for § 1.5

Practice problems:

1. Write an equation for the plane through P perpendicular to $\vec{N}$:

   (a) $P(2,-1,3)$, $\vec{N} = \vec{\imath} + \vec{\jmath} + \vec{k}$ (b) $P(1,1,1)$, $\vec{N} = 2\vec{\imath} - \vec{\jmath} + \vec{k}$

   (c) $P(3,2,1)$, $\vec{N} = \vec{\jmath}$

2. Find a point P on the given plane, and a vector normal to the plane:

   (a) $3x + y - 2z = 1$ (b) $x - 2y + 3z = 5$

   (c) $5x - 4y + z = 8$ (d) $z = 2x + 3y + 1$

   (e) $x = 5$

3. Find a parametrization of each plane below:

   (a) $2x + 3y - z = 4$ (b) $z = 4x + 5y + 1$

   (c) $x = 5$

4. Find an equation for the plane through the point $(2,-1,2)$

   (a) parallel to the plane $3x + 2y + z = 1$

(b) perpendicular to the line given by

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

5. Find the distance from the point $(3, 2, 1)$ to the plane $x - y + z = 5$.

6. Find the angle between $\mathcal{P}_1$ and $\mathcal{P}_2$:

   (a)

$$\begin{aligned} \mathcal{P}_1: &\quad 2x + y - z = 4 \\ \mathcal{P}_2: &\quad 2x - y + 3z = 3 \end{aligned}$$

   (b)

$$\begin{aligned} \mathcal{P}_1: &\quad 2x + 2y + 2\sqrt{6}z = 1 \\ \mathcal{P}_2: &\quad \sqrt{3}x + \sqrt{3}y + \sqrt{2}z = \sqrt{5} \end{aligned}$$

Theory problems:

7. (a) *Show:* If

$$\begin{aligned} \overrightarrow{p}(t) &= \overrightarrow{p}_0 + t\overrightarrow{v} \\ \overrightarrow{q}(t) &= \overrightarrow{p}_0 + t\overrightarrow{w} \end{aligned}$$

   are parametrizations of two distinct lines through a point P_0 in $\mathcal{P}$ (with position vector $\overrightarrow{p}_0$), then

$$\overrightarrow{p}(s, t) = \overrightarrow{p}_0 + s\overrightarrow{v} + t\overrightarrow{w}$$

   is a parametrization of the plane $\mathcal{P}$.

   (b) Suppose an equation for $\mathcal{P}$ is

$$Ax + By + Cz = D$$

with $C \neq 0$. *Show* that the intersections of $\mathcal{P}$ with the xz-plane and yz-plane are given by

$$z = -\frac{A}{C}x + \frac{D}{C}$$
$$z = -\frac{B}{C}y + \frac{D}{C}$$

and combine this with (a) to get a parametrization of $\mathcal{P}$.

(c) Apply this to the plane $x + 2y + 3z = 9$.

8. Find an equation for the plane $\mathcal{P}$ parametrized by

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

1.6 Cross Products

Oriented Areas in the Plane

The standard formula for the area of a triangle

$$\mathcal{A} = \frac{1}{2}bh, \tag{1.24}$$

where b is the "base" length and h is the "height", is not always convenient to apply. Often we are presented with either the lengths of the three sides or the coordinates of the vertices (from which these lengths are easily calculated); in either case we can take a convenient side as the *base*, but calculating the *height*—the perpendicular distance from the base to the opposite vertex—can require some work.

A famous formula for the area of a triangle in terms of the lengths of its sides is the second of two area formulas proved by Heron of Alexandria (*ca.* 75 AD) in his *Metrica*:

$$\mathcal{A} = \sqrt{s(s-a)(s-b)(s-c)} \tag{1.25}$$

where a, b and c are the lengths of the sides of the triangle, and s is the *semiperimeter*

$$s = \frac{1}{2}(a + b + c).$$

Equation (1.27) is known as **Heron's formula**, although it now seems clear from Arabic commentaries that it was already known to Archimedes of Syracuse (*ca.* 287-212 BC). In Exercise 15 and Exercise 16 we will consider both of the area formulas given in the *Metrica*; also, in Exercise 5 we will derive a vector formula for the area of a triangle based on the discussion of the distance from a point to a line on p. 51.

Here, however, we will concentrate on finding a formula for the area of a triangle in $\mathbb{R}^2$ in terms of the coordinates of its vertices. Suppose the vertices are $A(a_1, a_2)$, $B(b_1, b_2)$, and $C(c_1, c_2)$. Using the side AB as the base, we have

$$b = \left|\overrightarrow{AB}\right|$$

and, letting θ be the angle at vertex A,

$$h = \left|\overrightarrow{AC}\right| \sin\theta,$$

so

$$\mathcal{A}\left(\triangle ABC\right) = \frac{1}{2}\left|\overrightarrow{AB}\right|\left|\overrightarrow{AC}\right|\sin\theta.$$

To express this in terms of the coordinates of the vertices, note that

$$\overrightarrow{AB} = x_{AB}\overrightarrow{\imath} + y_{AB}\overrightarrow{\jmath}$$

where

$$\begin{aligned} x_{AB} &= b_1 - a_1 \\ y_{AB} &= b_2 - a_2 \end{aligned}$$

and similarly

$$\overrightarrow{AC} = x_{AC}\overrightarrow{\imath} + y_{AC}\overrightarrow{\jmath}.$$

Recall that any vector $\overrightarrow{v} = x\overrightarrow{\imath} + y\overrightarrow{\jmath}$ in the plane can also be written in "polar" form as

$$\overrightarrow{v} = \left|\overrightarrow{v}\right|\left(\cos\theta_v\overrightarrow{\imath} + \sin\theta_v\overrightarrow{\jmath}\right)$$

where θ_v is the counterclockwise angle between $\overrightarrow{v}$ and the horizontal vector $\overrightarrow{\imath}$. Thus,

$$\theta = \theta_2 - \theta_1$$

where θ_1 and θ_2 are the angles between $\overrightarrow{\imath}$ and each of the vectors $\overrightarrow{AB}$, $\overrightarrow{AC}$, and

$$\theta_2 > \theta_1.$$

But the formula for the sine of a sum of angles gives us

$$\sin\theta = \cos\theta_1 \sin\theta_2 - \cos\theta_2 \sin\theta_1.$$

Thus, if $\theta_{AC} > \theta_{AB}$ we have

$$\begin{aligned}\mathcal{A}\left(\triangle ABC\right) &= \frac{1}{2}\left|\overrightarrow{AB}\right|\left|\overrightarrow{AC}\right|\sin\theta \\ &= \frac{1}{2}\left|\overrightarrow{AB}\right|\left|\overrightarrow{AC}\right|(\cos\theta_{AB}\sin\theta_{AC} - \cos\theta_{AC}\sin\theta_{AB}) \\ &= \frac{1}{2}\left[(\left|\overrightarrow{AB}\right|\cos\theta_{AB})(\left|\overrightarrow{AC}\right|\sin\theta_{AC}) - (\left|\overrightarrow{AC}\right|\cos\theta_{AC})(\left|\overrightarrow{AB}\right|\sin\theta_{AB})\right] \\ &= \frac{1}{2}[x_{AB}y_{AC} - x_{AC}y_{AB}].\end{aligned}$$

The condition $\theta_{AC} > \theta_{AB}$ means that the direction of $\overrightarrow{AC}$ is a *counterclockwise* rotation (by an angle between 0 and π radians) from that of $\overrightarrow{AB}$; if the rotation from $\overrightarrow{AB}$ to $\overrightarrow{AC}$ is *clockwise*, then the two vectors trade places—or equivalently, the expression above gives us *minus* the area of $\triangle ABC$.

The expression in brackets is easier to remember using a “visual” notation. An array of four numbers

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$$

in two horizontal rows, with the entries vertically aligned in columns, is called a **2 × 2 matrix**[12]. The **determinant** of a 2×2 matrix is the product x_1y_2 of the *downward* diagonal minus the product x_2y_1 of the *upward* diagonal. We denote the determinant by replacing the brackets surrounding the array with vertical bars:[13]

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = x_1y_2 - x_2y_1.$$

It is also convenient to sometimes treat the determinant as a function of its rows, which we think of as vectors:

$$\overrightarrow{v}_i = x_i\overrightarrow{\imath} + y_i\overrightarrow{\jmath}, \quad i = 1, 2;$$

treated this way, the determinant will be denoted

$$\Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2\right) = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.$$

[12] pronounced *“two by two matrix”*

[13] When a matrix is given a letter name—say A—we name its determinant det A.

If we are simply given the coordinates of the vertices of a triangle in the plane, without a picture of the triangle, we can pick one of the vertices—call it A—and calculate the vectors to the other two vertices—call them B and C—and then take half the determinant. This will equal the area of the triangle *up to sign*:

$$\frac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = \sigma(A,B,C)\,\mathcal{A}\left(\triangle ABC\right), \tag{1.26}$$

where $\sigma(A,B,C) = \pm 1$ depending on the direction of rotation from $\overrightarrow{AB}$ to $\overrightarrow{AC}$. We refer to $\sigma(A,B,C)$ as the **orientation** of the triangle (so an **oriented triangle** is one whose vertices have been assigned a specific order) and the quantity $\sigma(A,B,C)\,\mathcal{A}\left(\triangle ABC\right)$ as the **signed area** of the oriented triangle. You should verify that the oriented triangle $\triangle ABC$ has **positive orientation** precisely if going from A to B to C and then back to A constitutes a *counterclockwise* transversal of its periphery, and a **negative orientation** if this traversal is *clockwise*. Thus the orientation is determined by the "cyclic order" of the vertices: a **cyclic permutation** (moving everything one space over, and putting the entry that falls off the end at the beginning) doesn't change the orientation:

$$\sigma(A,B,C) = \sigma(B,C,A) = \sigma(C,A,B)\,.$$

For example, the triangle with vertices $A(2,-3)$, $B(4,-2)$ and $C(3,-1)$, shown in Figure 1.36, has

$$\begin{aligned}\overrightarrow{AB} &= 2\overrightarrow{\imath} + \overrightarrow{\jmath} \\ \overrightarrow{AC} &= \overrightarrow{\imath} + 2\overrightarrow{\jmath}\end{aligned}$$

and its signed area is

$$\begin{aligned}\frac{1}{2}\begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} &= \frac{1}{2}[(2)(2) - (1)(1)] \\ &= \frac{1}{2}[4-1] \\ &= \frac{3}{2};\end{aligned}$$

you can verify from Figure 1.36 that the path $A \mapsto B \mapsto C \mapsto A$ traverses the triangle *counterclockwise*.

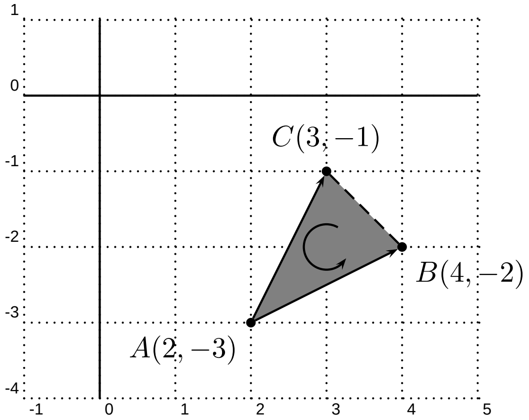


Figure 1.36: Oriented Triangle $\triangle ABC$, Positive Orientation

The triangle with vertices $A(-3,4)$, $B(-2,5)$ and $C(-1,3)$ (Figure 1.37) has

$$\begin{aligned}\overrightarrow{AB} &= -\overrightarrow{\imath} + \overrightarrow{\jmath}\\ \overrightarrow{AC} &= 2\overrightarrow{\imath} - \overrightarrow{\jmath}\end{aligned}$$

and its signed area is

$$\begin{aligned}\frac{1}{2}\begin{vmatrix} -1 & 1 \\ 2 & -1 \end{vmatrix} &= \frac{1}{2}[(-1)(-1)-(2)(1)]\\ &= \frac{1}{2}[1-2]\\ &= -\frac{1}{2};\end{aligned}$$

you can verify from Figure 1.37 that the path $A \mapsto B \mapsto C \mapsto A$ traverses the triangle *clockwise.*

Finally, the triangle with vertices $A(2,4)$, $B(4,5)$ and $C(3,3)$ (Figure 1.38) has

$$\begin{aligned}\overrightarrow{AB} &= 2\overrightarrow{\imath} + \overrightarrow{\jmath}\\ \overrightarrow{AC} &= \overrightarrow{\imath} - \overrightarrow{\jmath}\end{aligned}$$

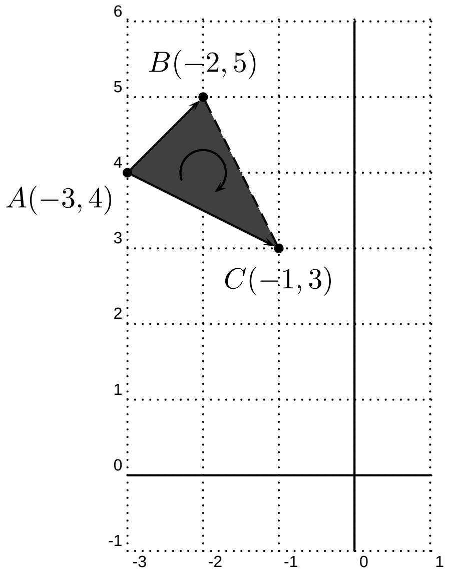


Figure 1.37: Oriented Triangle $\triangle ABC$, Negative Orientation

and its signed area is

$$\begin{aligned}\frac{1}{2}\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} &= \frac{1}{2}[(2)(-1)-(1)(1)] \\ &= \frac{1}{2}[-2-1] \\ &= -\frac{3}{2};\end{aligned}$$

you can verify from Figure 1.38 that the path $A \mapsto B \mapsto C \mapsto A$ traverses the triangle *clockwise*.

These ideas can be extended to polygons in the plane: for example, a quadrilateral with vertices A, B, C and D is positively (*resp.* negatively) oriented if the vertices in this order are consecutive in the counterclockwise (*resp.* clockwise) direction (Figure 1.39) and we can define its signed area as the area (*resp.* minus the area). By cutting the quadrilateral into two triangles with a diagonal, and using Equation (1.26) on each, we can calculate its signed area from the coordinates of its vertices. This will be explored in Exercises 9-13.

For the moment, though, we consider a very special case. Suppose we

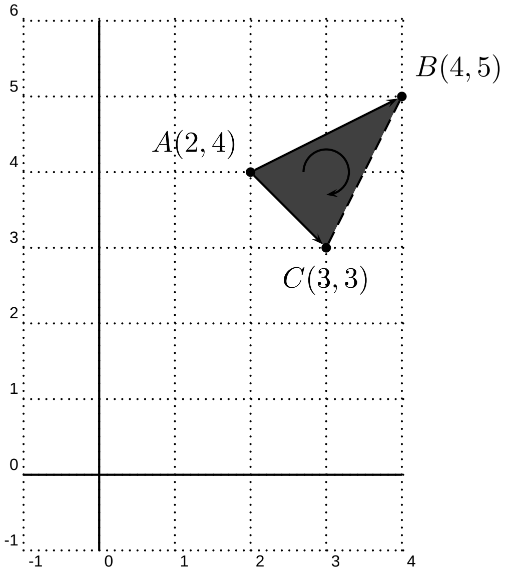


Figure 1.38: Oriented Triangle $\triangle ABC$, Negative Orientation

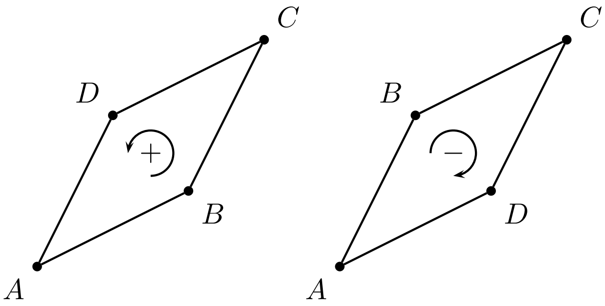


Figure 1.39: Oriented Quadrilaterals

have two nonzero vectors

$$\begin{aligned} \overrightarrow{v}_1 &= x_1\overrightarrow{\imath} + y_1\overrightarrow{\jmath} \\ \overrightarrow{v}_2 &= x_2\overrightarrow{\imath} + y_2\overrightarrow{\jmath}. \end{aligned}$$

Then the determinant using these rows

$$\Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2\right) = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.$$

can be interpreted geometrically as follows. Let $P(x_1, y_1)$ and $Q(x_2, y_2)$ be the points with position vectors $\overrightarrow{v}_1$ and $\overrightarrow{v}_2$, respectively, and let $R(x_1 + x_2, y_1 + y_2)$ be the point whose position vector is $\overrightarrow{v}_1 + \overrightarrow{v}_2$ (Figure 1.40). Then the signed area of $\triangle \mathcal{O}PQ$ equals $\frac{1}{2}\Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2\right)$; but note that the *parallelogram* $\mathcal{O}PRQ$ has the same orientation as the *triangles* $\triangle \mathcal{O}PQ$ and $\triangle PRQ$, and these two triangles $\triangle \mathcal{O}PQ$ and $\triangle PRQ$ are congruent, hence have the same area. Thus the signed area of the parallelogram $\mathcal{O}PRQ$ is twice the signed area of $\triangle \mathcal{O}PQ$; in other words

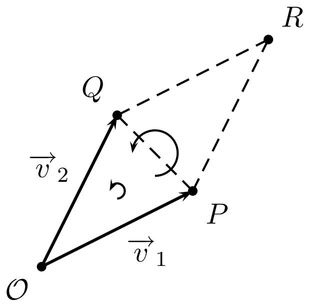


Figure 1.40: Proposition 1.6.1

Proposition 1.6.1. *The* 2×2 *determinant*

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$$

is the signed area of the parallelogram $\mathcal{O}PRQ$*, where*

$$\begin{aligned} \overrightarrow{\mathcal{O}P} &= x_1\overrightarrow{\imath} + y_1\overrightarrow{\jmath} \\ \overrightarrow{\mathcal{O}Q} &= x_2\overrightarrow{\imath} + y_2\overrightarrow{\jmath} \end{aligned}$$

and

$$\overrightarrow{\mathcal{O}R} = \overrightarrow{\mathcal{O}P} + \overrightarrow{\mathcal{O}Q}.$$

Let us note several properties of the determinant $\Delta\left(\overrightarrow{v}, \overrightarrow{w}\right)$ which make it a useful computational tool. The proof of each of these properties is a straightforward calculation (Exercise 6):

Proposition 1.6.2. *The 2×2 determinant $\Delta\left(\overrightarrow{v}, \overrightarrow{w}\right)$ has the following algebraic properties:*

1. *It is* ***additive*** *in each slot:*[14] *for any three vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{w} \in \mathbb{R}^2$*

$$\Delta\left(\overrightarrow{v}_1 + \overrightarrow{w}, \overrightarrow{v}_2\right) = \Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2\right) + \Delta\left(\overrightarrow{w}, \overrightarrow{v}_2\right)$$
$$\Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2 + \overrightarrow{w}\right) = \Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2\right) + \Delta\left(\overrightarrow{v}_1, \overrightarrow{w}\right).$$

2. *It is* ***homogeneous*** *in each slot: for any two vectors $\overrightarrow{v}_1, \overrightarrow{v}_2 \in \mathbb{R}^2$ and any scalar $r \in \mathbb{R}$*

$$\Delta\left(r\overrightarrow{v}_1, \overrightarrow{v}_2\right) = r\Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2\right) = \Delta\left(\overrightarrow{v}_1, r\overrightarrow{v}_2\right).$$

3. *It is* ***skew-symmetric****: for any two vectors $\overrightarrow{v}_1, \overrightarrow{v}_2 \in \mathbb{R}^2$*

$$\Delta\left(\overrightarrow{v}_2, \overrightarrow{v}_1\right) = -\Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2\right).$$

In particular,

Corollary 1.6.3. *A 2×2 determinant equals zero precisely if its rows are linearly dependent.*

Proof. If two vectors are linearly dependent, we can write both of them as scalar multiples of the same vector, and

$$\Delta\left(r\overrightarrow{v}, s\overrightarrow{v}\right) = rs\Delta\left(\overrightarrow{v}, \overrightarrow{v}\right) = \Delta\left(s\overrightarrow{v}, r\overrightarrow{v}\right) = -\Delta\left(r\overrightarrow{v}, s\overrightarrow{v}\right)$$

where the last equality comes from skew-symmetry. So $\Delta\left(r\overrightarrow{v}, s\overrightarrow{v}\right)$ equals its negative, and hence must equal zero.

To prove the reverse implication, write $\overrightarrow{v}_i = x_i\overrightarrow{\imath} + y_i\overrightarrow{\jmath}$, $i = 1, 2$, and suppose $\Delta\left(\overrightarrow{v}_1, \overrightarrow{v}_2\right) = 0$. This translates to

$$x_1y_2 - x_2y_1 = 0$$

[14] This is a kind of distributive law.

or

$$x_1y_2 = x_2y_1.$$

Assuming that $\overrightarrow{v}_1$ and $\overrightarrow{v}_2$ are both not vertical ($x_i \neq 0$ for $i = 1, 2$), we can conclude that

$$\frac{y_2}{x_2} = \frac{y_1}{x_1}$$

which means they are dependent. We leave it to you to show that if one of them *is* vertical (and the determinant is zero), then either the other is also vertical, or else one of them is the zero vector. □

Of course, Corollary 1.6.3 can also be proved on geometric grounds, using Proposition 1.6.1 (Exercise 7).

Oriented Areas in Space

Suppose now that A, B and C are three noncollinear points in $\mathbb{R}^3$. We can think of the ordered triple $(A < B < C)$ as defining an oriented triangle, and hence associate to it a "signed" area. But which sign should it have—positive or negative? The question is ill-posed, since the words "clockwise" and "counterclockwise" have no natural meaning in space: even when A, B and C all lie in the xy-plane, and have positive orientation in terms of the previous subsection, the motion from A to B to C will look counterclockwise only when viewed from *above* the plane; viewed from *underneath*, it will look *clockwise*. When the plane containing A, B and C is at some cockeyed angle, it is not at all clear which viewpoint is correct.

We deal with this by turning the tables:[15] the motion, instead of being inherently "clockwise" or "counterclockwise", determines which side of the plane yields a viewpoint from which the motion appears counterclockwise. We can think of this as replacing the *sign* $\sigma(A, B, C)$ with a *unit vector*, normal to the plane containing the three points and pointing toward the side of this plane from which the motion described by our order appears counterclockwise. One way to determine which of the two unit normals is correct is the **right-hand rule**: point the fingers of your right hand along the direction of motion; then your (right) thumb will point in the appropriate direction. In Figure 1.41 we sketch the triangle with vertices $A(2, -3, 4)$, $B(4, -2, 5)$, and $C(3, -1, 3)$; from our point of view (we are looking from moderately high in the first octant), the orientation appears counterclockwise.

[15] No pun intended! :-)

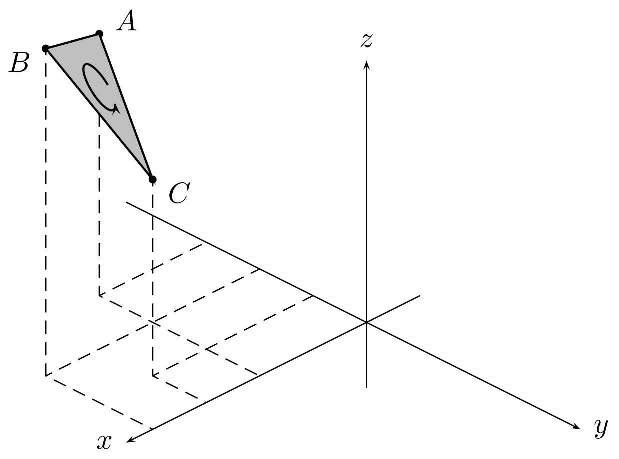


Figure 1.41: Oriented Triangle in $\mathbb{R}^3$

By interpreting $\sigma(A, B, C)$ as a unit normal vector, we associate to an oriented triangle $\triangle ABC \in \mathbb{R}^3$ an **oriented area**

$$\vec{\mathcal{A}}(\triangle ABC) = \overrightarrow{\sigma}(A, B, C)\,\mathcal{A}(\triangle ABC)$$

represented by a vector normal to the triangle whose length is the ordinary area of $\triangle ABC$. Note that for a triangle in the xy-plane, this means $\overrightarrow{\sigma}(ABC) = \sigma(ABC)\,\overrightarrow{k}$: the oriented area is the vector $\overrightarrow{k}$ times the signed area in our old sense. This interpretation can be applied as well to any oriented polygon contained in a plane in space.

In particular, by analogy with Proposition 1.6.1, we can define a function which assigns to a pair of vectors $\overrightarrow{v}, \overrightarrow{w} \in \mathbb{R}^3$ a new vector representing the oriented area of the parallelogram with two of its edges emanating from the origin along $\overrightarrow{v}$ and $\overrightarrow{w}$, and oriented in the direction of the first vector. This is called the **cross product**[16] of $\overrightarrow{v}$ and $\overrightarrow{w}$, and is denoted

$$\overrightarrow{v} \times \overrightarrow{w}.$$

[16] Also *vector product*, or *outer product*.

For example, the sides emanating from A in $\triangle ABC$ in Figure 1.41 are represented by

$$\begin{aligned}\vec{v} &= \overrightarrow{AB} = 2\vec{\imath} + \vec{\jmath} + \vec{k}\\ \vec{w} &= \overrightarrow{AC} = \vec{\imath} + 2\vec{\jmath} - \vec{k};\end{aligned}$$

these vectors, along with the direction of $\vec{v} \times \vec{w}$, are shown in Figure 1.42.

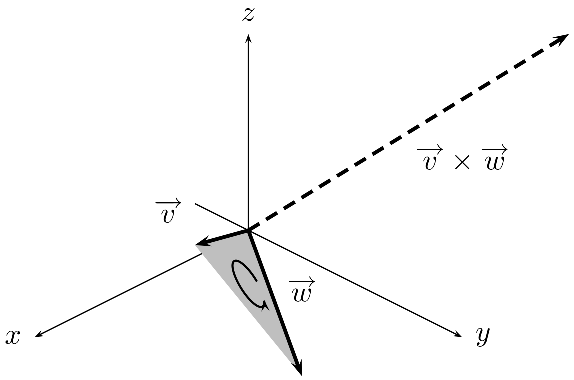


Figure 1.42: Direction of Cross Product

We stress that this product differs from the dot product in two essential ways: first, $\vec{v} \cdot \vec{w}$ is a *scalar*, but $\vec{v} \times \vec{w}$ is a *vector*; and second, the dot product is commutative ($\vec{w} \cdot \vec{v} = \vec{v} \cdot \vec{w}$), but the cross product is **anticommutative** ($\vec{w} \times \vec{v} = -\vec{v} \times \vec{w}$).

How do we calculate the components of the cross product $\vec{v} \times \vec{w}$ from the components of $\vec{v}$ and $\vec{w}$? To this end, we detour slightly and consider the projection of areas.

The (orthogonal) **projection** of points in $\mathbb{R}^3$ to a plane $\mathcal{P}'$ takes a point $P \in \mathbb{R}^3$ to the intersection with $\mathcal{P}'$ of the line through P perpendicular to $\mathcal{P}'$ (Figure 1.43). We denote this by

$$P' = \operatorname{proj}_{\mathcal{P}'} P.$$

Similarly, a vector $\vec{v}$ is projected onto the direction of the line where the plane containing $\vec{v}$ and the normal to $\mathcal{P}'$ meets $\mathcal{P}'$ (Figure 1.44).

Suppose $\triangle ABC$ is an oriented triangle in $\mathbb{R}^3$; its projection to $\mathcal{P}'$ is the oriented triangle $\triangle A'B'C'$, with vertices $A' = \operatorname{proj}_{\mathcal{P}'} A$, $B' = \operatorname{proj}_{\mathcal{P}'} B$, and $C' = \operatorname{proj}_{\mathcal{P}'} C$. What is the relation between the oriented areas of these two triangles?

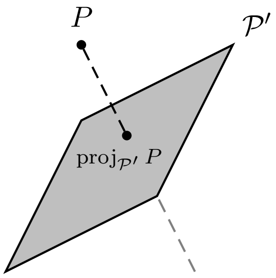


Figure 1.43: Projection of a Point P on the Plane $\mathcal{P}'$

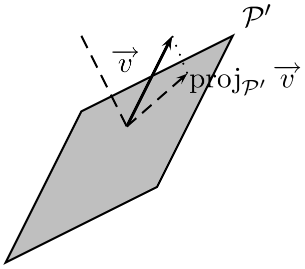


Figure 1.44: Projection of a Vector $\overrightarrow{v}$ on the Plane $\mathcal{P}'$

Let $\mathcal{P}$ be the plane containing $\triangle ABC$ and let $\overrightarrow{n}$ be the unit vector (normal to $\mathcal{P}$) such that

$$\vec{\mathcal{A}}\,(\triangle ABC) = \mathcal{A}\overrightarrow{n}$$

where $\mathcal{A}$ is the area of $\triangle ABC$. If the two planes $\mathcal{P}$ and $\mathcal{P}'$ are parallel, then $\triangle A'B'C'$ is a parallel translate of $\triangle ABC$, and the two oriented areas are the same. Suppose the two planes are not parallel, but meet at (acute) angle θ along a line ℓ (Figure 1.45).

Then a vector $\overrightarrow{v}_{\|}$ parallel to ℓ (and hence to both $\mathcal{P}$ and $\mathcal{P}'$) is unchanged by projection, while a vector $\overrightarrow{v}_{\perp}$ parallel to $\mathcal{P}$ but *perpendicular* to ℓ projects to a vector $\text{proj}_{\mathcal{P}'} \overrightarrow{v}_{\perp}$ parallel to $\mathcal{P}'$, also perpendicular to ℓ, with length

$$|\text{proj}_{\mathcal{P}'} \overrightarrow{v}_{\perp}| = |\overrightarrow{v}_{\perp}| \cos\theta.$$

The angle between these vectors is the same as between $\overrightarrow{n}$ and a unit vector $\overrightarrow{n}'$ normal to $\mathcal{P}'$; the oriented triangle $\triangle A'B'C'$ is traversed counterclockwise when viewed from the side of $\mathcal{P}'$ determined by $\overrightarrow{n}'$.

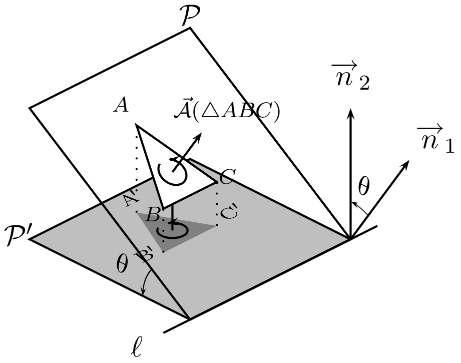


Figure 1.45: Projection of a Triangle

Furthermore, if $\triangle ABC$ has one side parallel to ℓ and another perpendicular to ℓ, then the same is true of $\triangle A'B'C'$; the sides parallel to ℓ have the same length, while projection scales the side perpendicular to ℓ—and hence the area—by a factor of $\cos\theta$. Since every triangle in $\mathcal{P}$ can be subdivided (using lines through the vertices parallel and perpendicular to ℓ) into triangles of this type, the area of *any* triangle $\triangle ABC$ is multiplied by $\cos\theta$ under projection. This means

$$\vec{\mathcal{A}}\left(\triangle A'B'C'\right) = (\mathcal{A}\cos\theta)\overrightarrow{n}'$$

which is easily seen to be the projection of $\vec{\mathcal{A}}(\triangle ABC)$ onto the direction normal to the plane $\mathcal{P}'$. We have shown

Proposition 1.6.4. *For any oriented triangle $\triangle ABC$ and any plane $\mathcal{P}'$ in $\mathbb{R}^3$, the oriented area of the projection $\triangle A'B'C'$ of $\triangle ABC$ onto $\mathcal{P}'$ (as a triangle) is the projection of the oriented area $\vec{\mathcal{A}}(\triangle ABC)$ (as a vector) onto the direction normal to $\mathcal{P}'$.*

Note in particular that when $\triangle ABC$ is *parallel* to $\mathcal{P}'$, its oriented area is *unchanged*, while if $\triangle ABC$ is *perpendicular* to $\mathcal{P}'$, its projection is a degenerate triangle with *zero* area.

As an example, let us consider the projections onto the coordinate planes of the triangle with vertices $A(2,-3,4)$, $B(4,-2,5)$, and $C(3,-1,3)$, which is the triangle we sketched in Figure 1.41. We reproduce this in Figure 1.46, showing the projections of $\triangle ABC$ on each of the coordinate axes

The projection onto the xy-plane has vertices $A(2,-3)$, $B(4,-2)$, and $C(3,-1)$, which is the triangle we sketched in Figure 1.36. This has signed

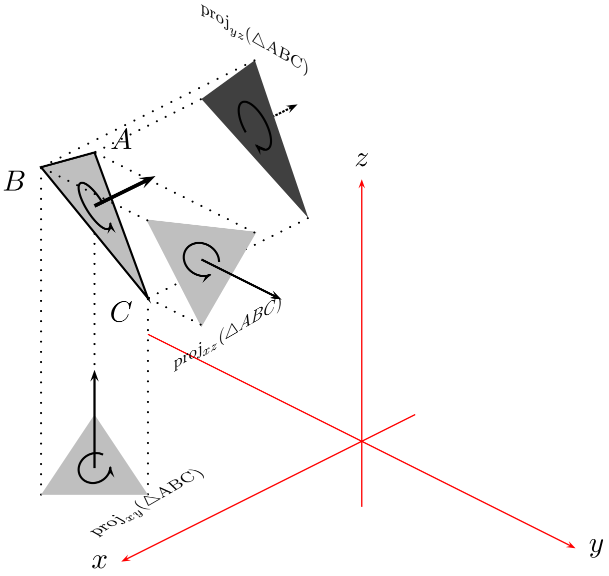

$\text{proj}_{yz}(\triangle ABC)$
A
B
z
C
$\text{proj}_{xz}(\triangle ABC)$
$\text{proj}_{xy}(\triangle ABC)$
x
y

Figure 1.46: Projections of $\triangle ABC$

area $3/2$, so its oriented area is

$$\frac{3}{2}\overrightarrow{k}$$

—that is, the area is $3/2$ and the orientation is counterclockwise when seen from *above* the xy-plane.

The projection onto the yz-plane has vertices $A(-3,4)$, $B(-2,5)$, and $C(-1,3)$ (Figure 1.37) and we saw that its signed area is $-1/2$. If we look at the yz-plane from the direction of the positive x-axis, then we see a "clockwise" triangle, so the oriented area is

$$-\frac{1}{2}\overrightarrow{\imath}$$

—it points in the direction of the *negative* x-axis.

Finally, the projection onto the xz-plane has vertices $A(2,4)$, $B(4,5)$, and $C(3,3)$. We sketched this in Figure 1.38, and calculated a negative signed area of $-3/2$. Note, however, that if we look at our triangle from the direction of the positive y-axis, we see a *counterclockwise* triangle. Why the discrepancy? The reason for this becomes clear if we take into account not just the triangle, but also the *axes* we see. In Figure 1.38, we sketched the triangle with positive abcissas pointing "east" and ordinates pointing "north", but in Figure 1.46 the positive x-axis points "west" from our point of view. In other words, the orientation of the x-axis and z-axis (in that order) looks *counterclockwise* only if we look from the direction of the *negative y-axis*. From *this* point of view–that is, the direction of $-\overrightarrow{\jmath}$ (which is the one we used to calculate the signed area)–the triangle looks negatively oriented, so the oriented area should be

$$\left(-\frac{3}{2}\right)(-\overrightarrow{\jmath}) = \frac{3}{2}\overrightarrow{\jmath}.$$

This agrees with the geometric observation based on Figure 1.42.

We have seen that the oriented area $\vec{\mathcal{A}}(\triangle ABC)$ has projections

$$\begin{aligned}
\mathrm{proj}_{\overrightarrow{k}}\vec{\mathcal{A}}(\triangle ABC) &= \frac{3}{2}\overrightarrow{k}\\
\mathrm{proj}_{\overrightarrow{\imath}}\vec{\mathcal{A}}(\triangle ABC) &= -\frac{1}{2}\overrightarrow{\imath}\\
\mathrm{proj}_{\overrightarrow{\jmath}}\vec{\mathcal{A}}(\triangle ABC) &= \frac{3}{2}\overrightarrow{\jmath}.
\end{aligned}$$

But these projections are simply the components of the vector, so we conclude that the oriented area $\vec{\mathcal{A}}(\triangle ABC)$ is

$$\vec{\mathcal{A}}(\triangle ABC) = -\frac{1}{2}\overrightarrow{\imath} + \frac{3}{2}\overrightarrow{\jmath} + \frac{3}{2}\overrightarrow{k}.$$

Looked at differently, the two sides of $\triangle ABC$ emanating from vertex A are represented by the vectors

$$\begin{aligned}
\overrightarrow{v} &= \overrightarrow{AB} = 2\overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}\\
\overrightarrow{w} &= \overrightarrow{AC} = \overrightarrow{\imath} + 2\overrightarrow{\jmath} - \overrightarrow{k}
\end{aligned}$$

and by definition

$$\begin{aligned}
\overrightarrow{v} \times \overrightarrow{w} &= 2\vec{\mathcal{A}}(\triangle ABC)\\
&= \overrightarrow{\imath}\begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} - \overrightarrow{\jmath}\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} + \overrightarrow{k}\begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix}.
\end{aligned}$$

The reasoning used in this example leads to the following general formula for the cross product of two vectors in $\mathbb{R}^3$ from their components.

Theorem 1.6.5. *The cross product of two vectors*

$$\begin{aligned} \overrightarrow{v} &= x_1\overrightarrow{\imath} + y_1\overrightarrow{\jmath} + z_1\overrightarrow{k} \\ \overrightarrow{w} &= x_2\overrightarrow{\imath} + y_2\overrightarrow{\jmath} + z_2\overrightarrow{k} \end{aligned}$$

is given by

$$\overrightarrow{v} \times \overrightarrow{w} = \overrightarrow{\imath}\begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} - \overrightarrow{\jmath}\begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix} + \overrightarrow{k}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.$$

Proof. Let $P(x_1, y_1, z_1)$ and $Q(x_2, y_2, z_2)$ be the points in $\mathbb{R}^3$ with position vectors $\overrightarrow{v}$ and $\overrightarrow{w}$, respectively. Then

$$\overrightarrow{v} \times \overrightarrow{w} = 2\vec{\mathcal{A}}(\triangle \mathcal{O}PQ) \qquad\qquad = a_1\overrightarrow{\imath} + a_2\overrightarrow{\jmath} + a_3\overrightarrow{k}.$$

The three components of $\vec{\mathcal{A}}(\triangle \mathcal{O}PQ)$ are its projections onto the three coordinate directions, and hence by Proposition 1.6.4 each represents the oriented area of the projection $\text{proj}_{\mathcal{P}} \triangle \mathcal{O}PQ$ of $\triangle \mathcal{O}PQ$ onto the plane $\mathcal{P}$ *perpendicular* to the corresponding vector.

Projection onto the plane perpendicular to a coordinate direction consists of taking the other two coordinates. For example, the direction of $\overrightarrow{k}$ is normal to the xy-plane, and the projection onto the xy-plane takes $P(x_1, y_1, z_1)$ onto $P(x_1, y_1)$.

Thus, the determinant

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$$

represents twice the signed area of $\triangle \mathcal{O}P_3Q_3$, the projection of $\triangle \mathcal{O}PQ$ onto the xy-plane, when viewed from above—that is, from the direction of $\overrightarrow{k}$—so the oriented area is given by

$$\begin{aligned} a_3\overrightarrow{k} &= 2\vec{\mathcal{A}}(\triangle \mathcal{O}P_3Q_3) \\ &= \overrightarrow{k}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}. \end{aligned}$$

Similarly,

$$\begin{aligned} a_1\overrightarrow{\imath} &= 2\vec{\mathcal{A}}(\triangle \mathcal{O}P_1Q_1) \\ &= \overrightarrow{\imath}\begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix}. \end{aligned}$$

Finally, noting that the direction from which the *positive* z-axis is *counterclockwise* from the positive x-axis is $-\overrightarrow{\jmath}$, we have

$$\begin{aligned} a_2\overrightarrow{\jmath} &= 2\vec{\mathcal{A}}(\triangle \mathcal{O}P_2Q_2) \\ &= -\overrightarrow{\jmath}\begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix}. \end{aligned}$$

Adding these yields the desired formula. □

In each projection, we used the 2×2 determinant obtained by omitting the coordinate along whose axis we were projecting. The resulting formula can be summarized in terms of the array of coordinates of $\overrightarrow{v}$ and $\overrightarrow{w}$

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{pmatrix}$$

by saying: the coefficient of the standard basis vector in a given coordinate direction is the 2×2 determinant obtained by eliminating the corresponding column from the above array, and multiplying by -1 for the second column.

We can make this even more "visual" by defining 3×3 determinants.

A **3 × 3 matrix** [17] is an array consisting of three rows of three entries each, vertically aligned in three columns. It is sometimes convenient to label the entries of an abstract 3×3 matrix using a single letter with a double index: the entry in the i^{th} *row* and j^{th} *column* of a matrix A is denoted [18] a_{ij}, giving the general form for a 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

We define the **determinant** of a 3×3 matrix as follows: for each entry a_{1j} in the first row, its **minor** is the 2×2 matrix A_{1j} obtained by deleting the

[17] Pronounced "3 by 3 matrix"

[18] Note that the *row* index precedes the *column* index: a_{ji} is in the j^{th} *row* and i^{th} *column*, a very different place in the matrix.

row and column containing our entry. Thus

$$A_{11} = \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & a_{22} & a_{23} \\ \cdot & a_{32} & a_{33} \end{pmatrix}$$
$$A_{12} = \begin{pmatrix} \cdot & \cdot & \cdot \\ a_{21} & \cdot & a_{23} \\ a_{31} & \cdot & a_{33} \end{pmatrix}$$
$$A_{13} = \begin{pmatrix} \cdot & \cdot & \cdot \\ a_{21} & a_{22} & \cdot \\ a_{31} & a_{32} & \cdot \end{pmatrix}.$$

Now, the 3×3 determinant of A can be expressed as the *alternating* sum of the *entries* of the first row times the *determinants of their minors*:

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} \\ &= \sum_{j=1}^{3} (-1)^{1+j} a_{1j} \det A_{1j}. \end{aligned}$$

For future reference, the numbers multiplying the first-row entries in the formula above are called the **cofactors** of these entries: the cofactor of a_{1j} is

$$\text{cofactor}(1j) := (-1)^{1+j} \det A_{1j}.$$

We shall see later that this formula usefully generalizes in several ways. For now, though, we see that, once we have mastered this formula, we can express the calculation of the cross product as

$$\overrightarrow{v} \times \overrightarrow{w} = \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

where

$$\begin{aligned} \overrightarrow{v} &= v_1 \overrightarrow{\imath} + v_2 \overrightarrow{\jmath} + v_3 \overrightarrow{k} \\ \overrightarrow{w} &= w_1 \overrightarrow{\imath} + w_2 \overrightarrow{\jmath} + w_3 \overrightarrow{k}. \end{aligned}$$

Exercises for § 1.6

Practice problems:

1. Calculate each determinant below:

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

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

   (c)
   $$\begin{vmatrix} -1 & 2 \\ 4 & -8 \end{vmatrix}$$

2. Sketch the triangle $\triangle ABC$ and indicate its orientation; find $\sigma(A, B, C)\mathcal{A}\,(\triangle ABC)$:

   (a) $A(0, 0),\ B(2, 1),\ C(1, 2)$
   (b) $A(1, 2),\ B(2, 0),\ C(3, 3)$
   (c) $A(2, 1),\ B(1, 3),\ C(3, 2)$

3. Calculate $\overrightarrow{v} \times \overrightarrow{w}$:

   (a) $\overrightarrow{v} = (1, 2, 3),\ \overrightarrow{w} = (3, 1, 2)$
   (b) $\overrightarrow{v} = (3, 1, 2),\ \overrightarrow{w} = (6, 5, 4)$
   (c) $\overrightarrow{v} = \overrightarrow{\imath},\ \overrightarrow{w} = \overrightarrow{\jmath}$
   (d) $\overrightarrow{v} = \overrightarrow{\imath},\ \overrightarrow{w} = \overrightarrow{k}$
   (e) $\overrightarrow{v} = 4\overrightarrow{\imath} - 3\overrightarrow{\jmath} + 7\overrightarrow{k},\ \overrightarrow{w} = -2\overrightarrow{\imath} - 5\overrightarrow{\jmath} + 4\overrightarrow{k}$

4. Find the oriented area vector $\vec{\mathcal{A}}\,(\triangle ABC)$ and calculate the area of the triangle:

   (a) $A = (0, 0, 0),\ B = (1, 2, 3),\ C = (3, 2, 1)$
   (b) $A = (1, 3, 2),\ B = (2, 3, 1),\ C = (3, 3, 2)$
   (c) $A = (2, -1, -4),\ B = (-1, 1, 0),\ C = (3, -3, -2)$

Theory problems:

5. Suppose that in $\triangle ABC$ the vector from B to A is $\overrightarrow{v}$ and that from B to C is $\overrightarrow{w}$. Use the vector formula for the distance from A to BC on p. 51 to prove that the area of the triangle is given by
   $$\mathcal{A}\,(\triangle ABC) = \frac{1}{2}\sqrt{(\overrightarrow{w} \cdot \overrightarrow{w})(\overrightarrow{v} \cdot \overrightarrow{v}) - (\overrightarrow{v} \cdot \overrightarrow{w})^2}.$$

6. Prove Proposition 1.6.2.

7. Use Proposition 1.6.1 to prove Corollary 1.6.3. (*Hint:* If the rows are linearly dependent, what does this say about the parallelogram $\mathcal{O}PRQ$?)

8. Show that the cross product is:

   (a) **skew-symmetric:**

   $$\overrightarrow{v} \times \overrightarrow{w} = -\overrightarrow{w} \times \overrightarrow{v}$$

   (b) **additive in each slot:**

   $$(\overrightarrow{v}_1 + \overrightarrow{v}_2) \times \overrightarrow{w} = (\overrightarrow{v}_1 \times \overrightarrow{w}) + (\overrightarrow{v}_2 \times \overrightarrow{w})$$

   (use skew-symmetry to take care of the other slot: this is a kind of distributive law)

   (c) **homogeneous in each slot:**

   $$(a\overrightarrow{v}) \times \overrightarrow{w} = a(\overrightarrow{v} \times \overrightarrow{w}) = \overrightarrow{v} \times (a\overrightarrow{w})$$

   (d) Conclude that the cross product is **bilinear:**

   $$(a_1\overrightarrow{w}_1 + a_2\overrightarrow{w}_2) \times \overrightarrow{v} = a_1(\overrightarrow{w}_1 \times \overrightarrow{v}) + a_2(\overrightarrow{w}_2 \times \overrightarrow{v})$$

   and, analogously

   $$\overrightarrow{v} \times (a_1\overrightarrow{w}_1 + a_2\overrightarrow{w}_2) = a_1(\overrightarrow{v} \times \overrightarrow{w}_1) + a_2(\overrightarrow{v} \times \overrightarrow{w}_2).$$

9. (a) Suppose A, B, and C lie on the line ℓ in $\mathbb{R}^2$, and that ℓ does not go through the origin.

   Explain why, if B is between A and C,

   $$\mathcal{A}(\triangle\mathcal{O}AB) + \mathcal{A}(\triangle\mathcal{O}BC) - \mathcal{A}(\triangle\mathcal{O}AC) = 0.$$

   (b) Show that

   $$\sigma(\mathcal{O}, A, B)\,\mathcal{A}(\triangle\mathcal{O}AB) + \sigma(\mathcal{O}, B, C)\,\mathcal{A}(\triangle\mathcal{O}BC) + \sigma(\mathcal{O}, C, A)\,\mathcal{A}(\triangle\mathcal{O}CA) = 0$$

   regardless of the order of A, B and C along the line.

(c) Show that the above is not necessarily true if we leave off the signs.

10. Show that the oriented area of a triangle can also be calculated as the cross product of the vectors obtained by moving along two successive edges:

$$\vec{\mathcal{A}}\,(\triangle ABC) = \overrightarrow{AB} \times \overrightarrow{BC}$$

(*Hint:* You may use Exercise 8.)

Challenge Problems:

Given a point D in the plane, and a directed line segment $\overrightarrow{AB}$, we can define the **area swept out** by the line DP as P moves from A to B along $\overrightarrow{AB}$ to be the signed area of the oriented triangle $[D, A, B]$. We can then extend this definition to the area swept out by DP as P moves along any broken-line path (*i.e.*, a path consisting of finitely many directed line segments) to be the sum of the areas swept out over each of the segments making up the path.

11. (a) *Show* that the area swept out by DP as P travels along an oriented triangle equals the signed area of the triangle: that is, show that

$$\sigma(ABC)\,\mathcal{A}\,(\triangle ABC) =$$
$$\sigma(DAB)\,\mathcal{A}\,(\triangle DAB) + \sigma(DBC)\,\mathcal{A}\,(\triangle DBC) + \sigma(DCA)\,\mathcal{A}\,(\triangle DCA)\,.$$

(*Hint:* This can be done geometrically. Consider three cases: D lies outside, inside, or on $\triangle ABC$. See Figure 1.47.)

(b) *Show* that the area swept out by $\mathcal{O}P$ as P moves along the line segment from (x_0, y_0) to (x_1, y_1) is

$$\frac{1}{2}\begin{vmatrix} x_0 & y_0 \\ x_1 & y_1 \end{vmatrix}.$$

(c) *Show* that if $\overrightarrow{v}_i = (x_i, y_i)$, $i = 0, \ldots, 3$ with $\overrightarrow{v}_0 = \overrightarrow{v}_3$ then the signed area of $[\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3]$ can be calculated as

$$\sigma(\overrightarrow{v}_1\overrightarrow{v}_2\overrightarrow{v}_3)\,\mathcal{A}\,(\triangle\overrightarrow{v}_1\overrightarrow{v}_2\overrightarrow{v}_3) = \frac{1}{2}\sum_{i=1}^{3}\begin{vmatrix} x_{i-1} & y_{i-1} \\ x_i & y_i \end{vmatrix}.$$

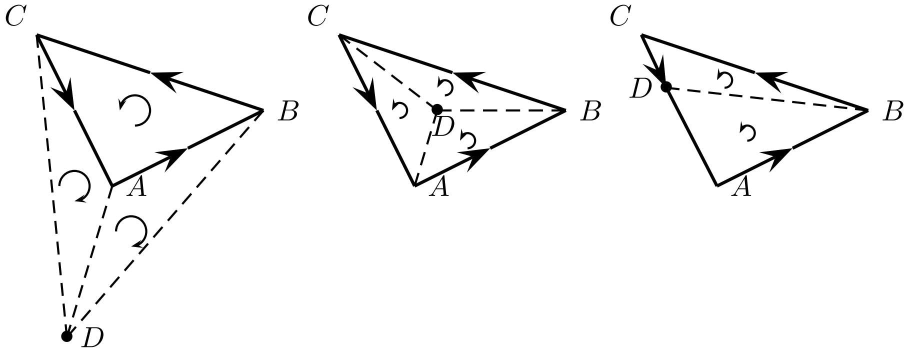


Figure 1.47: Area Swept Out by DP as P Traverses a Triangle

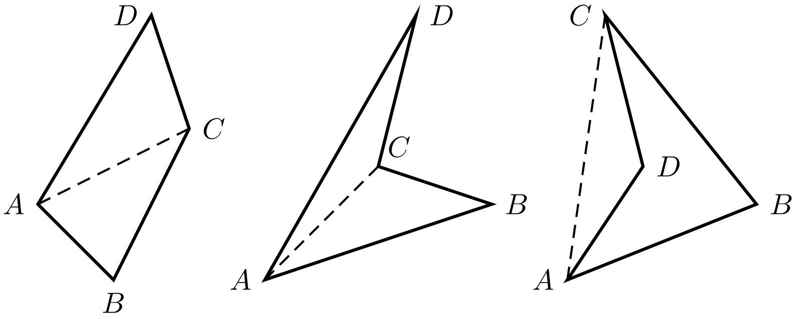


Figure 1.48: Signed Area of Quadrangles

12. (a) Consider the three quadrilaterals in Figure 1.48. In all three cases, the orientation of $\square[ABCD]$ and of $\triangle ABC$ is positive, but the orientation of $\triangle ACD$ is not necessarily positive. Show that in all three cases,

$$\mathcal{A}\left(\square ABCD\right) = \sigma(ABC)\,\mathcal{A}\left(\triangle ABC\right) + \sigma(ACD)\,\mathcal{A}\left(\triangle ACD\right).$$

(b) Use this to show that the signed area of a quadrilateral $\square[ABCD]$ is given by

$$\sigma(ABCD)\,\mathcal{A}\left(\square[ABCD]\right) = \frac{1}{2}\left\{(x_2-x_0)(y_1-y_3)+(x_1-x_3)(y_0-y_2)\right\}$$

where the coordinates of the vertices are

$$\begin{gathered} A(x_0, y_0) \\ B(x_1, y_1) \\ C(x_2, y_2) \\ D(x_3, y_3). \end{gathered}$$

Note that this is the same as

$$\frac{1}{2}\Delta\left(\overrightarrow{v}, \overrightarrow{w}\right)$$

where $\overrightarrow{v} = \overrightarrow{AC}$ and $\overrightarrow{w} = \overrightarrow{DA}$ are the diagonal vectors of the quadrilateral.

(c) What should be the (signed) area of the oriented quadrilateral $\square[ABCD]$ in Figure 1.49?

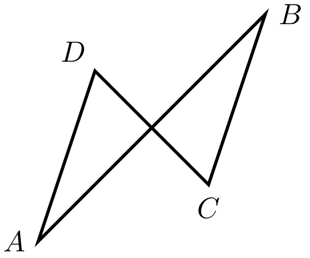


Figure 1.49: Signed Area of Quadrangles (2)

13. Show that the area swept out by a line DP as P travels along a closed polygonal path equals the signed area of the polygon: that is, suppose the vertices of a polygon in the plane, traversed in counterclockwise order, are

$$\vec{v}_i = (x_i, y_i), \quad i = 0, ..., n$$

with

$$\vec{v}_0 = \vec{v}_n.$$

Show that the (signed) area of the polygon is

$$\frac{1}{2}\sum_{i=1}^{n}\begin{vmatrix} x_{i-1} & y_{i-1} \\ x_i & y_i \end{vmatrix}.$$

14. Now extend the definition of the area swept out by a line to space, by replacing *signed area* (in the plane) with *oriented area* in space: that is, given three points $D, A, B \in \mathbb{R}^3$, the **area swept out** by the line DP as P moves from A to B along $\overrightarrow{AB}$ is defined to be the oriented area $\vec{\mathcal{A}}(\triangle DAB)$. *Show* that the oriented area of a triangle $\triangle ABC \subset \mathbb{R}^3$ in space equals the area swept out by the line DP as P traverses the triangle, for any point $D \in \mathbb{R}^3$. (*Hint:* Consider the projections on the coordinate planes, and use Exercise 11.)

History notes:

15. **Heron's First Formula:** The first area formula given by Heron in the *Metrica* is an application of the Law of Cosines, as given in Book II, Propositions 12 and 13 in the *Elements* . Given $\triangle ABC$, we denote the (lengths of the) side opposite each vertex using the corresponding lower case letter (see Figure 1.50).

    (a) **Obtuse Case:** Suppose the angle at C is obtuse. Extend BC to the foot of the perpendicular from A, at D. Prove Euclid's Proposition 11.12:

    $$c^2 = a^2 + b^2 + 2c \cdot CD.$$

    From this, prove Heron's formula in the obtuse case:

    $$\mathcal{A}(\triangle ABC) = \frac{a}{2}\sqrt{\frac{c^2 - (a^2 + b^2)}{2c}}$$

    (*Hint:* First find CD, then use the standard formula.)

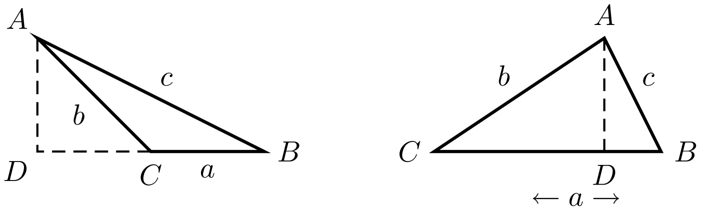


Figure 1.50: Propositions II.12-13: $c^2 = a^2 + b^2 \pm 2c \cdot CD$

(b) **Acute case:** Suppose the angle at C is acute. Let D be the foot of the perpendicular from A to BC. *Show* that

$$c^2 = a^2 + b^2 - 2c \cdot CD.$$

From this, prove Heron's formula in the acute case:

$$\mathcal{A}\,(\triangle ABC) = \frac{a}{2}\sqrt{\frac{(a^2+b^2)-c^2}{2c}}$$

16. **Heron's Second Formula:** Prove Heron's second (and more famous) formula for the area of a triangle:

$$\mathcal{A} = \sqrt{s(s-a)(s-b)(s-c)} \tag{1.27}$$

where a, b and c are the lengths of the sides of the triangle, and s is the *semiperimeter*

$$s = \frac{1}{2}(a+b+c).$$

Refer to Figure 1.51; we follow the exposition in [5, p. 186]:

The original triangle is $\triangle ABC$.

(a) Inscribe a circle inside $\triangle ABC$, touching the sides at D, E, and F. Denote the center of the circle by O; Note that

$$OE = OF = OD.$$

Show that

$$AE = AF$$
$$CE = CD$$
$$BD = BF.$$

(*Hint:* *e.g.*, the triangles $\triangle OAF$ and $\triangle OAE$ are similar—why?)

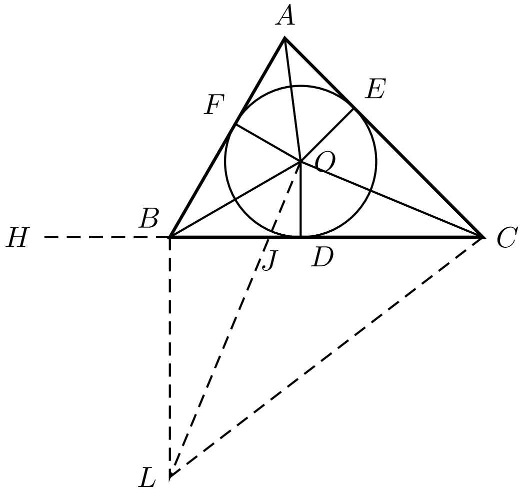


Figure 1.51: Heron's Formula

(b) *Show* that the area of $\triangle ABC$ equals $s \cdot OD$. (*Hint:* Consider $\triangle OBC$, $\triangle OAC$ and $\triangle OAB$.)

(c) Extend CB to H, so that $BH = AF$. *Show* that

$$s = CH.$$

(d) Let L be the intersection of the line through O perpendicular to OC with the line through B perpendicular to BC. *Show* that the points O, B, L and C all lie on a common circle. (*Hint:* Each of the triangles $\triangle CBL$ and $\triangle COL$ have right angles opposite their common edge CL, and the hypotenuse of a right triangle is a diameter of a circle containing the right angle.)

(e) It then follows by Proposition III.22 of the *Elements* (opposite angles of a quadrilateral inscribed in a circle sum to two right angles) that $\angle CLB + \angle COB$ equals two right angles.

Show that $\angle BOC + \angle AOF$ equals two right angles. (*Hint:* Each of the lines from O to a vertex of $\triangle ABC$ bisects the angle there.) It follows that

$$\angle CLB = \angle AOF.$$

(f) *Show* from this that $\triangle AOF$ and $\triangle CLB$ are similar.

(g) This leads to the proportions

$$\frac{BC}{BH} = \frac{BC}{AF} = \frac{BL}{OF} = \frac{BL}{OD} = \frac{BJ}{JD}.$$

Add one to both outside fractions to *show* that

$$\frac{CH}{BH} = \frac{BD}{JD}.$$

(h) Use this to show that

$$\frac{(CD)^2}{CH \cdot HB} = \frac{BD \cdot CD}{JD \cdot CD} = \frac{BD \cdot CD}{(OD)^2}.$$

Conclude that

$$(CH)^2(OD)^2 = CH \cdot HB \cdot BD \cdot DC.$$

(i) Explain how this proves Heron's formula.

1.7 Applications of Cross Products

In this section we explore some useful applications of cross products.

Equation of a Plane

The fact that $\overrightarrow{v} \times \overrightarrow{w}$ is perpendicular to both $\overrightarrow{v}$ and $\overrightarrow{w}$ can be used to find a "linear" equation for a plane, given three noncollinear points on it.

Remark 1.7.1. *If $\overrightarrow{v}$ and $\overrightarrow{w}$ are linearly independent vectors in $\mathbb{R}^3$, then any plane containing a line ℓ_v parallel to $\overrightarrow{v}$ and a line ℓ_w parallel to $\overrightarrow{w}$ has*

$$\overrightarrow{n} = \overrightarrow{v} \times \overrightarrow{w}$$

as a normal vector.

In particular, given a nondegenerate triangle $\triangle ABC$ in $\mathbb{R}^3$, an equation for the plane $\mathcal{P}$ containing this triangle is

$$\overrightarrow{n} \cdot (\overrightarrow{p} - \overrightarrow{p}_0) = 0 \tag{1.28}$$

where

$$\begin{aligned} \overrightarrow{p} &= x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k} \\ \overrightarrow{p}_0 &= \overrightarrow{\mathcal{O}A} \\ \overrightarrow{n} &= \overrightarrow{AB} \times \overrightarrow{AC}. \end{aligned}$$

For example, an equation for the plane $\mathcal{P}$ containing $\triangle PQR$ with vertices $P(1,-2,3)$, $Q(-2,4,-1)$ and $R(5,3,1)$ can be found using

$$\begin{aligned}
\overrightarrow{p}_0 &= \overrightarrow{\imath} - 2\overrightarrow{\jmath} + 3\overrightarrow{k} \\
\overrightarrow{PQ} &= -3\overrightarrow{\imath} + 6\overrightarrow{\jmath} - 4\overrightarrow{k} \\
\overrightarrow{PR} &= 4\overrightarrow{\imath} + 5\overrightarrow{\jmath} - 2\overrightarrow{k} \\
\overrightarrow{n} &= \overrightarrow{PQ} \times \overrightarrow{PR} \\
&= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ -3 & 6 & -4 \\ 4 & 5 & -2 \end{vmatrix} \\
&= \overrightarrow{\imath} \begin{vmatrix} 6 & -4 \\ 5 & -2 \end{vmatrix} - \overrightarrow{\jmath} \begin{vmatrix} -3 & -4 \\ 1 & -2 \end{vmatrix} + \overrightarrow{k} \begin{vmatrix} -3 & 6 \\ 4 & 5 \end{vmatrix} \\
&= \overrightarrow{\imath}(-12+20) - \overrightarrow{\jmath}(6+4) + \overrightarrow{k}(-15-10) \\
&= 8\overrightarrow{\imath} - 10\overrightarrow{\jmath} - 25\overrightarrow{k}
\end{aligned}$$

so the equation for $\mathcal{P}$ is

$$8(x-1) - 10(y+2) - 25(z-3) = 0$$

or

$$8x - 10y - 25z = -47.$$

As another example, consider the plane $\mathcal{P}'$ parametrized by

$$\begin{array}{rclrrr}
x &=& 3 & -2s & +t \\
y &=& -1 & +2s & -2t \\
z &=& & 3s & -t.
\end{array}$$

We can read off that

$$\overrightarrow{p}_0 = 3\overrightarrow{\imath} - \overrightarrow{\jmath}$$

is the position vector of $\overrightarrow{p}(0,0)$ (corresponding to $s=0$, $t=0$), and two vectors parallel to the plane are

$$\begin{aligned}
\overrightarrow{v}_s &= -2\overrightarrow{\imath} + \overrightarrow{\jmath} + 3\overrightarrow{k} \\
\overrightarrow{v}_t &= \overrightarrow{\imath} - 2\overrightarrow{\jmath} - \overrightarrow{k}.
\end{aligned}$$

Thus, a normal vector is

$$\begin{aligned}
\overrightarrow{n} &= \overrightarrow{v}_s \times \overrightarrow{v}_t \\
&= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ -2 & 1 & 3 \\ 1 & -2 & -1 \end{vmatrix} \\
&= \overrightarrow{\imath} \begin{vmatrix} 1 & 3 \\ -2 & -1 \end{vmatrix} - \overrightarrow{\jmath} \begin{vmatrix} -2 & 3 \\ 1 & -1 \end{vmatrix} + \overrightarrow{k} \begin{vmatrix} -2 & 1 \\ 1 & -2 \end{vmatrix} \\
&= 5\overrightarrow{\imath} + \overrightarrow{\jmath} + 3\overrightarrow{k}
\end{aligned}$$

and an equation for $\mathcal{P}'$ is

$$5(x-3) + 1(y+1) + 3(z-0) = 0$$

or

$$5x + y + 3z = 14.$$

Intersection of Planes

The line of intersection of two planes can be specified as the set of simultaneous solutions of two linear equations, one for each plane. How do we find a parametrization for this line?

Note that a linear equation for a plane

$$Ax + By + Cz = D$$

immediately gives us a normal vector

$$\overrightarrow{n} = A\overrightarrow{\imath} + B\overrightarrow{\jmath} + C\overrightarrow{k}.$$

If we are given two such equations

$$\begin{aligned}
A_1\overrightarrow{\imath} + B_1\overrightarrow{\jmath} + C_1\overrightarrow{k} &= D_1 \\
A_2\overrightarrow{\imath} + B_2\overrightarrow{\jmath} + C_2\overrightarrow{k} &= D_2
\end{aligned}$$

then the line of intersection ℓ (the locus of this *pair* of equations) is *perpendicular* to *both* normal vectors

$$\overrightarrow{n}_i = A_i\overrightarrow{\imath} + B_i\overrightarrow{\jmath} + C_i\overrightarrow{k} \quad i = 1, 2, 3$$

and hence *parallel* to their cross-product

$$\overrightarrow{v} = \overrightarrow{n}_1 \times \overrightarrow{n}_2.$$

Thus, given any one point $P_0(x_0, y_0, z_0)$ on ℓ (*i.e.*, one solution of the pair of equations) the line ℓ can be parametrized using P_0 as a basepoint and $\overrightarrow{v} = \overrightarrow{n}_1 \times \overrightarrow{n}_2$ as a direction vector.

For example, consider the two planes

$$\begin{aligned} 3x - 2y + z &= 1 \\ 2x + y - z &= 0. \end{aligned}$$

The first has normal vector

$$\overrightarrow{n}_1 = 3\overrightarrow{\imath} - 2\overrightarrow{\jmath} + \overrightarrow{k}$$

while the second has

$$\overrightarrow{n}_2 = 2\overrightarrow{\imath} + \overrightarrow{\jmath} - \overrightarrow{k}.$$

Thus, a direction vector for the intersection line is

$$\begin{aligned} \overrightarrow{v} &= \overrightarrow{n}_1 \times \overrightarrow{n}_2 \\ &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ 3 & -2 & 1 \\ 2 & 1 & -1 \end{vmatrix} \\ &= \overrightarrow{\imath} \begin{vmatrix} -2 & 1 \\ 1 & -1 \end{vmatrix} - \overrightarrow{\jmath} \begin{vmatrix} 3 & 1 \\ 2 & -1 \end{vmatrix} + \overrightarrow{k} \begin{vmatrix} 3 & -2 \\ 2 & 1 \end{vmatrix} \\ &= \overrightarrow{\imath} + 5\overrightarrow{\jmath} + 7\overrightarrow{k}. \end{aligned}$$

One point of intersection can be found by adding the equations to eliminate z

$$5x - y = 1$$

and, for example, picking

$$x = 1$$

which forces

$$y = 4.$$

Substituting back into either equation, we get

$$z = 6$$

so we can use $(1, 4, 6)$ as a basepoint; a parametrization of ℓ is

$$\overrightarrow{p}(t) = (\overrightarrow{\imath} + 4\overrightarrow{\jmath} + 6\overrightarrow{k}) + t(\overrightarrow{\imath} + 5\overrightarrow{\jmath} + 7\overrightarrow{k})$$

or

$$\begin{aligned} x &= 1 + t \\ y &= 4 + 5t \\ z &= 6 + 7t. \end{aligned}$$

If we try this when the two planes are parallel, we have linearly dependent normals, and their cross product is zero (Exercise 6 in § 1.6). In this case, the two left sides of the equations describing the planes are proportional: if the right sides have the same proportion, then we really have only one equation (the second is the first in disguise) and the two planes are the same, while if the right sides have a *different* proportion, the two equations are mutually contradictory—the planes are parallel, and have no intersection.

For example, the two equations

$$\begin{aligned} x - 2y + 3z &= 1 \\ -2x + 4y - 6z &= -2 \end{aligned}$$

are equivalent (the second is the first multiplied by -2) and describe a (single) plane, while

$$\begin{aligned} x - 2y + 3z &= 1 \\ -2x + 4y - 6z &= 0 \end{aligned}$$

are contradictory, and represent two parallel, nonintersecting planes.

Oriented Volumes

In common usage, a *cylinder* is the surface formed from two horizontal discs in space, one directly above the other, and of the same radius, by joining their boundaries with vertical line segments. Mathematicians generalize this, replacing the discs with horizontal copies of any plane region, and allowing the two copies to not be directly above one another (so the line segments joining their boundaries, while parallel to each other, need not be perpendicular to the two regions). Another way to say this is to define a (solid) **cylinder** on a given **base** (which is some region in a plane) to be formed by parallel line segments of equal length emanating from all points of the base (Figure 1.52). We will refer to a vector $\overrightarrow{v}$ representing these segments as a **generator** for the cylinder.

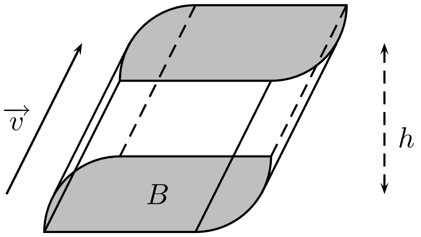


Figure 1.52: Cylinder with base B, generator $\overrightarrow{v}$, height h

Using Cavalieri's principle (*Calculus Deconstructed*, p. 365) it is fairly easy to see that the volume of a cylinder is the area of its base times its height (the perpendicular distance between the two planes containing the endpoints of the generating segments). Up to sign, this is given by orienting the base and taking the dot product of the generator with the oriented area of the base

$$V = \pm\overrightarrow{v} \cdot \vec{\mathcal{A}}(B)\,.$$

We can think of this dot product as the "signed volume" of the oriented cylinder, where the orientation of the cylinder is given by the direction of the generator together with the orientation of the base. The signed volume is positive (*resp.* negative) if $\overrightarrow{v}$ points toward the side of the base from which its orientation appears counterclockwise (*resp.* clockwise)—in other words, the orientation of the cylinder is positive if these data obey the right-hand rule. We will denote the signed volume of a cylinder C by $\overrightarrow{\mathcal{V}}(C)$.

A cylinder whose base is a parallelogram is called a **parallelepiped**: this has three quartets of parallel **edges**, which in pairs bound three pairs of parallel parallelograms,[19] called the **faces**. If the base parallelogram has sides represented by the vectors $\overrightarrow{w}_1$ and $\overrightarrow{w}_2$ and the generator is $\overrightarrow{v}$

[19] This tongue-twister was unintentional! :-)

(Figure 1.53) we denote the parallelepiped by $\square[\overrightarrow{v}, \overrightarrow{w}_1, \overrightarrow{w}_2]$. The oriented

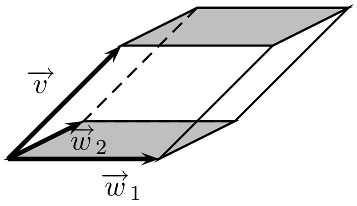


Figure 1.53: Parallelepiped

area of the base is

$$\vec{\mathcal{A}}(B) = \overrightarrow{w}_1 \times \overrightarrow{w}_2$$

so the signed volume is[20]

$$\overrightarrow{\mathcal{V}}(\square[\overrightarrow{v}, \overrightarrow{w}_1, \overrightarrow{w}_2]) = \overrightarrow{v} \cdot \vec{\mathcal{A}}(B) = \overrightarrow{v} \cdot (\overrightarrow{w}_1 \times \overrightarrow{w}_2)$$

(where $\overrightarrow{v}$ represents the third edge, or generator).

If the components of the "edge" vectors are

$$\begin{aligned}
\overrightarrow{v} &= a_{11}\overrightarrow{\imath} + a_{12}\overrightarrow{\jmath} + a_{13}\overrightarrow{k} \\
\overrightarrow{w}_1 &= a_{21}\overrightarrow{\imath} + a_{22}\overrightarrow{\jmath} + a_{23}\overrightarrow{k} \\
\overrightarrow{w}_2 &= a_{31}\overrightarrow{\imath} + a_{32}\overrightarrow{\jmath} + a_{33}\overrightarrow{k}
\end{aligned}$$

then

$$\begin{aligned}
\overrightarrow{w}_1 \times \overrightarrow{w}_2 &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\
&= \overrightarrow{\imath}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - \overrightarrow{\jmath}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + \overrightarrow{k}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}
\end{aligned}$$

so

$$\begin{aligned}
\overrightarrow{v} \cdot (\overrightarrow{w}_1 \times \overrightarrow{w}_2) &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.
\end{aligned}$$

[20] The last calculation in this equation is sometimes called the **triple scalar product** of $\overrightarrow{v}$, $\overrightarrow{w}_1$ and $\overrightarrow{w}_2$.

This gives us a geometric interpretation of a 3 × 3 (numerical) determinant:

Remark 1.7.2. *The* 3×3 *determinant*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

is the signed volume $\overrightarrow{\mathcal{V}}(\square[\overrightarrow{v}, \overrightarrow{w}_1, \overrightarrow{w}_2])$ *of the oriented parallelepiped* $\square[\overrightarrow{v}, \overrightarrow{w}_1, \overrightarrow{w}_2]$ *whose generator is the first row*

$$\overrightarrow{v} = a_{11}\overrightarrow{\imath} + a_{12}\overrightarrow{\jmath} + a_{13}\overrightarrow{k}$$

and whose base is the oriented parallelogram with edges represented by the other two rows

$$\begin{aligned} \overrightarrow{w}_1 &= a_{21}\overrightarrow{\imath} + a_{22}\overrightarrow{\jmath} + a_{23}\overrightarrow{k} \\ \overrightarrow{w}_2 &= a_{31}\overrightarrow{\imath} + a_{32}\overrightarrow{\jmath} + a_{33}\overrightarrow{k}. \end{aligned}$$

For example, the parallelepiped with base $\mathcal{O}PRQ$, with vertices the origin, $P(0,1,0)$, $Q(-1,1,0)$, and $R(-1,2,0)$ and generator $\overrightarrow{v} = \overrightarrow{\imath} - \overrightarrow{\jmath} + 2\overrightarrow{k}$ (Figure 1.54) has "top" face $OP'R'Q'$, with vertices $O(1,-1,2)$, $P'(1,0,2)$, $Q'(0,0,2)$ and $R'(0,1,2)$. Its signed volume is given by the 3×3 determinant

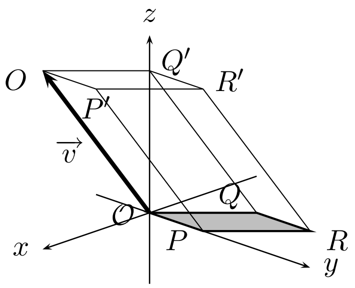


Figure 1.54: $\square\mathcal{O}PRQ$

whose rows are $\overrightarrow{v}$, $\overrightarrow{\mathcal{O}P}$ and $\overrightarrow{\mathcal{O}Q}$:

$$\begin{aligned}\overrightarrow{\mathcal{V}}(\square[\mathcal{O}PRQ]) &= \begin{vmatrix} 1 & -1 & 2 \\ 0 & 1 & 0 \\ -1 & 1 & 0 \end{vmatrix} \\ &= (1)\begin{vmatrix} 1 & 0 \\ 1 & 0 \end{vmatrix} - (-1)(1)\begin{vmatrix} 0 & 0 \\ -1 & 0 \end{vmatrix} + (2)(1)\begin{vmatrix} 0 & 1 \\ -1 & 1 \end{vmatrix} \\ &= (1)(0) - (-1)(0) + (2)(0+1) \\ &= 2.\end{aligned}$$

We see from Figure 1.54 that the vectors $\overrightarrow{\mathcal{O}P}$, $\overrightarrow{\mathcal{O}Q}$, $\overrightarrow{v}$ obey the right-hand rule, so have *positive* orientation.

Given any four points A, B, C, and D in $\mathbb{R}^3$, we can form a "pyramid" built on the triangle $\triangle ABC$, with a "peak" at D (Figure 1.55). The traditional name for such a solid is *tetrahedron*, but we will follow the terminology of combinatorial topology, calling this the **simplex**[21] with vertices A, B, C and D, and denote it $\triangle ABCD$; it is **oriented** when we pay attention to the order of the vertices. Just as for a triangle, the edges emanating from the vertex A are represented by the displacement vectors $\overrightarrow{AB}$, $\overrightarrow{AC}$, and $\overrightarrow{AD}$. The first two vectors determine the oriented "base" triangle $\triangle ABC$, and the simplex $\triangle ABCD$ is positively (*resp.* negatively) oriented if the orientation of $\triangle ABC$ is positive (*resp.* negative) when viewed from D, or equivalently if the dot product

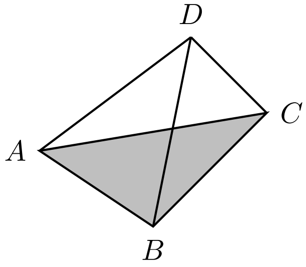


Figure 1.55: Simplex $\triangle ABCD$

$$\overrightarrow{AD} \cdot (\overrightarrow{AB} \times \overrightarrow{AC})$$

is positive (*resp.* negative).

[21] Actually, this is a **3-simplex**. In this terminology, a triangle is a **2-simplex** (it lies in a plane), and a line segment is a **1-simplex** (it lies on a line).

In Exercise **??**, we see that the parallelepiped $\square\mathcal{O}PRQ$ determined by the three vectors $\overrightarrow{AB}$, $\overrightarrow{AC}$, and $\overrightarrow{AD}$ can be subdivided into six simplices, all congruent to $\triangle ABCD$, and its orientation agrees with that of the simplex. Thus we have

Lemma 1.7.3. *The signed volume of the oriented simplex $\triangle ABCD$ is*

$$\begin{aligned}\overrightarrow{\mathcal{V}}(\triangle ABCD) &= \frac{1}{6}\overrightarrow{AD}\cdot(\overrightarrow{AB}\times\overrightarrow{AC}) \\ &= \frac{1}{6}\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}\end{aligned}$$

where

$$\begin{aligned}\overrightarrow{AB} &= a_{21}\overrightarrow{\imath} + a_{22}\overrightarrow{\jmath} + a_{23}\overrightarrow{k} \\ \overrightarrow{AC} &= a_{31}\overrightarrow{\imath} + a_{32}\overrightarrow{\jmath} + a_{33}\overrightarrow{k} \\ \overrightarrow{AD} &= a_{11}\overrightarrow{\imath} + a_{12}\overrightarrow{\jmath} + a_{13}\overrightarrow{k}.\end{aligned}$$

We can use this geometric interpretation (which is analogous to Proposition 1.6.1) to establish several algebraic properties of 3×3 determinants, analogous to those in the 2×2 case which we noted in § 1.6:

Remark 1.7.4. *The 3×3 determinant has the following properties:*

1. *It is* ***skew-symmetric*** *: Interchanging two rows of a 3×3 determinant reverses its sign (and leaves the absolute value unchanged).*
2. *It is* ***homogeneous*** *in each row: multiplying a single row by a scalar multiplies the determinant by that scalar.*
3. *It is* ***additive*** *in each row: Suppose two matrices (say A and B) agree in two rows (say, the two second rows are the same, and the two third rows are the same). Then the matrix with the same second and third rows, but with first row equal to the sum of the first rows of A and of B, has determinant $\det A + \det B$.*
4. *A 3×3 determinant* ***equals zero*** *precisely if its rows are linearly dependent.*

For the first item, note first that interchanging the two edges of the base reverses the sign of its oriented area and hence the sign of its oriented volume; if the first row is interchanged with one of the other two, you should check that this also reversed the orientation. Once we have the first item, we can assume in the second item that we are scaling the first row, and and in the second that A and B agree except in their first row(s). The additivity and homogeneity in this case follows from the fact that the oriented volume equals the oriented area of the base dotted with the first row. Finally, the last item follows from noting that zero determinant implies zero volume, which means the "height" measured off a plane containing the base is zero.

Rotations

So far, the physical quantities we have associated with vectors—forces, velocities—concern *displacements*. In effect, we have been talking about the motion of individual points, or the abstraction of such motion for larger bodies obtained by replacing each body with its center of mass. However, a complete description of the motion of solid bodies also involves *rotation*.

A rotation of 3-space about the z-axis is most easily described in cylindrical coordinates: a point P with cylindrical coordinates (r, θ, z), under a counterclockwise rotation (seen from above the xy-plane) by α radians does not change its r- or z- coordinates, but its θ- coordinate increases by α. Expressing this in rectangular coordinates, we see that the rotation about the z-axis by α radians counterclockwise (when seen from above) moves the point with rectangular coordinates (x, y, z), where

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

to the point

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

These new rectangular coordinates can be expressed in terms of the old ones, using the angle-summation formulas for sine and cosine, as

$$\begin{array}{rcll} x(\alpha) & = & x\cos\alpha & -y\sin\alpha \\ y(\alpha) & = & x\sin\alpha & +y\cos\alpha \\ z(\alpha) & = & z. & \end{array} \tag{1.29}$$

Under a steady rotation around the z-axis with angular velocity[22] $\dot{\alpha} = \omega$ *radians/sec*, the velocity $\overrightarrow{v}$ of our point is given by

$$\begin{aligned}
\dot{x} &= \left(\left.\frac{dx(\alpha)}{d\alpha}\right|_{\alpha=0}\right)\omega &&= (-x\sin 0 - y\cos 0)\omega &&= -y\omega \\
\dot{y} &== \left(\left.\frac{dy(\alpha)}{d\alpha}\right|_{\alpha=0}\right)\omega &&= (x\cos 0 - y\sin 0)\omega &&= x\omega \\
\dot{z} & &&= 0
\end{aligned}$$

which can also be expressed as

$$\begin{aligned}
\overrightarrow{v} &= -y\omega\overrightarrow{\imath} + x\omega\overrightarrow{\jmath} \\
&= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ 0 & 0 & \omega \\ x & y & z \end{vmatrix} \\
&= \omega\overrightarrow{k} \times \overrightarrow{p}
\end{aligned} \tag{1.30}$$

where $\overrightarrow{p} = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$ is the position vector of P.

When the rotation is about a different axis, the analogue of Equation (1.29) is rather complicated, but Equation (1.30) is relatively easy to carry over, on geometric grounds. Note first that the z coordinate does not affect the velocity in Equation (1.30): we could replace $\overrightarrow{p}$, which is the displacement $\overrightarrow{\mathcal{O}P}$ from the origin to our point, with the displacement $\overrightarrow{P_0P}$ from *any* point on the z-axis. Second, the vector $\omega\overrightarrow{k}$ can be characterized as a vector parallel to our axis of rotation whose length equals the angular velocity, where ω is positive if the rotation is counterclockwise when viewed from above. That is, we can regard the **angular velocity** as a vector $\overrightarrow{\omega}$ analogous to a oriented area: its *magnitude* is the angular speed, and its *direction* is normal to the planes invariant under the rotation (*i.e.*, planes perpendicular to the axis of rotation) in the direction from which the rotation is counterclockwise. These considerations easily yield

Remark 1.7.5. *The (spatial) velocity $\overrightarrow{v}$ of a point P under a steady rotation (about the axis ℓ) with angular velocity $\overrightarrow{\omega}$ is*

$$\overrightarrow{v} = \overrightarrow{\omega} \times \overrightarrow{P_0P} \tag{1.31}$$

where P_0 is an arbitrary point on ℓ, the axis of rotation.

Associated to the analysis of rotation of rigid bodies are the rotational analogues of momentum and force, called *moments*. Recall that the momentum of a (constant) mass m moving with velocity $\overrightarrow{v}$ is $m\overrightarrow{v}$; its **angular**

[22] We use a dot over a variable to indicate its time derivative.

momentum or **moment of momentum** about a point P_0 is defined to be $\overrightarrow{P_0P} \times m\overrightarrow{v}$. More generally, the **moment** about a point P_0 of any vector quantity $\overrightarrow{V}$ applied at a point P is defined to be

$$\overrightarrow{P_0P} \times \overrightarrow{V}.$$

For a rigid body, the "same" force applied at different positions on the body has different effects on its motion; in this context it is the *moment* of the force that is relevant. Newton's First Law of Motion [39, Law 1(p. 416)] is usually formulated as **conservation of momentum**: if the net force acting on a system of bodies is zero, then the (vector) sum of their momenta will not change with time: put differently, their center of mass will move with constant velocity. A second conservation law is **conservation of angular momentum**, which says that in addition the (vector) sum of the angular momenta about the center of mass will be constant. This net angular momentum specifies an axis (through the center of mass) and a rotation about that axis. For a rigid body, the motion can be decomposed into these two parts: the *displacement* motion of its center of mass, and its *rotation* about this axis through the (moving) center of mass.

Exercises for § 1.7

Practice problems:

1. Find an equation for the plane $\mathcal{P}$ described in each case:

   (a) $\mathcal{P}$ goes through $(1,2,3)$ and contains lines parallel to each of the vectors

   $$\overrightarrow{v} = (3,1,2)$$

   and

   $$\overrightarrow{w} = (1,0,2).$$

   (b) $\mathcal{P}$ contains the three points

   $$\begin{aligned}&P(4,5,6)\\&Q(3,2,7)\\&R(5,1,1).\end{aligned}$$

(c) $\mathcal{P}$ contains $P(3,1,4)$ and the line

$$\begin{array}{rcrr} x &=& 1 & +t \\ y &=& -2 & +2t \\ z &=& 3 & -t. \end{array}$$

(d) $\mathcal{P}$ is parametrized by

$$\begin{array}{rcrrr} x &=& 1 & -2s & +3t \\ y &=& 2 & -s & +t \\ z &=& -2 & +s & +t. \end{array}$$

2. Give a parametrization of each plane $\mathcal{P}$ described below:

(a) $\mathcal{P}$ contains the three points

$$\begin{array}{l} P(3,-1,2) \\ Q(2,1,-1) \\ R(8,3,1). \end{array}$$

(b) $\mathcal{P}$ contains the lines

$$\ell_1 : \left\{ \begin{array}{rlr} x & = -2 & +t \\ y & = 1 & -2t \\ z & = 4 & +t \end{array} \right.$$

$$\ell_2 : \left\{ \begin{array}{rlr} x & = -1 & +2t \\ y & = -1 & +t \\ z & = 5 & -3t \end{array} \right. .$$

(c) $\mathcal{P}$ meets the plane $3x + y + z = 8$ in the line

$$\begin{array}{rlr} x & = 1 & -t \\ y & = 2 & +t \\ z & = -3 & +2t \end{array}$$

and is perpendicular to it.

(d) $\mathcal{P}$ is the locus of

$$2x - 3y + 4z = 3.$$

3. (a) Find a line in the plane $3x + 7y + z = 29$ which is perpendicular to the line

$$\begin{array}{rlr} x & = 1 & -2t \\ y & = 3 & +t \\ z & = 5 & -t. \end{array}$$

(b) Find the line in the plane $x+y+z=0$ which meets the line ℓ given by

$$\begin{array}{lll} x & =-5 & +3t \\ y & =4 & -2t \\ z & =1 & -t \end{array}$$

at the point $(-2,2,0)$ and is perpendicular to ℓ.

4. Parametrize the line described in each case below:

(a) ℓ is the intersection of the planes

$$\begin{aligned} \mathcal{P}_1: &\quad 5x-2y+3z=0 \\ \mathcal{P}_2: &\quad 2x+2y+z=3. \end{aligned}$$

(b) ℓ is the intersection of the planes parametrized as follows:

$$\mathcal{P}_1: \left\{ \begin{array}{ccccc} x & = & 1 & +2s & +3t \\ y & = & 2 & -s & +t \\ z & = & 3 & +s & -2t \end{array} \right.$$

$$\mathcal{P}_2: \left\{ \begin{array}{ccccc} x & = & 1 & +s & -t \\ y & = & 2 & +2s & +3t \\ z & = & 3 & -3s & -t \end{array} \right.$$

5. Find the volume of each parallelepiped described below:

(a) The origin is a vertex, and the three vertices joined to the origin by an edge are

$$\begin{array}{l} P(1,-3,2) \\ Q(2,3,-1) \\ R(3,2,1). \end{array}$$

(b) The faces of the parallelepiped lie on the planes

$$\begin{array}{l} z=0 \\ z=1 \\ z=2y \\ z=2y-1 \\ z=x \\ z=x+1. \end{array}$$

6. Determine the orientation and volume of the simplex $\triangle ABCD$ whose vertices are

$$\begin{aligned}&A(1,-1,1)\\&B(2,0,1)\\&C(2,-2,1)\\&D(1,-1,0)\end{aligned}$$

7. The plane $x+y+z=3$ is continuously rotated about the line

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

(which is perpendicular to the plane and meets it at the point $P_0(1,1,1)$). If the point $P(2,2,-1)$ has velocity $\overrightarrow{v}=\overrightarrow{\imath}-\overrightarrow{\jmath}$, what is its angular momentum about the line?

Challenge problems:

8. Suppose $\overrightarrow{v}_0, \overrightarrow{v}_1, ..., \overrightarrow{v}_n = \overrightarrow{v}_0$ are the vertices of an oriented polygon in the plane, traversed in order around the circumference. Show that the sum of the moments of the vectors $v_i - v_{i-1}$, $i = 1, \ldots, n$, about the origin is twice the area of the polygon. (*Hint:* Compare problem 9 in the previous section.)

9. Consider the "prism" E bounded below by the xy-plane ($z=0$), above by the plane $z=1$, and on the sides by the three vertical planes $x=0$ (the yz-plane), $y=0$ (the xz-plane), and $x+y=1$ (see Figure 1.56).

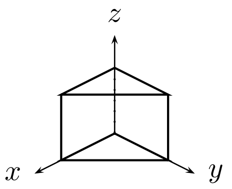


Figure 1.56: The Prism E

(a) Show that E consists of all points in $\mathbb{R}^3$ which simultaneously satisfy the inequalities

$$\begin{aligned} x &\geq 0 \\ y &\geq 0 \\ x+y &\leq 1 \\ 0 \leq z &\leq 1. \end{aligned}$$

(b) Show that the six vertices of E are

$$\begin{gathered} P_0(0,0,0) \\ P_1(1,0,0) \\ P_2(0,1,0) \\ Q_0(0,0,1) \\ Q_1(1,0,1) \\ Q_2(0,1,1). \end{gathered}$$

(Note that in this numbering, Q_i is directly above P_i.)

(c) Now consider the three oriented simplices

$$\begin{aligned} \triangle_1 &= \triangle P_0P_1P_2Q_0 \\ \triangle_2 &= \triangle P_1P_2Q_0Q_1 \\ \triangle_3 &= \triangle P_2Q_0Q_1Q_2. \end{aligned}$$

Show that

i. $\triangle_1$ consists of all points in E which also satisfy

$$x+y+z \leq 1.$$

ii. $\triangle_2$ consists of all points in E which also satisfy

$$x+y+z \geq 1 \text{ and } y+z \leq 1$$

iii. $\triangle_3$ consists of all points in E which also satisfy

$$y+z \geq 1.$$

(d) Show that each of the pairs of simplices $\triangle_1$ and $\triangle_2$ (*resp.* $\triangle_2$ and $\triangle_3$) meets along a common face, while $\triangle_1$ and $\triangle_3$ meet only at P_2.

(e) Show that each of these simplices has volume $\frac{1}{6}$.

2

Curves and Vector-Valued Functions of One Variable

2.1 Conic Sections

We begin this chapter by looking at the *conic sections*, which were regarded by the Greeks as the simplest curves after the straight line and circle. A major source of information about classical Greek mathematics is Pappus of Alexandria (*ca.* 300 AD), a formidable geometer of the late third century AD.[1] In his *Mathematical Collection*[2] he surveyed the work of his predecessors; many of these works have been lost. He classified mathematical problems according to the kinds of loci (curves) required for their solution:

- **planar** problems can be solved using circles and straight lines, or **planar loci**. These are often called **compass and straightedge constructions**;

- **solid** problems involve the intersection of a plane with a cone (**solid loci**, or **conic sections**);

- **linear** problems[3] are those involving other loci, such as **spirals** (see

[1] The work of Pappus is sometimes taken to mark the end of the classical Greek tradition in mathematics.

[2] Parts of this survive in a twelfth-century copy.

[3] *Caution*: this is *not* the modern meaning of "linear"!

p. 145 and Exercises 6-7), **quadratrices** (Exercise 8) and **conchoids** (Exercise 9).

One classic problem is that of **duplicating the cube**: given a cube, we are to construct a second cube whose volume is twice that of the first (or some other specified multiple). Hippocrates of Chios (460-380 BC) reduced this [25, p. 131], [31, p. 41] to the problem of **two mean proportionals**: given line segments a and b, to construct two other segments, y and x, whose lengths satisfy

$$|a| : |y| = |y| : |x| = |x| : |b|\,.$$

Early solutions of this problem [25, pp. 154-170] used "linear" loci, but two solutions by Menaechmus (*ca.* 350 BC), a follower of Plato, appear to be the first investigation and use of conic sections. The impossibility of duplicating the cube by compass and straightedge was first proved in the nineteenth century, using some deep algebraic and analytic results.

In this section, we will summarize two approaches to the conic sections. First, we discuss briefly the way these curves arise from intersecting a cone with a plane, the classic approach of Apollonius of Perga (*ca.*262-190 BC). Second, we discuss the focus-directrix property, which was not mentioned by Apollonius, but appeared some six hundred years later in the work of Pappus of Alexandria (*ca.* 300 AD)—who however seems to have been summarizing lost work by Euclid and his contemporaries, a generation before Apollonius. Our treatment here will focus on results; more detailed arguments are in Appendix A and Appendix B, respectively. At the end of this section, we will also note the relation of this geometric work to the analysis of quadratic curves in the plane.

Conics according to Apollonius

Pappus referred to two works on conic sections, by Euclid and Aristaeus the Elder (*ca.* 320 BC), which preceded him by six centuries. These works have been lost,[4] but in any case they were quickly eclipsed by the work of Apollonius of Perga (*ca.* 262-*ca.* 190 BC), *Conics* in eight books, recognized by his contemporaries as the definitive work on the subject.[5] Here, we give a

[4] Pappus refers to the "still surviving" *Solid Loci* of Aristaeus, but the *Conics* of Euclid were apparently already lost by the time of Pappus.

[5] The first four books of Apollonius' *Conics* have survived in a Greek edition with commentaries by Eutocius (*ca.* 520 AD), and the next three survived in an Arabic translation of Eutocius' edition by Thabit ibn Qurra (826-901); the eighth book is lost. A modern translation of Books I-IV is [41]. An extensive detailed and scholarly examination of the *Conics* has recently been published by Fried and Unguru [15].

simplified and anachronistic version of the basic ideas in Book I, bowlderizing [25, pp. 355-9].

Conical Surface: Start with a horizontal circle $\mathcal{C}$; on the vertical line through the center of $\mathcal{C}$ (the **axis**[6]) pick a point A distinct from the center of $\mathcal{C}$. The union of the lines through A intersecting $\mathcal{C}$ (the **generators**) is a surface $\mathcal{K}$ consisting of two cones joined at their common vertex (Figure 2.1). If we put the origin at A, the axis coincides with the z-axis, and $\mathcal{K}$ is the locus of the equation in rectangular coordinates

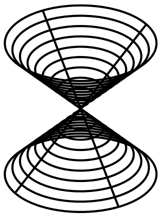


Figure 2.1: Conical Surface $\mathcal{K}$

$$z^2 = m^2(x^2 + y^2) \tag{2.1}$$

where

$$m = \cot \alpha$$

is the cotangent of the angle α between the axis and the generators.

Horizontal Sections: A horizontal plane $\mathcal{H}$ not containing A intersects $\mathcal{K}$ in a circle centered on the axis. The yz-plane intersects $\mathcal{H}$ in a line which meets this circle at two points, B and C; clearly the segment BC is a diameter of the circle. Given a point Q on this circle distinct from B and C (Figure 2.2), the line through Q parallel to the x-axis intersects the circle in a second point R, and the segment QR is bisected at V, the intersection of QR with the yz-plane. A basic property of circles, implicit in Prop. 13, Book VI of Euclid's *Elements* [27, vol. 2, p. 216] and equivalent to the equation of a circle in rectangular coordinates (Exercise 3), is

> *The product of the segments on a chord equals the product of the segments on the diameter perpendicular to it.*

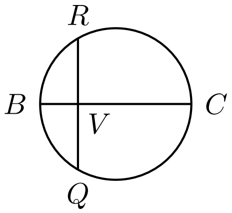


Figure 2.2: *Elements*, Book VI, Prop. 13

In Figure 2.2, this means

$$|QV|^2 = |QV| \cdot |VR| = |BV| \cdot |VC| . \tag{2.2}$$

Conic Sections: Now consider the intersection of a plane $\mathcal{P}$ with the conical surface $\mathcal{K}$.

If $\mathcal{P}$ contains the origin A, there are three possible forms for the intersection $\mathcal{P} \cap \mathcal{K}$:

- just the origin if $\mathcal{P}$ is horizontal or is tilted not too far off the horizontal;
- a single generator if $\mathcal{P}$ is tangent to the cone, and
- a pair of generators otherwise.

These are rather uninteresting.

To classify the more interesting intersections when $\mathcal{P}$ does *not* contain the origin A, recall that $\mathcal{P} \cap \mathcal{K}$ is a circle when $\mathcal{P}$ is horizontal; so suppose that $\mathcal{P}$ is any *non*-horizontal plane *not* containing A, and let γ be the intersection of $\mathcal{P}$ with $\mathcal{K}$. Rotating our picture about the axis if necessary, we can assume that $\mathcal{P}$ intersects any horizontal plane in a line parallel to the x-axis. The yz-plane intersects $\mathcal{P}$ in a line that meets γ in one or two points; we label the first P and the second (if it exists) P'; these are the **vertices** of γ (Figure 2.3). Given a point Q on γ distinct from the vertices, let $\mathcal{H}$ be the horizontal plane through Q, and define the points R, V, B and C as in Figure 2.2. The line segments QV and PV are, respectively, the **ordinate** and **abcissa**.

There are three possibilities:

[6] Apollonius allows the axis to be *oblique*—not necessarily normal to the plane of $\mathcal{C}$.

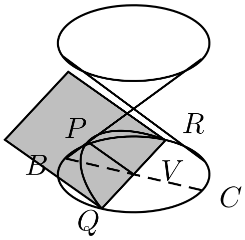


Figure 2.3: Conic Section

- **Parabolas:** If PV is parallel to AC, then P is the only vertex of γ. We wish to relate the square of the ordinate, $|QV|^2 = |QV| \cdot |VR|$, to the abcissa $|PV|$. By Equation (2.2), the square of the ordinate equals $|BV| \cdot |VC|$. Apollonius constructs a line segment PL perpendicular to the abcissa PV, called the **orthia** [7]: he then formulates the relation between the square of the ordinate and the abcissa [8] as equality of area between the rectangle LPV and a square with side $|QV|$ (recall Equation (2.2)).

$$|QV|^2 = |PL|\,|PV|\,. \tag{2.3}$$

In a terminology going back to the Pythagoreans, this says that the square on the ordinate is equal to the rectangle *applied* to PL, with width equal to the abcissa. Accordingly, Apollonius calls this curve a **parabola** (the Greek word for "application" is παραβολή) [25, p. 359].

If we take rectangular coordinates in $\mathcal{P}$ with the origin at P and axes parallel to QV ($y = |QV|$) and PV ($x = |PV|$), then denoting the length of the orthia PL by p, we obtain the equation for the rectangular coordinates of Q

$$y^2 = px. \tag{2.4}$$

The coefficient p above is called the **parameter of ordinates** for γ.

[7] the Latin translation of this term is **latus rectum**, although this term has come to mean a slightly different quantity, the parameter of ordinates.

[8] Details of a proof are in Appendix A

- **Ellipses:** If PV is not parallel to AC, then the line PV (extended) meets the line AB (extended) at a second vertex P'. If ϕ denotes the (acute) angle between $\mathcal{P}$ and a horizontal plane $\mathcal{H}$, then V lies between P and P' if $0 \leq \phi < \frac{\pi}{2} - \alpha$ and P lies between V and P' if $\frac{\pi}{2} - \alpha < \phi \leq \frac{\pi}{2}$.

  In the first case, in contrast to the case of the parabola, the ratio of $|QV|^2$ to $|PV|$ depends on the point Q on the curve γ. To understand it, we again form the "orthia" of γ as a line segment PL perpendicular to PV with length p.

  Now let S be the intersection of LP' with the line through V parallel to PL (Figure 2.4).

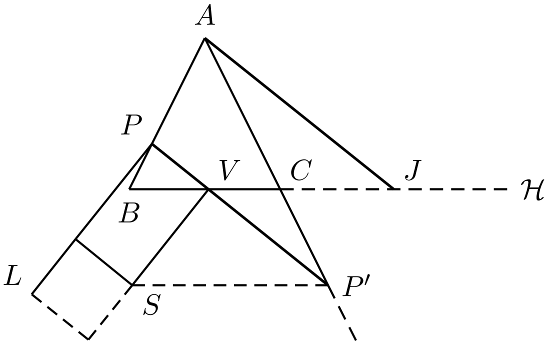


Figure 2.4: Definition of S

One derives the equation (see Appendix A)

$$|QV|^2 = |VS| \cdot |PV| \,. \tag{2.5}$$

This is like Equation (2.3), but $|PL|$ is replaced by the shorter length $|VS|$; in the Pythagorean terminology, the square on the ordinate is equal to the rectangle with width equal to the abcissa applied to the segment VS, *falling short* of PL. The Greek for "falling short" is ἔλλειψιζ, and Apollonius calls γ an **ellipse** in this case.

To obtain the rectangular equation of the ellipse, we set

$$d = \left|PP'\right|$$

(the **diameter**) and derive as the equation of the ellipse

$$y^2 = |VS|\, x = p\left(1 - \frac{x}{d}\right) x = px - \frac{p}{d}x^2. \tag{2.6}$$

- **Hyperbolas:** In the final case, when $\frac{\pi}{2} - \alpha < \phi \leq \frac{\pi}{2}$, The same arguments as in the ellipse case yield Equation (2.5), but this time the segment VS *exceeds* PL; the Greek for "excess" is ὑπερβολή, and γ is called a **hyperbola**.

  A verbatim repetition of the calculation leading to Equation (2.6) leads to its hyperbolic analogue,

  $$y^2 = px + \frac{p}{d}x^2. \tag{2.7}$$

  P lies between V and P'.

The Focus-Directrix Property

Pappus, in a section of the *Collection* headed "Lemmas to the *Surface Loci*[9] of Euclid", proves the following ([25, p. 153]):

Lemma 2.1.1. *If the distance of a point from a fixed point be in a given ratio to its distance from a fixed straight line, the locus of the point is a conic section, which is an ellipse, a parabola, or a hyperbola according as the ratio is less than, equal to, or greater than, unity.*

The fixed point is called the **focus**, the line is the **directrix**, and the ratio is called the **eccentricity** of the conic section. This *focus-directrix property* of conics is not mentioned by Apollonius, but Heath deduces from the way it is treated by Pappus that this lemma must have been stated without proof, and regarded as well-known, by Euclid. We outline a proof in Appendix B.

The focus-directrix characterization of conic sections can be turned into an equation. This approach—treating a curve as the locus of an equation in the rectangular coordinates—was introduced in the early seventeenth century by René Descartes (1596-1650) and Pierre de Fermat (1601-1665). Here, we sketch how this characterization leads to equations for the conic sections.

To fix ideas, let us place the y-axis along the directrix and the focus at a point $F(k,0)$ on the x-axis. The distance of a generic point $P(x,y)$ from the y-axis is $|x|$, while its distance from the focus is

$$|FP| = \sqrt{(x-k)^2 + y^2}.$$

[9] This work, like Euclid's *Conics*, is lost, and little information about its contents can be gleaned from Pappus.

Thus the focus-directrix property can be written

$$\frac{|FP|}{|x|} = e$$

where e is the eccentricity. Multiplying through by $|x|$ and squaring both sides leads to the equation of degree two

$$(x-k)^2 + y^2 = e^2x^2.$$

which can be rewritten

$$(1-e^2)x^2 - 2kx + y^2 = -k^2. \tag{2.8}$$

Parabolas

When $e = 1$, the x^2-term drops out, and we have

$$y^2 = 2kx - k^2 = 2k\left(x - \frac{k}{2}\right). \tag{2.9}$$

Now, we change coordinates, moving the y-axis $k/2$ units to the right. The effect of this on the equation is a bit counter-intuitive. To understand this, let us fix a point P whose coordinates *before* the move were (x, y); let us for a moment denote its coordinates *after* the move by (X, Y). Clearly, since nothing moves up or down, $Y = y$. However, the new origin is $k/2$ units to the right of the old one—or looking at it another way, the *new* origin was, in the *old* coordinate system, at $(x, y) = (k/2, 0)$, and is now (in the *new* coordinate system) at $(X, Y) = (0, 0)$; in particular, $X = x - k/2$. But this effect applies to all points. So if a point is on our parabola—that is, if its *old* coordinates satisfy Equation (2.9), then rewriting this in terms of the *new* coordinates we get $Y^2 = 2kX$.

Switching back to using lower-case x and y for our coordinates *after* the move, and setting

$$p = 2k,$$

we recover Equation (2.4)

$$y^2 = px$$

as the equation of the parabola with directrix

$$x = -\frac{p}{4}$$

and focus

$$F(\frac{p}{4}, 0).$$

This is sketched in Figure 2.5.

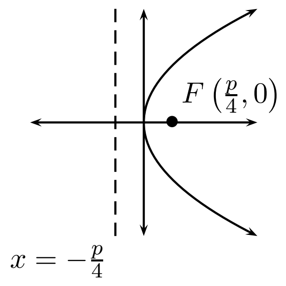


Figure 2.5: The Parabola $y^2 = px$, $p > 0$

By interchanging x and y and setting $p = 1/a$, we get the more familiar equation

$$y = ax^2 \tag{2.10}$$

which represents a parabola with focus $F(0, 1/4a)$ and directrix $y = -1/4a$. (Figure 2.6)

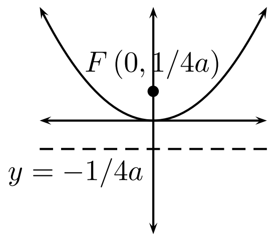


Figure 2.6: The Parabola $y = ax^2$, $a > 0$

Ellipses

When $e \neq 1$, completing the square in Equation (2.8) and moving the y-axis to the right $k/(1-e^2)$ units, we obtain

$$(1-e^2)x^2 + y^2 = \frac{k^2e^2}{1-e^2} \tag{2.11}$$

as the equation of the conic section with eccentricity $e \neq 1$, directrix where

$$x = -\frac{k}{1-e^2},$$

and focus

$$F(\frac{ke^2}{e^2-1}, 0) = F(\frac{-ke^2}{1-e^2}, 0).$$

Noting that the x-coordinate of the focus is e^2 times the constant in the equation of the directrix, let us consider the case when the focus is at $F(-ae, 0)$ and the directrix is $x = -a/e$: that is, let us set

$$a = \frac{ke}{1-e^2}.$$

This is positive provided $0 < e < 1$, the case of the ellipse.

If we divide both sides of Equation (2.11) by its right-hand side, we recognize the resulting coefficient for x^2 as $1/a^2$ and get the equation

$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1-e^2)} = 1. \tag{2.12}$$

In the case $e < 1$, we can define a new constant b by

$$b = a\sqrt{1-e^2},$$

so that (2.12) becomes

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{2.13}$$

This is the equation of the ellipse with focus $F(-ae, 0)$ and directrix $x = -a/e$, where

$$e = \sqrt{1 - \left(\frac{b}{a}\right)^2}. \tag{2.14}$$

Let us briefly note a few features of this curve:

- First, since x and y each enter Equation (2.13) only as their squares, replacing x with $-x$ (or y with $-y$) does not change the equation: this means the curve is invariant under **reflection across the *y*-axis**. In particular, this gives us a second focus/directrix pair for the curve: $F(ae, 0)$ and $x = a/e$.

- Second, it is clear that the ellipse is bounded: in fact the curve has x-intercepts $(\pm a, 0)$ and y-intercepts $(0, \pm b)$. In the case $a > b$ the distance $2a$ (*resp.* $2b$) between the x-intercepts (*resp.* y-intercepts) is called the **major axis** (*resp.* **minor axis**); the corresponding numbers a and b are the **semimajor axis** and the **semiminor axis**,and the x-intercepts are sometimes called the **vertices** of the ellipse. When $a < b$, the names are interchanged. Equation (2.13) with $b > a$ can be regarded as obtained from a version with $b < a$ by interchanging x with y. Geometrically, this means that when $b > a$ the foci are on the y-axis instead of the x-axis (and the directrices are horizontal).

- Third, if we know the semimajor axis a and the semiminor axis b, then the distance from the origin to the two foci is given by the formula (when $a > b$)

  $$c = |ae| = \sqrt{a^2 - b^2}$$

  A consequence of this formula is that the hypotenuse of the right triangle formed by a y-intercept, the origin, and a focus has length a. From this we have another characterization of an ellipse (Exercise 4):

  > *The sum of the distances of any point on the ellipse to the two foci equals the major axis.*

- Finally, note that when $a = b$, Equation (2.14) forces $e = 0$. Equation (2.13) is then the equation of a circle with radius a and center at the origin; however, the analysis which led us to Equation (2.13) is no longer valid—in particular the equation from which we started becomes $x^2 + y^2 = 0$, whose locus is just the origin. However, we can think of this as a virtual "limiting case" where the two foci are located at the origin and the directrices are "at infinity".

This information is illustrated in Figure 2.7

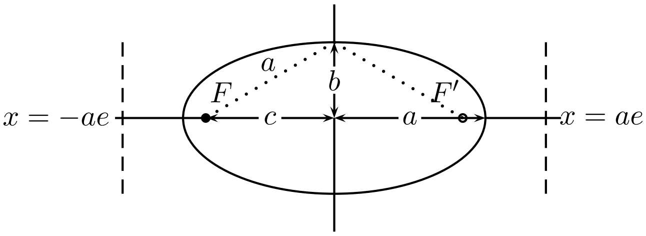


Figure 2.7: The ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, $a > b > 0$

Hyperbolas:

When $e > 1$, then the coefficient of y^2 in Equation (2.12) is *negative*; in this case we define the number $b > 0$ by

$$b = a\sqrt{e^2 - 1}$$

and obtain the hyperbolic analogue of Equation (2.13)

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{2.15}$$

as the equation of the hyperbola with focus $F(-ae, 0)$ and directrix $x = a/e$, where

$$e = \sqrt{1 + \left(\frac{b}{a}\right)^2}.$$

This curve shares some features with the ellipse, but is dramatically different in other respects:

- As in Equation (2.13), x and y each enter Equation (2.15) only as their squares, so replacing x with $-x$ (or y with $-y$) does not change the equation: this means the curve is invariant under reflection across the y-axis. In particular, this gives us a second focus/directrix pair for the curve: $F(ae, 0)$ and $x = a/e$.

- Equation (2.15) forces $|x| \geq a$: thus there are no y-intercepts: the curve has two separate branches, one opening to the right from $(a, 0)$ and the other opening to the left from $(-a, 0)$; these are the **vertices** of the hyperbola. The distance $2a$ between the vertices is called the **transverse axis** of the hyperbola. As $|x|$ grows, so does $|y|$: in each branch, y ranges over all values, with x decreasing to $x = a$ in the right branch (increasing to $x = -a$ in the left branch) as y decreases from ∞ to zero and then increasing (*resp.* decreasing) again as y passes zero and goes to $-\infty$.

- This time, there is no *a priori* inequality between a and b.

  Replacing the number 1 on the right of Equation (2.15) with -1

  $$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1; \tag{2.16}$$

  amounts to interchanging the roles of x (*resp.* a) and y (*resp.* b):

  $$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1 \tag{2.17}$$

  The locus of this equation is a curve whose branches open up and down from $(0, b)$ and $(0, -b)$ respectively; their foci are at $(0, be)$ (*resp.* $(0, -be)$) and their directrices are $y = b/e$ (*resp.* $y = -b/e$), where the eccentricity is

  $$e = \sqrt{1 + \left(\frac{a}{b}\right)^2}.$$

- The equation obtained by replacing the number 1 with 0 on the right of Equation (2.15)

  $$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0 \tag{2.18}$$

  has as its locus the two lines $y/x = \pm b/a$. It is straightforward to check that as a point $P(x, y)$ moves along the hyperbola with $x \to \pm\infty$, the ratio y/x tends to $\pm b/a$, so the distance of P from one of these lines goes to zero. These lines are called the **asymptotes** of the hyperbola.

- The distance from the origin to the two foci is given by the formula

  $$c = |ae| = \sqrt{a^2 + b^2}$$

  A consequence of this formula is another characterization of a a hyperbola (Exercise 5):

  > *The (absolute value of the) difference between the distances of any point on the hyperbola to the two foci equals the transverse axis.*

This information is illustrated in Figure 2.8.

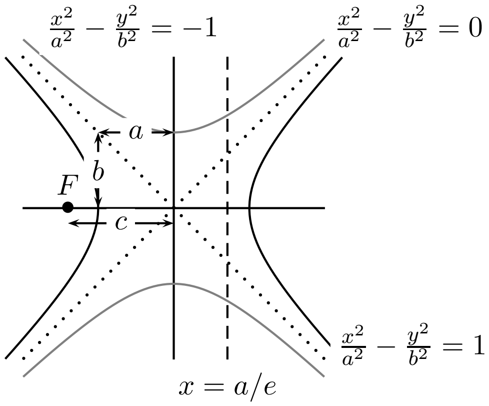

$\frac{x^2}{a^2} - \frac{y^2}{b^2} = -1$
$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$
a
b
F
c
$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$
$x = a/e$

Figure 2.8: Hyperbolas and asymptotes

Moving loci

In the model equations we have obtained for parabolas, ellipses and hyperbolas in this section, the origin and the two coordinate axes play special roles with respect to the geometry of the locus. For the parabola given by Equation (2.10), the origin is the *vertex*, the point of closest approach to the directrix, and the y-axis is an axis of symmetry for the parabola, while the x-axis is a kind of boundary which the curve can touch but never crosses. For the ellipse given by Equation (2.13), the coordinate axes are both axes of symmetry, containing the major and minor axes, and the origin is their intersection (the *center* of the ellipse). For the hyperbola given by Equation (2.15), the coordinate axes are again both axes of symmetry, and the origin is their intersection, as well as the intersection of the asymptotes (the *center* of the hyperbola).

Suppose we want to move one of these loci to a new location: that is, we want to displace the locus (without rotation) so that the special point given by the origin for the model equation moves to (α, β). Any such motion is accomplished by replacing x with x plus a constant and y with y plus another constant inside the equation; we need to do this in such a way that substituting $x = \alpha$ and $y = \beta$ into the *new* equation leads to the same calculation as substituting $x = 0$ and $y = 0$ into the *old* equation. It may seem wrong that this requires replacing x with $x - \alpha$ and y with $y - \beta$ in the old equation; to convince ourselves that it is right, let us consider a few simple examples.

First, the substitution

$$\begin{aligned} x &\mapsto x-1 \\ y &\mapsto y-2 \end{aligned}$$

into the model parabola equation

$$y = x^2$$

leads to the equation

$$y-2 = (x-1)^2;$$

we note that in the new equation, substitution of the point $(1,2)$ leads to the equation $0 = 0$, and furthermore no point lies below the horizontal line through this point, $y - 2 = 0$: we have displaced the parabola so as to move its vertex from the origin to the point $(1,2)$ (Figure 2.9).

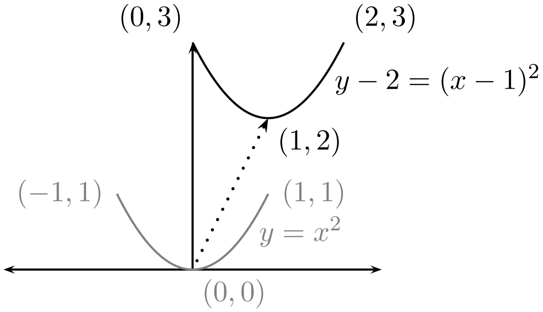


Figure 2.9: Displacing a parabola

Second, to move the ellipse

$$\frac{x^2}{4} + \frac{y^2}{1} = 1$$

so that its center moves to $(-2,2)$, we perform the substitution

$$\begin{aligned} x &\mapsto x-(-2) = x+2 \\ y &\mapsto y-2 \end{aligned}$$

(Figure 2.10)

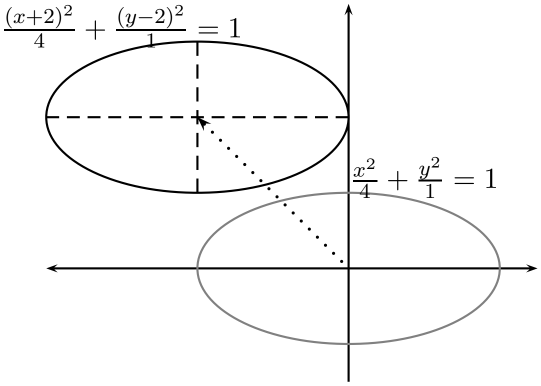

$\frac{(x+2)^2}{4} + \frac{(y-2)^2}{1} = 1$
$\frac{x^2}{4} + \frac{y^2}{1} = 1$

Figure 2.10: Displacing an ellipse

We can also reflect a locus about a coordinate axis. Since our model ellipses and hyperbolas are symmetric about these axes, this has no effect on the curve. However, while the model parabola given by Equation (2.10) is symmetric about the y-axis, it opens *up*; we can reverse this, making it open *down*, by replacing y with $-y$, or equivalently replacing the positive coefficient p with its negative. For example, when $p = 1$ this leads to the equation

$$y = -x^2$$

whose locus opens *down*: it is the reflection of our original parabola $y = x^2$ about the x-axis (Figure 2.11).

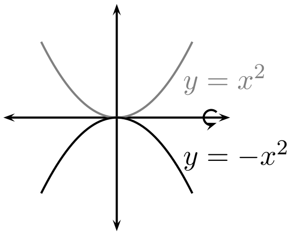


Figure 2.11: Reflecting a parabola about the x-axis

Finally, we can interchange the two variables; this effects a reflection about the diagonal line $y = x$. We have seen the effect of this on an ellipse and hyperbola. For a parabola, the interchange $x \leftrightarrow y$ takes the parabola

$y = x^2$, which opens along the positive y-axis (*i.e.*, *up*), to the parabola $x = y^2$, which opens along the positive x-axis (*i.e.*, *to the right*) (Figure 2.12), and the parabola $y = -x^2$, which opens along the negative y-axis (*i.e.*, *down*), to the parabola $x = -y^2$, which opens along the negative x-axis (*i.e.*, *to the left*).

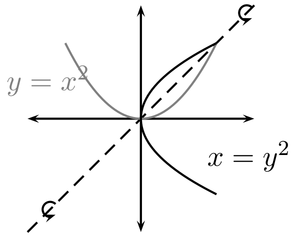


Figure 2.12: Reflecting a parabola about the diagonal

We shall see later § 3.10 that, with a few degenerate exceptions, every quadratic equation has as its locus one of the conic sections discussed here.

Exercises for § 2.1

Practice problems:

1. Identify each of the following curves as a circle, ellipse, hyperbola, parabola, or degenerate locus. For a parabola, determine the axis of symmetry and vertex. For a hyperbola, determine the vertices, asymptotes and center. For an ellipse (*resp.* circle), determine the center and semimajor and semiminor axes (*resp.* radius).

   (a) $y^2 = x + 2y$

   (b) $4x^2 + 4x + 4y^2 - 12y = 15$

   (c) $4x^2 + 4x + y^2 + 6y = 15$

   (d) $x^2 - 10x - y^2 - 6y - 2 = 0$

2. Determine the focus, directrix and eccentricity of each conic section below:

   (a) $2x^2 - 4x - y = 0$

   (b) $4y^2 - 16y + x + 16 = 0$

   (c) $4x^2 - 8x + 9y^2 + 36y + 4 = 0$

(d) $x^2 + 4x - 16y^2 + 32y + 4 = 0$

Theory problems:

3. **Show** that Equation (2.2) (the statement of Prop. 13, Book VI of the *Elements*) is equivalent to the standard equation for a circle.

4. **Show** that the sum of the distances from a point on an ellipse to its two foci equals the major axis. (You may assume the equation is in standard form.) This is sometimes called the *Gardener's characterization* of an ellipse: explain how one can construct an ellipse using a piece of string.

5. **Show** that the (absolute value of the difference between the distances from a point on a hyperbola to its two foci equals the transverse axis. (You may assume the equation is in standard form.)

History notes:

Spiral of Archimedes: Archimedes in his work *On Spirals* [3], studied the curve with polar equation $r = a\theta$ (a a positive constant) (see p. 145).

6. **Quadrature of the Circle:** According to Heath [24, vol. 1, p. 230] and Eves [13, p. 84], Archimedes is said to have used the spiral to construct a square whose area equals that of a given circle. This was one of the three classical problems (along with trisecting the angle and duplicating the cube) which the Greeks realized could not be solved by ruler-and-compass constructions [24, vol 1, pp. 218ff], although a proof of this impossibility was not given until the nineteenth century. However, a number of constructions using other curves (*not* constructible by compass and straightedge) were given. Our exposition of Archimedes' approach follows [13, p. 84].

   (a) The area of a circle is equal, by Archimedes' result in *Measurement of a Circle* [2, Prop. 1, p. 91], to half the product of its radius and its circumference. **Show** that the the ray perpendicular to the initial position of the ray generating the spiral cuts the spiral in a segment whose length is one-fourth of the circumference of the circle of radius a.

   (b) Use this to **show** that the side s of a square whose area equals that of the circle of radius a is the **mean proportional** between the circumference of the circle and the length of this segment.

(The mean proportional between A and B is the number M such that $A : M = M : B$.)

7. **Trisection of an Angle:** Proposition 12 in *On Spirals* [3, p. 166] gives an immediate construction for trisecting a given angle. Again, I follow Eves [13, p. 85]: given a spiral and a given angle $\angle AOB = \theta$, draw a spiral starting with the generating ray along OA, and let P be its intersection with the spiral. Now divide OP into three equal parts OP_1, $]Ps1P_2$, and P_2P. **Show** that $\angle P_1 0P = \angle P_1OP_2 = \angle P_2OP = \frac{\theta}{3}$. Note that a similar argument allows division of an angle into arbitrarily many equal parts.

8. **The Quadratrix of Hippias:** Pappus describes the construction of a curve he calls the *quadratrix*, which can be used for the quadrature of the circle as well as trisection of an angle. He ascribes it to Nicomedes (*ca.* 280-210 BC), but Proclus (411-485), a later commentator on Euclid and Greek geometry as important as Pappus, ascribes its invention to Hippias of Elis (*ca.* 460-400 BC), and Heath trusts him more than Pappus on this score (see [24, vol. 1, pp. 225-226]). The construction is as follows [13, p. 95]: the radius OX of a circle rotates through a quarter-turn (with constant angular speed) from position OC to position OA, while in the same time interval a line BD parallel to OA undergoes a parallel displacement (again with constant speed) from going through C to containing OA. The *quadratrix* is the locus of the intersection of the two during this motion (except for the final moment, when they coincide).

   (a) Assuming the circle has center at the origin and radius a and the final position of the radius OA is along the positive x-axis, **show** that the equation of the quadratrix in polar coordinates is

   $$\pi r \sin\theta = 2a\theta.$$

   (b) **Show** that if P is on the arc of the circle in the first quadrant, then the angle $\angle POA$ can be trisected as follows: let F be the intersection of OP with the quadratrix, and let FH be the vertical line segment to the x-axis. If F' is one-third the way from H to F along this segment, and $F'L$ is a horizontal segment with L on the quadratrix, then **show** that $\angle LOA = \frac{1}{3}\angle POA$.

   (c) i. **Show** that if the quadratrix intersects OA at G, then $OG = \frac{2a}{\pi}$. (You can use calculus here: in the proof by Dinostratus (*ca.* 390-320 BC), it is done by contradiction, using only Euclidean geometry.)

ii. Conclude from this that

$$\overset{\frown}{CA} : OA = OA : OG.$$

iii. **Show** how, in light of this, we can construct a line segment equal in length to $\overset{\frown}{CA}$.

iv. **Show** that a rectangle with one side equal to twice this line segment and the other equal to a has the same area as the circle of radius a.

v. Given a rectangle with sides of length w and h, show that the side s of a square with the same area satisfies $w : s = s : h$. The construction of a segment of length s given segments of respective lengths w and h is given in Proposition 13, Book VI of Euclid's *Elements*.

9. **The Conchoid of Nicomedes:** Nicomedes (*ca.* 280-210 BC) constructed the following curve: Fix a point O and a line L not going through O, and fix a length ℓ. Now, for each ray through O, let Q be its intersection with L and let P be further out along the ray so that QP has length a.

(a) **Show** that if O is the origin and L is the horizontal line at height b, then the equation of the conchoid in polar coordinates is

$$r = a + b \csc \theta.$$

(b) **Show** that the equation of the same conchoid in rectangular coordinates is

$$(y - b)^2(x^2 + y^2) = a^2y^2.$$

(c) **Trisecting an angle with the conchoid:** [25, p. 148] Consider the following configuration (see Figure 2.13): Given a rectangle $BCAF$, suppose that the line FA is extended to E in such a way that the line AD cuts off from BE a segment DE of length precisely $2AB$. Now let AG bisect DE. Then $AB = DG = GE$; **show** that these are also equal to AG. (*Hint:* $\angle DAE$ is a right angle.) Conclude that $\angle ABG = \angle AGB$ and $\angle GAE = \angle GEA$; use this to **show** that $\angle GBA = 2\angle AEG$. (*Hint:* external angles.)

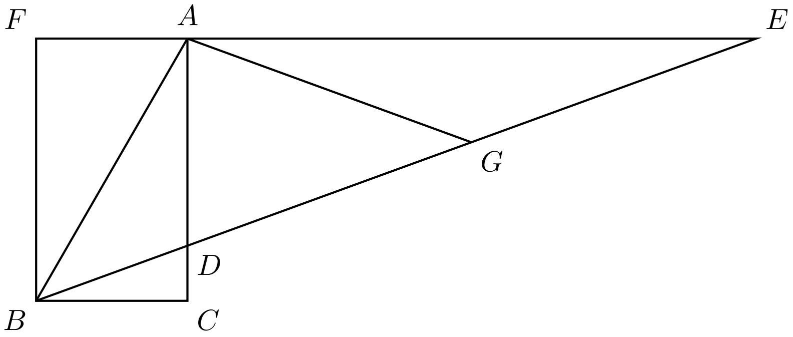


Figure 2.13: Trisecting an Angle

Finally, **show** that $\angle GBC = \angle AEG$, and use this to **show** that $\angle ABC = 3\angle GBC$.

How do we use this to trisect an angle? Given an angle, draw it as $\angle ABC$ where AC is perpendicular to BC. Now using B in place of) and the line AC in place of L, with $a = 2AB$, carry out the construction of the conchoid. **Show** that E is the intersection of the line through A parallel to BC with the conchoid. But then we have constructed the angle $\angle GBA$ to be one-third of the given angle.

Challenge problem:

10. **Show** that the locus of the equation

$$xy = 1$$

is a hyperbola. (*Hint:* consider a different coordinate system, using the diagonal and anti-diagonal as axes.)

2.2 Parametrized Curves

Parametrized Curves in the Plane

There are two distinct ways of specifying a curve in the plane. In classical geometric studies, a curve is given in a static way, either as the intersection of the plane with another surface (like the conical surface in Apollonius) or by a geometric condition (like fixing the distance from a point or the focus-directrix property in Euclid and Pappus). This approach reached its modern version in the seventeenth century with Descartes' and Fermat's formulation of a curve as the *locus of an equation* in the coordinates of a point. A second and equally important source of curves is dynamic in nature: a curve can be generated as the *path of a moving point.* This is the fundamental viewpoint in Newton's *Principia* (as well as the work of Newton's older contemporary Christian Huygens (1629-1695)), but "mechanical" constructions of curves also go back to antiquity, for example in "Archimedes' spiral" (p. 145).

We have seen in the case of lines how these two approaches interact: for example, the intersection of two lines is easier to find as the simultaneous solution of their equations, but a parametrized version more naturally encodes intrinsic geometric properties like the "direction" of a line. We have also seen that when one goes from lines in the plane to lines in space, the static formulation becomes unwieldy, requiring *two* equations, while—especially with the language of vectors—the dynamic formulation extends quite naturally. For this reason, we will adopt the dynamic approach as our primary way to specify a curve.

We can think of the position of a point moving in the plane as a **vector-valued function** assigning to each $t \in I$ the vector $\overrightarrow{p}(t)$; this point of view is signified by the notation

$$\overrightarrow{p}\colon \mathbb{R} \to \mathbb{R}^2$$

indicating that the function $\overrightarrow{p}$ takes real numbers as input and produces vectors in $\mathbb{R}^2$ as output. If we want to be explicit about the domain I we write

$$\overrightarrow{p}\colon I \to \mathbb{R}^2.$$

The component functions of a vector-valued function

$$\overrightarrow{p}(t) = (x(t)\,, y(t))$$

are simply the (changing) coordinates of the moving point; thus a vector-valued function $\overrightarrow{p}\colon \mathbb{R} \to \mathbb{R}^2$ is the same thing as a pair of (ordinary, real-valued) functions.

We have seen how to parametrize a line in the plane. Some other standard parametrizations of curves in the plane are:

Circle: A circle in the plane with center at the origin and radius $R > 0$ is the locus of the equation

$$x^2 + y^2 = R^2.$$

A natural way to locate a point on this circle is to give the angle that the radius through the point makes with the positive x-axis; equivalently, we can think of the circle as given by the equation $r = R$ in polar coordinates, so that the point is specified by the polar coordinate θ. Translating back to rectangular coordinates we have

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

and the parametrization of the circle is given by the vector-valued function

$$\overrightarrow{p}(\theta) = (R\cos\theta, R\sin\theta).$$

As θ goes through the values from 0 to 2π, $\overrightarrow{p}(\theta)$ traverses the circle once counterclockwise; if we allow *all* real values for θ, $\overrightarrow{p}(\theta)$ continues to travel counterclockwise around the circle, making a full circuit every time θ increases by 2π. Note that if we interchange the two formulas for x and y, we get another parametrization

$$\overrightarrow{q}(t) = (R\sin\theta, R\cos\theta)$$

which traverses the circle *clockwise*.

We can displace this circle, to put its center at any specified point $C(c_1, c_2)$, by adding the (constant) position vector of the desired center C to $\overrightarrow{p}(\theta)$ (or $\overrightarrow{q}(t)$):

$$\begin{aligned} \overrightarrow{r}(\theta) &= (R\cos\theta, R\sin\theta) + (c_1, c_2) \\ &= (c_1 + R\cos\theta, c_2 + R\sin\theta). \end{aligned}$$

Ellipse: The "model equation" for an ellipse with center at the origin (Equation (2.13) in § 2.1)

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

looks just like the equation for a circle of radius 1 centered at the origin, but with x (*resp.* y)) replaced by x/a (*resp.* y/b), so we can parametrize this locus via

$$\begin{aligned} \frac{x}{a} &= \cos\theta \\ \frac{y}{b} &= \sin\theta \end{aligned}$$

or

$$\overrightarrow{p}(\theta) = (a\cos\theta, b\sin\theta).$$

To understand the geometric significance of the parameter θ in this case (Figure 2.14), imagine a pair of circles centered at the origin, one circumscribed (with radius the semi-major axis a), the other inscribed (with radius the semi-minor axis b) in the ellipse.

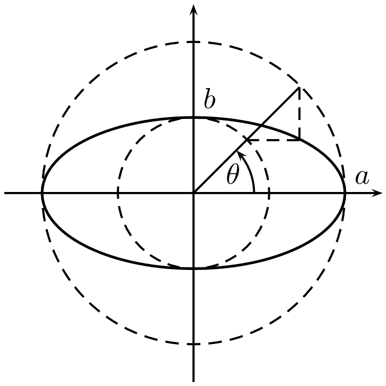


Figure 2.14: Parametrization of an Ellipse

Draw a ray at angle θ with the positive x-axis; the point $\overrightarrow{p}(\theta)$ is the intersection of two lines—one vertical, the other horizontal—through the intersections of the ray with the two circles. Again, the ellipse is traversed once counterclockwise as θ varies by 2π.

Again, by adding a constant displacement vector, we can move the ellipse so that its center is at (c_1, c_2):

$$\begin{aligned}\overrightarrow{r}(\theta) &= (a\cos\theta, b\sin\theta) + (c_1, c_2)\\ &= (c_1 + a\cos\theta, c_2 + b\sin\theta).\end{aligned}$$

Hyperbola: The "model equation" for a hyperbola (Equation (2.15) and Equation (2.16) in § 2.1)

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

can be parametrized as follows. The substitution

$$\begin{aligned}\frac{x}{a} &= \frac{e^t \pm e^{-t}}{2}\\ \frac{y}{b} &= \frac{e^t \mp e^{-t}}{2}\end{aligned}$$

yields

$$\begin{aligned}\left(\frac{x}{a}\right)^2 &= \frac{e^{2t}}{4} \pm 2\left(\frac{e^t}{2}\right)\left(\frac{e^{-t}}{2}\right) + \frac{e^{-2t}}{4}\\ &= \frac{e^{2t}}{4} \pm \frac{1}{2} + \frac{e^{-2t}}{4}\end{aligned}$$

and similarly

$$\begin{aligned}\left(\frac{y}{b}\right)^2 &= \frac{e^{2t}}{4} \mp 2\left(\frac{e^t}{2}\right)\left(\frac{e^{-t}}{2}\right) + \frac{e^{-2t}}{4}\\ &= \frac{e^{2t}}{4} \mp \frac{1}{2} + \frac{e^{-2t}}{4}\end{aligned}$$

so

$$\begin{aligned}\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 &= \pm\frac{1}{2} - \left(\mp\frac{1}{2}\right)\\ &= \pm 1.\end{aligned}$$

The functions

$$\begin{cases}\cosh t &= \frac{e^t+e^{-t}}{2}\\ \sinh t &= \frac{e^t-e^{-t}}{2}\end{cases} \tag{2.19}$$

are known, respectively, as the **hyperbolic cosine** and **hyperbolic sine** of t. Using Euler's formula (*Calculus Deconstructed*, p. 475), they can be interpreted in terms of the sine and cosine of an imaginary multiple of t, and satisfy variants of the usual trigonometric identities (Exercise 6):

$$\begin{aligned}\cosh t &= \frac{1}{2}\cos it\\ \sinh t &= \frac{1}{2i}\sin it.\end{aligned}$$

We see that

$$\overrightarrow{p}(t) = (a\cosh t, b\sinh t) \qquad -\infty < t < \infty$$

gives a curve satisfying

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

However, note that $\cosh t$ is always positive (in fact, $\cosh t \geq 1$ for all t), so this parametrizes only the "right branch" of the hyperbola; the "left branch" is parametrized by

$$\overrightarrow{p}(t) = (-a\cosh t, b\sinh t) \qquad -\infty < t < \infty.$$

Similarly, the two branches of

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

are parametrized by

$$\overrightarrow{p}(t) = (a\sinh t, \pm b\cosh t) \qquad -\infty < t < \infty.$$

Parabolas: The model equation for a parabola with horizontal directrix (Equation (2.10) in § 2.1)

$$y = ax^2$$

is easily parametrized using x as the parameter:

$$\begin{aligned} x &= t \\ y &= at^2 \end{aligned}$$

which leads to

$$\overrightarrow{p}(t) = \left(t, at^2\right) \qquad -\infty < t < \infty.$$

This last example illustrates how to parametrize a whole class of curves. The equation for a parabola gives one of the coordinates as an explicit function of the other—that is, the curve is represented as the graph of a function.

Remark 2.2.1. *If a curve is expressed as the graph of a function*

$$y = f(x)$$

then using the independent variable as our parameter, we can parametrize the curve as

$$\overrightarrow{p}(t) = (t, f(t)).$$

The circle $x^2 + y^2 = 1$ consists of two graphs: if we solve for y as a function of x, we obtain

$$y = \pm\sqrt{1 - x^2}, \quad -1 \leq x \leq 1.$$

The graph of the positive root is the upper semicircle, and this can be parametrized by

$$\begin{aligned} x(t) &= t \\ y(t) &= \sqrt{1 - t^2} \end{aligned}$$

or

$$\overrightarrow{p}(t) = (t, \sqrt{1 - t^2}), \quad t \in [-1, 1] \,.$$

Note, however, that in this parametrization, the upper semicircle is traversed *clockwise*; to get a *counterclockwise* motion, we replace t with its negative:

$$\overrightarrow{q}(t) = (-t, \sqrt{1 - t^2}), \quad t \in [-1, 1] \,.$$

The lower semicircle, traversed counterclockwise, is the graph of the negative root:

$$\overrightarrow{p}(t) = (t, -\sqrt{1 - t^2}, \quad t \in [-1, 1] \,.$$

The general relation between a plane curve, given as the locus of an equation, and its possible parametrizations will be clarified by means of the Implicit Function Theorem in Chapter 3.

Analyzing a Curve from a Parametrization

The examples in the preceding section all went from a static expression of a curve as the locus of an equation to a dynamic description as the image of a vector-valued function. The converse process can be difficult, but given a function $\overrightarrow{p}: \mathbb{R} \to \mathbb{R}^2$, we can try to "trace out" the path as the point moves.

As an example, consider the function $\overrightarrow{p}: \mathbb{R} \to \mathbb{R}^2$ defined by

$$\begin{aligned} x(t) &= t^3 \\ y(t) &= t^2 \end{aligned}$$

with domain $(-\infty, \infty)$. We note that $y(t) \geq 0$, with equality only for $t = 0$, so the curve lies in the upper half-plane. Note also that $x(t)$ takes each real value once, and that since $x(t)$ is an *odd* function and $y(t)$ is an *even* function, the curve is symmetric across the y-axis. Finally, we might note that the two functions are related by

$$(y(t))^3 = (x(t))^2$$

or

$$y(t) = (x(t))^{2/3}$$

so the curve is the graph of the function $x^{2/3}$—that is, it is the locus of the equation

$$y = x^{2/3}.$$

This is shown in Figure 2.15: as t goes from $-\infty$ to ∞, the point moves to the right, "bouncing" off the origin at $t = 0$.

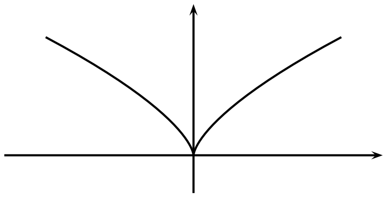


Figure 2.15: The curve $y^3 = x^2$

A large class of curves can be given as the graph of an equation in *polar coordinates.* Usually, this takes the form

$$r = f(\theta)\,.$$

Using the relation between polar and rectangular coordinates, this can be parametrized as

$$\overrightarrow{p}(\theta) = (f(\theta)\cos\theta, f(\theta)\sin\theta).$$

We consider a few examples.

The polar equation

$$r = \sin\theta$$

describes a curve which starts at the origin when $\theta = 0$; as θ increases, so does r until it reaches a maximum at $t = \frac{\pi}{2}$ (when $\overrightarrow{p}\left(\frac{\pi}{2}\right) = (0,1)$) and then decreases, with $r = 0$ again at $\theta = \pi$ ($\overrightarrow{p}(\pi) = (-1,0)$). For $\pi < \theta < 2\pi$, r is negative, and by examining the geometry of this, we see that the actual points $\overrightarrow{p}(\theta)$ trace out the same curve as was already traced out for $0 < \theta < \pi$. The curve is shown in Figure 2.16. In this case, we can recover an equation in rectangular coordinates for our curve: multiplying both sides of

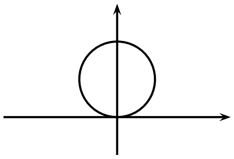


Figure 2.16: The curve $r = \sin\theta$

$$r = \sin\theta$$

by r, we obtain

$$r^2 = r\sin\theta$$

and then using the identities $r^2 = x^2 + y^2$ and $y = r\sin\theta$, we can write

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

which, after completing the square, can be rewritten as

$$x^2 + \left(y - \frac{1}{2}\right)^2 = \frac{1}{4}.$$

We recognize this as the equation of a circle centered at $(0, \frac{1}{2})$ with radius $\frac{1}{2}$.

The polar equation

$$r = \sin 2\theta$$

may appear to be an innocent variation on the preceding, but it turns out to be quite different. Again the curve begins at the origin when $\theta = 0$ and r increases with θ, but this time it reaches its maximum $r = 1$ when $\theta = \frac{\pi}{4}$, which is to say along the diagonal ($\overrightarrow{p}\left(\frac{\pi}{4}\right) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$), and then decreases, hitting $r = 0$ and hence the origin when $\theta = \frac{\pi}{2}$. Then r turns negative, which means that as θ goes from $\frac{\pi}{2}$ to π, the point $\overrightarrow{p}(\theta)$ lies in the fourth quadrant ($x > 0$, $y < 0$); for $\pi < \theta < \frac{3\pi}{2}$, r is again positive, and the point makes a "loop" in the third quadrant, and finally for $\frac{3\pi}{2} < \theta < 2\pi$, it traverses a loop in the second quadrant. After that, it traces out the same curve all over again. This curve is sometimes called a **four-petal rose** (Figure 2.17). Again, it is possible to express this curve as the locus of an equation in rectangular coordinates via mutliplication by r. However, it is slightly more complicated: if we multiply

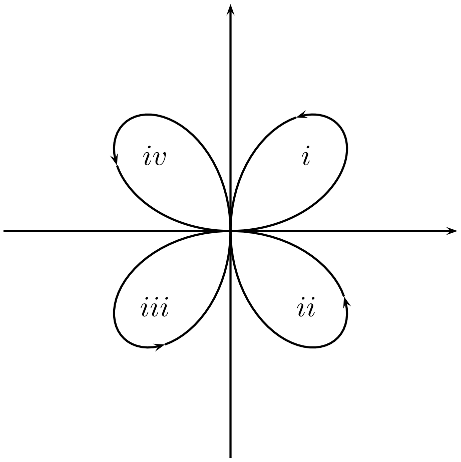


Figure 2.17: Four-petal Rose $r = \sin 2\theta$

$$r = \sin 2\theta$$

by r, we obtain

$$r^2 = r \sin 2\theta$$

whose left side is easy to interpret as $x^2 + y^2$, but whose right side is not so obvious. If we recall the identity $\sin 2\theta = 2 \sin\theta \cos\theta$, we see that

$$r \sin 2\theta = 2r \sin\theta \cos\theta$$

but to turn the right side into a recognizable expression in x and y we need to multiply through by r again; this yields

$$r^3 = 2r^2 \sin\theta r \cos\theta$$

or

$$\left(x^2 + y^2\right)^{3/2} = 2xy.$$

While this *is* an equation in rectangular coordinates, it is *not* particularly informative about our curve.

Polar equations of the form

$$r = \sin n\theta$$

define curves known as “roses”: it turns out that when n is *even* (as in the preceding example) there are $2n$ “petals”, traversed as θ goes over an interval of length 2π, but when n is *odd*—as for example $n = 1$, which was the previous example—then there are n “petals”, traversed as θ goes over an interval of length π.

A different kind of example is provided by the polar equation

$$r = a\theta$$

where $a > 0$ is a constant, which was (in different language, of course) studied by Archimedes of Syracuse (*ca.*287-212 BC) in his work *On Spirals* [3] and is sometimes known as the **spiral of Archimedes**. Here is his own description (as translated by Heath [26, p. 154]):

> *If a straight line of which one extremity remains fixed be made to revolve at a uniform rate in a plane until it returns to the position from which it started, and if, at the same time as the straight line revolves, a point move at a uniform rate along the straight line, starting from the fixed extremity, the point will describe a spiral in the plane.*

Of course, Archimedes is describing the above curve for the variation of θ from 0 to 2π. If we continue to increase θ beyond 2π, the curve continues to spiral out, as illustrated in Figure 2.18. If we include negative values of θ, we get another spiral, going clockwise instead of counterclockwise (Figure 2.19) It is difficult to see how to write down an equation in x and y with this locus.

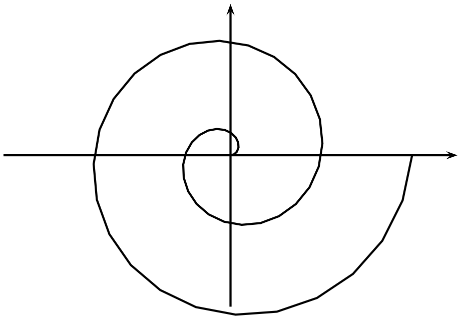


Figure 2.18: The Spiral of Archimedes, $r = \theta$, $\theta \geq 0$

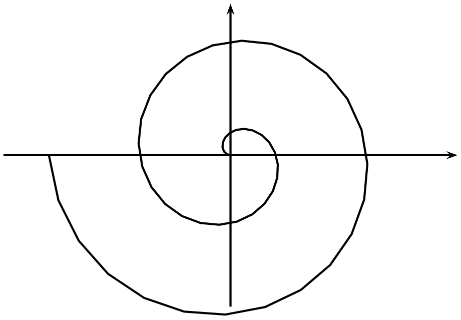


Figure 2.19: $r = \theta$, $\theta < 0$

Finally, we consider the **cycloid**, which can be described as the path of a point on the rim of a wheel which rolls along a line (Figure 2.20). Let R be the radius of the wheel, and assume that at the beginning the point is located on the line—which we take to be the ξ_x—at the origin, so the center of the wheel is at $(0, R)$. We take as our parameter the (clockwise) angle θ which the radius to the point makes with the downward vertical, that is, the amount by which the wheel has turned from its initial position. When the wheel turns θ radians, its center travels $R\theta$ units to the right, so the position of the *center* of the wheel corresponding to a given value of θ is

$$\begin{aligned}\overrightarrow{c}(\theta) &= R\overrightarrow{\jmath} + (R\theta)\overrightarrow{\imath} \\ &= (R\theta, R).\end{aligned}$$

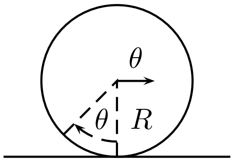


Figure 2.20: Turning Wheel

At that moment, the radial vector $\overrightarrow{r}(\theta)$ from the center of the wheel to the point on the rim is

$$\overrightarrow{r}(\theta) = -R(\sin\theta\overrightarrow{\imath} + \cos\theta\overrightarrow{\jmath})$$

and so the position vector of the point is

$$\begin{aligned}\overrightarrow{p}(\theta) &= \overrightarrow{c}(\theta) + \overrightarrow{r}(\theta)\\ &= (R\theta - R\sin\theta, R - R\cos\theta)\end{aligned}$$

or

$$\begin{aligned}x(\theta) &= R(\theta - \sin\theta)\\ y(\theta) &= R(1 - \cos\theta).\end{aligned}$$

The curve is sketched in Figure 2.21.

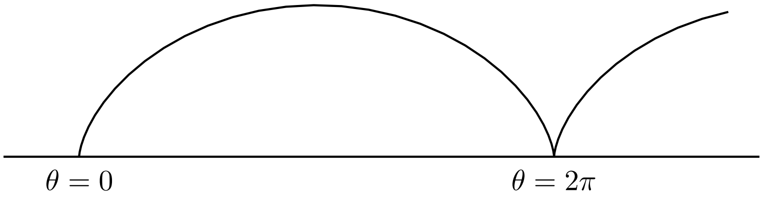


Figure 2.21: Cycloid

Curves in Space

As we have seen in the case of lines, when we go from curves in the plane to curves in space, the static formulation of a curve as the locus of an equation must be replaced by the more complicated idea of the locus of a *pair* of equations. By contrast, the dynamic view of a curve as the path of a moving point—especially when we use the language of vectors—extends very naturally to curves in space. We shall adopt this latter approach to specifying a curve in space.

The position vector of a point in space has three components, so the (changing) position of a moving point is specified by a function whose values are vectors in $\mathbb{R}^3$, which we denote by $\overrightarrow{p}\colon\mathbb{R}\to\mathbb{R}^3$; this can be regarded as a *triple* of functions:

$$\begin{aligned} x &= x(t) \\ y &= y(t) \\ z &= z(t) \end{aligned}$$

or

$$\overrightarrow{p}(t) = (x(t)\,, y(t)\,, z(t)).$$

As before, it is important to distinguish the *vector-valued function* $\overrightarrow{p}(t)$, which specifies the motion of a point, from the *path* traced out by the point. Of course the same *path* can be traced out by different *motions*; the *curve* parametrized by the function $\overrightarrow{p}(t)$ is the **range** (or **image**) of the function:

$$\mathcal{C} = \{\overrightarrow{p}(t) \mid t \in \operatorname{domain}(\overrightarrow{p})\}.$$

When we are given a vector-valued function $\overrightarrow{p}\colon\mathbb{R}\to\mathbb{R}^3$, we can try to analyze the motion by considering its projection on the coordinate planes. As an example, consider the function defined by

$$\begin{aligned} x(t) &= \cos 2\pi t \\ y(t) &= \sin 2\pi t \\ z(t) &= t \end{aligned}$$

which describes a point whose projection on the xy-plane moves counterclockwise in a circle of radius 1 about the origin; as this projection circulates around the circle, the point itself rises in such a way that during a complete "turn" around the circle, the "rise" is one unit. The "corkscrew" curve traced out by this motion is called a **helix** (Figure 2.22).

While this can be considered as the locus of the pair of equations

$$\begin{aligned} x &= \cos 2\pi z \\ y &= \sin 2\pi z \end{aligned}$$

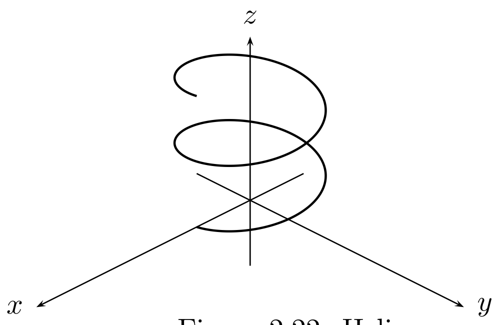


Figure 2.22: Helix

such a description gives us far less insight into the curve than the parametrized version.

As another example, let us parametrize the locus of the pair of equations

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

which, geometrically, is the intersection of the vertical cylinder

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

with the plane

$$y + z = 0.$$

The projection of the cylinder on the xy-plane is easily parametrized by

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

and then substitution into the equation of the plane gives us

$$z = -\sin t.$$

Thus, this curve can be described by the function $\overrightarrow{p}: \mathbb{R} \to \mathbb{R}^3$

$$\overrightarrow{p}(t) = (\cos t, \sin t, -\sin t).$$

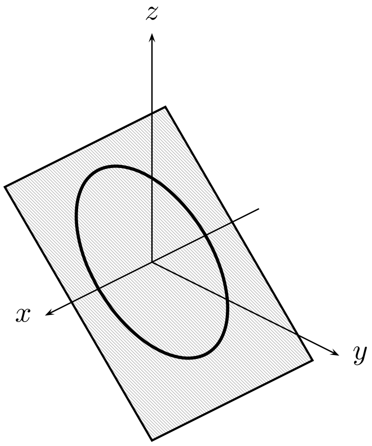


Figure 2.23: Intersection of the Cylinder $x^2+y^2=1$ and the Plane $y+z=0$.

It is shown in Figure 2.23. Note that it is an *ellipse*, not a circle (for example, it intersects the x-axis in a line of length 2, but it intersects the yz-plane in the points $(0,\pm 1,\mp 1)$, which are distance $\sqrt{2}$ apart).

How would we parametrize a *circle* in the plane $y+z=0$, centered at the origin? One way is to set up a rectangular coordinate system (much like we did for conic sections) given by

$$\begin{aligned} X &= x \\ Y &= y\sqrt{2} \end{aligned}$$

which gives the distance from the yz-plane and the x-axis. The translation back is

$$\begin{aligned} x &= X \\ y &= \frac{1}{\sqrt{2}}Y \\ z &= -\frac{1}{\sqrt{2}}Y. \end{aligned}$$

Then a circle of radius 1 centered at the origin but lying in the plane is given by the parametrization

$$\begin{aligned} X &= \cos t \\ Y &= \sin t \end{aligned}$$

and the translation of this to space coordinates is

$$\begin{aligned} x &= \cos t \\ y &= \frac{1}{\sqrt{2}} \sin t \\ z &= -\frac{1}{\sqrt{2}} \sin t \end{aligned}$$

or

$$\overrightarrow{p}(t) = (\cos t, \frac{1}{\sqrt{2}} \sin t, -\frac{1}{\sqrt{2}} \sin t).$$

This is sketched in Figure 2.24.

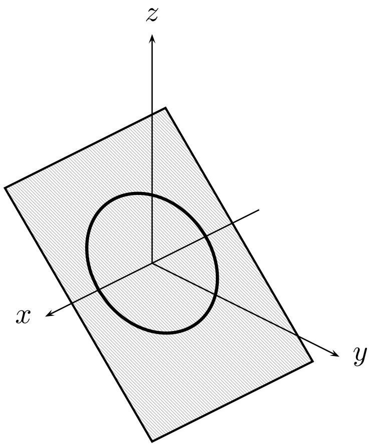


Figure 2.24: Circle of radius 1 about the Origin in the Plane $y + z = 0$.

Exercises for § 2.2

Practice problems:

1. Parametrize each plane curve below, indicating an interval of parameter values over which the curve is traversed once:

   (a) The circle of radius 5 with center $(2, 3)$.

   (b) The ellipse centered at $(1, 2)$ with horizontal semimajor axis 3 and vertical semiminor axis 1.

   (c) The upper branch of the hyperbola $y^2 - x^2 = 4$.

   (d) The lower branch of the hyperbola $4y^2 - x^2 = 1$.

2. Sketch the curve traced out by each function $\overrightarrow{p}: \mathbb{R} \to \mathbb{R}^2$:

   (a) $\overrightarrow{p}(t) = (t, \sin t)$

   (b) $\overrightarrow{p}(t) = (\cos t, t)$

   (c) $\overrightarrow{p}(t) = (3\cos t, \sin t)$

   (d) $\overrightarrow{p}(t) = (t\cos t, t\sin t)$

   (e) $\overrightarrow{p}(t) = (t + \sin t, t + \cos t)$

3. Sketch the curve given by the polar equation:

   (a) $r = 3\cos\theta$

   (b) $r = \sin 3\theta$

   (c) $r = \sin 4\theta$

   (d) $r = 1 - \cos\theta$

   (e) $r = 2\cos 2\theta$

4. Parametrize each of the curves in $\mathbb{R}^3$ described below:

   (a) The intersection of the plane $x + y + z = 1$ with the cylinder $y^2 + z^2 = 1$

   (b) The circle of radius 1, centered at $(1, 1, 1)$, and lying in the plane $x + y + z = 3$.

   (c) A curve lying on the cone $z = \sqrt{x^2 + y^2}$ which rotates about the z-axis while rising in such a way that in one rotation it rises 2 units. (*Hint:* Think cylindrical.)

   (d) The great circle[10] on the sphere of radius 1 about the origin which goes through the points $(1, 0, 0)$ and $(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$.

[10] A **great circle** on a sphere is a circle whose center is the center of the sphere.

Theory problems:

5. Using the definition of the hyperbolic cosine and sine (Equation (2.19)), prove that they satisfy the identities:

   (a)
$$\cosh^2 t - \sinh^2 t = 1.$$

   (b)
$$\cosh^2 t = \frac{1}{2}(1 + \cosh 2t)$$

   (c)
$$\sinh^2 t = \frac{1}{2}(1 - \cosh 2t)$$

Challenge problem:

6. Using Euler's formula
$$e^{a+bi} = e^a(\cos b + i \sin t)$$
   prove the identities
$$\cosh t = \frac{1}{2}\cos it$$
$$\sinh t = \frac{1}{2i}\sin it.$$

7. (a) A wheel of radius 1 in the plane, rotating counterclockwise with angular velocity ω_1 rotations per second, is attached to the end of a stick of length 3 whose other end is fixed at the origin, and which itself is rotating counterclockwise with angular velocity ω_2 rotations per second. Parametrize the motion of a point on the rim of the wheel.

   (b) A wheel of radius 1 in the plane rolls along the outer edge of the disc of radius 3 centered at the origin. Parametrize the motion of a point on the rim.

8. A vertical plane $\mathcal{P}$ through the z-axis makes an angle θ radians with the xz-plane counterclockwise (seen from above). The **torus** $\mathcal{T}$ consists of all points in $\mathbb{R}^3$ at distance 1 from the circle $x^2 + y^2 = 9$, $z = 0$ in the xy-plane. Parametrize the intersection $\mathcal{P} \cap \mathcal{T}$ of these surfaces. (*Hint:* It is a circle.)

9. Parametrize the path in space of a point on the wheel of a unicycle of radius b which is ridden along a circular path of radius a centered at the origin. (*Hint:* Note that the plane of the unicycle is vertical and contains, at any moment, the line tangent to the path at the point of contact with the wheel. Note also that as the wheel turns, it travels along the path a distance given by the amount of rotation (in radians) times the radius of the wheel.)

2.3 Calculus of Vector-Valued Functions

To apply methods of calculus to curves in $\mathbb{R}^2$ or $\mathbb{R}^3$ or equivalently to their parametrizations via vector-valued functions, we must first reformulate the basic notion of convergence, as well as differentiation and integration, in these contexts.

Convergence of Sequences of Points

The convergence of sequences of points $\{\overrightarrow{p}_i\}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$ is a natural extension of the corresponding idea for numbers, or points on the line $\mathbb{R}$. We will state everything in terms of $\mathbb{R}^3$, but the corresponding statements and/or proofs for $\mathbb{R}^2$ are easy modifications of the $\mathbb{R}^3$ versions.

Before formulating a geometric definition of convergence, we note a few properties of the distance function on $\mathbb{R}^3$. The first property will allow us to use estimates on coordinates to obtain estimates on distances, and *vice-versa*.

Lemma 2.3.1. *Suppose $P, Q \in \mathbb{R}^3$ have respective (rectangular) coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2). Let*

$$\delta := \max(|\triangle x|, |\triangle y|, |\triangle z|)$$

(where $\triangle x := x_2 - x_1$, etc.)

Then

$$\delta \leq \operatorname{dist}(P, Q) \leq \delta\sqrt{3}. \tag{2.20}$$

Proof. On one hand, clearly each of $(\triangle x)^2$, $(\triangle y)^2$ and $(\triangle z)^2$ is less than or equal to their sum $(\triangle x)^2 + (\triangle y)^2 + (\triangle z)^2$, since all three are non-negative. Thus

$$\delta^2 \leq \operatorname{dist}(P, Q)^2$$

and taking square roots,

$$\delta \leq \operatorname{dist}(P, Q).$$

On the other hand,

$$(\triangle x)^2 + (\triangle y)^2 + (\triangle z)^2 \leq \delta^2 + \delta^2 + \delta^2 = 3\delta^2$$

and taking square roots we have

$$\operatorname{dist}(P, Q) \leq \delta\sqrt{3}.$$

□

In particular, we clearly have

$$\operatorname{dist}(P, Q) = 0 \iff P = Q. \tag{2.21}$$

The next important property is proved by a calculation which you do in Exercise 3.

Lemma 2.3.2 (Triangle Inequality)**.** *For any three points $P, Q, R \in \mathbb{R}^3$,*

$$\operatorname{dist}(P, Q) \leq \operatorname{dist}(P, R) + \operatorname{dist}(R, Q). \tag{2.22}$$

With these properties in hand, we consider the notion of convergence for a sequence $\{\overrightarrow{p}_i\}$ of points $\overrightarrow{p}_i \in \mathbb{R}^3$. The definition is an almost verbatim translation of the corresponding notion for sequences of numbers (*i.e.*, of points in $\mathbb{R}$) (*Calculus Deconstructed*, Dfn. 2.2.2).

Definition 2.3.3. *A sequence of points $\overrightarrow{p}_i \in \mathbb{R}^3$ **converges** to a point $L \in \mathbb{R}^3$ if for every desired accuracy $\varepsilon > 0$ there exists a place N in the sequence such that every later point of the sequence approximates L with accuracy ε:*

$$i > N \text{ guarantees } \operatorname{dist}(\overrightarrow{p}_i, L) < \varepsilon.$$

We will write

$$\overrightarrow{p}_i \to L$$

in this case.

An immediate corollary of the triangle inequality is the uniqueness of limits:

Corollary 2.3.4. *If a sequence $\{\overrightarrow{p}_i\}$ converges to L and also to L', then $L = L'$.*

Proof. For any $\varepsilon > 0$, we can find integers N and N' so that $\operatorname{dist}(\overrightarrow{p}_i, L) < \varepsilon$ for every $i > N$ and also $\operatorname{dist}(\overrightarrow{p}_i, L') < \varepsilon$ for every $i > N'$. Pick an index i beyond both N and N', and estimate the distance from L to L' as follows:

$$\begin{aligned}\operatorname{dist}(L, L') &\leq \operatorname{dist}(L, \overrightarrow{p}_i) + \operatorname{dist}(\overrightarrow{p}_i, L') \\ &< \varepsilon + \varepsilon = 2\varepsilon.\end{aligned}$$

But this says that $\operatorname{dist}(L, L')$ is less than any positive number and hence equals zero, so $L = L'$ by Equation (2.21). □

As a result of Corollary 2.3.4, if $\overrightarrow{p}_i \to L$ we can refer to L as *the* **limit** of the sequence, and write

$$L = \lim \overrightarrow{p}_i.$$

A sequence is **convergent** if it has a limit, and **divergent** if it has none.

The next result lets us relate convergence of points to convergence of their coordinates.

Lemma 2.3.5. *Suppose $\{\overrightarrow{p}_i\}$ is a sequence of points in $\mathbb{R}^3$ with respective coordinates (x_i, y_i, z_i) and $L \in \mathbb{R}^3$ has coordinates (ℓ_1, ℓ_2, ℓ_3). Then the following are equivalent:*

1. *$\overrightarrow{p}_i \to L$ (in $\mathbb{R}^3$);*

2. *$x_i \to \ell_1$, $y_i \to \ell_2$, and $z_i \to \ell_3$ (in $\mathbb{R}$).*

Proof. (1) ⇒ (2)**:** Suppose $\overrightarrow{p}_i \to L$. Given $\varepsilon > 0$, we can find N so that $i > N$ guarantees $\operatorname{dist}(\overrightarrow{p}_i, L) < \varepsilon$. But then by Lemma 2.3.1

$$\max(|x_i - \ell_1|, |y_i - \ell_2|, |z_i - \ell_3|) < \varepsilon,$$

showing that each of the coordinate sequences converges to the corresponding coordinate of L.

(2) ⇒ (1)**:** Suppose $x_i \to \ell_1$, $y_i \to \ell_2$, and $z_i \to \ell_3$. Given $\varepsilon > 0$, we can find

$$\begin{aligned}&N_1 \text{ so that } i > N_1 \text{ guarantees } |x_i - \ell_1| < \frac{\varepsilon}{\sqrt{3}} \\ &N_2 \text{ so that } i > N_2 \text{ guarantees } |y_i - \ell_2| < \frac{\varepsilon}{\sqrt{3}} \\ &N_3 \text{ so that } i > N_3 \text{ guarantees } |z_i - \ell_3| < \frac{\varepsilon}{\sqrt{3}}.\end{aligned}$$

Let $L \in \mathbb{R}^3$ be the point with rectangular coordinates (ℓ_1, ℓ_2, ℓ_3). Setting $N = \max(N_1, N_2, N_3)$, we see that

$$i > N \text{ guarantees } \delta := \max(|x_i - \ell_1|, |y_i - \ell_2|, |z_i - \ell_3|) < \frac{\varepsilon}{\sqrt{3}}$$

and hence by Lemma 2.3.1

$$i > N \text{ guarantees } \operatorname{dist}(\overrightarrow{p}_i, L) < \sqrt{3}\frac{\varepsilon}{\sqrt{3}} = \varepsilon,$$

so $\overrightarrow{p}_i \to L$.

□

As in $\mathbb{R}$, we say a sequence $\{\overrightarrow{p}_i\}$ of points is **bounded** if there is a finite upper bound on the distance of all the points in the sequence from the origin—that is,

$$\sup\{\operatorname{dist}(\overrightarrow{p}_i, \mathcal{O})\} < \infty.$$

An easy analogue of a basic property of sequences of numbers is the following, whose proof we leave to you (Exercise 4):

Remark 2.3.6. *Every convergent sequence is bounded.*

A major difference between sequences of numbers and sequences of points in $\mathbb{R}^3$ is that there is no natural way to compare two points: a statement like "$P < Q$" does not make sense for points in space. As a result, there is no natural way to speak of monotone sequences, and correspondingly we cannot think about, for example, the maximum or supremum of a (bounded) sequence of points. What we *can* do, however, is to think about the maximum or supremum of a sequence of *numbers* associated to a sequence of points—we have already seen an instance of this in the definition of boundedness for a sequence.

One consequence of the lack of natural inequalities between points is that we cannot translate the Completeness Axiom (*Calculus Deconstructed*, Axiom 2.3.2) directly to $\mathbb{R}^3$. However, the Bolzano-Weierstrass Theorem (*Calculus Deconstructed*, Prop. 2.3.8), which is an effective substitute for the Completeness Axiom, can easily be extended from sequences of numbers to sequences of points:

Proposition 2.3.7 (Bolzano-Weierstrass Theorem)**.** *Every bounded sequence of points in $\mathbb{R}^3$ has a convergent subsequence.*

Proof. Suppose M is an upper bound on distances from the origin:

$$\operatorname{dist}(\overrightarrow{p}_i, \mathcal{O}) < M \text{ for all } i.$$

In particular, by Lemma 2.3.1,

$$|x_i| < M \text{ for all } i$$

so the first coordinates of our points form a bounded sequence of numbers. But then the Bolzano-Weierstrass Theorem in $\mathbb{R}$ says that we can pick a convergent subsequence of these numbers, say

$$x_{i_k} \to \ell_1 \in \mathbb{R}.$$

Now, consider the (sub)sequence of points $\{\overrightarrow{p}_{i_k}\}$, and look at their *second* coordinates. We have

$$|y_{i_k}| < M \text{ for all } i_k$$

and hence passing to a sub-(sub)sequence, we have

$$y_{i'_k} \to \ell_2 \in \mathbb{R}.$$

Note that passing to a subsequence does not affect convergence of the *first* coordinates:

$$x_{i'_k} \to \ell_1.$$

In a similar way, the *third* coordinates are bounded

$$\left|z_{i'_k}\right| < M \text{ for all } i'_k$$

and hence we can find a convergent sub-(sub-sub)sequence $\{\overrightarrow{p}_{i''_k}\}$ for which the *third* coordinates converge as well:

$$z_{i''_k} \to \ell_3 \in \mathbb{R}.$$

Since passing to subsequences has not hurt the convergence of the first and second coordinates

$$x_{i''_k} \to \ell_1$$
$$y_{i''_k} \to \ell_2$$

we see that

$$\overrightarrow{p}_{i''_k} \to L$$

where $L \in \mathbb{R}^3$ is the point with rectangular coordinates (ℓ_1, ℓ_2, ℓ_3). □

In the exercises, you will check a number of features of convergence (and divergence) which carry over from sequences of numbers to sequences of points.

Continuity of Vector-Valued Functions

Using the notion of convergence formulated in the previous subsection, the notion of continuity for real-valued functions extends naturally to vector-valued functions.

Definition 2.3.8. $\overrightarrow{f}:\mathbb{R}\rightarrow\mathbb{R}^3$ *is* ***continuous*** *on* $D \subset \mathbb{R}$ *if for every convergent sequence* $t_i \rightarrow t$ *in* D *the sequence of points* $\overrightarrow{f}(t_i)$ *converges to* $\overrightarrow{f}(t)$*.*

Every function from $\mathbb{R}$ to $\mathbb{R}^3$ can be expressed as

$$\overrightarrow{f}(t) = (f_1(t)\,, f_2(t)\,, f_3(t))$$

or

$$\overrightarrow{f}(t) = f_1(t)\,\overrightarrow{\imath} + f_2(t)\,\overrightarrow{\jmath} + f_3(t)\,\overrightarrow{k}$$

where $f_1(t)$, $f_2(t)$ and $f_3(t)$, the **component functions** of $\overrightarrow{f}(t)$, are ordinary (real-valued) functions. Using Lemma 2.3.5, it is easy to connect continuity of $\overrightarrow{f}(t)$ with continuity of its components:

Remark 2.3.9. *A function* $\overrightarrow{f}:\mathbb{R}\rightarrow\mathbb{R}^3$ *is continuous on* $D \subset \mathbb{R}$ *precisely if each of its components* $f_1(t)$*,* $f_2(t)$*,* $f_3(t)$ *is continuous on* D*.*

A related notion, that of limits, is an equally natural generalization of the single-variable idea:

Definition 2.3.10. $\overrightarrow{f}:\mathbb{R}\rightarrow\mathbb{R}^3$ ***converges*** *to* $\overrightarrow{L} \in \mathbb{R}^3$ *as* $t \rightarrow t_0$ *if* t_0 *is an accumulation point of the domain of* $\overrightarrow{f}(t)$ *and for every sequence* $\{t_i\}$ *in the domain of* $\overrightarrow{f}$ *which converges to, but is distinct from,* t_0*, the sequence of points* $p_i = \overrightarrow{f}(t_i)$ *converges to* $\overrightarrow{L}$*.*

We write

$$\overrightarrow{f}(t) \rightarrow \overrightarrow{L} \text{ as } t \rightarrow t_0$$

or

$$\overrightarrow{L} = \lim_{t\rightarrow t_0} \overrightarrow{f}(t)$$

when this holds.

Again, convergence of $\overrightarrow{f}$ relates immediately to convergence of its components:

Remark 2.3.11. *$\overrightarrow{f}\colon\mathbb{R}\to\mathbb{R}^3$ converges to $\overrightarrow{L}$ as $t\to t_0$ precisely when the components of $\overrightarrow{f}$ converge to the components of $\overrightarrow{L}$ as $t\to t_0$.*

If any of the component functions diverges as $t\to t_0$, then so does $\overrightarrow{f}(t)$.

The following algebraic properties of limits are easy to check (Exercise 7):

Proposition 2.3.12. *Suppose $\overrightarrow{f},\overrightarrow{g}\colon\mathbb{R}\to\mathbb{R}^3$ satisfy*

$$\overrightarrow{L}_f = \lim_{t\to t_0}\overrightarrow{f}(t)$$
$$\overrightarrow{L}_g = \lim_{t\to t_0}\overrightarrow{g}(t)$$

and $r\colon\mathbb{R}\to\mathbb{R}$ satisfies

$$L_r = \lim_{t\to t_0} r(t).$$

Then

1. $\lim_{t\to t_0}\left[\overrightarrow{f}(t)\pm\overrightarrow{g}(t)\right] = \overrightarrow{L}_f \pm \overrightarrow{L}_g$
2. $\lim_{t\to t_0} r(t)\overrightarrow{f}(t) = L_r\overrightarrow{L}_f$
3. $\lim_{t\to t_0}\left[\overrightarrow{f}(t)\cdot\overrightarrow{g}(t)\right] = \overrightarrow{L}_f \cdot \overrightarrow{L}_g$
4. $\lim_{t\to t_0}\left[\overrightarrow{f}(t)\times\overrightarrow{g}(t)\right] = \overrightarrow{L}_f \times \overrightarrow{L}_g.$

Derivatives of Vector-Valued Functions

When we think of a function $\overrightarrow{f}\colon\mathbb{R}\to\mathbb{R}^3$ as describing a moving point, it is natural to ask about its velocity, acceleration and so on. For this, we need to extend the notion of differentiation. We shall often use the Newtonian "dot" notation for the derivative of a vector-valued function interchangeably with "prime".

Definition 2.3.13. *The **derivative** of the function $\overrightarrow{f}\colon\mathbb{R}\to\mathbb{R}^3$ at an interior point t_0 of its domain is the limit*

$$\dot{\overrightarrow{f}}(t_0) = \vec{f}'(t_0) = \left.\frac{d}{dt}\right|_{t=t_0}\left[\overrightarrow{f}\right] = \lim_{h\to 0}\frac{1}{h}\left[\overrightarrow{f}(t_0+h)-\overrightarrow{f}(t_0)\right]$$

provided it exists. (If not, the function is not differentiable at $t=t_0$.)

Again, using Lemma 2.3.5, we connect this with differentiation of the component functions:

Remark 2.3.14. *The vector-valued function*

$$\overrightarrow{f}(t) = (x(t), y(t), z(t))$$

is differentiable at $t = t_0$ precisely if all of its component functions are differentiable at $t = t_0$, and then

$$\vec{f}'(t_0) = (x'(t_0), y'(t_0), z'(t_0)).$$

In particular, every differentiable vector-valued function is continuous.

When $\overrightarrow{p}(t)$ describes a moving point, then its derivative is referred to as the **velocity** of $\overrightarrow{p}(t)$

$$\overrightarrow{v}(t_0) = \dot{\overrightarrow{p}}(t_0)$$

and the derivative of velocity is **acceleration**

$$\overrightarrow{a}(t_0) = \dot{\overrightarrow{v}}(t_0) = \ddot{\overrightarrow{p}}(t_0).$$

The *magnitude* of the velocity is the **speed**, sometimes denoted

$$\frac{d\mathfrak{s}}{dt} = \|\overrightarrow{v}(t)\|.$$

Note the distinction between *velocity*, which has a direction (and hence is a vector) and *speed*, which has no direction (and is a scalar).

For example, the point moving along the helix

$$\overrightarrow{p}(t) = (\cos 2\pi t, \sin 2\pi t, t)$$

has velocity

$$\overrightarrow{v}(t) = \dot{\overrightarrow{p}}(t) = (-2\pi \sin 2\pi t, 2\pi \cos 2\pi t, 1)$$

speed

$$\frac{d\mathfrak{s}}{dt} = \sqrt{4\pi^2 + 1}$$

and acceleration

$$\overrightarrow{a}(t) = \dot{\overrightarrow{v}}(t) = (-4\pi^2 \cos 2\pi t, -4\pi^2 \sin 2\pi t, 0).$$

The relation of derivatives to vector algebra is analogous to the situation for real-valued functions.

Theorem 2.3.15. *Suppose the vector-valued functions $\overrightarrow{f}, \overrightarrow{g}: I \to \mathbb{R}^3$ are differentiable on I. Then the following are also differentiable:*

Linear Combinations: *for any real constants $\alpha, \beta \in \mathbb{R}$, the function*

$$\alpha \overrightarrow{f}(t) + \beta \overrightarrow{g}(t)$$

is differentiable on I, and

$$\frac{d}{dt}\left[\alpha \overrightarrow{f}(t) + \beta \overrightarrow{g}(t)\right] = \alpha \vec{f}'(t) + \beta \vec{g}'(t)\,.$$

Products: [11]

- *The product with any differentiable real-valued function $\alpha(t)$ on I is differentiable on I:*

$$\frac{d}{dt}\left[\alpha(t)\overrightarrow{f}(t)\right] = \alpha'(t)\overrightarrow{f}(t) + \alpha(t)\vec{f}'(t).$$

- *The dot product (resp. cross product) of two differentiable vector-valued functions on I is differentiable on I:*

$$\frac{d}{dt}\left[\overrightarrow{f}(t) \cdot \overrightarrow{g}(t)\right] = \vec{f}'(t) \cdot \overrightarrow{g}(t) + \overrightarrow{f}(t) \cdot \vec{g}'(t)$$
$$\frac{d}{dt}\left[\overrightarrow{f}(t) \times \overrightarrow{g}(t)\right] = \vec{f}'(t) \times \overrightarrow{g}(t) + \overrightarrow{f}(t) \times \vec{g}'(t)\,.$$

Compositions: [12] *If $\mathfrak{t}(s)$ is differentiable function on J and takes values in I, then the composition $(\overrightarrow{f} \circ \mathfrak{t})(s)$ is differentiable on J:*

$$\frac{d}{dt}\left[\overrightarrow{f}(\mathfrak{t}(s))\right] = \frac{d\overrightarrow{f}}{dt}\frac{dt}{ds} = \vec{f}'(\mathfrak{t}(s))\frac{d}{ds}\left[\mathfrak{t}(s)\right].$$

Proof. The proofs of differentiability of linear combinations, as well as products or compositions with real-valued functions, are most easily done by looking directly at the corresponding coordinate expressions. For example, to prove differentiability of $\alpha(t)\overrightarrow{f}(t)$, write

$$\overrightarrow{f}(t) = (x(t)\,, y(t)\,, z(t));$$

[11] These are the **product rules** or **Leibniz formulas** for vector-valued functions of one variable.

[12] This is a **chain rule** for curves

then

$$\alpha(t)\overrightarrow{f}(t) = (\alpha(t)x(t)\,, \alpha(t)y(t)\,, \alpha(t)z(t)).$$

The ordinary product rule applied to each coordinate in the last expression, combined with Remark 2.3.14, immediately gives us the second statement. The proof of the first statement, which is similar, is left to you (Exercise 8a).

The product rules for dot and cross products are handled more efficiently using vector language. We will prove the product rule for the dot product and leave the cross product version to you (Exercise 8b). In effect, we mimic the proof of the product rule for real-valued functions. Write

$$\begin{aligned}
&\overrightarrow{f}(t_0+h)\cdot\overrightarrow{g}(t_0+h)-\overrightarrow{f}(t_0)\cdot\overrightarrow{g}(t_0)\\
&\qquad=\overrightarrow{f}(t_0+h)\cdot\overrightarrow{g}(t_0+h)-\overrightarrow{f}(t_0+h)\cdot\overrightarrow{g}(t_0)\\
&\qquad\quad+\overrightarrow{f}(t_0+h)\cdot\overrightarrow{g}(t_0)-\overrightarrow{f}(t_0)\cdot\overrightarrow{g}(t_0)\\
&\qquad=\overrightarrow{f}(t_0+h)\cdot\left(\overrightarrow{g}(t_0+h)-\overrightarrow{g}(t_0)\right)\\
&\qquad\quad+\left(\overrightarrow{f}(t_0+h)-\overrightarrow{f}(t_0)\right)\cdot\overrightarrow{g}(t_0)\,;
\end{aligned}$$

then using Proposition 2.3.12 we see that

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=t_0}\left[\overrightarrow{f}(t)\cdot\overrightarrow{g}(t)\right] &= \lim_{h\to 0}\left[\overrightarrow{f}(t_0+h)\cdot\left(\frac{\overrightarrow{g}(t_0+h)-\overrightarrow{g}(t_0)}{h}\right)\right]\\
&\quad+\lim_{h\to 0}\left[\left(\frac{\overrightarrow{f}(t_0+h)-\overrightarrow{f}(t_0)}{h}\right)\cdot\overrightarrow{g}(t_0)\right]\\
&=\overrightarrow{f}(t_0)\cdot\vec{g}'(t_0)+\vec{f}'(t_0)\cdot\overrightarrow{g}(t_0)\,.
\end{aligned}$$

□

An interesting and useful corollary of this is

Corollary 2.3.16. *Suppose* $\overrightarrow{f}:\mathbb{R}\to\mathbb{R}^3$ *is differentiable, and let*

$$\rho(t) := \left\|\overrightarrow{f}(t)\right\|.$$

Then $\rho^2(t)$ *is differentiable, and*

1. $\dfrac{d}{dt}\left[\rho^2(t)\right] = 2\overrightarrow{f}(t)\cdot\vec{f}'(t)$.

2. *$\rho(t)$ is constant precisely if $\overrightarrow{f}(t)$ is always perpendicular to its derivative.*

3. *If $\rho(t_0) \neq 0$, then $\rho(t)$ is differentiable at $t = t_0$, and $\rho'(t_0)$ equals the component of $\vec{f}'(t_0)$ in the direction of $\overrightarrow{f}(t_0)$:*

$$\rho'(t_0) = \frac{\vec{f}'(t_0) \cdot \overrightarrow{f}(t_0)}{\left\|\overrightarrow{f}(t_0)\right\|}. \tag{2.23}$$

Proof. Since $\rho^2(t) = \overrightarrow{f}(t) \cdot \overrightarrow{f}(t)$, the first statement is a special case of the Product Rule for dot products.

To see the second statement, note that $\rho(t)$ is constant precisely if $\rho^2(t)$ is constant, and this occurs precisely if the right-hand product in the first statement is zero.

Finally, the third statement follows from the Chain Rule applied to

$$\rho(t) = \sqrt{\rho^2(t)}$$

so that (using the first statement)

$$\begin{aligned}\rho'(t_0) &= \frac{1}{2\rho(t_0)} \left.\frac{d}{dt}\right|_{t=t_0} \left[\rho^2(t)\right] \\ &= \frac{\overrightarrow{f}(t_0)}{\rho(t_0)} \cdot \vec{f}'(t_0) \\ &= \overrightarrow{u} \cdot \vec{f}'(t_0)\end{aligned}$$

where

$$\overrightarrow{u} = \frac{\overrightarrow{f}(t_0)}{\left\|\overrightarrow{f}(t_0)\right\|}$$

is the unit vector in the direction of $\overrightarrow{f}(t_0)$. □

Integration of Vector-Valued Functions

Integration also extends to vector-valued functions componentwise. Given $\overrightarrow{f}\colon [a,b] \to \mathbb{R}^3$ and a partition $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}$ of $[a,b]$, we can't form upper or lower sums, since the "sup" and "inf" of $\overrightarrow{f}(t)$ over I_j don't make sense. However we *can* form (vector-valued) Riemann sums

$$\mathcal{R}(\mathcal{P}, \overrightarrow{f}, \{t_j^*\}) = \sum_{j=1}^{n} \overrightarrow{f}(t_j^*) \bigtriangleup t_j$$

and ask what happens to these Riemann sums for a sequence of partitions whose mesh size goes to zero. If all such sequences have a common (vector) limit, we call it the **definite integral** of $\overrightarrow{f}(t)$ over $[a,b]$. It is natural (and straightforward to verify, using Lemma 2.3.5) that this happens precisely if each of the component functions $f_i(t)$, $i = 1, 2, 3$ is integrable over $[a,b]$, and then

$$\int_a^b \overrightarrow{f}(t)\; dt = \left(\int_a^b f_1(t)\; dt, \int_a^b f_2(t)\; dt, \int_a^b f_3(t)\; dt \right).$$

A direct consequence of this and the Fundamental Theorem of Calculus is that the integral of (vector) velocity is the net (vector) displacement:

Lemma 2.3.17. *If $\overrightarrow{v}(t) = \dot{\overrightarrow{p}}(t)$ is continuous on $[a,b]$, then*

$$\int_a^b \overrightarrow{v}(t)\; dt = \overrightarrow{p}(b) - \overrightarrow{p}(a)\,.$$

The proof of this is outlined in Exercise 10.

In Appendix C, we discuss the way in which the calculus of vector-valued functions can be used to reproduce Newton's derivation of the Law of Universal Gravitation from Kepler's Laws of Planetary Motion.

Exercises for § 2.3

Practice problems:

1. For each sequence $\{\overrightarrow{p}_n\}$ below, find the limit, or show that none exists.

   (a) $\left(\frac{1}{n}, \frac{n}{n+1}\right)$

   (b) $(\cos(\frac{\pi}{n}), \sin(\frac{n\pi}{n+1}))$

   (c) $(\sin(\frac{1}{n}), \cos(n))$

   (d) $(e^{-n}, n^{1/n})$

   (e) $\left(\frac{n}{n+1}, \frac{n}{2n+1}, \frac{2n}{n+1}\right)$

(f) $\left(\frac{n}{n+1}, \frac{n}{n^2+1}, \frac{n^2}{n+1}\right)$

(g) $(\sin\frac{n\pi}{n+1}, \cos\frac{n\pi}{n+1}, \tan\frac{n\pi}{n+1})$

(h) $(\frac{1}{n}\ln n, \frac{1}{\sqrt{n^2+1}}, \frac{1}{n}\ln\sqrt{n^2+1})$

(i) $(x_1, y_1, z_1) = (1, 0, 0), \quad (x_{n+1}, y_{n+1}, z_{n+1}) = (y_n, z_n, 1 - \frac{x_n}{n})$

(j) $(x_1, y_1, z_1) = (1, 2, 3), \quad (x_{n+1}, y_{n+1}, z_{n+1}) = (x_n + \frac{1}{2}y_n, y_n + \frac{1}{2}z_n, \frac{1}{2}z_n)$

2. An **accumulation point** of a sequence $\{\overrightarrow{p}_i\}$ of points is any limit point of any subsequence. Find all the accumulation points of each sequence below.

(a) $\left(\frac{1}{n}, \frac{(-1)^n n}{n+1}\right)$

(b) $\left(\frac{n}{n+1}\cos n, \frac{n}{n+1}\sin n\right)$

(c) $\left(\frac{n}{n+1}, \frac{(-1)^n n}{2n+1}, (-1)^n\frac{2n}{n+1}\right)$

(d) $\left(\frac{n}{n+1}\cos\frac{n\pi}{2}, \frac{n}{n+1}\sin\frac{n\pi}{2}, \frac{2n}{n+1}\right)$

Theory problems:

3. Prove the **triangle inequality**

$$\operatorname{dist}(P, Q) \leq \operatorname{dist}(P, R) + \operatorname{dist}(R, Q)$$

(a) in $\mathbb{R}^2$;

(b) in $\mathbb{R}^3$.

(*Hint:* Replace each distance with its definition. Square both sides of the inequality and expand, cancelling terms that appear on both sides, and then rearrange so that the single square root is on on one side; then square again and move all terms to the same side of the equals sign (with zero on the other). Why is the given quantity non-negative?

You may find it useful to introduce some notation for differences of coordinates, for example

$$\begin{aligned}\triangle x_1 &= x_2 - x_1\\ \triangle x_2 &= x_3 - x_2;\end{aligned}$$

note that then

$$\triangle x_1 + \triangle x_2 = x_3 - x_1.$$

)

4. Show that if $\overrightarrow{p}_i \to L$ in $\mathbb{R}^3$, then $\{\overrightarrow{p}_i\}$ is bounded.

5. Suppose $\{\overrightarrow{p}_i\}$ is a sequence of points in $\mathbb{R}^3$ for which the distances between consecutive points form a convergent series:

$$\sum_0^\infty \operatorname{dist}(\overrightarrow{p}_i, \overrightarrow{p}_{i+1}) < \infty.$$

   (a) Show that the sequence $\{\overrightarrow{p}_i\}$ is bounded. (*Hint:* Use the triangle inequality)

   (b) Show that the sequence is **Cauchy**—that is, for every $\varepsilon > 0$ there exists N so that $i, j > N$ guarantees $\operatorname{dist}(\overrightarrow{p}_i, \overrightarrow{p}_j) < \varepsilon$. (*Hint:* see (*Calculus Deconstructed*, Exercise 2.5.9))

   (c) Show that the sequence is convergent.

6. This problem concerns some properties of accumulation points (Exercise 2).

   (a) Show that a sequence with at least two distinct accumulation points diverges.

   (b) Show that a *bounded* sequence has at least one accumulation point.

   (c) Give an example of a sequence with *no* accumulation points.

   (d) Show that a *bounded* sequence with *exactly one* accumulation point converges to that point.

7. Prove Proposition 2.3.12.

8. Prove the following parts of Theorem 2.3.15:

(a) $\frac{d}{dt}\left[\alpha\overrightarrow{f}(t)+\beta\overrightarrow{g}(t)\right]=\alpha\vec{f}'(t)+\beta\vec{g}'(t)$

(b) $\frac{d}{dt}\left[\overrightarrow{f}(t)\times\overrightarrow{g}(t)\right]=\vec{f}'(t)\times\overrightarrow{g}(t)+\overrightarrow{f}(t)\times\vec{g}'(t)$

9. Prove that the moment of velocity about the origin is constant if and only if the acceleration is radial (*i.e.,* parallel to the position vector).

10. Prove Lemma 2.3.17. (*Hint:* Look at each component separately.)

Challenge problem:

11. (David Bressoud) A missile travelling at constant speed is homing in on a target at the origin. Due to an error in its circuitry, it is consistently misdirected by a constant angle α. Find its path. Show that if $|\alpha|<\frac{\pi}{2}$ then it will eventually hit its target, taking $\frac{1}{\cos\alpha}$ times as long as if it were correctly aimed.

2.4 Regular Curves

What, exactly, is a "curve"?

In § 2.2 we saw two ways to answer this question: a static one, summarized by describing a curve in the plane as the locus of an equation, and a kinetic one, describing a curve as the image of a vector-valued function, visualized as the path of a moving point. We noted that the static approach can become unwieldy when we try to apply it to curves in space—these are more fruitfully approached kinematically. As we shall see, the kinematic formulation of curves in the plane as well as in space is a more natural setting for applying calculus to curves. However, there is a difference between a motion and a path; in particular, a given path can be traced out by many different motions. So part of our mission here is to understand which properties of a vector-valued function encode *geometric* properties of the path it traces out, which are themselves independent of the function we use to describe it.

This section is a bit philosophical and a bit technical: we seek to come up with a clearer idea of what we mean by a *curve*, based on the kinematic notion of a *parametrization*. At the same time, we will try to understand which vector-valued functions should qualify as parametrizations. We have left more than the usual number of proofs, which can be quite technical, to the exercises; at first reading, these can be skipped in order to get the overall thrust of the theory here.

Consider, for example, the vector-valued function $\overrightarrow{p}\colon\mathbb{R}\to\mathbb{R}^2$ defined by

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

which is to say $\overrightarrow{p}(t) = (\cos t, \cos t)$. As t increases from $t = 0$ to $t = \pi$, this function traces out the line segment from $(1, 1)$ to $(-1, -1)$. Once t exceeds π, the motion traces back across this segment, so in a sense the motion becomes redundant.

Arcs

We would like, ideally, to have a parametrization of a curve $\mathcal{C}$ which traces it out only once—that is, we want each point of the line segment to correspond to a *single* parameter value. A (vector-valued) function is **one-to-one**[13] if distinct inputs yield distinct outputs:

$$t \neq t' \Rightarrow \overrightarrow{p}(t) \neq \overrightarrow{p}(t')\,.$$

A *real*-valued continuous function $f(t)$, defined on an interval I, is one-to-one precisely if it is strictly monotone on I.[14] Note however, that the component functions of a *vector*-valued function need not be monotone: for a given pair $t \neq t'$ of parameter values, the corresponding function values are distinct *vectors* if *at least one* of the component functions disagrees at the two values—the others can agree there. For example, neither the cosine nor the sine is one-to-one on the interval $\left[0, \frac{3\pi}{2}\right]$, but for any pair of distinct angles in this interval, at least one of them is different, and so the vector-valued function $\overrightarrow{p}(t) = (\cos t, \sin t)$ gives a one-to-one vector-valued function on the interval $\left[0, \frac{3\pi}{2}\right]$.

As we shall see, not every curve can be parametrized by a one-to-one vector-valued function, but those that can are the "simplest" curves, which we shall call *arcs*:

Definition 2.4.1. *A (continuous)* ***arc*** *is the image*

$$\mathcal{C} = \{\overrightarrow{p}(t) \mid t \in I\}$$

of a continuous one-to-one vector-valued function $\overrightarrow{p}$ *on a closed interval* $I = [t_0, t_1]$.

[13] A synonym for one-to-one, derived from the French literature, is **injective**.

[14] This is a consequence of the Intermediate Value Theorem: Exercise 6.

Any continuous, one-to-one vector-valued function whose image is the arc $\mathcal{C}$ is a **parametrization** of $\mathcal{C}$.

An easy example of an arc is the graph of a function (Exercise 2):

Remark 2.4.2. *If $f(x)$ is a continuous (real-valued) function on a closed interval I, then its graph—the locus of the equation $y = f(x)$, $x \in I$—is an arc in the plane.*

We can picture an arc $\mathcal{C}$ as a "bent" (and/or "twisted") copy of the interval I over which it is parametrized. Any particular parametrization $\overrightarrow{p}$ of $\mathcal{C}$ can be thought of as a coordinate system on $\mathcal{C}$, which locates a particular point P by giving the corresponding parameter value $t \in I$ (*i.e.*, $P = \overrightarrow{p}(t)$). This coordinate system is continuous, in two senses. First, the continuity of $\overrightarrow{p}$ guarantees that a convergent sequence of coordinates $t_i \in I$ yields a convergent sequence of points $\overrightarrow{p}(t_i) \in \mathcal{C}$. However, for an arc the converse also holds (Exercise 3):

Remark 2.4.3. *If a sequence of points on an arc converges, then the corresponding parameter values converge (to the parameter value giving the limit point).*

A consequence of this (Exercise 7) is that any two parametrizations $\overrightarrow{p}(t)$, $t \in I$ and $\overrightarrow{q}(s)$, $s \in J$ of the same arc $\mathcal{C}$ differ by a substitution: that is, the equation

$$\overrightarrow{p}(t) = \overrightarrow{q}(s)$$

can be solved for t in terms of s:

$$t = \mathfrak{t}(s)$$

As an example, we compare the two parametrizations we had of the upper semicircle of radius 1 centered at the origin in the plane

$$x^2 + y^2 = 1, \quad y \geq 0.$$

One parametrization came from the fact that this is the graph of a function

$$y = \sqrt{1 - x^2}, \quad -1 \leq x \leq 1$$

or

$$\overrightarrow{p}(x) = (x, \sqrt{1 - x^2}), \quad x \in [-1, 1]$$

while the other came from polar coordinates

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

with $0 \leq \theta \leq \pi$, or

$$\overrightarrow{q}(\theta) = (\cos\theta, \sin\theta), \quad \theta \in [0, \pi]\,.$$

In this case, the first equation in the definition of $\overrightarrow{q}$ already gives us $\mathfrak{t}(s)$, when $s = \theta$ and $t = x$

$$\mathfrak{t}(\theta) = \cos\theta$$

(which is strictly decreasing on $[0, \pi]$), while interchanging the role of $\overrightarrow{p}$ and $\overrightarrow{q}$ gives

$$\mathfrak{t}(x) = \arccos x.$$

We will refer to the switch from one parametrization of an arc $\mathcal{C}$ to another as a **reparametrization** of the arc $\mathcal{C}$, and to the function $\mathfrak{t}(s)$ as the **reparametrization function** associated to the reparametrization.

Notice that a parametrization of an arc $\mathcal{C}$ determines $\mathcal{C}$ as a set of points, but it also determines a *direction* along $\mathcal{C}$. For example, the first parametrization of the semicircle has points moving *clockwise* (as x increases) while the second has them moving *counterclockwise* (as θ increases). The fact that these two parametrizations lead to opposite directions is reflected in the fact that the reparametrization function $\mathfrak{t}$ is decreasing: as s increases, $t = \mathfrak{t}(s)$ decreases. We refer to this as a **direction-reversing** reparametrization. When $\mathfrak{t}$ is increasing, the reparametrization is **direction-preserving**. A direction-preserving reparametrization results in a motion which is speeded up or slowed down relative to the original parametrization.

Regular Curves

As noted earlier, not every curve can be parametrized by a continuous, one-to-one vector-valued function on a closed interval. For example, any traversal of the circle $x^2 + y^2 = 1$ in the plane must eventually come back to its starting position: to avoid hitting the same point twice (and so contradicting the one-to-one condition) we need to stop before reaching it, but stopping short of it means we haven't filled out the whole circle:[15] our only choice is to stop *exactly* when we reach the starting point, but excluding it, which means our domain is a *half-open* interval. So we need to give up at least one of the two conditions defining an arc: either we must stop expecting the parametrization to be one-to-one, or we must stop expecting to define it on a closed interval.

In fact, we will abandon both requirements, but at the same time by introducing derivatives into the picture, we will salvage some features of the definition of an arc.

So far, we have only required that our vector-valued function be continuous, but certainly if we want to use calculus we need some differentiability as well. We will say that a vector-valued function $\overrightarrow{p}$ is **continuously differentiable** if each coordinate function has a continuous derivative, or equivalently, if the velocity $\overrightarrow{v}(t) = \vec{p}'(t)$ exists for each parameter value t and defines a continuous vector-valued function.

While we can't expect a parametrization to be one-to-one, we also want to avoid the "backtracking" behavior we saw in the example at the start of the section: we want, intuitively, to continue tracing a curve in the same direction. Now, if we are moving along a one-dimensional object and suddenly want to reverse direction, then (if the velocity is continuous) we must stop, at least momentarily: the velocity can only reverse direction by passing through the zero vector. Thus, if we assume that the speed $\|\overrightarrow{v}(t)\|$ is never zero (or equivalently, that $\overrightarrow{v}(t) \neq \overrightarrow{0}$ for all t) then we avoid backtracking: we take this as a definition.

Definition 2.4.4. *A vector-valued function $\overrightarrow{p}$ is **regular** if it is continuously differentiable, and its speed is always positive—or equivalently its velocity is never the zero vector.*

*A **regular curve** is the image*

$$\mathcal{C} = \{\overrightarrow{p}(t) \mid t \in I\}$$

*of a regular vector-valued function $\overrightarrow{p}$ on an interval I; $\overrightarrow{p}$ is a **parametrization** of $\mathcal{C}$.*

The lack of backtracking by the parametrization of a regular curve is encoded in the following observation (Exercise 4):

[15] In other words, the corresponding vector-valued function does not map its domain **onto** the whole circle; in the corresponding French-inspred terminology, the function is not **surjective**.

Remark 2.4.5. *If $\overrightarrow{p}$ is a regular vector-valued function on the interval I, then it is* ***locally one-to-one****: for each $t \in I$, there is some positive distance $\varepsilon > 0$ such that any two distinct parameter values $t_1 \neq t_2 \in I$, both with $\|t_i - t\| < \varepsilon$ ($i = 1, 2$), yield distinct points of $\mathcal{C}$ ($\overrightarrow{p}(t_1) \neq \overrightarrow{p}(t_2)$).*

This says that a regular curve $\mathcal{C}$ is made up of arcs, in the sense that every point of $\mathcal{C}$ is contained in some arc which is, in turn, contained in $\mathcal{C}$. It can be shown (Exercise 8) that a regular curve can always be partitioned into arcs that abut at their endpoints.

Tangent Lines

Recall that the line tangent at a point P to the graph of a differentiable function $f(x)$ is defined geometrically as the limiting position of the secant line—the line through $P = (x_0, f(x_0))$ and a nearby point $P' = (x_0 + \triangle x, f(x_0 + \triangle x))$—as P' goes to P. We note that differentiability of a vector-valued function does not guarantee that its image has a tangent line in this sense: for example, the graph of $f(x) = |x|$, which has a "corner" at $x = 0$, is the image of the vector-valued function $\overrightarrow{p}(t) = (t^3, t^2\,|t|)$, which is continuously differentiable at $t = 0$. However, when the velocity is nonvanishing we avoid this phenomenon.

Although the "limiting position" of a line in the plane can be described in terms of its slope, for a curve in space we need to make a vector formulation: a direction in space can be specified by any nonzero vector, but to standardize matters we adopt the convention that it is specified by a *unit* vector $\overrightarrow{u}$. Given two points $P = \overrightarrow{p}(t_0)$ and $P' = \overrightarrow{p}(t_0 + \triangle t)$, the unit vector pointing from P to P' (which gives the direction of the secant line PP') is

$$\overrightarrow{u}(t) = \frac{1}{\left\|\overrightarrow{PP'}\right\|}\overrightarrow{PP'}$$

or

$$\overrightarrow{u}(t) = \frac{\overrightarrow{p}(t_0 + \triangle t) - \overrightarrow{p}(t_0)}{\|\overrightarrow{p}(t_0 + \triangle t) - \overrightarrow{p}(t_0)\|}.$$

Then we can define the **unit tangent vector** to $\mathcal{C}$ at $P = \overrightarrow{p}(t_0)$ as the one-sided limit[16]

$$\overrightarrow{T}(\overrightarrow{p}(t_0)) = \lim_{\triangle t \to 0^+} \overrightarrow{u}(t)\,. \tag{2.24}$$

[16] As we see (Exercise 5b), the limit from the other side is the negative of this one.

A reasonably easy calculation (Exercise 5a) shows that this limit exists and is given by

$$\overrightarrow{T}(\overrightarrow{p}(t_0)) = \frac{1}{\|\vec{p}'(t_0)\|}\vec{p}'(t_0)\,. \tag{2.25}$$

This definition is in terms of a specific parametrization of $\mathcal{C}$. When $\mathcal{C}$ is a regular *arc* (that is, the parametrization is one-to-one on a closed interval), then we can use the fact that, given two parametrizations $\overrightarrow{p}(t)$ and $\overrightarrow{q}(s)$ of $\mathcal{C}$, $\overrightarrow{u}(t)$ can be described using either parametrization, and these descriptions are related by the reparametrization function $\mathfrak{t}$, to show (Exercise 9)

Proposition 2.4.6. *If $\overrightarrow{p}(t)$ and $\overrightarrow{q}(s)$ are both regular parametrizations of the arc $\mathcal{C}$, with reparametrization function $t = \mathfrak{t}(s)$, then $\mathfrak{t}$ is differentiable, and the unit tangent vector at each point of $\mathcal{C}$ as calculated from $\overrightarrow{p}$ and $\overrightarrow{q}$ is the same up to a scalar factor of ± 1. More precisely:*

1. *If the reparametrization is direction-preserving, then*

$$\overrightarrow{T}(\overrightarrow{p}(t_0)) = \overrightarrow{T}(\overrightarrow{q}(s_0)) \tag{2.26}$$

and

$$\mathfrak{t}'(s) = \frac{dt}{ds} = \frac{\|\vec{q}'(s)\|}{\|\vec{p}'(t)\|}, \tag{2.27}$$

while

2. *If the reparametrization is direction-reversing, then*

$$\overrightarrow{T}(\overrightarrow{p}(t_0)) = -\overrightarrow{T}(\overrightarrow{q}(s_0)) \tag{2.28}$$

and

$$\mathfrak{t}'(s) = \frac{dt}{ds} = -\frac{\|\vec{q}'(s)\|}{\|\vec{p}'(t)\|}. \tag{2.29}$$

In view of Proposition 2.4.6, the **tangent line** at any point of an arc $\mathcal{C}$ can be defined as the line through the point with direction vector the unit tangent $\overrightarrow{T}$, calculated from *any* regular parametrization.

We note that the tangent line can be calculated from any regular parametrization of $\mathcal{C}$ by using the velocity vector in place of the unit tangent vector as a direction vector: that is, if $\overrightarrow{p}(t)$ is a regular vector-valued function, then the tangent line at $\overrightarrow{p}(t_0)$ is the line with parametric equation

$$\overrightarrow{\ell}(t) = \overrightarrow{p}(t_0) + t \cdot \dot{\overrightarrow{p}}(t_0)\,. \tag{2.30}$$

If we examine the component functions of this parametrization

$$\begin{aligned}\ell_1'(t) &= x(t_0) + tx'(t_0)\\ \ell_2'(t) &= y(t_0) + ty'(t_0)\\ \ell_3'(t) &= z(t_0) + tz'(t_0)\end{aligned}$$

we note that each is a polynomial of degree one whose slope is the derivative of the corresponding component of $\overrightarrow{p}(t)$ and whose value at $t = 0$ agrees with the value of the appropriate component of $\overrightarrow{p}(t)$ at $t = t_0$. We can shift our parameter along the line to put it "in sync" with $\overrightarrow{p}(t)$, so that it also hits the point $\overrightarrow{p}(t_0)$ at time $t = t_0$, by replacing t in the formulas above with $t - t_0$. We then see that the shifted coordinate functions agree with the first-order Taylor polynomials for the components of $\overrightarrow{p}(t)$ at $t = t_0$:

$$\begin{aligned}\ell_1'(t - t_0) &= x(t_0) + x'(t_0)(t - t_0) = T_{t_0}x(t)\\ \ell_2'(t - t_0) &= y(t_0) + y'(t_0)(t - t_0) = T_{t_0}y(t)\\ \ell_3'(t - t_0) &= z(t_0) + z'(t_0)(t - t_0) = T_{t_0}z(t)\end{aligned}$$

and so we can use the notation for the parametrization above

$$\overrightarrow{\ell}(t - t_0) = T_{t_0}\overrightarrow{p}(t).$$

In kinematic terms, this represents how a point moving according to $\overrightarrow{p}(t)$ for $t < t_0$ would move for $t \geq t_0$ (according to Newton's First Law of Motion) if any forces (which kept the point moving as described for $t < t_0$) were "turned off "at $t = t_0$.

As an example, consider the helix, which is an arc,

$$\overrightarrow{p}(t) = (\cos 2\pi t, \sin 2\pi t, t)$$

with velocity

$$\overrightarrow{v}(t) = \dot{\overrightarrow{p}}(t) = (-2\pi \sin 2\pi t, 2\pi \cos 2\pi t, 1)$$

and speed

$$\frac{ds}{dt} = \sqrt{4\pi^2 + 1}.$$

Let us find the unit tangent vector and tangent line at $P(1, 0, 1) = \overrightarrow{p}(1)$: the velocity vector is

$$\overrightarrow{v}(1) = = (0, 2\pi, 1)$$

and the unit tangent vector is

$$\begin{aligned}\vec{T}(P) &= \frac{1}{\|\vec{v}(1)\|}\vec{v}(1)\\ &= \left(\frac{-2\pi}{\sqrt{4\pi^2+1}}, 0, \frac{1}{\sqrt{4\pi^2+1}}\right);\end{aligned}$$

the tangent line has parametrization

$$\begin{aligned}T_1\vec{p}(t) &= \vec{p}(1) + \vec{v}(1)\,(t-1)\\ &= (1,0,1) + \left(\frac{-2\pi}{\sqrt{4\pi^2+1}}, 0, \frac{1}{\sqrt{4\pi^2+1}}\right)(t-1)\end{aligned}$$

or

$$\begin{aligned}T_1x(t) &= 1 - \frac{2\pi(t-1)}{\sqrt{4\pi^2+1}}\\ T_1y(t) &= 0\\ T_1z(t) &= 1 + \frac{(t-1)}{\sqrt{4\pi^2+1}}.\end{aligned}$$

The “kinematic” parametrization $T_{t_0}\vec{p}(t)$ of the tangent line which we get by putting the tangent motion “in sync” with the motion given by the parametrization of our curve displays an important property, analogous to one enjoyed by the Taylor polynomial of degree one for a real-valued function $f(x)$. Recall that $T_{x_0}f = f(x_0) + f'(x_0)\,(x-x_0)$ is the unique polynomial of degree one which has **first-order contact** with the graph $y = f(x)$ at the point $(x_0, f(x_0))$:

$$\left|f(x) - T_{x_0}f(x)\right| = \mathfrak{o}(|x-x_0|)$$

which is to say

$$\frac{\left|f(x) - T_{x_0}f(x)\right|}{|x-x_0|} \to 0 \text{ as } x \to x_0.$$

In particular, this applies to each of the component functions of our parametrization, and from this it is easy to see that in fact the vector-valued function $T_{t_0}\vec{p}(t)$ has first-order contact with $\vec{p}(t)$ at $t = t_0$:

$$\frac{\left\|\vec{p}(t) - T_{t_0}\vec{p}(t)\right\|}{|t-t_0|} \to 0 \text{ as } t \to t_0$$

or, in "little oh" notation,

$$\left\|\overrightarrow{p}(t) - T_{t_0}\overrightarrow{p}(t)\right\| = \mathfrak{o}(|t - t_0|).$$

This approximation property of the derivative is the key to its usefulness in analyzing curves and functions.

While a regular *arc* has a well-defined tangent line at each point,[17] a regular curve can have several different tangent lines at points where it crosses itself: for example in a figure-eight or the "rose" in § 2.2.

There is a second ambiguity that can arise as a result of self-crossings. If we get away momentarily from regular parametrizations and consider only continuous, locally one-to-one parametrizations of a curve, then two different parametrizations might no longer be related by a continuous reparametrization function. For example, the parametrization of the "rose" in § 2.2 given by the polar equation $r = \cos 2\theta$ goes around the "petals" in a certain order and direction; it is easy geometrically (and a little harder analytically) to imagine other parametrizations in which the arcs are gone around in a different order and/or in different directions, for example the parametrization sketched in Figure 2.25, in which the directions are the same, but the order of traversal is changed (think of the rose as a circle, for which the four points on the axes have been "pinched" into the origin). The point set is the same, but the traversal is radically different.

The "pinched circle" traversal cannot be achieved by a regular parametrization, because it ignores the two tangent lines (the coordinate axes) as it goes through the origin. However, using the numbering in Figure 2.25, it is easy to see that the petal numbered *i* (traversed counterclockwise) can be followed, in a regular parametrization, by either *iv* traversed counterclockwise (as it is in the standard parametrization given by $r = \sin 2\theta$), or by *iii* traversed clockwise, and similarly *ii* (counterclockwise) can be followed by either *i* (counterclockwise) or *iv* (clockwise). In all of these cases, the new parametrization changes the important property we have of deciding which points are "between" certain others. Thus, in dealing with general regular curves, we need to tie a general curve $\mathcal{C}$, regarded as a point set, to a limited class of possible parametrizations: we take the existence of a reparametrization function $\mathfrak{t}$ as a definition.

Definition 2.4.7. *Given a curve $\mathcal{C}$ parametrized by the regular function $\overrightarrow{p}$, a* ***reparametrization*** *is a vector-valued function $\overrightarrow{q}$ which can be expressed*

[17] The one-sided limit defining the unit tangent vector at the ending point of an arc is not well defined, but we can use a direction-reversing reparametrization to define its negative.

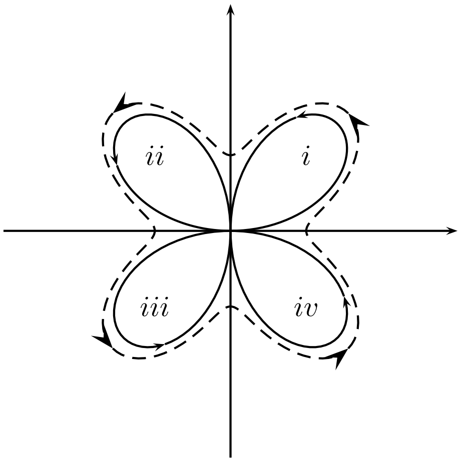


Figure 2.25: Four-petal Rose as a Pinched Circle

in terms of $\overrightarrow{p}(t)$ as

$$\overrightarrow{q}(s) = \overrightarrow{p}(\mathfrak{t}(s))$$

*where $\mathfrak{t}(s)$ is a continuously differentiable function from the domain of $\overrightarrow{q}$ onto that of $\overrightarrow{p}$. We say the reparametrization is **direction-preserving** (resp. **direction-reversing**) if $\mathfrak{t}$ is an increasing (resp. decreasing) function.*

*We shall refer to $\mathcal{C}$ as a **parametrized curve** when it is associated with a given parametrization (or a given family of possible parametrizations related as above).*

Note that the function $\mathfrak{r}(t) = (a+b) - t$ is a strictly decreasing function from $[a,b]$ to itself that interchanges the endpoints; composing this with a parametrization $\overrightarrow{p}$ of a curve with domain $[a,b]$ is a direction-reversing reparametrization.

In Appendix D, we explore further the way that the calculus of vector-valued functions gives a way to study and classify curves in the plane and in space.

Exercises for § 2.4

Practice problems:

1. For each pair of vector-valued functions $\overrightarrow{p}(t)$ and $\overrightarrow{q}(t)$ below, find a recalibration function $\mathfrak{t}(s)$ so that $\overrightarrow{q}(s)=\overrightarrow{p}(\mathfrak{t}(s))$ and another so that $\overrightarrow{p}(s)=\overrightarrow{q}(\mathfrak{t}(s))$:

(a)

$$\begin{aligned}\overrightarrow{p}(t) &= (t,t) \quad -1 \leq t \leq 1\\ \overrightarrow{q}(t) &= (\cos t, \cos t) \quad 0 \leq t \leq \pi\end{aligned}$$

(b)

$$\begin{aligned}\overrightarrow{p}(t) &= (t,e^{t}) \quad -\infty < t < \infty\\ \overrightarrow{q}(t) &= (\ln t, t) \quad 0 < t < \infty\end{aligned}$$

(c)

$$\begin{aligned}\overrightarrow{p}(t) &= (\cos t, \sin t) \quad -\infty < t < \infty\\ \overrightarrow{q}(t) &= (\sin t, \cos t) \quad -\infty < t < \infty\end{aligned}$$

(d)

$$\begin{aligned}\overrightarrow{p}(t) &= (\cos t, \sin t, \sin 2t) \quad -\frac{\pi}{2} \leq t \leq \frac{\pi}{2}\\ \overrightarrow{q}(t) &= (\sqrt{1-t^2}, t, t\sqrt{4-4t^2}) \quad -1 \leq t \leq 1\end{aligned}$$

Theory problems:

2. Prove Remark 2.4.2. (*Hint:* Use the input to the function as the parameter.)

3. (a) Suppose $\overrightarrow{p}: I \to \mathbb{R}^3$ is a continuous, one-to-one vector-valued function on the closed interval $I = [a,b]$ with image the arc $\mathcal{C}$, and suppose $\overrightarrow{p}(t_i) \to \overrightarrow{p}(t_0) \in \mathcal{C}$. *Show* that $t_i \to t_0$ in I, thus proving Remark 2.4.3. (*Hint:* Show that the sequence t_i must have at lest one accumulation point t_* in I, and that for *every* such accumulation point t_*, we must have $\overrightarrow{p}(t_*) = \overrightarrow{p}(t_0)$. Then use the fact that $\overrightarrow{p}$ is one-to-one to conclude that the *only* accumulation point of t_i is $t_* = t_0$. But a bounded sequence with exactly one accumulation point must converge to that point.)

   (b) Show that this property fails for the parametrization of the circle by $\overrightarrow{p}(\theta) = (\cos\theta, \sin\theta)$, $0 \leq \theta < 2\pi$.

4. Prove Remark 2.4.5. (*Hint:* If $\overrightarrow{p}(t) \neq \overrightarrow{0}$, then some coordinate is strictly increasing on a nontrivial interval containing t, and this guarantees that $\overrightarrow{p}$ is one-to-one on that interval.)

5. (a) *Show* that if $\overrightarrow{p}$ is a regular vector-valued function, the limit in Equation (2.24) exists and is given by Equation (2.25). (*Hint:* Write

$$\begin{aligned}\overrightarrow{u}(t) &= \frac{\overrightarrow{p}(t_0 + \triangle t) - \overrightarrow{p}(t_0)}{\|\overrightarrow{p}(t_0 + \triangle t) - \overrightarrow{p}(t_0)\|} \\ &= \frac{\overrightarrow{p}(t_0 + \triangle t) - \overrightarrow{p}(t_0)}{\triangle t} \bigg/ \frac{\|\overrightarrow{p}(t_0 + \triangle t) - \overrightarrow{p}(t_0)\|}{\triangle t} \\ &= \frac{\overrightarrow{p}(t_0 + \triangle t) - \overrightarrow{p}(t_0)}{\triangle t} \bigg/ \left\| \frac{\overrightarrow{p}(t_0 + \triangle t) - \overrightarrow{p}(t_0)}{\triangle t} \right\| .\end{aligned}$$

)

(b) *Show* that

$$\lim_{\triangle t \to 0^-} \overrightarrow{u}(t) = - \lim_{\triangle t \to 0^+} \overrightarrow{u}(t) .$$

Challenge problems:

6. Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is continuous on $\mathbb{R}$. *Show*:

   (a) f is one-to-one if and only if it is strictly monotone.

   (*Hint:* One direction is trivial. For the other direction, use the Intermediate Value Theorem: what does it mean to *not* be monotone?)

   (b) If f is *locally* one-to-one, then it is *globally* one-to-one.

   (c) Give an example of a function $f(x)$ which is one-to-one on $[-1, 1]$ but is *not* strictly monotone on $[-1, 1]$.

7. Suppose $\overrightarrow{p}(t)$, $t \in I$ and $\overrightarrow{q}(s)$, $s \in J$ are both continuous, one-to-one vector-valued functions parametrizing the same arc $\mathcal{C}$. *Show:*

   (a) For each $s \in J$, the equation

$$\overrightarrow{p}(t) = \overrightarrow{q}(s)$$

   has a unique solution $t = \mathfrak{t}(s)$, and so defines a function $\mathfrak{t}\colon J \to I$;

   (b) The function $\mathfrak{t}$ is continuous and strictly monotone. (*Hint:* Use Exercise 3 for continuity, and show that the function $\mathfrak{t}$ is one-to-one.)

8. Suppose $\overrightarrow{p}(t)$, $t \in I$ is a regular vector-valued function defined on the interval I. In this exercise, we show that I can be partitioned into closed intervals on each of which $\overrightarrow{p}$ is one-to-one.

   First we show that if $I = [a, b]$ is a closed interval then there exist finitely many points

   $$a = t_0 < t_1 < \cdots < t_n = b$$

   such that the restriction of $\overrightarrow{p}$ to each of the intervals $[t_{i-1}, t_i]$ is one-to-one:

   (a) Assume that $\overrightarrow{p}$ isn't already one-to-one. Then for each $t \in I$ there exists $\varepsilon(t) > 0$ such that $\overrightarrow{p}$ is one-to-one on the interval $(t - \varepsilon(t)\,, t + \varepsilon(t))$. *Show* that if we pick $\varepsilon(t)$ to be the largest such value, then there exist $t' \neq t'' \in I$ with $|t - t'| \leq \varepsilon(t)$ and $|t - t''| \leq \varepsilon(t)$ (with at least one equality) such that $\overrightarrow{p}(t') = \overrightarrow{p}(t''))$.

   (b) Now consider $\varepsilon = \inf_{t \in I} \varepsilon(t)$. If $\varepsilon > 0$ then *Show* that any partition for which $|t_i - t_{i-1}| < \varepsilon$ for each i will work.

   (c) Prove by contradiction that $\varepsilon > 0$: if not, there exist pairs $t'_i \neq t''_i$ with $|t'_i - t''_i| < \frac{1}{i}$ $\overrightarrow{p}(t'_i) = \overrightarrow{p}(t''_i)$. By going to a subsequence, assume $t'_i \to t_0$. *Show* that also $t''_i \to t_0$. But then $\overrightarrow{p}$ is not locally one-to-one at t_0.

   (d) Now, let I be any interval (open, closed, or half-open). Show that there is a bisequence of points t_i, $i \in \mathbb{Z}$ such that $t_{i-1} < t_i$ and $\overrightarrow{p}$ is one-to-one on $[t_{i-1}, t_i]$ (*Hint:* By the preceding, we can partition any closed bounded subinterval of I into finitely many arcs. Use this to show that all of I can be partitioned into a possibly infinite collection of abutting arcs.)

9. Prove Proposition 2.4.6 as follows:

   (a) If the reparametrization is direction-preserving, then Equation (2.26) follows from the fact that if $s_i \to s_0^+$ then

   $$t_i = \mathfrak{t}(s_i) \to \mathfrak{t}(s_0) = t_0^+.$$

   (b) In the direction-reversing case, $s_i \to s_0^+$ yields

   $$t_i = \mathfrak{t}(s_i) \to \mathfrak{t}(s_0) = t_0^-$$

   and this in combination with Exercise 5b gives Equation (2.28).

(c) To prove Equation (2.27) and Equation (2.29), it is enough to show that $\mathfrak{t}$ is differentiable. To this end, show that if the velocity is nonzero, then at least one of the coordinates of $\overrightarrow{p}$ has a nonvanishing derivative near $\overrightarrow{p}(t_0)$, and hence has a differentiable inverse; then express $\mathfrak{t}$ as a composition of differentiable functions. The required formulas then follow from the Chain Rule.

History note:

10. **Bolzano's curve:** A version of the following was constructed by Bernhard Bolzano (1781-1848) in the 1830's; a fuller study is given in (*Calculus Deconstructed*, §4.11).

(a) Start with the following: suppose we have an affine function f, defined over the interval $[a, b]$, with $f(a) = c$ and $f(b) = d$; thus its graph is the straight line segment from (a, c) to (b, d). Construct a new, piecewise-affine function $\bar{f}$ by keeping the endpoint values, but interchanging the values at the points one-third and two-thirds of the way across (see Figure 2.26).

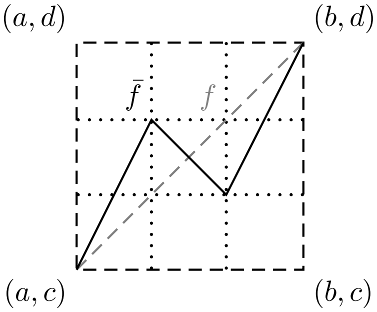


Figure 2.26: The basic construction

Thus, originally

$$f(x) = c + m(x - a)$$

where

$$m = \frac{d - c}{b - a},$$

and in particular

$$\begin{aligned} f(a) &= c \\ f(a_1) &= \frac{2a+b}{3} = \frac{2c+d}{3} \\ f(a_2) &= \frac{a+2b}{3} = \frac{c+2d}{3} \\ f(b) &= d. \end{aligned}$$

Now, $\bar{f}$ is defined by

$$\begin{aligned} \bar{f}(a) &= c \\ \bar{f}(a) &= c \\ \bar{f}(a_1) &= c_1 = \frac{c+2d}{3} \\ \bar{f}(a_2) &= c_2 = \frac{2c+d}{3} \\ \bar{f}(b) &= d \end{aligned}$$

and $\bar{f}$ is affine on each of the intervals $I_1 = [a, a_1]$, $I_2 = [a_1, a_2]$, and $I_3 = [a_2, b]$. **Show** that the slopes m_j of the graph of $\bar{f}$ on I_j satisfy

$$\begin{aligned} m_1 = m_3 &= 2m \\ m_2 &= -m. \end{aligned}$$

(b) Now, we construct a sequence of functions f_k on $[0,1]$ via the recursive definition

$$\begin{aligned} f_0 &= id \\ f_{k+1} &= \bar{f}_k. \end{aligned}$$

Show that

$$|f_k(x) - f_{k+1}(x)| \leq \left(\frac{2}{3}\right)^{k+1} \tag{2.31}$$

for all $x \in [0,1]$. This implies that for each $x \in [0,1]$,

$$f(x) := \lim f_k(x)$$

is well-defined for each $x \in [0,1]$. We shall accept without proof the fact that Equation (2.31) (which implies a property called *uniform convergence*) also guarantees that f is continuous on $[0,1]$. Thus its graph is a continuous curve—in fact, it is an arc.

(c) **Show** that if x_0 is a **triadic rational** (that is, it has the form $x_0 = \frac{p}{3^j}$ for some j) then $f_{k+1}(x_0) = f_k(x_0)$ for k sufficiently large, and hence this is the value $f(x_0)$. In particular, **show** that f has a local extremum at each triadic rational. (*Hint:* x_0 is a local extremum for all f_k once k is sufficiently large; furthermore, once this happens, the sign of the slope on either side does not change, and its absolute value is increasing with k.)

This shows that f has infinitely many local extrema—in fact, between any two points of $[0, 1]$ there is a local maximum (and a local minimum); in other words, the curve has infinitely many "corners". It can be shown (see (*Calculus Deconstructed*, §4.11)) that the function f, while it is continuous on $[0, 1]$, is not differentiable at any point of the interval. In Exercise 5 in § 2.5, we will also see that this curve has infinite "length".

2.5 Integration along Curves

Arclength

How long is a curve? While it is clear that the length of a straight line segment is the distance between its endpoints, a rigorous notion of the "length" for more general curves is not so easy to formulate. We will formulate a geometric notion of length for arcs, essentially a modernization of the method used by Archimedes of Syracuse (*ca.* 287-212 BC) to measure the circumference of a circle. Archimedes realized that the length of an inscribed (*resp.* circumscribed) polygon is a natural lower (*resp.* upper) bound on the circumference, and also that by using polygons with many sides, the difference between these two bounds could be made as small as possible. Via a proof by contradiction he proved in [2] that the area of a circle is the same as that of a triangle whose base equals the circumference and whose height equals the radius. By using regular polygons with 96 sides, he was able to establish that π, defined as the ratio of the circumference to the diameter, is between $\frac{22}{7}$ and $\frac{221}{71}$. Archimedes didn't worry about whether the length of the circumference makes sense; this was taken to be self-evident. His argument about the lengths of polygons providing bounds for the circumference was based on a set of axioms concerning *convex* curves; this was needed most for the use of the *circumscribed* polygons as an *upper* bound. The fact that *inscribed* polygons give a *lower* bound follows from the much simpler assumption, which *we* take as self-evident, that the shortest curve between two points is a straight line segment.

Suppose that $\mathcal{C}$ is an arc parametrized by $\overrightarrow{p}\colon\mathbb{R}\to\mathbb{R}^3$ and let $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}$ be a partition of the domain of $\overrightarrow{p}$. The sum

$$\ell\,(\mathcal{P}, \overrightarrow{p}) = \sum_{j=1}^{n} \|\overrightarrow{p}\,(t_j) - \overrightarrow{p}\,(t_{j-1})\|$$

is the sum of the straight-line distances between successive points along $\mathcal{C}$. It is clear that any reasonable notion of the "length" of $\mathcal{C}$ at least equals $\ell\,(\mathcal{P}, \overrightarrow{p})$. We would also think intuitively that a partition with small mesh size should give a good approximation to the "true" length of $\mathcal{C}$. We therefore say $\mathcal{C}$ is **rectifiable** if the values of $\ell\,(\mathcal{P}, \overrightarrow{p})$ among all partitions are bounded, and then we define the **arclength** of $\mathcal{C}$ to be

$$\mathfrak{s}\,(\mathcal{C}) = \sup_{\mathcal{P}} \ell\,(\mathcal{P}, \overrightarrow{p})\,.$$

Not every curve is rectifiable. Two examples of *non*-rectifiable curves are the the graph of Bolzano's nowhere-differentiable function, constructed in Exercise 10 in § 2.4 (see Exercise 5) and the graph of $y = x \sin \frac{1}{x}$ (see Exercise 4). In such cases, there exist partitions $\mathcal{P}$ for which $\ell\,(\mathcal{P}, \overrightarrow{p})$ is arbitrarily high.

We need to show that the arclength $\mathfrak{s}\,(\mathcal{C})$ does not depend on the parametrization we use to construct $\mathcal{C}$. Suppose $\overrightarrow{p}, \overrightarrow{q}\colon\mathbb{R}\to\mathbb{R}^3$ are two one-to-one continuous functions with the same image $\mathcal{C}$. As noted in § 2.4, we can find a strictly monotone recalibration function $\mathfrak{t}(s)$ from the domain of $\overrightarrow{q}$ to the domain of $\overrightarrow{p}$ so that

$$\overrightarrow{p}\,(\mathfrak{t}(s)) = \overrightarrow{q}\,(s)$$

for all parameter values of $\overrightarrow{q}$. If $\mathcal{P}$ is a partition of the domain of $\overrightarrow{p}$, then there is a unique sequence of parameter values for $\overrightarrow{q}$ defined by

$$\mathfrak{t}(s_j) = t_j;$$

this sequence is either strictly increasing (if $\mathfrak{t}(s)\uparrow$) or strictly decreasing (if $\mathfrak{t}(s)\downarrow$); renumbering if necessary in the latter case, we see that the s_j form a partition $\mathcal{P}_s$ of the domain of $\overrightarrow{q}$, with the same succession as the t_j; in particular,

$$\ell\,(\mathcal{P}_s, \overrightarrow{q}) = \ell\,(\mathcal{P}, \overrightarrow{p})$$

and so the supremum of $\ell\,(\mathcal{P}', \overrightarrow{q})$ over *all* partitions $\mathcal{P}'$ of the domain of $\overrightarrow{q}$ is at least the same as that over partitions of the domain of $\overrightarrow{p}$. Reversing the roles of the two parametrizations, we see that the two suprema are actually the same.

This formulation of arclength has the advantage of being clearly based on the geometry of the curve, rather than the parametrization we use to construct it. However, as a tool for computing the arclength, it is as useful (or useless) as the definition of the definite integral via Riemann sums is for calculating definite integrals. Fortunately, for regular curves, we can use definite integrals to calculate arclength.

Theorem 2.5.1. *Every regular arc is rectifiable, and if $\overrightarrow{p}\colon [a,b]\to\mathbb{R}^3$ is a regular one-to-one function with image $\mathcal{C}$, then*

$$\mathfrak{s}\left(\mathcal{C}\right) = \int_a^b \left\|\dot{\overrightarrow{p}}\left(t\right)\right\| \, dt.$$

This can be understood as saying that *the length of a regular curve is the integral of its speed*, which agrees with our understanding (for real-valued functions representing motion along an axis) that the integral of speed is the total distance traveled. If we consider the function $\mathfrak{s}\left(t\right)$ giving the arclength (or distance travelled) between the starting point and the point $\overrightarrow{p}\left(t\right)$, then our notation for the speed is naturally suggested by applying the Fundamental Theorem of Calculus to the formula above

$$\begin{aligned}\frac{d}{dt}\left[\mathfrak{s}\left(t\right)\right] &= \frac{d}{dt}\int_a^t \left\|\dot{\overrightarrow{p}}\left(t\right)\right\| \, dt \\ &= \frac{d}{dt}\left\|\dot{\overrightarrow{p}}\left(t\right)\right\|\end{aligned}$$

or

$$\frac{d\mathfrak{s}}{dt} = \left\|\dot{\overrightarrow{p}}\left(t\right)\right\|.$$

Before proving Theorem 2.5.1, we establish a technical estimate.

Lemma 2.5.2. *Suppose $\overrightarrow{p}\colon [a,b]\to\mathbb{R}^3$ is a regular, one-to-one function and that $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}$ is a partition of $[a,b]$ such that the speed varies by less than $\delta > 0$ over each component interval I_j:*

$$\left|\left\|\dot{\overrightarrow{p}}\left(t\right)\right\| - \left\|\dot{\overrightarrow{p}}\left(t'\right)\right\|\right| < \delta \text{ whenever } t_{j-1} \le t, t' \le t_j.$$

Then

$$\left|\int_a^b \left\|\dot{\overrightarrow{p}}\left(t\right)\right\| \, dt - \ell\left(\mathcal{P}, \overrightarrow{p}\right)\right| < 3\delta(b-a).$$

Proof. For the moment, let us fix a component interval $I_j = [t_{j-1}, t_j]$ of $\mathcal{P}$. Applying the Mean Value Theorem to each of the component functions, there exist parameter values s_i, $i = 1, 2, 3$ such that

$$\begin{aligned} x(t_j) - x(t_{j-1}) &= \dot{x}(s_1)\delta_j \\ y(t_j) - y(t_{j-1}) &= \dot{y}(s_2)\delta_j \\ z(t_j) - z(t_{j-1}) &= \dot{z}(s_3)\delta_j. \end{aligned}$$

Then the vector

$$\vec{v}_j = (\dot{x}(s_1), \dot{y}(s_2), \dot{z}(s_3))$$

satisfies

$$\vec{p}(t_j) - \vec{p}(t_{j-1}) = \vec{v}_j \delta_j$$

and hence

$$\|\vec{p}(t_j) - \vec{p}(t_{j-1})\| = \|\vec{v}_j\| \, \delta_j.$$

But also, for any $t \in I_j$,

$$\begin{aligned} \left\| \dot{\vec{p}}(t) - \vec{v}_j \right\| &\leq |\dot{x}(t) - \dot{x}(s_1)| + |\dot{y}(t) - \dot{y}(s_2)| + |\dot{z}(t) - \dot{z}(s_3)| \\ &< 3\delta \end{aligned}$$

Now, an easy application of the Triangle Inequality says that the lengths of two vectors differ by at most the length of their difference; using the above this gives us

$$\left| \left\| \dot{\vec{p}}(t) \right\| - \|\vec{v}_j\| \right| < 3\delta \text{ for all } t \in I_j.$$

Picking $\vec{v}_j$, $j = 1, \ldots, n$ as above, we get

$$\begin{aligned} \left| \int_a^b \left\| \dot{\vec{p}}(t) \right\| \, dt - \ell(\mathcal{P}, \vec{p}) \right| &= \left| \sum_{j=1}^n \left(\int_{I_j} \left\| \dot{\vec{p}}(t) \right\| \, dt \right) - \|\vec{v}_j\| \, \delta_j \right| \\ &\leq \sum_{j=1}^n \int_{I_j} \left| \left\| \dot{\vec{p}}(t) \right\| - \|\vec{v}_j\| \right| \, dt \\ &< 3\delta \sum_{j=1}^n \delta_j \\ &= 3\delta(b - a). \end{aligned}$$

□

Proof of Theorem 2.5.1. We will use the fact that since the speed is continuous on the closed interval $[a, b]$, it is uniformly continuous, which means that given any $\delta > 0$, we can find $\mu > 0$ so that it varies by at most δ over any subinterval of $[a, b]$ of length μ or less. Put differently, this says that the hypotheses of Lemma 2.5.2 are satisfied by any partition of mesh size μ or less. We will also use the easy observation that refining the partition raises (or at least does not lower) the "length estimate" $\ell(\mathcal{P}, \overrightarrow{p})$ associated to the partition.

Suppose now that $\mathcal{P}_k$ is a sequence of partitions of $[a, b]$ for which $\ell_k = \ell(\mathcal{P}_k, \overrightarrow{p}) \uparrow \mathfrak{s}(\mathcal{C})$ (which, *a priori* may be infinite). Without loss of generality, we can assume (refining each partition if necessary) that the mesh size of $\mathcal{P}_k$ goes to zero monotonically. Given $\varepsilon > 0$, we set

$$\delta = \frac{\varepsilon}{3(b-a)}$$

and find $\mu > 0$ such that every partition with mesh size μ or less satisfies the hypotheses of Lemma 2.5.2; eventually, $\mathcal{P}_k$ satisfies $\operatorname{mesh}(\mathcal{P}_k) < \mu$, so

$$\left| \int_a^b \left\| \dot{\overrightarrow{p}}(t) \right\| \, dt - \ell(\mathcal{P}_k, \overrightarrow{p}) \right| < 3\delta(b-a) = \varepsilon.$$

This shows first that the numbers ℓ_k converge to $\int_a^b \left\| \dot{\overrightarrow{p}}(t) \right\| dt$—but by assumption, $\lim \ell_k = \mathfrak{s}(\mathcal{C})$, so we are done. □

The content of Theorem 2.5.1 is encoded in a notational device: given a regular parametrization $\overrightarrow{p}: \mathbb{R} \to \mathbb{R}^3$ of the curve $\mathcal{C}$, we define the **differential of arclength**, denoted $\boldsymbol{d\mathfrak{s}}$, to be the formal expression

$$d\mathfrak{s} := \left\| \dot{\overrightarrow{p}}(t) \right\| \, dt = \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2 + \dot{z}(t)^2} \, dt.$$

This may seem a bit mysterious at first, but we will find it very useful; using this, the content of Theorem 2.5.1 can be written

$$\mathfrak{s}(\mathcal{C}) = \int_a^b d\mathfrak{s}.$$

As an example, let us use this formalism to find the length of the helix parametrized by

$$\begin{aligned} x(t) &= \cos 2\pi t \\ y(t) &= \sin 2\pi t \\ z(t) &= t \end{aligned}$$

or

$$\overrightarrow{p}(t) = (\cos 2\pi t, \sin 2\pi t, t) \quad 0 \leq t \leq 2:$$

we have

$$\begin{aligned} \dot{x}(t) &= -2\pi \sin 2\pi t \\ \dot{y}(t) &= 2\pi \cos 2\pi t \\ \dot{z}(t) &= 1 \end{aligned}$$

so

$$\begin{aligned} d\mathfrak{s} &= \sqrt{(\dot{x}(t))^2 + (\dot{y}(t))^2 + (\dot{z}(t))^2}\, dt \\ &= \sqrt{(-2\pi \sin 2\pi t)^2 + (2\pi \cos 2\pi t)^2 +)(1)^2} \\ &= \sqrt{4\pi^2 + 1}\, dt \end{aligned}$$

and

$$\begin{aligned} \mathfrak{s}\left(\mathcal{C}\right) &= \int_0^2 d\mathfrak{s} \\ &= \int_0^2 \sqrt{4\pi^2 + 1}\, dt \\ &= 2\sqrt{4\pi^2 + 1}. \end{aligned}$$

The arclength of the parabola

$$y = x^2$$

between $(0, 0)$ and $(\frac{1}{2}, \frac{1}{4})$ can be calculated using x as a parameter

$$\overrightarrow{p}(x) = (x, x^2) \quad 0 \leq t \leq 1$$

with

$$\begin{aligned} d\mathfrak{s} &= \sqrt{1 + (2x)^2}\, dx \\ &= \sqrt{1 + 4x^2}\, dx. \end{aligned}$$

Thus

$$\mathfrak{s}\left(\mathcal{C}\right) = \int_0^{\frac{1}{2}} \sqrt{1 + 4x^2}\, dx$$

which is best done using the trigonometric substitution

$$\begin{aligned}
2x &= \tan\theta \\
2\,dx &= \sec^2\theta\,d\theta \\
dx &= \frac{1}{2}\sec^2\theta\,d\theta \\
\sqrt{1+4x^2} &= \sec\theta \\
x = 0 &\leftrightarrow \theta = 0 \\
x = \frac{1}{2} &\leftrightarrow \theta = \frac{\pi}{4}
\end{aligned}$$

and

$$\begin{aligned}
\int_0^2 \sqrt{1+4x^2}\,dx &= \int_0^{\pi/4} (\sec\theta)(\frac{1}{2}\sec^2\theta\,d\theta) \\
&= \int_0^{\pi/4} \frac{1}{2}\sec^3\theta\,d\theta
\end{aligned}$$

integration by parts (or cheating and looking it up in a table) yields

$$\begin{aligned}
\int_0^{\pi/4} \frac{1}{2}\sec^3\theta\,d\theta &= \frac{1}{2}\left(\frac{1}{2}\sec\theta\tan\theta + \frac{1}{2}\ln|\sec\theta+\tan\theta|\right)_0^{\pi/4} \\
&= \frac{1}{4}(\sqrt{2}+\ln(1+\sqrt{2}).
\end{aligned}$$

As another example, let us use this formalism to compute the circumference of a circle. The circle is not an arc, but the domain of the standard parametrization

$$\overrightarrow{p}(t) = (\cos t, \sin t) \quad 0 \leq t \leq 2\pi$$

can be partitioned via

$$\mathcal{P} = \{0, \pi, 2\pi\}$$

into two semicircles, $\mathcal{C}_i$, $i = 1, 2$ which meet only at the endpoints; it is natural then to say

$$\mathfrak{s}(\mathcal{C}) = \mathfrak{s}(\mathcal{C}_1) + \mathfrak{s}(\mathcal{C}_2).$$

We can calculate that

$$\begin{aligned} \dot{x}(t) &= -\sin t \\ \dot{y}(t) &= \cos t \end{aligned}$$

so

$$\begin{aligned} d\mathfrak{s} &= \sqrt{(-\sin t)^2 + (\cos t)^2}\, dt \\ &= dt \end{aligned}$$

and thus

$$\begin{aligned} \mathfrak{s}\left(\mathcal{C}\right) &= \mathfrak{s}\left(\mathcal{C}_1\right) + \mathfrak{s}\left(\mathcal{C}_2\right) \\ &= \int_0^{\pi} dt + \int_{\pi}^{2\pi} dt \\ &= \int_0^{2\pi} dt \\ &= 2\pi. \end{aligned}$$

The example of the circle illustrates the way that we can go from the definition of arclength for an *arc* to arclength for a general *curve*. By Corollary **??**, any parametrized curve $\mathcal{C}$ can be partitioned into arcs $\mathcal{C}_k$, and the arclength of $\mathcal{C}$ is in a natural way the sum of the arclengths of the arcs:

$$\mathfrak{s}\left(\mathcal{C}\right) = \sum_k \mathfrak{s}\left(\mathcal{C}_k\right);$$

when the curve is parametrized over a closed interval, this is a finite sum, but it can be an infinite (positive) series when the domain is an open interval. Notice that a reparametrization of $\mathcal{C}$ is related to the original one via a strictly monotone, continuous function, and this associates to every partition of the original domain a partition of the reparametrized domain involving the same segments of the curve, and hence having the same value of $\ell\left(\mathcal{P}, .\right)$ Furthermore, when the parametrization is regular, the sum above can be rewritten as a single (possibly improper) integral. This shows

Remark 2.5.3. *The arclength of a parametrized curve $\mathcal{C}$ does not change under reparametrization. If the curve is regular, then the arclength is given* *by the integral of the speed (possibly improper if the domain is open)*

$$\begin{aligned}\mathfrak{s}\left(\mathcal{C}\right) &= \int_a^b d\mathfrak{s} \\ &= \int_a^b \left(\frac{d\mathfrak{s}}{dt}\right) dt \\ &= \int_a^b \left\|\dot{\vec{p}}\,(t)\right\| dt.\end{aligned}$$

for any regular parametrization of $\mathcal{C}$.

In retrospect, this justifies our notation for speed, and also fits our intuitive notion that the length of a curve $\mathcal{C}$ is the distance travelled by a point as it traverses $\mathcal{C}$ once.

As a final example, we calculate the arclength of one "arch" of the cycloid

$$\begin{aligned}x &= \theta - \sin\theta \\ y &= 1 - \cos\theta\end{aligned}$$

or

$$\vec{p}\,(\theta) = (\theta - \sin\theta, 1 - \cos\theta), \quad 0 \leq \theta \leq 2\pi.$$

Differentiating, we get

$$\vec{v}\,(\theta) = (1 - \cos\theta, \sin\theta)$$

so

$$\begin{aligned}d\mathfrak{s} &= \sqrt{(1-\cos\theta)^2 + \sin^2\theta}\,d\theta \\ &= \sqrt{2 - 2\cos\theta}\,d\theta.\end{aligned}$$

The arclength integral

$$\mathfrak{s}\left(\mathcal{C}\right) = \int_0^{2\pi} \sqrt{2 - 2\cos\theta}\,d\theta$$

can be rewritten, multiplying and dividing the integrand by $\sqrt{1+\cos\theta}$, as

$$= \sqrt{2}\int_0^{2\pi} \frac{\sqrt{1-\cos^2\theta}}{\sqrt{1+\cos\theta}}\,d\theta$$

which suggests the substitution

$$\begin{aligned} u &= 1+\cos\theta \\ du &= -\sin\theta \end{aligned}$$

since the numerator of the integrand looks like $\sin\theta$. However, there is a pitfall here: the numerator *does* equal $\sqrt{\sin^2\theta}$, but this equals $\sin\theta$ only when $\sin\theta \geq 0$, which is to say over the first half of the curve, $0 \leq \theta \leq \pi$; for the second half, it equals $-\sin\theta$. Therefore, we break the integral into two

$$\begin{aligned} \sqrt{2}\int_0^{2\pi}\sqrt{1-\cos\theta}\,d\theta &= \sqrt{2}\int_0^{\pi}\frac{\sin\theta\,d\theta}{\sqrt{1+\cos\theta}} - \sqrt{2}\int_\pi^{2\pi}\frac{\sin\theta\,d\theta}{\sqrt{1+\cos\theta}} \\ &= \sqrt{2}\int_2^0 -u^{-1/2}\,du - \sqrt{2}\int_0^2 u^{-1/2}\,du \\ &= 2\sqrt{2}\int_0^2 u^{-1/2}\,du \\ &= 4\sqrt{2}u^{1/2}\Big|_0^2 \\ &= 8. \end{aligned}$$

We note one technical point here: strictly speaking, the parametrization of the cycloid is not regular: while it is continuously differentiable, the velocity vector is zero at the ends of the arch. To get around this problem, we can think of this as an improper integral, taking the limit of the arclength of the curve $\overrightarrow{p}(\theta)$, $\varepsilon \leq \theta \leq 2\pi - \varepsilon$ as $\varepsilon \to 0$. The principle here (similar, for example, to the hypotheses of the Mean Value Theorem) is that the velocity can vanish at an endpoint of an arc in Theorem 2.5.1, or more generally that it can vanish at a set of isolated points of the curve[18] and the integral formula still holds, provided we don't "back up" after that.

Integrating a Function along a Curve (Path Integrals)

Suppose we have a wire which is shaped like an arc, but has variable thickness, and hence variable density. If we know the density at each point along the arc, how do we find the total mass? If the arc happens to be an interval along the x-axis, then we simply define a function $f(x)$ whose value at each point is the density, and integrate. We would like to carry out a similar process along an arc or, more generally, along a curve.

[18] With a little thought, we see that it can even vanish on a nontrivial closed interval.

Our abstract setup is this: we have an arc, $\mathcal{C}$, parametrized by the (continuous, one-to-one) vector-valued function $\overrightarrow{p}(t)$, $a \leq t \leq b$, and we have a (real-valued) function which assigns to each point $\overrightarrow{p}$ of $\mathcal{C}$ a number $f(\overrightarrow{p})$; we want to integrate f along $\mathcal{C}$. The process is a natural combination of the Riemann integral with the arclength calculation of § 2.5. Just as for arclength, we begin by partitioning $\mathcal{C}$ via a partition of the domain $[a,b]$ of our parametrization

$$\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\}.$$

For a small mesh size, the arclength of $\mathcal{C}$ between successive points $\overrightarrow{p}(t_j)$ is well approximated by

$$\triangle\mathfrak{s}_j = \|\overrightarrow{p}(t_{j-1}) - \overrightarrow{p}(t_j)\|$$

and we can form lower and upper sums

$$\mathcal{L}(\mathcal{P},f) = \sum_{j=1}^{n} \inf_{t\in I_j} f(\overrightarrow{p}(t))\,\triangle\mathfrak{s}_j$$

$$\mathcal{U}(\mathcal{P},f) = \sum_{j=1}^{n} \sup_{t\in I_j} f(\overrightarrow{p}(t))\,\triangle\mathfrak{s}_j.$$

As in the usual theory of the Riemann integral, we have for any partition $\mathcal{P}$ that

$$\mathcal{L}(\mathcal{P},f) \leq \mathcal{U}(\mathcal{P},f);$$

it is less clear that refining a partition lowers $\mathcal{U}(\mathcal{P},f)$ (although it clearly does increase $\mathcal{L}(\mathcal{P},f)$), since the quantity $\ell\,(\mathcal{P},\overrightarrow{p})$ increases under refinement. However, if the arc is rectifiable, we can modify the upper sum by using $\mathfrak{s}\,(\overrightarrow{p}(I_j))$ in place of $\triangle\mathfrak{s}_j$; denoting this by

$$\mathcal{U}^*(\mathcal{P},f) = \sum_{j=1}^{n} \sup_{t\in I_j} f(\overrightarrow{p}(t))\,\mathfrak{s}\,(\overrightarrow{p}(I_j))$$

we have, for any two partitions $\mathcal{P}_i$, $i = 1, 2$,

$$\mathcal{L}(\mathcal{P}_1,f) \leq \mathcal{U}^*(\mathcal{P}_2,f)$$

We will say the function $f(\overrightarrow{p})$ is **integrable** over the arc $\mathcal{C}$ if

$$\sup_{\mathcal{P}} \mathcal{L}(\mathcal{P},f) = \inf_{\mathcal{P}} \mathcal{U}^*(\mathcal{P},f)$$

and in this case the common value is called the **path integral** or **integral with respect to arclength** of f along the arc $\mathcal{C}$, denoted

$$\int_{\mathcal{C}} f \, d\mathfrak{s}.$$

As in the case of the usual Riemann integral, we can show that if f is integrable over $\mathcal{C}$ then for any sequence $\mathcal{P}_k$ of partitions of $[a,b]$ with $\text{mesh}(\mathcal{P}_k) \to 0$, the Riemann sums using any sample points $t_j^* \in I_j$ converge to the integral:

$$\mathcal{R}(\mathcal{P}_k, f, \{t_j^*\}) = \sum_{j=1}^{n} f\left(t_j^*\right) \triangle\mathfrak{s}_j \to \int_{\mathcal{C}} f \, d\mathfrak{s}.$$

It is easy to see that the following analogue of Remark 2.5.3 holds for path integrals:

Remark 2.5.4. *The path integral of a function over a parametrized curve is unchanged by reparametrization; when the parametrization $\overrightarrow{p}\colon\mathbb{R}\to\mathbb{R}^3$ is regular, we have*

$$\int_{\mathcal{C}} f \, d\mathfrak{s} = \int_a^b f(\overrightarrow{p}(t)) \left\| \dot{\overrightarrow{p}}(t) \right\| \, dt.$$

As two examples, let us take $\mathcal{C}$ to be the parabola $y = x^2$ between $(0,0)$ and $(1,1)$, and compute the two integrals

$$\int_{\mathcal{C}} x \, d\mathfrak{s}$$
$$\int_{\mathcal{C}} y \, d\mathfrak{s}.$$

To compute the first integral, we use the standard parametrization in terms of x

$$\overrightarrow{p}(x) = (x, x^2), \quad 0 \le x \le 1;$$

then the element of arclength is given by

$$\begin{aligned} d\mathfrak{s} &= \sqrt{(\dot{x})^2 + (\dot{y})^2} \, dx \\ &= \sqrt{1 + (2x)^2} \, dx \\ &= \sqrt{1 + 4x^2} \, dx \end{aligned}$$

so

$$\int_{\mathcal{C}} x\,d\mathfrak{s} = \int_0^1 (x)(\sqrt{1+4x^2}\,dx)$$

which we can do using the substitution

$$\begin{aligned} u &= 1+4x^2 \\ du &= 8x\,dx \\ x\,dx &= \frac{1}{8}du \\ x = 0 &\leftrightarrow u = 1 \\ x = 1 &\leftrightarrow u = 5 \end{aligned}$$

and

$$\begin{aligned} \int_0^1 x\sqrt{1+4x^2}\,dx &= \frac{1}{8}\int_1^5 u^{1/2}\,du \\ &= \frac{1}{12}u^{3/2}\Big|_1^5 \\ &= \frac{5\sqrt{5}-1}{12}. \end{aligned}$$

If we try to use the same parametrization to find the second integral, we have

$$\begin{aligned} \int_{\mathcal{C}} y\,d\mathfrak{s} &= \int_{\mathcal{C}} x^2\,d\mathfrak{s} \\ &= \int_0^1 x^2\sqrt{1+4x^2}\,dx \end{aligned}$$

which, while not impossible, is a lot harder. However, we can also express $\mathcal{C}$ as the graph of $x = \sqrt{y}$ and parametrize in terms of y; this yields

$$\begin{aligned} d\mathfrak{s} &= \sqrt{\left(\frac{1}{2\sqrt{y}}\right)^2 + 1}\,dy \\ &= \sqrt{\frac{1}{4y} + 1}\,dy \end{aligned}$$

so

$$\begin{aligned}\int_{\mathcal{C}} y\,d\mathfrak{s} &= \int_0^1 y\sqrt{\frac{1}{4y}+1}\,dy\\ &= \int_0^1 \sqrt{\frac{y}{4}+y^2}\,dy\end{aligned}$$

which, completing the square,

$$\begin{aligned}&= \int_0^1 \sqrt{\left(y+\frac{1}{8}\right)^2-\frac{1}{64}}\,dy\\ &= \frac{1}{8}\int_0^1 \sqrt{(8y+1)^2-1}\,dy\end{aligned}$$

and the substitution

$$8y+1=\sec\theta$$

changes this into

$$\begin{aligned}\frac{1}{64}\int_0^{\operatorname{arcsec}9}(sec^3\theta-\sec\theta)\,d\theta &= \frac{1}{128}\left(\tan\theta\sec\theta-\ln(\sec\theta+\tan\theta)\right)_0^{\operatorname{arcsec}9}\\ &= \frac{9\sqrt{5}}{32}-\frac{1}{128}\ln(9+4\sqrt{5}).\end{aligned}$$

Exercises for § 2.5

Practice problems:

1. Set up an integral expressing the arc length of each curve below. Do not attempt to integrate.

   (a) $y=x^n, \quad 0\le x\le 1$

   (b) $y=e^x, \quad 0\le x\le 1$

   (c) $y=\ln x, \quad 1\le x\le e$

   (d) $y=\sin x, \quad 0\le x\le \pi$

   (e) $\left\{\begin{array}{rcl} x &=& a\cos\theta\\ y &=& b\sin\theta\end{array}\right., \quad 0\le\theta\le 2\pi$

   (f) $\left\{\begin{array}{rcl} x &=& e^t+e^{-t}\\ y &=& e^t-e^{-t}\end{array}\right., \quad -1\le t\le 1$

2. Find the length of each curve below.

(a) $y = x^{3/2}, \quad 0 \le x \le 1$

(b) $y = x^{2/3}, \quad 0 \le x \le 1$

(c) $y = \dfrac{x^3}{3} + \dfrac{1}{4x}, \quad 1 \le x \le 2$

(d) $y = \displaystyle\int_1^x \sqrt{t^4 - 1}\, dt, \quad 1 \le x \le 2$

(e) $\begin{cases} x &= \sin^3 t \\ y &= \cos^3 t \end{cases}, \quad 0 \le t \le \dfrac{\pi}{4}$

(f) $\begin{cases} x &= 9t^2 \\ y &= 4t^3 \\ z &= t^4 \end{cases}, \quad 0 \le t \le 1$

(g) $\begin{cases} x &= 8t^3 \\ y &= 15t^4 \\ z &= 15t^5 \end{cases}, \quad 0 \le t \le 1$

(h) $\begin{cases} x &= t^2 \\ y &= \ln t \\ z &= 2t \end{cases}, \quad 1 \le t \le 2$

(i) $\begin{cases} x &= \sin\theta \\ y &= \theta + \cos\theta \end{cases}, \quad 0 \le \theta \le \dfrac{\pi}{2}$

(j) $\begin{cases} x &= 3t \\ y &= 4t\sin t \\ z &= 4t\cos t \end{cases}, \quad 0 \le t \le \dfrac{5}{4}$

3. Calculate $\int_{\mathcal{C}} f\, d\mathfrak{s}$:

(a) $f(x,y) = 36x^3$, $\mathcal{C}$ is $y = x^3$ from $(0,0)$ to $(1,1)$.

(b) $f(x,y) = 32x^5$, $\mathcal{C}$ is $y = x^4$ from $(0,0)$ to $(1,1)$.

(c) $f(x,y) = x^2 + y^2$, $\mathcal{C}$ is $y = 2x$ from $(0,0)$ to $(1,2)$.

(d) $f(x,y) = 4(x + \sqrt{y})$, $\mathcal{C}$ is $y = x^2$ from $(0,0)$ to $(1,1)$.

(e) $f(x,y) = x^2$, $\mathcal{C}$ is the upper half circle $x^2 + y^2 = 1$, $y \ge 0$.

(f) $f(x,y) = x^2 + y^2$, $\mathcal{C}$ is given in parametric form as

$$\begin{cases} x &= t \\ y &= \sqrt{1 - t^2} \end{cases}, \quad 0 \le t \le 1.$$

(g) $f(x,y) = (1 - x^2)^{3/2}$, $\mathcal{C}$ is upper half of the circle $x^2 + y^2 = 1$.

(h) $f(x,y) = x^3 + y^3$, $\mathcal{C}$ is given in parametric form as

$$\begin{cases} x &= 2\cos t \\ y &= 2\sin t \end{cases}, \quad 0 \le t \le \pi.$$

(i) $f(x,y) = xy$, $\mathcal{C}$ is $y = x^2$ from $(0,0)$ to $(1,1)$.

(j) $f(x,y,z) = xy$, $\mathcal{C}$ is given in parametric form as

$$\begin{cases} x &= \cos t \\ y &= \sin t \\ z &= t \end{cases}, \quad 0 \le t \le \pi.$$

(k) $f(x,y,z) = x^2y$, $\mathcal{C}$ is given in parametric form as

$$\begin{cases} x &= \cos t \\ y &= \sin t \\ z &= t \end{cases}, \quad 0 \le t \le \pi.$$

(l) $f(x,y,z) = z$, $\mathcal{C}$ is given in parametric form as

$$\begin{cases} x &= \cos t \\ y &= \sin t \\ z &= t \end{cases}, \quad 0 \le t \le \pi.$$

(m) $f(x,y,z) = 4y$, $\mathcal{C}$ is given in parametric form as

$$\begin{cases} x &= t \\ y &= 2t \\ z &= t^2 \end{cases}, \quad 0 \le t \le 1.$$

(n) $f(x,y,z) = x^2 - y^2 + z^2$, $\mathcal{C}$ is given in parametric form as

$$\begin{cases} x &= \cos t \\ y &= \sin t \\ z &= 3t \end{cases}, \quad 0 \le t \le \pi.$$

(o) $f(x,y,z) = 4x + 16z$, $\mathcal{C}$ is given in parametric form as

$$\begin{cases} x &= 2t \\ y &= t^2 \\ z &= \frac{4t^3}{9} \end{cases}, \quad 0 \le t \le 3.$$

Theory problems:

4. Consider the graph of the function

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{for } x > 0, \\ 0 & \text{for } x = 0 \end{cases}$$

over the interval $[0,1]$.

(a) **Show** that

$$|f(x)| \leq |x|$$

with equality at 0 and the points

$$x_k := \frac{2}{(2k-1)\pi}, \quad k = 1, \ldots.$$

(b) **Show** that f is continuous. (*Hint:* the issue is $x = 0$) . Thus, its graph is a curve. Note that f is differentiable except at $x = 0$.

(c) Consider the piecewise linear approximation to this curve (albeit with infinitely many pieces) consisting of joining $(x_k, f(x_k))$ to $(x_{k+1}, f(x_{k+1}))$ with straight line segments: note that at one of these points, $f(x) = x$ while at the other $f(x) = -x$. **Show** that the line segment joining the points on the curve corresponding to $x = x_k$ and $x = x_{k+1}$ has length at least

$$\begin{aligned}\triangle \mathfrak{s}_k &= |f(x_{k+1}) - f(x_k)| \\ &= x_{k+1} + x_k \\ &= \frac{2}{(2k+1)\pi} + \frac{2}{(2k-1)\pi} \\ &= \frac{2}{\pi}\left(\frac{4k}{4k^2-1}\right).\end{aligned}$$

(d) **Show** that the sum

$$\sum_{k=1}^{\infty} \triangle \mathfrak{s}_k$$

diverges.

(e) Thus, if we take (for example) the piecewise linear approximations to the curve obtained by taking the straight line segments as above to some finite value of k and then join the last point to $(0, 0)$, their lengths will also diverge as the finite value increases. Thus, there exist partitions of the curve whose total lengths are arbitrarily large, and the curve is not rectifiable.

Challenge problem:

5. **Bolzano's curve (continued):** We continue here our study of the curve described in Exercise 10 in § **??**; we keep the notation of that exercise.

(a) **Show** that the slope of each straight piece of the graph of f_k has the form $m = \pm 2^n$ for some integer $0 \leq n \leq k$. Note that each interval over which f_k is affine has length 3^{-k}.

(b) **Show** that if two line segments start at a common endpoint and end on a vertical line, and their slopes are 2^n and 2^{n+1} respectively, then the ratio of the second to the first length is

$$\frac{m_2}{m_1} = \sqrt{\frac{1+2^{n+1}}{1+2^n}}$$

Show that this quantity is non-decreasing, and that therefore it is always at least equal to $\sqrt{5/3}$.

(c) Use this to show that the ratio of the lengths of the graphs of f_{k+1} and f_k are bounded below by $2\sqrt{5}/3\sqrt{3} + 1/3 \geq 1.19$.

(d) How does this show that the graph of f is non-rectifieable?

3
Differential Calculus for Real-Valued Functions of Several Variables

In this chapter and in Chapter 5, we consider functions whose input involves several variables—or equivalently, whose input is a vector—and whose output is a real number.

We shall restrict ourselves to functions of two or three variables, where the vector point of view can be interpreted geometrically.

A function of two (*resp.* three) variables can be viewed in two slightly different ways, reflected in two different notations.

We can think of the input as three separate variables; often it will be convenient to use subscript notation x_i (instead of x, y, z) for these variables, so we can write

$$f(x,y) = f(x_1,x_2)$$

in the case of two variables and

$$f(x,y,z) = f(x_1,x_2,x_3)$$

in the case of three variables.

Alternatively, we can think of the input as a single vector $\overrightarrow{x}$ formed from listing the variables in order:

$$\overrightarrow{x} = (x,y) = (x_1,x_2)$$

or

$$\overrightarrow{x} = (x, y, z) = (x_1, x_2, x_3)$$

and simply write our function as

$$f(\overrightarrow{x}) \, .$$

A third notation which is sometimes useful is that of mappings: we write

$$f \colon \mathbb{R}^n \to \mathbb{R}$$

(with $n = 2$ or $n = 3$) to indicate that f has inputs coming from $\mathbb{R}^n$ and produces outputs that are real numbers.[1]

In much of our expositon we will deal explicitly with the case of three variables, with the understanding that in the case of two variables one simply ignores the third variable; conversely, we will in some cases concentrate on the case of two variables and if necessary indicate how to incorporate the third variable.

Much of what we will do has a natural extension to any number of input variables, and we shall occasionally comment on this.

In this chapter, we consider the definition and use of derivatives in this context.

3.1 Continuity and Limits

Recall from § 2.3 that *a sequence of vectors converges if it converges coordinatewise* Using this notion, we can define continuity of a real-valued function of *three (or two)* variables $f(\overrightarrow{x})$ by analogy to the definition for real-valued functions $f(x)$ of *one* variable:

Definition 3.1.1. *A real-valued function $f(\overrightarrow{x})$ is* ***continuous*** *on a subset $D \subset \mathbb{R}^3$ of its domain if whenever the inputs converge in D (as points in $\mathbb{R}^3$) the corresponding outputs also converge (as numbers):*

$$\overrightarrow{x}_k \to \overrightarrow{x}_0 \Rightarrow f(\overrightarrow{x}_k) \to f(\overrightarrow{x}_0) \, .$$

It is easy, using this definition and basic properties of convergence for sequences of numbers, to verify the following analogues of properties of continuous functions of one variable. First, the composition of continuous functions is continuous (Exercise 3):

[1] When the domain is an explicit subset $D \subset \mathbb{R}^n$ we will write $f \colon D \to \mathbb{R}$.

Remark 3.1.2. *Suppose $f(\overrightarrow{x})$ is continuous on $D \subset \mathbb{R}^3$.*

1. *If $g\colon\mathbb{R}\to\mathbb{R}$ is continuous on $G \subset \mathbb{R}$ and $f(\overrightarrow{x}) \in G$ for every $\overrightarrow{x} = (x, y, z) \in D$, then the composition $g \circ f\colon\mathbb{R}^3\to\mathbb{R}$, defined by*

$$(g \circ f)(\overrightarrow{x}) = g(f(\overrightarrow{x}))$$

i.e.,

$$(g \circ f)\,(x_1, \dots, x_n) = g(f(x, y, z))$$

is continuous on D.

2. *If $\overrightarrow{g}\colon\mathbb{R}\to\mathbb{R}^3$ is continuous on $[a, b]$ and $\overrightarrow{g}(t) \in D$ for every $t \in [a, b]$, then $f \circ \overrightarrow{g}\colon\mathbb{R}\to\mathbb{R}$, defined by*

$$(f \circ \overrightarrow{g})(t) = f(\overrightarrow{g}(t))$$

i.e.,

$$(f \circ \overrightarrow{g})(t) = f(g_1(t)\,, g_2(t)\,, g_3(t))$$

is continuous on $[a, b]$

Second, functions defined by reasonable formulas are continuous where they are defined:

Lemma 3.1.3. *If $f(x, y, z)$ is defined by a formula composed of arithmetic operations, powers, roots, exponentials, logarithms and trigonometric functions applied to the various components of the input, then $f(x, y, z)$ is continuous where it is defined.*

Proof. Consider the functions on $\mathbb{R}^2$

$$\begin{aligned} add(x_1, x_2) &= x_1 + x_2 \\ sub(x_1, x_2) &= x_1 - x_2 \\ mul(x_1, x_2) &= x_1 x_2 \\ div(x_1, x_2) &= \frac{x_1}{x_2}; \end{aligned}$$

each of the first three is continuous on $\mathbb{R}^2$, and the last is continuous off the x_2-axis, because of the basic laws about arithmetic of convergent sequences (*Calculus Deconstructed*, Theorem 2.4.1).

But then application of Remark 3.1.2 to these and powers, roots, exponentials, logarithms and trigonometric functions (which are all continuous where defined) yields the lemma. □

Remark 3.1.2 can also be used to get a weak analogue of the Intermediate Value Theorem (*Calculus Deconstructed*, Theorem 3.2.1). Recall that this says, for $f\colon \mathbb{R} \to \mathbb{R}$ continuous on $[a, b]$, that if $f(a) = A$ and $f(b) = B$ then for every C between A and B the equation $f(x) = C$ has at least one solution between a and b. Since the notion of a point in the plane or in space being "between" two others doesn't really make sense, there isn't really a direct analogue of the Intermediate Value Theorem, either for $\overrightarrow{f}\colon \mathbb{R} \to \mathbb{R}^3$ or for $f\colon \mathbb{R}^3 \to \mathbb{R}$. However, we can do the following: Given two points $\overrightarrow{a}, \overrightarrow{b} \in \mathbb{R}^3$, we define a **path** from $\overrightarrow{a}$ to $\overrightarrow{b}$ to be the image of any locally one-to-one continuous function $\overrightarrow{p}\colon \mathbb{R} \to \mathbb{R}^3$, parametrized so that $\overrightarrow{p}(a) = \overrightarrow{a}$ and $\overrightarrow{p}(b) = \overrightarrow{b}$. Then we can talk about points "between" $\overrightarrow{a}$ and $\overrightarrow{b}$ *along this curve.*

Proposition 3.1.4. *If $f\colon \mathbb{R}^3 \to \mathbb{R}$ is continuous on a set $D \subset \mathbb{R}^3$ and $\overrightarrow{a}$ and $\overrightarrow{b}$ are points of D that can be joined by a path in D, then for every number C between $f(\overrightarrow{a})$ and $f\left(\overrightarrow{b}\right)$ the equation*

$$f(\overrightarrow{x}) = C$$

has at least one solution between $\overrightarrow{a}$ and $\overrightarrow{b}$ along any path in D which joins the two points.

The proof of this is a simple application of Remark 3.1.2 to $f \circ \overrightarrow{p}$ (Exercise 4).

For example, if $f(\overrightarrow{x})$ is continuous on $\mathbb{R}^3$ and $f(\overrightarrow{a})$ is positive while $f\left(\overrightarrow{b}\right)$ is negative, then the function must equal zero somewhere on any path from $\overrightarrow{a}$ to $\overrightarrow{b}$.

To study discontinuities for a real-valued function of one variable, we defined the limit of a function at a point. In this context, we always ignored the value of the function *at* the point in question, looking only at the values at points *nearby*. The old definition carries over verbatim:

Definition 3.1.5. *Suppose the function $f(\overrightarrow{x})$ is defined on a set $D \subset \mathbb{R}^3$ and $\overrightarrow{x}_0$ is an accumulation point[2] of D; we say that the function* ***converges*** *to $L \in \mathbb{R}$ as $\overrightarrow{x}$ goes to $\overrightarrow{x}_0$ if whenever $\{\overrightarrow{x}_k\}$ is a sequence of points in D, all distinct from $\overrightarrow{x}_0$, which converges to $\overrightarrow{x}_0$, the corresponding sequence of*

[2] A point $\overrightarrow{x}_0$ is an **accumulation point** of the set $D \subset \mathbb{R}^3$ if there exists a sequence of points in D, all distinct from $\overrightarrow{x}_0$, which converge to $\overrightarrow{x}_0$.

values of $f(x_0)$ *converges to* L*:*

$$\begin{aligned}\overrightarrow{x}_0 \neq \overrightarrow{x}_k \to \overrightarrow{x}_0 \Rightarrow \\ f(\overrightarrow{x}_k) \to L.\end{aligned}$$

The same arguments that worked before show that a function converges to at most one number at any given point, so we can speak of "the" **limit** of the function at $\overrightarrow{x} = \overrightarrow{x}_0$, denoted

$$L = \lim_{\overrightarrow{x} \to \overrightarrow{x}_0} f(\overrightarrow{x})\,.$$

For functions of one variable, we could consider "one-sided" limits, and this often helped us understand (ordinary, two-sided) limits. Of course, this idea does not really work for functions of more than one variable, since the "right" and "left" sides of a point in the plane or space don't make much sense. We might be tempted instead to probe the limit of a function at a point in the plane by considering what happens along a line through the point: that is, we might think that a function has a limit at a point if it has a limit along some line (or even *every* line) through the point. The following example shows the folly of this point of view: consider the function defined for $\overrightarrow{x} \neq \overrightarrow{0} \in \mathbb{R}^2$ by

$$f(x,y) = \frac{xy}{x^2+y^2}, \quad (x,y) \neq (0,0)$$

If we look at the values of the function along a line through $\overrightarrow{x}_0 = \overrightarrow{0}$ of slope m,

$$y = mx,$$

we see that the values of $f(\overrightarrow{x})$ at points on this line are

$$\begin{aligned} f(x,mx) &= \frac{(x)(mx)}{x^2+m^2x^2} \\ &= \frac{m}{1+m^2}. \end{aligned}$$

This shows that for any sequence of points approaching the origin along a given line, the corresponding values of $f(\overrightarrow{x})$ converge—but *the limit they converge to varies with the (slope of) the line*, so the limit $\lim_{\overrightarrow{x} \to \overrightarrow{x}_0} f(\overrightarrow{x})$ does not exist.

Actually, the situation is even worse: if we consider another function defined on the plane except the origin by

$$f(x,y) = \frac{x^2y}{x^4+y^2}, \quad (x,y) \neq (0,0)$$

then along a line $y = mx$ through the origin, we have the values

$$\begin{aligned} f(x,mx) &= \frac{(x^2)(mx)}{x^4+(mx)^2} \\ &= \frac{mx^3}{x^2(x^2+m^2)} \\ &= \frac{mx}{x^2+m^2} \\ &\to 0 \end{aligned}$$

as $\overrightarrow{x} \to \overrightarrow{0}$.[3] Thus, the limit along every line through the origin exists *and equals zero*. We might conclude that the function converges to zero as $\overrightarrow{x}$ goes to $\overrightarrow{0}$. However, along the *parabola* $y = mx^2$ we see a different behavior:

$$\begin{aligned} f\left(x,mx^2\right) &= \frac{(x)^2(mx^2)}{x^4+(mx^2)^2} \\ &= \frac{m}{1+m^2} \end{aligned}$$

so the limit along a *parabola* depends on which parabola we use to approach the origin. In fact, we really need to require that the limit of the function along *every curve* through the origin is the same. This is even harder to think about than looking at every *sequence* converging to $\overrightarrow{0}$.

The definition of limits in terms of δ's and ε's, which we downplayed in the context of single variable calculus, is a much more useful tool in the context of functions of several variables.

Remark 3.1.6. *(ε-δ Definition of limit:)*
For a function $f(\overrightarrow{x})$ defined on a set $D \subset \mathbb{R}^3$ with $\overrightarrow{x}_0$ an accumulation point of D, the following conditions are equivalent:

1. *For every sequence $\{\overrightarrow{x}_k\}$ of points in D distinct from $\overrightarrow{x}_0$,*

$$f(\overrightarrow{x}_k) \to L;$$

[3] The preceding argument assumes $m \neq 0$. What happens if $m = 0$?

2. *For every $\varepsilon > 0$ there exists $\delta > 0$ such that for points $\overrightarrow{x} \in D$*

$$0 < \operatorname{dist}(\overrightarrow{x}, \overrightarrow{x}_0) < \delta \textit{ guarantees } |f(\overrightarrow{x}) - L| < \varepsilon.$$

The ε-δ formulation can sometimes be awkward to apply, but for finding limits of functions of two variables at the origin in $\mathbb{R}^2$, we can sometimes use a related trick, based on polar coordinates. To see how it works, consider the example

$$f(x,y) = \frac{x^3}{x^2+y^2}, \quad (x,y) \neq (0,0).$$

If we express this in the polar coordinates of (x,y)

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

we have

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

Now, the distance of (x,y) from the origin is r, so convergence to a limit at the origin would mean that by making

$$r < \delta$$

we can insure that

$$|f(x,y) - L| < \varepsilon;$$

in other words, we want to know whether $r\cos^3\theta$ approaches a limit as $r \to 0$, regardless of the behavior of θ. But this is clear: since

$$\left|\cos^3\theta\right| \leq 1,$$

any sequence of points $\overrightarrow{p}_i$ with respective polar coordinates (r_i, θ_i) satisfies

$$r_i\cos^3\theta_i \to 0$$

and so

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

How does this play out for our earlier examples? To see what happens in the first example above

$$f(x,y) = \frac{xy}{x^2+y^2}, \quad (x,y) \neq (0,0)$$

write it in terms of polar coordinates:

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

We see that the limiting behavior of the function very much depends on the behavior of θ; thus the function **diverges** as $(x,y) \to (0,0)$.

This trick is not universally useful. The second example

$$f(x,y) = \frac{x^2y}{x^4+y^2}$$

is expressed in polar coordinates by

$$f(r\cos\theta, r\sin\theta) = \frac{r^3\cos^2\theta\sin\theta}{r^4\cos^4\theta + r^2\sin^2\theta}$$

where things don't cancel quite so nicely. We can try to pull out an r^2 factor, to get

$$= \frac{r(\cos^2\theta\sin\theta)}{r^2\cos^4\theta + \sin^2\theta}.$$

If $\sin\theta$ stays bounded away from zero, then this goes to zero as $r \to 0$ (why?), but it is less clear what happens when r and $\sin\theta$ *both* tend to zero.

Recall that a function is **continuous at a point** x_0 in its domain if

$$\lim_{x\to x_0} f(x) = f(x_0)\,.$$

This carries over verbatim to functions of several variables: a function $f:\mathbb{R}^3 \to \mathbb{R}$ is continuous at a point $\overrightarrow{x}_0$ in its domain if

$$\lim_{\overrightarrow{x}\to\overrightarrow{x}_0} f(\overrightarrow{x}) = f(\overrightarrow{x}_0).$$

If a function has a limit at $\overrightarrow{x}_0$ but fails to be continuous at $\overrightarrow{x}_0$ either because the limit as $\overrightarrow{x} \to \overrightarrow{x}_0$ differs from the value at $\overrightarrow{x} = \overrightarrow{x}_0$, or because $f(\overrightarrow{x}_0)$ is undefined, then we can restore continuity at $\overrightarrow{x} = \overrightarrow{x}_0$ simply by redefining the function at $\overrightarrow{x} = \overrightarrow{x}_0$ to equal its limit there; we call this a **removable discontinuity**. If on the other hand the limit as $\overrightarrow{x} \to \overrightarrow{x}_0$ fails to exist, there is no way (short of major revisionism) of getting the function to be continuous at $\overrightarrow{x} = \overrightarrow{x}_0$, and we have an **essential discontinuity**.

Our divergent examples above show that the behavior of a rational function (a ratio of polynomials) in several variables near a zero of its denominator can be much more complicated than for one variable, if the discontinuity is essential.

Exercises for § 3.1

Practice problems:

1. For each function below, find its limit as $(x, y) \to (0, 0)$:

   (a) $\dfrac{\sin(x^2+y^2)}{x^2+y^2}$ (b) $\dfrac{x^2}{\sqrt{x^2+y^2}}$ (c) $\dfrac{x^2}{x^2+y^2}$

   (d) $\dfrac{2x^2y}{x^2+y^2}$ (e) $e^x y$ (f) $\dfrac{(x+y)^2-(x-y)^2}{xy}$

   (g) $\dfrac{x^3-y^3}{x^2+y^2}$ (h) $\dfrac{\sin(xy)}{y}$ (i) $\dfrac{e^{xy}-1}{y}$

   (j) $\dfrac{\cos(xy)-1}{x^2y^2}$ (k) $\dfrac{xy}{x^2+y^2+2}$ (l) $\dfrac{(x-y)^2}{x^2+y^2}$

2. Find the limit of each function as $(x, y, z) \to (0, 0, 0)$:

   (a) $\dfrac{2x^2y\cos z}{x^2+y^2}$ (b) $\dfrac{xyz}{x^2+y^2+z^2}$

Theory problems:

3. Prove Remark 3.1.2.

4. Prove Proposition 3.1.4.

3.2 Linear and Affine Functions

So far we have seen the derivative in two settings. For a real-valued function $f(x)$ of one variable, the derivative $f'(x_0)$ at a point x_0 first comes up as a number, which turns out to be the slope of the tangent line. This in turn is the line which best approximates the graph $y = f(x)$ near the point, in the sense that it is the graph of the polynomial of degree one, $T_{x_0}f = f(x_0) + f'(x_0)\,(x - x_0)$, which has **first-order contact** with the curve at the point $(x_0, f(x_0))$:

$$\left|f(x) - T_{x_0}f(x)\right| = \mathfrak{o}(|x - x_0|)$$

or

$$\frac{\left|f(x) - T_{x_0}f(x)\right|}{|x - x_0|} \to 0 \text{ as } x \to x_0.$$

If we look back at the construction of the derivative $\vec{p}'(t)$ of a vector-valued function $\overrightarrow{p}\colon\mathbb{R}\to\mathbb{R}^3$ in § **??**, we see a similar phenomenon: $\vec{p}'(t_0) = \overrightarrow{v}(t_0)$ is the direction vector for a parametrization of the tangent line, and the resulting vector-valued function, $T_{t_0}\overrightarrow{p}(t) = \overrightarrow{p}(t_0) + \overrightarrow{v}(t_0)\,(t - t_0)$, expresses how the point would move if the constraints keeping it on the curve traced out by $\overrightarrow{p}(t)$ were removed after $t = t_0$. In complete analogy to the real-valued case, $T_{t_0}\overrightarrow{p}(t)$ has first-order contact with $\overrightarrow{p}(t)$ at $t = t_0$:

$$\frac{\left\|\overrightarrow{p}(t) - T_{t_0}\overrightarrow{p}(t)\right\|}{|t - t_0|} \to 0 \text{ as } t \to t_0$$

or, in "little oh" notation,

$$\left\|\overrightarrow{p}(t) - T_{t_0}\overrightarrow{p}(t)\right\| = \mathfrak{o}(|t - t_0|).$$

It is really this last approximation property of the derivative in both cases that is at the heart of the way we use derivatives. So it would be useful to find an analogous formulation for derivatives in the case of a real-valued function $f(\overrightarrow{x})$ of a *vector* variable. This section is devoted to formulating what kind of approximation we are looking for (the analogue of having a parametrization of a line in the vector-valued case); then in the next section we will see how this gives us the right kind of approximation to $f(\overrightarrow{x})$.

Linearity

In both of the cases reviewed above, the tangent approximation to a function (real- or vector-valued) is given by polynomials of degree one in the variable. Analogously, in trying to approximate a function $f(x_1, x_2, x_3)$ of 3 variables, we would expect to look for a polynomial of degree one in these variables:

$$p(x_1, x_2, x_3) = a_1x_1 + a_2x_2 + a_3x_3 + c$$

where the coefficients a_i, $i = 1, 2, 3$ and c are real constants. To formulate this in vector terms, we begin by ignoring the constant term (which in the case of our earlier approximations is just the value of the function being approximated, at the time of approximation). A degree one polynomial with zero constant term (also called a **homogeneous polynomial** of degree one)

$$h(x_1, x_2, x_3) = a_1x_1 + a_2x_2 + a_3x_3$$

has two important properties:

Scaling: If we multiply each variable by some common real number α, the value of the function is multiplied by α:

$$\begin{aligned} h(\alpha x_1, \alpha x_2, \alpha x_3) &= (a_1)(\alpha x_1) + (a_2)(\alpha x_2) + (a_3)(\alpha x_3) \\ &= (\alpha)(a_1x_1 + a_2x_2 + a_3x_3) \\ &= \alpha \cdot h(x_1, x_2, x_3)\,; \end{aligned}$$

in vector terms, this can be written

$$h(\alpha\overrightarrow{x}) = \alpha h(\overrightarrow{x})\,.$$

This property is often referred to as **homogeneity of degree one**.

Additivity: If the value of each variable is a sum of two values, the value of the function is the same as its value over the first summands plus its value over the second ones:

$$\begin{aligned} &h(x_1 + y_1, x_2 + y_2, x_3 + y_3) \\ &\qquad = (a_1)(x_1 + y_1) + (a_2)(x_2 + y_2) + (a_3)(x_3 + y_3) \\ &\qquad = (a_1x_1 + a_2x_2 + a_3x_3) + (a_1y_1 + a_2y_2 + a_3y_3) \\ &\qquad\qquad = h(x_1, x_2, x_3) + h(y_1, y_2, y_3) \end{aligned}$$

or in vector terms

$$h(\overrightarrow{x} + \overrightarrow{y}) = h(\overrightarrow{x}) + h(\overrightarrow{y})\,.$$

These two properties together can be summarized by saying that $h(\overrightarrow{x})$ **respects linear combinations**: for any two vectors $\overrightarrow{x}$ and $\overrightarrow{y}$ and any two numbers α and β,

$$h(\alpha\overrightarrow{x}+\beta\overrightarrow{y})=\alpha h(\overrightarrow{x})+\beta h(\overrightarrow{y}).$$

A function which respects linear combinations is called a **linear function**.

The preceding discussion shows that every homogeneous polynomial of degree one is a linear function.

Recall that the **standard basis** for $\mathbb{R}^3$ is the collection $\overrightarrow{\imath}$, $\overrightarrow{\jmath}$, $\overrightarrow{k}$ of unit vectors along the three positive coordinate axes; we will find it useful to replace the "alphabetical" notation for the standard basis with an indexed one:

$$\begin{aligned}\overrightarrow{e}_1&=\overrightarrow{\imath}\\ \overrightarrow{e}_2&=\overrightarrow{\jmath}\\ \overrightarrow{e}_3&=\overrightarrow{k}.\end{aligned}$$

The basic property[4] of the standard basis is that every vector $\overrightarrow{x}\in\mathbb{R}^3$ is, in a standard way, a linear combination of these specific vectors:

$$\overrightarrow{x}=(x,y,z)=x\overrightarrow{\imath}+y\overrightarrow{\jmath}+z\overrightarrow{k}$$

or

$$(x_1,x_2,x_3)=x_1\overrightarrow{e}_1+x_2\overrightarrow{e}_2+x_3\overrightarrow{e}_3.$$

Then combining this with the fact that linear functions respect linear combinations, we easily see (Exercise 8) that all linear functions are homogeneous polynomials in the coordinates of their input:

Remark 3.2.1. *Every linear function $\ell\colon\mathbb{R}^3\to\mathbb{R}$ is determined by its effect on the elements of the standard basis for $\mathbb{R}^3$: if*

$$\ell(\overrightarrow{e}_i)=a_i,\qquad \textit{for } i=1,2,3$$

then $\ell(x_1,x_2,x_3)$ is the degree one homogeneous polynomial

$$\ell(x_1,x_2,x_3)=a_1x_1+a_2x_2+a_3x_3.$$

[4] No pun intended

Matrix Representation of Linear Functions

We are now going to set up what will at first look like an unnecessary complication of the picture above, but in time it will open the door to appropriate generalizations. The essential data concerning a linear function (*a.k.a.* a homogeneous polynomial of degree one) is the set of values taken by ℓ on the standard basis of $\mathbb{R}^3$:

$$a_i = \ell(\overrightarrow{e}_i)\,, \quad i = 1, 2, 3.$$

We shall form these numbers into a 1×3 matrix (a **row matrix**), called the **matrix representative** of ℓ:

$$[\ell] = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}.$$

We shall also create a 3×1 matrix (a **column matrix**) whose entries are the components of the vector $\overrightarrow{x}$, called the **coordinate matrix** of $\overrightarrow{x}$:

$$[\overrightarrow{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

We then define the **product** of a row with a column as the result of substituting the entries of the column into the homogeneous polynomial whose coefficients are the entries of the row; equivalently, we match the i^{th} entry of the row with the i^{th} entry of the column, multiply each matched pair, and add:

$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = a_1x_1 + a_2x_2 + a_3x_3.$$

Of course, in this language, we are representing the linear function $\ell\colon \mathbb{R}^3 \to \mathbb{R}$ as the product of its matrix representative with the coordinate matrix of the input

$$\ell(\overrightarrow{x}) = [\ell]\,[\overrightarrow{x}]\,.$$

Another way to think of this representation is to associate, to any **row**, a **vector** $\overrightarrow{a}$ (just put commas between the entries of the row matrix), and then to notice that the product of the *row* with the *coordinate matrix* of $\overrightarrow{x}$ is the same as the *dot product* of the *vector* $\overrightarrow{a}$ with $\overrightarrow{x}$:

$$\begin{aligned} \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= (a_1, a_2, a_3) \cdot (x_1, x_2, x_3) \\ &= \overrightarrow{a} \cdot \overrightarrow{x}\,. \end{aligned}$$

Thus we see that there are three ways to think of the action of the linear function $\ell\colon \mathbb{R}^3 \to \mathbb{R}$ on a vector $\overrightarrow{x} \in \mathbb{R}^3$:

- Substitute the components of $\overrightarrow{x}$ into a homogeneous polynomial of degree one, whose coefficients are the values of ℓ on the standard basis;
- Multiply the coordinate matrix of $\overrightarrow{x}$ by the matrix representative of ℓ;
- Take the dot product of the vector $\overrightarrow{a}$ (obtained from the row matrix $[\ell]$ by introducing commas) with the vector $\overrightarrow{x}$.

Affine Functions

Finally, we introduce one more piece of terminology: an **affine function** is the sum of a constant and a linear function:

$$\phi(\overrightarrow{x}) = c + \ell(\overrightarrow{x})\,.$$

In other words, an affine function is the same thing as a polynomial of degree one (with no homogeneity conditions—that is, without any restriction on the constant term).

Note that if $\phi(\overrightarrow{x}) = c + \ell(\overrightarrow{x})$ is an affine function, then for any two vectors $\overrightarrow{x}$ and $\overrightarrow{y}$,

$$\begin{aligned}\phi(\overrightarrow{y}) - \phi(\overrightarrow{x}) &= \ell(\overrightarrow{y}) - \ell(\overrightarrow{x}) \\ &= \ell(\overrightarrow{y} - \overrightarrow{x})\,;\end{aligned}$$

setting

$$\triangle\overrightarrow{x} = \overrightarrow{y} - \overrightarrow{x},$$

so that

$$\overrightarrow{y} = \overrightarrow{x} + \triangle\overrightarrow{x}$$

we can write

$$\phi(\overrightarrow{x} + \triangle\overrightarrow{x}) = \phi(\overrightarrow{x}) + \ell(\triangle\overrightarrow{x})\,. \tag{3.1}$$

Remark 3.2.2. *Given any $\overrightarrow{x}_0 \in \mathbb{R}^3$, the affine function $\phi\colon \mathbb{R}^3 \to \mathbb{R}$ can be written in the form of Equation (3.1), as its value at $\overrightarrow{x}_0$ plus a linear function of the displacement from $\overrightarrow{x}_0$:*

$$\phi(\overrightarrow{x}_0 + \triangle\overrightarrow{x}) = \phi(\overrightarrow{x}_0) + \ell(\triangle\overrightarrow{x})$$

or, stated differently,

$$\phi(\overrightarrow{x}_0 + \triangle\overrightarrow{x}) - \phi(\overrightarrow{x}_0) = \ell(\triangle\overrightarrow{x})\,;$$

the displacement of $\phi(\overrightarrow{x})$ from $\phi(\overrightarrow{x}_0)$ is a linear function of the displacement of $\overrightarrow{x}$ from $\overrightarrow{x}_0$.

In light of this observation, we can use Remark 3.2.1 to determine an affine function from its value at a point $\overrightarrow{x}_0$ together with its values at the points $\overrightarrow{x}_0 + \overrightarrow{e}_j$ obtained by displacing the original point in a direction parallel to one of the coordinate axes. A brief calculation shows that

$$\phi(\overrightarrow{x}_0 + \triangle\overrightarrow{x}) = a_0 + \sum_{j=1}^{3} a_j \triangle x_j \tag{3.2}$$

where

$$\begin{aligned} \triangle\overrightarrow{x} &= (\triangle x_1, \triangle x_2, \triangle x_3) \\ a_0 &= \phi(\overrightarrow{x}_0) \end{aligned}$$

and for $j = 1, 2, 3$

$$a_j = \phi(\overrightarrow{x}_0 + \overrightarrow{e}_j) - \phi(\overrightarrow{x}_0)\,.$$

Exercises for § 3.2

Practice problems:

1. For each linear function $\ell(\overrightarrow{x})$ below, you are given the values on the standard basis. For each function, find $\ell(1, -2, 3)$ and $\ell(2, 3, -1)$.

   (a) $\ell(\overrightarrow{\imath}) = 2$, $\ell(\overrightarrow{\jmath}) = -1$, $\ell\left(\overrightarrow{k}\right) = 1$.

   (b) $\ell(\overrightarrow{\imath}) = 1$, $\ell(\overrightarrow{\jmath}) = 1$, $\ell\left(\overrightarrow{k}\right) = 1$.

   (c) $\ell(\overrightarrow{\imath}) = 3$, $\ell(\overrightarrow{\jmath}) = 4$, $\ell\left(\overrightarrow{k}\right) = -5$.

2. Is there a linear function $\ell\colon \mathbb{R}^3 \to \mathbb{R}$ for which

$$\begin{aligned} \ell(1, 1, 1) &= 0 \\ \ell(1, -1, 2) &= 1 \\ \ell(2, 0, 3) &= 2? \end{aligned}$$

   Why or why not? Is there an *affine* function with these values? If so, give one. Are there others?

3. If $\ell: \mathbb{R}^3 \to \mathbb{R}$ is linear and

$$\begin{aligned}\ell(1,1,1) &= 3\\ \ell(1,2,0) &= 5\\ \ell(0,1,2) &= 2\end{aligned}$$

then

   (a) Find $\ell(\overrightarrow{\imath})$, $\ell(\overrightarrow{\jmath})$, and $\ell\left(\overrightarrow{k}\right)$.

   (b) Express $\ell(x, y, z)$ as a homogeneous polynomial.

   (c) Express $\ell(\overrightarrow{x})$ as a matrix multiplication.

   (d) Express $\ell(\overrightarrow{x})$ as a dot product.

4. Consider the affine function $\phi: \mathbb{R}^3 \to \mathbb{R}$ given by the polynomial

$$3x - 2y + z + 5.$$

Express $\phi(\overrightarrow{x})$ in the form given by Remark 3.2.2, when $\overrightarrow{x}_0$ is each of the vectors given below:

   (a) $\overrightarrow{x}_0 = (1, 2, 1)$

   (b) $\overrightarrow{x}_0 = (-1, 2, 1)$

   (c) $\overrightarrow{x}_0 = (2, 1, 1)$.

Theory problems:

5. **Show** that an affine function $f: \mathbb{R}^2 \to \mathbb{R}$ is determined by its values on the vertices of any nondegenerate triangle.

6. Suppose $\overrightarrow{p}(s, t)$ is a parametrization of a plane in $\mathbb{R}^3$, and $f: \mathbb{R}^3 \to \mathbb{R}$ is linear. Show that $f \circ \overrightarrow{p}: \mathbb{R}^2 \to \mathbb{R}$ is an affine function.

7. A *level set* of a function is the set of points where the function takes a particular value. Show that any level set of an affine function on $\mathbb{R}^2$ is a line, and a level set of an affine function on $\mathbb{R}^3$ is a plane. When does the line/plane go through the origin?

8. Prove Remark 3.2.1.

9. Prove Remark 3.2.2.

10. Carry out the calculation that establishes Equation (3.2).

3.3 Derivatives

In this section we carry out the program outlined at the beginning of § 3.2, trying to formulate the derivative of a real-valued function of several variables $f(\vec{x})$ in terms of an affine function making first-order contact with $f(\vec{x})$.

Definition 3.3.1. *A real-valued function of n variables $f\colon\mathbb{R}^3\to\mathbb{R}$ is **differentiable** at $\vec{x}_0\in\mathbb{R}^3$ if f is defined for all $\vec{x}$ sufficiently near $\vec{x}_0$ and there exists an affine function $T_{\vec{x}_0}f(\vec{x})\colon\mathbb{R}^3\to\mathbb{R}$ which has first-order contact with $f(\vec{x})$ at $\vec{x}=\vec{x}_0$:*

$$\left|f(\vec{x})-T_{\vec{x}_0}f(\vec{x})\right|=\mathfrak{o}(\|\vec{x}-\vec{x}_0\|) \tag{3.3}$$

which is to say

$$\lim_{\vec{x}\to\vec{x}_0}\frac{\left|f(\vec{x})-T_{\vec{x}_0}f(\vec{x})\right|}{\|\vec{x}-\vec{x}_0\|}=0. \tag{3.4}$$

*When such an affine function exists, we call it the **linearization** of $f(\vec{x})$ or the **linear approximation** to $f(\vec{x})$, at $\vec{x}=\vec{x}_0$.*[5]

Since functions with first-order contact must agree at the point of contact, we know that

$$T_{\vec{x}_0}f(\vec{x}_0)=f(\vec{x}_0);$$

then Remark 3.2.2 tells us that

$$T_{\vec{x}_0}f(\vec{x}_0+\triangle\vec{x})=f(x_0)+\ell(\triangle\vec{x}). \tag{3.5}$$

Furthermore, since $\ell(\triangle\vec{x})$ is a polynomial, it is continuous, so that

$$\lim_{\triangle\vec{x}\to\vec{0}}\ell(\triangle\vec{x})=0$$

and

$$\begin{aligned}\lim_{\vec{x}\to\vec{x}_0}\left(f(\vec{x})-T_{\vec{x}_0}f(\vec{x})\right)&=\lim_{\vec{x}\to\vec{x}_0}[f(\vec{x})-f(\vec{x}_0)-\ell(\triangle\vec{x})]\\&=\lim_{\vec{x}\to\vec{x}_0}[f(\vec{x})-f(\vec{x}_0)]-\lim_{\triangle\vec{x}\to\vec{0}}\ell(\triangle\vec{x})\\&=\lim_{\vec{x}\to\vec{x}_0}[f(\vec{x})-f(\vec{x}_0)].\end{aligned}$$

But since the denominator in Equation (3.4) goes to zero, so must the numerator, which says that the limit above is zero. This shows

[5] Properly speaking, it *should* be called the **affine approximation**.

Remark 3.3.2. *If $f(\overrightarrow{x})$ is differentiable at $\overrightarrow{x} = \overrightarrow{x}_0$ then it is continuous there.*

To calculate the "linear part" $\ell(\triangle\overrightarrow{x})$ of $T_{\overrightarrow{x}_0} f(\overrightarrow{x})$ (if it exists), we consider the action of $f(\overrightarrow{x})$ along the line through $\overrightarrow{x}_0$ with a given direction vector $\overrightarrow{v}$: this is parametrized by

$$\overrightarrow{p}(t) = \overrightarrow{x}_0 + t\overrightarrow{v}$$

and the restriction of $f(\overrightarrow{x})$ to this line is given by the composition

$$f(\overrightarrow{p}(t)) = f(\overrightarrow{x}_0 + t\overrightarrow{v}) .$$

Then setting $\triangle\overrightarrow{x} = t\overrightarrow{v}$ in Equation (3.5) we have

$$\begin{aligned} T_{\overrightarrow{x}_0} f(\overrightarrow{x}_0 + t\overrightarrow{v}) &= f(\overrightarrow{x}_0) + \ell(t\overrightarrow{v}) \\ &= f(\overrightarrow{x}_0) + t\ell(\overrightarrow{v}) . \end{aligned}$$

Equation (3.4) then says that, if we let $t \to 0$,

$$\frac{|f(\overrightarrow{x}_0 + t\overrightarrow{v}) - f(\overrightarrow{x}_0) - t\ell(\overrightarrow{v})|}{\|t\overrightarrow{v}\|} \to 0,$$

from which it follows that

$$\ell(\overrightarrow{v}) = \lim_{t\to 0} \frac{1}{t} (f(\overrightarrow{x}_0 + t\overrightarrow{v}) - f(\overrightarrow{x}_0)) .$$

This formula shows that, *if it exists*, the affine approximation to $f(\overrightarrow{x})$ at $\overrightarrow{x} = \overrightarrow{x}_0$ is unique; we call the "linear part" $\ell(\triangle\overrightarrow{x})$ of $T_{\overrightarrow{x}_0} f(\overrightarrow{x})$ the **derivative** or **differential** of $f(\overrightarrow{x})$ at $\overrightarrow{x} = \overrightarrow{x}_0$, and denote it $d_{\overrightarrow{x}_0} f$. Note that this equation can also be interpreted in terms of the derivative at $t = 0$ of the composite function $f(\overrightarrow{p}(t))$:

$$\begin{aligned} d_{\overrightarrow{x}_0} f(\overrightarrow{v}) &= \lim_{t\to 0} \frac{1}{t} (f(\overrightarrow{x}_0 + t\overrightarrow{v}) - f(\overrightarrow{x}_0)) \\ &= \left.\frac{d}{dt}\right|_{t=0} [f(\overrightarrow{x}_0 + t\overrightarrow{v})] . \end{aligned} \tag{3.6}$$

For example, if

$$f(x, y) = x^2 - xy$$

and

$$\begin{aligned} \overrightarrow{x}_0 &= (3,1) \\ \overrightarrow{v} &= (v_1, v_2) \end{aligned}$$

then, using the limit formula,

$$\begin{aligned} d_{(3,1)}f((v_1,v_2)) &= \lim_{t\to 0}\frac{1}{t}\left[f(3+v_1t, 1+v_2t) - f(3,1)\right] \\ &= \lim_{t\to 0}\frac{1}{t}\left[(3+v_1t)^2 - (3+v_1t)(1+v_2t) - 6\right] \\ &= \lim_{t\to 0}\frac{1}{t}\left[(9+6v_1t+t^2v_1^2) - (3+v_1t+3v_2t+t^2v_1v_2) - 6\right] \\ &= \lim_{t\to 0}\left[(5v_1-3v_2) + t(v_1^2 - v_1v_2)\right] \\ &= 5v_1 - 3v_2 \end{aligned}$$

or, alternatively, we could use the differentiation formula:

$$\begin{aligned} d_{(3,1)}f((v_1,v_2)) &= \left.\frac{d}{dt}\right|_{t=0}\left[f(3+v_1t, 1+v_2t)\right] \\ &= \left.\frac{d}{dt}\right|_{t=0}\left[6 + (5v_1-3v_2)t + (v_1^2 - v_1v_2)t^2\right] \\ &= (5v_1 - 3v_2) + 2(v_1^2 - v_1v_2)\cdot 0 \\ &= (5v_1 - 3v_2). \end{aligned}$$

Partial Derivatives

Equation (3.6), combined with Remark 3.2.1, gives us a way of expressing the differential $d_{\overrightarrow{x}_0}f(\overrightarrow{v})$ as a homogeneous polynomial in the components of $\overrightarrow{v}$. The quantity given by Equation (3.6) when $\overrightarrow{v} = \overrightarrow{e}_j$ is an element of the standard basis for $\mathbb{R}^3$, is called a **partial derivative**. It corresponds to moving through $\overrightarrow{x}_0$ parallel to one of the coordinate axes with unit speed—that is, the motion parametrized by

$$\overrightarrow{p}_j(t) = \overrightarrow{x}_0 + t\overrightarrow{e}_j :$$

Definition 3.3.3. *The j^{th} **partial derivative** (or **partial with respect to x_j**) of a function $f(x_1,x_2,x_3)$ of three variables at $\overrightarrow{x} = \overrightarrow{x}_0$ is the derivative*[6]

[6] The symbol $\frac{\partial f}{\partial x_j}$ is pronounced as if the ∂'s were d's.

(if it exists) of the function $(f \circ \overrightarrow{p}_j)(t)$ obtained by fixing all variables except the j^{th} at their values at $\overrightarrow{x}_0$, and letting x_j vary:

$$\begin{aligned} f_{x_j}(\overrightarrow{x}_0) := \frac{\partial f}{\partial x_j}(\overrightarrow{x}_0) &= \left.\frac{d}{dt}\right|_{t=0} [f(\overrightarrow{p}_j(t))] \\ &= \left.\frac{d}{dt}\right|_{t=0} [f(\overrightarrow{x}_0 + t\overrightarrow{e}_j)] \end{aligned}$$

or

$$\begin{aligned} f_x(x,y,z) &:= \frac{\partial f}{\partial x}(x,y,z) = \lim_{t\to 0} \frac{1}{t}\left[f(x+t,y,z) - f(x,y,z)\right] \\ f_y(x,y,z) &:= \frac{\partial f}{\partial y}(x,y,z) = \lim_{t\to 0} \frac{1}{t}\left[f(x,y+t,z) - f(x,y,z)\right] \\ f_z(x,y,z) &:= \frac{\partial f}{\partial z}(x,y,z) = \lim_{t\to 0} \frac{1}{t}\left[f(x,y,z+t) - f(x,y,z)\right]. \end{aligned}$$

In practice, partial derivatives are easy to calculate: we just differentiate, treating all but one of the variables as a constant. For example, if

$$f(x,y) = x^2y + 3x + 4y$$

then $\frac{\partial f}{\partial x}$, the partial with respect to x, is obtained by treating y as the name of some constant:

$$f_x(x,y) := \frac{\partial f}{\partial x}(x,y) = 2xy + 3$$

while the partial with respect to y is found by treating x as a constant:

$$f_y(x,y) := \frac{\partial f}{\partial y}(x,y) = x^2 + 4;$$

similarly, if

$$g(x,y,z) = \sin 2x \cos y + xyz^2$$

then

$$\begin{aligned} g_x(x,y,z) &:= \frac{\partial g}{\partial x}(x,y,z) = 2\cos 2x \cos y + yz^2 \\ g_y(x,y,z) &:= \frac{\partial g}{\partial y}(x,y,z) = -\sin 2x \sin y + xz^2 \\ g_z(x,y,z) &:= \frac{\partial g}{\partial z}(x,y,z) = 2xyz. \end{aligned}$$

Remark 3.2.1 tells us that the differential of f, being linear, is determined by the partials of f:

$$\begin{aligned} d_{\overrightarrow{x}_0} f(\overrightarrow{v}) &= \left(\frac{\partial f}{\partial x_1}(\overrightarrow{x}_0)\right) v_1 + \left(\frac{\partial f}{\partial x_2}(\overrightarrow{x}_0)\right) v_2 + \left(\frac{\partial f}{\partial x_3}(\overrightarrow{x}_0)\right) v_3 \\ &= \sum_{j=1}^{3} \frac{\partial f}{\partial x_j} v_j. \end{aligned}$$

So far, we have avoided the issue of existence: all our formulas above assume that $f(\overrightarrow{x})$ is differentiable at $\overrightarrow{x} = \overrightarrow{x}_0$. Since the partial derivatives of a function are essentially derivatives as we know them from single-variable calculus, it is usually pretty easy to determine whether they exist and if so to calculate them formally. However, the existence of the *partials* is not by itself a guarantee that the function is *differentiable*. For example, the function we considered in § 3.1

$$f(x) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{at } (0,0) \end{cases}$$

has the constant value zero along both axes, so certainly its two partials at the origin exist and equal zero

$$\begin{aligned} \frac{\partial f}{\partial x}(0,0) &= 0 \\ \frac{\partial f}{\partial y}(0,0) &= 0 \end{aligned}$$

but if we try to calculate $d_{(0,0)} f(\overrightarrow{v})$ for the vector $\overrightarrow{v} = (1, m)$ using Equation (3.6),

$$d_{(0,0)} f(1, m) = \lim_{t \to 0} \frac{1}{t} \left(f(t, mt) - f(0,0)\right)$$

then, since along the line $y = mx$ the function has a constant value

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

—which is nonzero if m is—but $f(0,0) = 0$, we see that the limit above does not exist:

$$\lim_{t \to 0} \frac{1}{t} \left(f(t, mt) - f(0,0)\right) = \lim_{t \to 0} \frac{1}{t} \left(\frac{m}{1+m^2}\right)$$

diverges, and the differential cannot be evaluated along the vector $\overrightarrow{v} = \overrightarrow{\imath} + m\overrightarrow{\jmath}$ if $m \neq 0$. In fact, we saw before that this function is not continuous at the origin, which already contradicts differentiability, by Remark 3.3.2.

Another example, this time one which *is* continuous at the origin, is

$$f(x,y) = \begin{cases} \frac{2xy}{\sqrt{x^2+y^2}} & \text{if } (x,y) \neq (0,0), \\ 0 & \text{at } (0,0) \end{cases}.$$

This function is better understood when expressed in polar coordinates, where it takes the form

$$\begin{aligned} f(r\cos\theta, r\sin\theta) &= \frac{2r^2\cos\theta\sin\theta}{r} \\ &= 2r\cos\theta\sin\theta \\ &= r\sin 2\theta. \end{aligned}$$

From this we see that along the line making angle θ with the x-axis, $f(x,y)$ is a constant ($\sin 2\theta$) times the distance from the origin: geometrically, the graph of $f(x,y)$ over this line is itself a line through the origin of slope $m = \sin 2\theta$. Along the two coordinate axes, this slope is zero, but for example along the line $y = x$ ($\theta = \pi/4$), the slope is $\sin\pi/2 = 1$. So this time, the function defined by Equation (3.6) (without asking about differentiability) exists at the origin, but *it is not linear* (since again it is zero on each of the standard basis elements $\overrightarrow{\imath}$ and $\overrightarrow{\jmath}$).

A third example is defined by a straightforward formula (no "cases"):

$$f(x,y) = x^{1/3}y^{1/3}.$$

Again, the function is constant along the coordinate axes, so both partials are zero. However, if we try to evaluate the limit in Equation (3.6) using any vector not pointing along the axes, we get

$$d_{(0,0)}f(\alpha\overrightarrow{\imath} + \beta\overrightarrow{\jmath}) = \left.\frac{d}{dt}\right|_{t=0}\left[\alpha^{1/3}\beta^{1/3}t^{2/3}\right];$$

since $t^{2/3}$ is definitely not differentiable at $t = 0$, the required linear map $d_{(0,0)}f(\alpha\overrightarrow{\imath} + \beta\overrightarrow{\jmath})$ cannot exist.

From all of this, we see that having the partials at a point $\overrightarrow{x}_0$ is not enough to guarantee differentiability of $f(\overrightarrow{x})$ at $\overrightarrow{x} = \overrightarrow{x}_0$. It is not even enough to also have partials at every point near $\overrightarrow{x}_0$—all our examples above have this property. However, a slight tweaking of this last condition *does* guarantee differentiability. We call $f(\overrightarrow{x})$ **continuously differentiable** at $\overrightarrow{x}_0$ if all the partial derivatives exist for every point near $\overrightarrow{x}_0$ (including $\overrightarrow{x}_0$ itself), *and are continuous at* $\overrightarrow{x}_0$. Then we can assert

Theorem 3.3.4. *If $f(\overrightarrow{x})$ is continuously differentiable at $\overrightarrow{x}_0$, then it is differentiable there.*

Proof. For notational convenience, we concentrate on the case of a function of two variables; the modification of this proof to the case of three variables is straightforward (Exercise 11).

We know that, if it exists, the linearization of $f(\overrightarrow{x})$ at $\overrightarrow{x} = \overrightarrow{x}_0 = (x, y)$ is determined by the partials to be

$$T_{\overrightarrow{x}_0}f(x + \triangle x, y + \triangle y) = f(x, y) + \frac{\partial f}{\partial x}(x, y)\triangle x + \frac{\partial f}{\partial y}(x, y)\triangle y; \tag{3.7}$$

so we need to show that

$$\frac{1}{\|(\triangle x, \triangle y)\|}\left|f(x + \triangle x, y + \triangle y) - \left(f(x, y) + \frac{\partial f}{\partial x}(x, y)\triangle x + \frac{\partial f}{\partial y}(x, y)\triangle y\right)\right| \to 0 \tag{3.8}$$

as $(\triangle x, \triangle y) \to 0$. When we remove the parentheses inside the absolute value, we have an expression whose first two terms are $f(x + \triangle x, y + \triangle y) - f(x, y)$; we rewrite this as follows. By adding and subtracting the value of f at a point that shares one coordinate with each of these two points—say $f(x, y + \triangle y)$—we can write

$$\begin{aligned} f(x + \triangle x, y + \triangle y) - f(x, y) &= (f(x + \triangle x, y + \triangle y) - f(x, y + \triangle y)) \\ &\quad + (f(x, y + \triangle y) - f(x, y)) \end{aligned}$$

and then proceed to analyze each of the two quantities in parentheses. Note that the first quantity is the difference between the values of f along a horizontal line segment, which can be parametrized by

$$\overrightarrow{p}(t) = (x + t\triangle x, y + \triangle y), \quad 0 \le t \le 1;$$

the composite function

$$\begin{aligned} g(t) &= f(\overrightarrow{p}(t)) \\ &= f(x + t\triangle x, y + \triangle y) \end{aligned}$$

is an ordinary function of one variable, whose derivative is related to a partial derivative of f (Exercise 6):

$$g'(t) = \frac{\partial f}{\partial x}\left(x + t\triangle x, y + \triangle y\right)\triangle x.$$

Thus, we can apply the Mean Value Theorem to conclude that there is a value $t = t_1$ between 0 and 1 for which

$$g(1) - g(0) = g'(t_1)\,.$$

Letting $t_1\triangle x = \delta_1$, we can write

$$\begin{aligned} f(x+\triangle x, y+\triangle y) - f(x, y+\triangle y) &= g(1) - g(0) \\ &= g'(t_1) \\ &= \frac{\partial f}{\partial x}\left(x+\delta_1, y+\triangle y\right)\triangle x \end{aligned}$$

where

$$|\delta_1| \leq |\triangle x|\,.$$

A similar argument applied to the second term in parentheses (Exercise 6) yields

$$f(x, y+\triangle y) - f(x, y) = \frac{\partial f}{\partial y}\left(x, y+\delta_2\right)\triangle y$$

where

$$|\delta_2| \leq |\triangle y|\,.$$

This allows us to rewrite the quantity inside the absolute value in Equation (3.8) as

$$\begin{aligned} &f(x+\triangle x, y+\triangle y) - \left(f(x,y) + \frac{\partial f}{\partial x}(x,y)\triangle x + \frac{\partial f}{\partial y}(x,y)\triangle y\right) \\ &= \left(f(x+\triangle x, y+\triangle y) - f(x,y)\right) - \left(\frac{\partial f}{\partial x}(x,y)\triangle x + \frac{\partial f}{\partial y}(x,y)\triangle y\right) \\ &= \left(\frac{\partial f}{\partial x}(x+\delta_1, y+\triangle y)\triangle x + \frac{\partial f}{\partial y}(x, y+\delta_2)\triangle y\right) - \left(\frac{\partial f}{\partial x}(x,y)\triangle x + \frac{\partial f}{\partial y}(x,y)\triangle y\right) \\ &= \left(\frac{\partial f}{\partial x}(x+\delta_1, y+\triangle y) - \frac{\partial f}{\partial x}(x,y)\right)\triangle x + \left(\frac{\partial f}{\partial y}(x, y+\delta_2) - \frac{\partial f}{\partial y}(x,y)\right)\triangle y. \end{aligned}$$

Now, we want to show that this quantity, divided by $\|(\triangle x, \triangle y)\| = \sqrt{\triangle x^2 + \triangle y^2}$, goes to zero as $(\triangle x, \triangle y) \to (0, 0)$. Clearly,

$$\frac{|\triangle x|}{\sqrt{\triangle x^2 + \triangle y^2}} \leq 1$$
$$\frac{|\triangle y|}{\sqrt{\triangle x^2 + \triangle y^2}} \leq 1,$$

so it suffices to show that each of the quantities in parentheses goes to zero. But as $(\triangle x, \triangle y) \to 0$, all of the quantities $\triangle x$, $\triangle y$, δ_1 and δ_2 go to zero, which means that all of the points at which we are evaluating the partials are tending to $\overrightarrow{x}_0 = (x, y)$; in particular, the difference inside each pair of (large) parentheses is going to zero. Since each such quantity is being multiplied by a bounded quantity ($\triangle x / \sqrt{\triangle x^2 + \triangle y^2}$, or $\triangle y / \sqrt{\triangle x^2 + \triangle y^2}$), the whole mess goes to zero.

This proves our assertion, that the affine function $T_{\overrightarrow{x}_0} f(\overrightarrow{x})$ as defined by Equation (3.7) has first-order contact with $f(\overrightarrow{x})$ at $\overrightarrow{x} = \overrightarrow{x}_0$. □

This result ensures that functions defined by algebraic or analytic expressions such as polynomials (in two or three variables) or combinations of trigonometric, exponential, logarithmic functions and roots are generally differentiable, since by the formal rules of differentiation the partials are again of this type, and hence are continuous wherever they are defined; the only difficulties arise in cases where differentiation introduces a denominator which becomes zero at the point in question.

The Gradient and Directional Derivatives

Recall from § 3.2 that a linear function can be viewed in three different ways: as a homogeneous *polynomial* of degree one, as multiplication of the coordinate matrix by its *matrix representative*, and as the *dot product* of the input with a fixed vector. We have seen that when $f(\overrightarrow{x})$ is differentiable at $\overrightarrow{x} = \overrightarrow{x}_0$, then the coefficients of the differential $d_{\overrightarrow{x}_0} f(\overrightarrow{v})$, as a polynomial in the entries of $\overrightarrow{v}$, are the partial derivatives of f at $\overrightarrow{x} = \overrightarrow{x}_0$; this tells us that the matrix representative of $d_{\overrightarrow{x}_0} f$ is

$$\left[d_{\overrightarrow{x}_0} f \right] = \left[\frac{\partial f}{\partial x} (\overrightarrow{x}_0), \frac{\partial f}{\partial y} (\overrightarrow{x}_0), \frac{\partial f}{\partial z} (\overrightarrow{x}_0) \right].$$

This matrix is sometimes referred to as the **Jacobian** of f at $\overrightarrow{x} = \overrightarrow{x}_0$, and denoted Jf. Equivalently, when we regard this row as a vector, we get the

gradient[7] of f:

$$\vec{\nabla} f(\vec{x}_0) = \left(\frac{\partial f}{\partial x}(\vec{x}_0), \frac{\partial f}{\partial y}(\vec{x}_0), \frac{\partial f}{\partial z}(\vec{x}_0)\right).$$

That is, *the gradient is the vector whose entries are the partials.* It is worth reiterating how these two objects represent the differential:

$$\begin{aligned} d_{\vec{x}_0} f(\vec{v}) &= (Jf(\vec{x}_0))\,[\vec{v}] \qquad \text{(Matrix product)} \qquad (3.9)\\ &= \vec{\nabla} f(\vec{x}_0)\cdot\vec{v} \qquad \text{(Dot Product).} \qquad (3.10) \end{aligned}$$

These ways of representing the differential carry essentially the same information. However, the gradient in particular has a nice geometric interpretation.

Recall that we represent a direction in the plane or space by means of a unit vector $\vec{u}$. When the differential is applied to such a vector, the resulting number is called the **directional derivative** of the function at the point. From Equation (3.6), we see that *the directional derivative gives the rate at which $f(\vec{x})$ changes as we move in the direction $\vec{u}$ at speed one.* In the plane, a unit vector has the form

$$\vec{u}_\alpha = (\cos\alpha)\vec{\imath} + (\sin\alpha)\vec{\jmath}$$

where α is the angle our direction makes with the x-axis. In this case, the directional derivative in the direction given by α is

$$d_{\vec{x}_0} f(\vec{u}_\alpha) = \frac{\partial f}{\partial x}(\vec{x}_0)\cos\alpha + \frac{\partial f}{\partial y}(\vec{x}_0)\sin\alpha.$$

Equation (3.10) tells us that the directional derivative in the direction of the unit vector $\vec{u}$ is the dot product

$$d_{\vec{x}_0} f(\vec{u}) = \vec{\nabla} f\cdot\vec{u}$$

which is related to the angle θ between the two vectors, so also

$$\begin{aligned} \vec{\nabla} f\cdot\vec{u} &= \left\|\vec{\nabla} f(\vec{x}_0)\right\| \|\vec{u}\|\cos\theta\\ &= \left\|\vec{\nabla} f(\vec{x}_0)\right\|\cos\theta \end{aligned}$$

since $\vec{u}$ is a unit vector. Now, $\cos\theta$ reaches its *maximum* value, which is 1, when $\theta = 0$, which is to say when $\vec{u}$ points in the direction of $\vec{\nabla}\vec{x}_0$, and its *minimum* value of -1 when $\vec{u}$ points in the *opposite* direction. This gives us a geometric interpretation of the gradient, which will prove very useful.

[7] The symbol $\vec{\nabla} f$ is pronounced "grad f"; another notation for the gradient is grad f.

Remark 3.3.5. *The gradient vector* $\vec{\nabla} f(\vec{x}_0)$ *points in the* direction *in which the directional derivative has its highest value, known as the* ***direction of steepest ascent****, and its* length *is the* value *of the directional derivative in that direction.*

As an example, consider the function

$$f(x,y) = 49 - x^2 - y^2$$

at the point

$$\vec{x}_0 = (4,1).$$

The graph of this function is an elliptic paraboloid opening down; that is, it can be viewed as a hill whose peak is above the origin, at height $f(0,0) = 49$. The gradient of this function is

$$\vec{\nabla} f(x,y) = (-2x)\vec{\imath} + (-6x)\vec{\jmath};$$

at the point $(4,1)$,

$$\vec{\nabla} f(4,1) = -8\vec{\imath} - 6\vec{\jmath}$$

has length

$$\begin{aligned} \left\|\vec{\nabla} f(4,1)\right\| &= \sqrt{8^2 + 6^2} \\ &= 10 \end{aligned}$$

and the unit vector parallel to $\vec{\nabla} f(4,1)$ is

$$\vec{u} = -\frac{4}{5}\vec{\imath} - \frac{3}{5}\vec{\jmath}.$$

This means that at the point 4 units east and one unit north of the peak, a climber who wishes to gain height as fast as possible should move in the direction given on the map by $\vec{u}$; by moving in this direction, the climber will be ascending at 10 units per unit of horizontal motion from an initial height of $f(4,1) = 30$. Alternatively, if a stream flowing down the mountain passes the point 4 units east and one unit north of the peak, its direction of flow on the map will be in the opposite direction, the *direction of steepest descent.*

The analogue of Remark 3.3.5 for a function of three variables holds for the same reasons. Note that in either case, the gradient "lives" in the domain of the function; thus, although the *graph* of a function of two variables is a surface in space, its gradient vector at any point is a vector in the plane.

Chain Rules

For two differentiable real-valued functions of a (single) real variable, the Chain Rule tells us that the derivative of the composition is the product of the derivatives of the two functions:

$$\left.\frac{d}{dt}\right|_{t=t_0} [f \circ g] = f'(g(t_0)) \cdot g'(t_0)\,.$$

Similarly, if g is a differentiable real-valued function of a real variable and $\overrightarrow{f}$ is a differentiable vector-valued function of a real variable, the composition $\overrightarrow{f} \circ g$ is another vector-valued function, whose derivative is the product of the derivative of $\overrightarrow{f}$ (*i.e.*, , its velocity) and the derivative of g:

$$\left.\frac{d}{dt}\right|_{t=t_0} \left[\overrightarrow{f} \circ g\right] = \vec{f}'(g(t_0)) \cdot g'(t_0)\,.$$

We would now like to turn to the case when $\overrightarrow{g}$ is a vector-valued function of a real variable, and f is a real-valued function of a vector variable, so that their composition $f \circ \overrightarrow{g}$ is a real-valued function of a real variable. We have already seen that if $\overrightarrow{g}$ is steady motion along a straight line, then the derivative of the composition is the same as the action of the differential of f on the derivative (*i.e.*, the velocity) of $\overrightarrow{g}$. We would like to say that this is true in general. For ease of formulating our result, we shall use the notation $\overrightarrow{p}(t)$ in place of $\overrightarrow{g}(t)$, $\overrightarrow{v}$ for the velocity of $\overrightarrow{p}(t)$ at $t = t_0$, and the representation of $d_{\overrightarrow{x}_0} f(\overrightarrow{v})$ as $\overrightarrow{\nabla} f(\overrightarrow{x}_0) \cdot \overrightarrow{v}$.

Proposition 3.3.6 (Chain Rule for $\mathbb{R} \to \mathbb{R}^3 \to \mathbb{R}$)**.** *Suppose $f{:}\,\mathbb{R}^3 \to \mathbb{R}$ is differentiable at $\overrightarrow{x} = \overrightarrow{x}_0$ and $\overrightarrow{p}{:}\,\mathbb{R} \to \mathbb{R}^3$ is a vector-valued function which is differentiable at $t = t_0$, where $\overrightarrow{p}(t_0) = \overrightarrow{x}_0$.*

Then the composite function $(f \circ \overrightarrow{p}){:}\,\mathbb{R} \to \mathbb{R}$, $(f \circ \overrightarrow{p})(t) = f(\overrightarrow{p}(t))$ is differentiable at $t = t_0$, and

$$\left.\frac{d}{dt}\right|_{t=t_0} [f \circ \overrightarrow{p}] = \overrightarrow{\nabla} f(\overrightarrow{x}_0) \cdot \overrightarrow{v}$$

where

$$\vec{v} = \dot{\vec{p}}(t_0)$$

is the velocity with which the curve passes $\vec{x}_0$ at $t = t_0$.

Proof. For the purpose of this proof, it will be convenient to write the condition that $f(\vec{x})$ and $T_{\vec{x}_0}f(\vec{x})$ have first-order contact at $\vec{x} = \vec{x}_0$ in a somewhat different form. If we set

$$\varepsilon = \frac{\left|f(\vec{x}_0 + \triangle\vec{x}) - T_{\vec{x}_0}f(\vec{x}_0 + \triangle\vec{x})\right|}{\|\triangle\vec{x}\|}$$

where $\triangle\vec{x} = \vec{x} - \vec{x}_0$, then Equation (3.4) can be rewritten in the form

$$f(\vec{x}_0 + \triangle\vec{x}) = T_{\vec{x}_0}f(\vec{x}_0 + \triangle\vec{x}) + \|\triangle\vec{x}\|\,\varepsilon,$$
$$\text{where } \varepsilon \to 0 \text{ as } \triangle\vec{x} \to \vec{0}.$$

If we substitute into this the expression for the affine approximation

$$T_{\vec{x}_0}f(\vec{x}_0 + \triangle\vec{x}) = f(\vec{x}_0) + d_{\vec{x}_0}f(\triangle\vec{x})$$

we obtain the following version of Equation (3.4):

$$f(\vec{x}_0 + \triangle\vec{x}) - f(\vec{x}_0) = d_{\vec{x}_0}f(\triangle\vec{x}) + \|\triangle\vec{x}\|\,\varepsilon,$$
$$\text{where } \varepsilon \to 0 \text{ as } \triangle\vec{x} \to \vec{0}.$$

Using the representation of $d_{\vec{x}_0}f(\triangle\vec{x})$ as a dot product, we can rewrite this in the form

$$f(\vec{x}_0 + \triangle\vec{x}) - f(\vec{x}_0) = \vec{\nabla}f(\vec{x}_0)\cdot\triangle\vec{x} + \|\triangle\vec{x}\|\,\varepsilon,$$
$$\text{where } \varepsilon \to 0 \text{ as } \triangle\vec{x} \to \vec{0}.$$

In a similar way, we can write the analogous statement for $\vec{p}(t)$, using $\vec{v} = \dot{\vec{p}}(t_0)$:

$$\vec{p}(t_0 + \triangle t) - \vec{p}(t_0) = \vec{v}\triangle t + |\triangle t|\,\vec{\delta},$$
$$\text{where } \vec{\delta} \to \vec{0} \text{ as } \triangle t \to 0.$$

Now, we consider the variation of the composition $f(\vec{p}(t))$ as t varies from $t = t_0$ to $t = t_0 + \triangle t$:

$$\begin{aligned} f(\vec{p}(t_0 + \triangle t)) - f(\vec{p}(t_0)) &= \vec{\nabla}f(\vec{x}_0)\cdot\left(\vec{v}\triangle t + |\triangle t|\,\vec{\delta}\right) + \|\triangle\vec{x}\|\,\varepsilon \\ &= \vec{\nabla}f(\vec{x}_0)\cdot(\vec{v}\triangle t) + |\triangle t|\,\vec{\nabla}f(\vec{x}_0)\cdot\vec{\delta} + \|\triangle\vec{x}\|\,\varepsilon. \end{aligned}$$

Subtracting the first term on the right from both sides, we can write

$$\begin{aligned} f(\overrightarrow{p}(t_0+\triangle t)) - f(\overrightarrow{p}(t_0)) - \overrightarrow{\nabla} f(\overrightarrow{x}_0)\cdot(\overrightarrow{v}\triangle t) \\ = (\triangle t)\overrightarrow{\nabla} f(\overrightarrow{x}_0)\cdot\overrightarrow{\delta} + |\triangle t|\left\|\frac{\triangle\overrightarrow{x}}{\triangle t}\right\|\varepsilon. \end{aligned}$$

Taking the absolute value of both sides and dividing by $|\triangle t|$, we get

$$\begin{aligned} \frac{1}{|\triangle t|}\left| f(\overrightarrow{p}(t_0+\triangle t)) - f(\overrightarrow{p}(t_0)) - \overrightarrow{\nabla} f(\overrightarrow{x}_0)\cdot(\triangle t\overrightarrow{v})\right| \\ = \left|\overrightarrow{\nabla} f(\overrightarrow{x}_0)\cdot\overrightarrow{\delta} \pm \left\|\frac{\triangle\overrightarrow{x}}{\triangle t}\right\|\varepsilon\right| \\ \leq \left\|\overrightarrow{\nabla} f(\overrightarrow{x}_0)\right\|\left\|\overrightarrow{\delta}\right\| + \left\|\frac{\triangle\overrightarrow{x}}{\triangle t}\right\||\varepsilon|. \end{aligned}$$

In the first term above, the first factor is fixed and the second goes to zero as $\triangle t\to 0$, while in the second term, the first factor is bounded (since $\triangle\overrightarrow{x}/\triangle t$ converges to $\overrightarrow{v}$) and the second goes to zero. Thus, the whole mess goes to zero, proving that the affine function inside the absolute value in the numerator on the left above represents the linearization of the composition, as required. □

An important aspect of Proposition 3.3.6 (perhaps *the* important aspect) is that the rate of change of a function applied to a moving point depends *only* on the gradient of the function and the velocity of the moving point at the given moment, *not* on how the motion might be accelerating, etc.

For example, consider the distance from a moving point $\overrightarrow{p}(t)$ to the point $(1,2)$: the distance from (x,y) to $(1,2)$ is given by

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

with gradient

$$\overrightarrow{\nabla} f(x,y)=\frac{(x-1)}{\sqrt{(x-1)^2+(y-1)^2}}\overrightarrow{\imath}+\frac{(y-2)}{\sqrt{(x-1)^2+(y-1)^2}}\overrightarrow{\jmath}.$$

If at a given moment our point has position

$$\overrightarrow{p}(t_0)=(5,-3)$$

and velocity

$$\overrightarrow{v}(t_0) = -2\overrightarrow{\imath} - 3\overrightarrow{\jmath}$$

then regardless of acceleration and so on, the rate at which its distance from $(1, 2)$ is changing is given by

$$\begin{aligned}\left.\frac{d}{dt}\right|_{t=t_0} [f(\overrightarrow{p}(t))] &= \overrightarrow{\nabla} f(5, -3) \cdot \overrightarrow{v}(t_0) \\ &= \left(\frac{4}{5}\overrightarrow{\imath} - \frac{3}{5}\overrightarrow{\jmath}\right) \cdot (-2\overrightarrow{\imath} - 3\overrightarrow{\jmath}) \\ &= -\frac{8}{5} + \frac{9}{5} \\ &= \frac{1}{5}.\end{aligned}$$

The other kind of chain rule that can arise is when we compose a real-valued function f of a vector variable with a real-valued function g of a real variable:

Proposition 3.3.7 (Chain Rule $\mathbb{R}^3 \to \mathbb{R} \to \mathbb{R}$)**.** *Suppose $f(\overrightarrow{x})$ is a real-valued function of three variables, differentiable at $v_= \overrightarrow{x}_0$, and $g(y)$ is a real-valued function of a real variable, differentiable at $y = y_0 = f(\overrightarrow{x}_0)$.*

Then the composition $(g \circ f)(\overrightarrow{x})$ is differentiable at $\overrightarrow{x} = \overrightarrow{x}_0$, and

$$\overrightarrow{\nabla}(g \circ f)(\overrightarrow{x}_0) = g'(y_0)\, \overrightarrow{\nabla} f(\overrightarrow{x}_0)\,.$$

Proof. This is formally very similar to the preceding proof. Let

$$\triangle y = f(\overrightarrow{x}_0 + \triangle \overrightarrow{x}) - f(\overrightarrow{x}_0)$$

then

$$\triangle y = \overrightarrow{\nabla} f(\overrightarrow{x}_0) \cdot \triangle \overrightarrow{x} + \delta \left\|\triangle \overrightarrow{x}\right\|$$

where $\delta \to 0$ as $\triangle \overrightarrow{x} \to \overrightarrow{0}$. Note for future reference that

$$|\triangle y| \leq \left(\left\|\overrightarrow{\nabla} f(\overrightarrow{x}_0)\right\| + \delta\right) \left\|\triangle \overrightarrow{x}\right\|.$$

Now,

$$\begin{aligned}g(f(\overrightarrow{x}_0 + \triangle \overrightarrow{x})) &= g(y_0 + \triangle y) \\ &= g(y_0) + g'(y_0)\, \triangle y + \varepsilon\, |\triangle y|\end{aligned}$$

where $\varepsilon \to 0$ as $\triangle y \to 0$. From this we can conclude

$$g(f(\overrightarrow{x}_0 + \triangle\overrightarrow{x})) - g(f(\overrightarrow{x}_0)) - \left[g'(y_0)\,\overrightarrow{\nabla} f(\overrightarrow{x}_0)\right] \\ = g'(y_0)\,\delta\,\|\triangle\overrightarrow{x}\| + \varepsilon\,|\triangle y|\,.$$

Taking absolute values and dividing by $\|\triangle\overrightarrow{x}\|$, we have

$$\frac{1}{\triangle\overrightarrow{x}}\left|g(f(\overrightarrow{x}_0 + \triangle\overrightarrow{x})) - g(f(\overrightarrow{x}_0)) - \left[g'(y_0)\,\overrightarrow{\nabla} f(\overrightarrow{x}_0)\right]\right| \\ \leq \left|g'(y_0)\right|\,|\delta| + \varepsilon\frac{|\triangle y|}{\|\triangle\overrightarrow{x}\|} \\ = \left|g'(y_0)\right|\,|\delta| + \varepsilon\left(\left\|\overrightarrow{\nabla} f(\overrightarrow{x}_0)\right\| + \delta\right).$$

Both terms consist of a bounded quantity times one that goes to zero as $\triangle\overrightarrow{x} \to \overrightarrow{0}$, and we are done. □

Finally, we note that, as a corollary of Proposition 3.3.7, we get a formula for the partial derivatives of the composite function $g \circ f$:

$$\frac{\partial g \circ f}{\partial x_i}(\overrightarrow{x}_0) = g'(y_0)\,\frac{\partial f}{\partial x_i}(\overrightarrow{x}_0)\,. \tag{3.11}$$

For example, suppose we consider the function that expresses the rectangular coordinate y in terms of spherical coordinates:

$$f(\rho, \phi, \theta) = \rho \sin\phi \sin\theta;$$

its gradient is

$$\overrightarrow{\nabla} f(\rho, \phi, \theta) = (\sin\phi\sin\theta, \rho\cos\phi\sin\theta, \rho\sin\phi\cos\theta).$$

Suppose further that we are interested in

$$\begin{aligned} z &= g(y) \\ &= \ln y. \end{aligned}$$

To calculate the partial derivatives of z with respect to the spherical coordinates when

$$\begin{aligned} \rho &= 2 \\ \phi &= \frac{\pi}{4} \\ \theta &= \frac{\pi}{3} \end{aligned}$$

we calculate the value and gradient of f at this point:

$$\begin{aligned} f\left(2,\frac{\pi}{4},\frac{\pi}{3}\right) &= (2)\left(\frac{1}{\sqrt{2}}\right)(\frac{1}{2}) \\ &= \frac{1}{\sqrt{2}} \\ \overrightarrow{\nabla} f\left(2,\frac{\pi}{4},\frac{\pi}{3}\right) &= \left(\frac{\sqrt{3}}{2\sqrt{2}},\frac{\sqrt{3}}{\sqrt{2}},\frac{1}{\sqrt{2}}\right) \end{aligned}$$

or

$$\begin{aligned} \frac{\partial f}{\partial \rho} &= \frac{\sqrt{3}}{2\sqrt{2}} \\ \frac{\partial f}{\partial \phi} &= \frac{\sqrt{3}}{\sqrt{2}} \\ \frac{\partial f}{\partial \theta} &= \frac{1}{\sqrt{2}}. \end{aligned}$$

The value and derivative of $g(y)$ at $y = f\left(2,\frac{\pi}{4},\frac{\pi}{3}\right) = \frac{1}{\sqrt{2}}$ are

$$\begin{aligned} g\left(\frac{1}{\sqrt{2}}\right) &= -\frac{1}{2}\ln 2; \\ g'\left(\frac{1}{\sqrt{2}}\right) &= \sqrt{2} \end{aligned}$$

and from this we get

$$\begin{aligned}
\frac{\partial z}{\partial \rho} &= g'\left(\frac{1}{\sqrt{2}}\right)\frac{\partial f}{\partial \rho}\\
&= \sqrt{2}\left(\frac{\sqrt{3}}{2\sqrt{2}}\right)\\
&= \frac{\sqrt{3}}{2}\\
\frac{\partial z}{\partial \phi} &= g'\left(\frac{1}{\sqrt{2}}\right)\frac{\partial f}{\partial \phi}\\
&= \sqrt{2}\left(\frac{\sqrt{3}}{\sqrt{2}}\right)\\
&= \sqrt{3}\\
\frac{\partial z}{\partial \theta} &= g'\left(\frac{1}{\sqrt{2}}\right)\frac{\partial f}{\partial \theta}\\
&= \sqrt{2}\left(\frac{1}{\sqrt{2}}\right)\\
&= 1.
\end{aligned}$$

Note that this formula could have been found directly, using Definition 3.3.3 (Exercise 7): the substantive part of the proof above was to show that the composite function is differentiable.

Approximation and Estimation

Just as for functions of one variable, the linearization of a function can be used to get "quick and dirty" estimates of the value of a function when the input is close to one where the exact value is known.

For example, consider the function

$$f(x,y) = \sqrt{x^2 + 5xy + y^2};$$

you can check that $f(3,1) = 5$; what is $f(2.9, 1.2)$? We calculate the partial derivatives at $(3,1)$:

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y) &= \frac{2x + 5y}{2\sqrt{x^2 + 5xy + y^2}}\\
\frac{\partial f}{\partial y}(x,y) &= \frac{5x + 2y}{2\sqrt{x^2 + 5xy + y^2}}
\end{aligned}$$

so

$$\begin{aligned}\frac{\partial f}{\partial x}(3,1) &= \frac{11}{10} = 1.1\\ \frac{\partial f}{\partial y}(3,1) &= \frac{17}{10} = 1.7;\end{aligned}$$

since

$$(2.9, 1.2) = (3,1) + (-0.1, 0.2)$$

we use

$$\begin{aligned}\triangle x &= -0.1,\\ \triangle y &= 0.2\end{aligned}$$

to calculate the linearization

$$\begin{aligned}T_{(3,1)}f(2.9,1.2) &= f(3,1) + \frac{\partial f}{\partial x}(3,1)\,\triangle x + \frac{\partial f}{\partial y}(3,1)\,\triangle y\\ &= 5 + (1.1)(-0.1) + (1.7)(0.2)\\ &= 5 - 0.11 + 0.34\\ &= 5.23.\end{aligned}$$

This is an easy calculation, but the answer is only an estimate; by comparison, a calculator "calculation" of $f(2.9, 1.2)$ gives $\sqrt{27.25} \approx 5.220$.

As a second example, we consider the accuracy of the result of the calculation of a quantity whose inputs are only known approximately. Suppose, for example, that we have measured the height of a rectangular box as 2 feet, with an accuracy of $\pm 0.1ft$, and its a base as 5×10 feet, with an accuracy in each dimension of $\pm 0.2ft$. We calculate the volume as $100ft^3$; how accurate is this? Here we are interested in how far the actual value of $f(x,y,z) = xyz$ can vary from $f(5,10,2) = 100$ when x and y can vary by at most $\triangle x = \triangle y = \pm 0.2$ and z can vary by at most $\triangle z = \pm 0.1$. The best estimate of this is the differential:

$$\begin{aligned}f(x,y,z) &= xyz\\ \frac{\partial f}{\partial x}(x,y,z) &= yz\\ \frac{\partial f}{\partial y}(x,y,z) &= xz\\ \frac{\partial f}{\partial z}(x,y,z) &= xy\end{aligned}$$

and at our point

$$\begin{aligned}\frac{\partial f}{\partial x}(5,10,2) &= 20\\ \frac{\partial f}{\partial y}(5,10,2) &= 10\\ \frac{\partial f}{\partial z}(5,10,2) &= 50\end{aligned}$$

so the differential is

$$d_{(5,10,2)}f(\triangle x, \triangle y, \triangle z) = 20\triangle x + 10\triangle y + 50\triangle z$$

which is at most

$$\begin{aligned}(20)(0.2) + (10)(0.2) + (50)(0.1) &= 4 + 2 + 5\\ &= 11.\end{aligned}$$

We conclude that the figure of 100 cubic feet is correct to within ±11 cubic feet.

Exercises for § 3.3

Practice problems:

1. Find all the partial derivatives of each function below:

   (a) $f(x,y) = x^2y - 2xy^2$ (b) $f(x,y) = x\cos y + y\sin x$

   (c) $f(x,y) = e^x\cos y + y\tan x$ (d) $f(x,y) = (x+1)^2y^2 - x^2(y-1)^2$

   (e) $f(x,y,z) = x^2y^3z$ (f) $f(x,y,z) = \dfrac{xy + xz + yz}{xyz}$

2. For each function below, find its derivative $d_{\overrightarrow{a}}f(\triangle\overrightarrow{x})$, the linearization $T_{\overrightarrow{a}}f(\overrightarrow{x})$, and the gradient $\operatorname{grad} f(\overrightarrow{a}) = \overrightarrow{\nabla}f(\overrightarrow{a})$ at the given point $\overrightarrow{a}$.

   (a) $f(x,y) = x^2 + 4xy + 4y^2, \quad \overrightarrow{a} = (1,-2)$

   (b) $f(x,y) = \cos(x^2+y), \quad \overrightarrow{a} = (\sqrt{\pi}, \frac{\pi}{3})$

   (c) $f(x,y) = \sqrt{x^2+y^2}, \quad \overrightarrow{a} = (1,-1)$

   (d) $f(x,y) = x\cos y - y\cos x, \quad \overrightarrow{a} = (\frac{\pi}{2}, -\frac{\pi}{2})$

(e) $f(x,y,z) = xy + xz + yz, \quad \overrightarrow{a} = (1,-2,3)$

(f) $f(x,y,z) = (x+y)^2 - (x-y)^2 + 2xyz, \quad \overrightarrow{a} = (1,2,1)$

3. (a) Use the linearization of $f(x,y) = \sqrt{xy}$ at $\overrightarrow{a} = (9,4)$ to find an approximation to $\sqrt{(8.9)(4.2)}$. (Give your approximation to four decimals.)

   (b) A cylindrical tin can is $h = 3$ inches tall and its base has radius $r = 2$ inches. If the can is made of tin that is 0.01 inches thick, use the differential of $V(r,h) = \pi r^2 h$ to estimate the total volume of tin in the can.

4. If two resistors with respective resistance R_1 and R_2 are hooked up in parallel, the net resistance R is related to R_1 and R_2 by

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}.$$

   (a) Show that the differential of $R = R(R_1, R_2)$, as a function of the two resistances, is given by

$$dR = \left(\frac{R}{R_1}\right)^2 \triangle R_1 + \left(\frac{R}{R_2}\right)^2 \triangle R_2.$$

   (b) If we know $R_1 = 150$ ohms and $R_2 = 400$ ohms, both with a possible error of 10%, what is the net resistance, and what is the possible error?

5. A moving point starts at location $(1,2)$ and moves with a fixed speed; in which of the following directions is the sum of its distances from $(-1,0)$ and $(1,0)$ increasing the fastest?

$$\begin{aligned}
&\overrightarrow{v}_1 \text{ is parallel to } \overrightarrow{\imath}\\
&\overrightarrow{v}_2 \text{ is parallel to } \overrightarrow{\jmath}\\
&\overrightarrow{v}_3 \text{ is parallel to } \overrightarrow{\imath} + \overrightarrow{\jmath}\\
&\overrightarrow{v}_4 \text{ is parallel to } \overrightarrow{\jmath} - \overrightarrow{\imath}.
\end{aligned}$$

   In what direction (among *all* possible directions) will this sum increase the fastest?

Theory problems:

6. Fill in the following details in the proof of Theorem 3.3.4:

   (a) Show that if $f(x,y)$ is differentiable at (x,y) and $g(t)$ is defined by
   $$g(t) = f(x + t\triangle x, y + \triangle y)$$
   then g is differentiable at $t = 0$ and
   $$g'(t) = \frac{\partial f}{\partial x}(x + t\triangle x, y + \triangle y)\,\triangle x.$$

   (b) Show that we can write
   $$f(x, y + \triangle y) - f(x,y) = \frac{\partial f}{\partial y}(x, y + \delta_2)\,\triangle y$$
   where
   $$|\delta_2| \leq |\triangle y|\,.$$

7. (a) Use Proposition 3.3.7 to prove Equation (3.11).

   (b) Use Definition 3.3.3 to prove Equation (3.11) directly.

8. Show that if $f(x,y)$ and $g(x,y)$ are both differentiable real-valued functions of two variables, then so is their product
   $$h(x,y) = f(x,y)\,g(x,y)$$
   and the following Leibniz formula holds:
   $$\overrightarrow{\nabla} h = f\overrightarrow{\nabla} g + g\overrightarrow{\nabla} f.$$

Challenge problem:

9. Show that the if $f(x,y) = g(ax + by)$ where $g(t)$ is a differentiable function of one variable, then for every point (x,y) in the plane with equation $ax + by = c$ (for some constant c), $\overrightarrow{\nabla} f$ is perpendicular to this plane.

10. (a) Show that if $f(x,y)$ is a function whose value depends only on the product xy then
   $$x\frac{\partial f}{\partial x} = y\frac{\partial f}{\partial y}.$$

(b) Is the converse true? That is, suppose $f(x, y)$ is a function satisfying the condition above on its partials. Can it be expressed as a function of the product

$$f(x, y) = g(xy)$$

for some real-valued function $g(t)$ of a real variable? (*Hint:* First, consider two points in the same quadrant, and join them with a path on which the product xy is constant. Not that this cannot be done if the points are in different quadrants.)

11. Adapt the proof of Theorem 3.3.4 given in this section for functions of two variables to get a proof for functions of three variables.

3.4 Level Sets

A **level set** of a function f is any subset of its domain of the form

$$\mathcal{L}(f, c) := \{\overrightarrow{x} \mid f(\overrightarrow{x}) = c\}$$

where $c \in \mathbb{R}$ is some constant. This is nothing other than the solution set of the equation in two or three variables

$$f(x, y) = c$$

or

$$f(x, y, z) = c.$$

For a function of two variables, we expect this set to be a curve in the plane and for three variables we expect a surface in space.

Level Curves and Implicit Differentiation

For a function of two variables, there is another way to think about the level set, which in this case is called a **level curve**: the *graph* of $f(x, y)$ is the locus of the equation

$$z = f(x, y)$$

which is a surface in space, and $\mathcal{L}(f, c)$ is found by intersecting this surface with the horizontal plane $z = c$, and then projecting the resulting curve onto the xy-plane. Of course, this is a "generic" picture: if for example the function itself happens to be constant, then its level set is the xy-plane for one value, and the empty set for all others. We can cook up other examples for which the level set is quite exotic. However, for many functions, the level sets really are curves.

For example, the level curves of a non-constant affine function are parallel straight lines.

The level curves of the function

$$f(x,y)=x^2+y^2$$

are concentric circles centered at the origin for $c>0$, just the origin for $c=0$, and the empty set for $c<0$.

For the function

$$f(x,y)=\frac{x^2}{4}+y^2$$

the level sets $\mathcal{L}(f,c)$ for $c>0$ are the ellipses centered at the origin

$$\frac{x^2}{4c^2}+\frac{y^2}{c^2}=1$$

which all have the same eccentricity. For $c=0$, we again get just the origin, and for $c<0$ the empty set.

The level curves of the function

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

are hyperbolas in general: for $c=a^2>0$, $\mathcal{L}(f,c)$ is the hyperbola

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

which “opens” left and right, and for $c=-a^2<0$ we have

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

which “opens” up and down. For $c=0$ we have the common asymptotes of all these hyperbolas.

We would like to establish criteria for when a level set of a function $f(x,y)$ will be a regular curve. This requires in particular that the curve have a well-defined tangent line. We have often found the slope of the tangent to the locus of an equation via implicit differentiation: for example to find the slope of the tangent to the ellipse

$$x^2 + 4y^2 = 8 \tag{3.12}$$

at the point $(2, -1)$, we think of y as a function of x and differentiate both sides to obtain

$$2x + 8y\frac{dy}{dx} = 0; \tag{3.13}$$

then substituting $x = 2$ and $y = -1$ yields

$$4 - 8\frac{dy}{dx} = 0$$

which we can solve for dy/dx:

$$\frac{dy}{dx} = \frac{4}{8} = \frac{1}{2}.$$

However, the process can break down: at the point $(2\sqrt{2}, 0)$, substitution into (3.13) yields

$$4\sqrt{2} + 0\frac{dy}{dx} = 0$$

which has no solutions. Of course, here we can instead differentiate (3.12) treating x as a function of y, to get

$$2x\frac{dx}{dy} + 8y = 0$$

which upon substituting $x = 2\sqrt{2}$, $y = 0$ yields

$$4\sqrt{2}\frac{dx}{dy} + 0 = 0$$

which *does* have a solution,

$$\frac{dx}{dy} = 0.$$

In this case, we can see the reason for our difficulty by explicitly solving the original equation (3.12) for y in terms of x: near $(2, -1)$, y can be expressed as the function of x

$$\begin{aligned} y &= -\sqrt{\frac{8 - x^2}{4}} \\ &= -\sqrt{2 - \frac{x^2}{4}}. \end{aligned}$$

(We need the minus sign to get $y = -1$ when $x = 2$.) Note that this solution is *local*: near $(2, 1)$ we would need to use the positive root. Near $(2\sqrt{2}, 0)$, we cannot solve for y in terms of x, because the "vertical line test" fails: for any x-value slightly below $x = 2\sqrt{2}$, there are two distinct points with this abcissa (corresponding to the two signs for the square root). However, near this point, the "*horizontal* line test" works: to each y-value near $y = 0$, there corresponds a unique x-value near $x = 2\sqrt{2}$ yielding a point on the ellipse, given by

$$x = \sqrt{8 - 4y^2}.$$

While we are able in this particular case to determine what works and what doesn't, in other situations an explicit solution for one variable in terms of the other is not so easy. For example, the curve

$$x^3 + xy + y^3 = 13 \tag{3.14}$$

contains the point $(3, -2)$. We cannot easily solve this for y in terms of x, but implicit differentiation yields

$$3x^2 + y + x\frac{dy}{dx} + 3y^2\frac{dy}{dx} = 0 \tag{3.15}$$

and substituting $x = 3$, $y = -2$ we get the equation

$$27 - 2 + 3\frac{dy}{dx} + 12\frac{dy}{dx} = 0$$

which is easily solved for dy/dx:

$$\begin{aligned} 15\frac{dy}{dx} &= -25 \\ \frac{dy}{dx} &= -\frac{5}{3}. \end{aligned}$$

It seems that we have found the slope of the line tangent to the locus of Equation (3.14) at the point $(3, -2)$; but how do we know that this line even exists?

A clue to what is going on can be found by recasting the process of implicit differentiation in terms of level curves. Suppose that near the point (x_0, y_0) on the level set $\mathcal{L}(f, c)$

$$f(x, y) = c \tag{3.16}$$

we can (in principle) solve for y in terms of x:

$$y = \phi(x)\,.$$

Then the graph of this function can be parametrized as

$$\overrightarrow{p}(x) = (x, \phi(x)).$$

Since this function is a solution of Equation (3.16) for y in terms of x, its graph lies on $\mathcal{L}(f, c)$:

$$f(\overrightarrow{p}(x)) = f(x, \phi(x)) = c.$$

Applying the Chain Rule to the composition, we can differentiate both sides of this to get

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0$$

and, *provided the derivative $\partial f/\partial y$ is not zero*, we can solve this for $\phi'(x) = dy/dx$:

$$\phi'(x) = \frac{dy}{dx} = -\frac{\partial f/\partial x}{\partial f/\partial y}.$$

This process breaks down if $\partial f/\partial y = 0$: either there are no solutions, if $\partial f/\partial x \neq 0$, or, if $\partial f/\partial x = 0$, the equation tells us nothing about the slope.

Of course, as we have seen, even when $\partial f/\partial y$ is zero, all is not lost, for if $\partial f/\partial x$ is nonzero, then we can interchange the roles of y and x, solving for the derivative of x as a function of y. So the issue seems to be: is at least *one* of the partials nonzero? If so, we seem to have a perfectly reasonable way to calculate the direction of a line tangent to the level curve at that point. All that remains is to establish our original assumption—that one of the variables can be expressed as a function of the other—as valid. This is what the Implicit Function Theorem does for us.

We want to single out points for which at least one partial is nonzero, or what is the same, at which the gradient is a nonzero vector. Note that to even talk about the gradient or partials, we need to assume that $f(x, y)$ is defined not just at the point in question, but at all points nearby: such a point is called an **interior point** of the domain.

Definition 3.4.1. *Suppose $f(x,y)$ is a differentiable function of two variables. An interior point $\overrightarrow{x}$ of the domain of f is a* ***regular point*** *if*

$$\overrightarrow{\nabla} f(\overrightarrow{x}) \neq \overrightarrow{0},$$

that is, at least one *partial derivative at $\overrightarrow{x}$ is nonzero. $\overrightarrow{x}$ is a* ***critical point*** *of $f(x,y)$ if*

$$\frac{\partial f}{\partial x}(\overrightarrow{x}) = 0 = \frac{\partial f}{\partial y}(\overrightarrow{x}).$$

Our result will be a *local* one, describing the set of solutions to the equation $f(x,y) = c$ *near* a given solution. Our earlier examples showed completely reasonable curves with the exception (in each case except the affine one) of the origin: for the first two functions, the level "curve" corresponding to $c = 0$ is a single point, while for the last function, it crosses itself at the origin. These were all cases in which the origin was a critical point of $f(x,y)$, where we already know that the formal process of implicit differentiation fails; we can only expect to get a reasonable result near *regular points* of $f(x,y)$.

The following result will reappear in § 4.4, in more a elaborate form; it is a fundamental fact about regular points of functions.[8]

Theorem 3.4.2 (Implicit Function Theorem for $\mathbb{R}^2 \to \mathbb{R}$)**.** *The level set of a continuously differentiable function $f{:}\,\mathbb{R}^2 \to \mathbb{R}$ can be expressed near each of its regular points as the graph of a function.*

Specifically, suppose

$$f(x_0, y_0) = c$$

and

$$\frac{\partial f}{\partial y}(x_0, y_0) \neq 0.$$

Then there exists a rectangle

$$R = [x_0 - \delta_1, x_0 + \delta_1] \times [y_0 - \delta_2, y_0 + \delta_2]$$

centered at $\overrightarrow{x}_0 = (x_0, y_0)$(where $\delta_1, \delta_2 > 0$), such that the intersection of $\mathcal{L}(f,c)$ with R is the graph of a C^1 function $\phi(x)$, defined on $[x_0 - \delta_1, x_0 + \delta_1]$ and taking values in $[y_0 - \delta_2, y_0 + \delta_2]$.

[8] For a detailed study of the Implicit Function Theorem in its many incarnations, including some history, and the proof on which the one we give is modeled, see [32].

In other words, if $(x, y) \in R$, *(i.e.,* $|x - x_0| \leq \delta_1$ *and* $|y - y_0| \leq \delta_2$*), then*

$$f(x, y) = c \iff \phi(x) = y. \tag{3.17}$$

Furthermore, at any point $x \in (x_0 - \delta_1, x_0 + \delta_1)$, *the derivative of* $\phi(x)$ *is*

$$\frac{d\phi}{dx} = -\left[\frac{\partial f}{\partial x}(x, \phi(x))\right] \Big/ \left[\frac{\partial f}{\partial y}(x, \phi(x))\right]. \tag{3.18}$$

Proof. The proof will be in two parts.

First we show that Equation (3.17) determines a well-defined function $\phi(x)$:

For notational convenience, we assume without loss of generality that

$$f(x_0, y_0) = 0$$

(that is, $c = 0$), and

$$\frac{\partial f}{\partial y}(x_0, y_0) > 0.$$

Since $f(x, y)$ is continuous, we know that $\frac{\partial f}{\partial y}(\overrightarrow{x}) > 0$ at all points $\overrightarrow{x} = (x, y)$ sufficiently near $\overrightarrow{x}_0$, say for $|x - x_0| \leq \delta$ and $|y - y_0| \leq \delta_2$. For any $a \in [x - \delta, x + \delta]$, consider the function of y obtained by fixing the value of x at $x = a$:

$$g_a(y) = f(a, y);$$

then

$$g_a'(y) = \frac{\partial f}{\partial y}(a, y) > 0$$

so $g_a(y)$ is strictly increasing on $[y - \delta_2, y + \delta_2]$. In particular, when $a = x_0$,

$$g_{x_0}(y_0 - \delta_2) < 0 < g_{x_0}(y_0 + \delta_2)$$

and we can pick $\delta_1 > 0$ ($\delta_1 \leq \delta$) so that

$$g_a(y_0 - \delta_2) < 0 < g_a(y_0 + \delta_2)$$

for each $a \in [x_0 - \delta_1, x_0 + \delta_1]$. The Intermediate Value Theorem insures that for each such a there is *at least one* $y \in [y_0 - \delta_2, y_0 + \delta_2]$ for which

$g_a(y) = f(a,y) = 0$, and the fact that $g_a(y)$ is strictly increasing insures that there is *precisely one*. Writing x in place of a, we see that the definition

$$\phi(x) = y \iff f(a,y) = 0 \quad \text{and } |y - y_0| < \delta_2$$

gives a well-defined function $\phi(x)$ on $[x_0 - \delta_1, x_0 + \delta_1]$ satisfying Equation (3.17).

Second we show that this function satisfies Equation (3.18).

We fix

$$(x,y) = (x, \phi(x))$$

in our rectangle and consider another point

$$(x + \triangle x, y + \triangle y) = (x + \triangle x, \phi(x + \triangle x))$$

on the graph of $\phi(x)$.

Since f is differentiable,

$$f(x + \triangle x, y + \triangle y) - f(x,y) = \triangle x \frac{\partial f}{\partial x}(x,y) + \triangle y \frac{\partial f}{\partial y}(x,y) + \|(\triangle x, \triangle y)\| \varepsilon$$

where $\varepsilon \to 0$ as $(\triangle x, \triangle y) \to (0,0)$.

Since both points lie on the graph of $\phi(x)$, and hence on the same level set of f, the left side of this equation is zero:

$$0 = \triangle x \frac{\partial f}{\partial x}(x,y) + \triangle y \frac{\partial f}{\partial y}(x,y) + \|(\triangle x, \triangle y)\| \varepsilon. \tag{3.19}$$

We will exploit this equation in two ways. For notational convenience, we will drop reference to where a partial is being taken: *for the rest of this proof,*

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial x}(x,y)$$
$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial y}(x,y)$$

where $\overrightarrow{x} = (x,y)$ is the point at which we are trying to prove differentiability of ϕ.

Moving the first two terms to the left side, dividing by $(\triangle x)(\frac{\partial f}{\partial y})$, and taking absolute values, we have

$$\begin{aligned}\left|\frac{\triangle y}{\triangle x}+\frac{\partial f/\partial x}{\partial f/\partial y}\right| & \\ &= \frac{|\varepsilon|}{|\partial f/\partial y|}\frac{\|(\triangle x,\triangle y)\|}{|\triangle x|} \\ &\leq \frac{|\varepsilon|}{|\partial f/\partial y|}\left[1+\left|\frac{\triangle y}{\triangle x}\right|\right] \qquad (3.20)\end{aligned}$$

(since $\|(\triangle x,\triangle y)\| \leq |\triangle x|+|\triangle y|$). To complete the proof, we need to find an upper bound for $\left|1+\frac{\triangle y}{\triangle x}\right|$ on the right side.

To this end, we come back to Equation (3.19), this time moving just the second term to the left, and then take absolute values, using the triangle inequality (and $\|(\triangle x,\triangle y)\| \leq |\triangle x|+|\triangle y|$):

$$|\triangle y|\left|\frac{\partial f}{\partial y}\right| \leq |\triangle x|\left|\frac{\partial f}{\partial x}\right| + |\varepsilon|\,|\triangle x| + |\varepsilon|\,|\triangle y|\,.$$

Gathering the terms involving $\triangle x$ on the left and those involving $\triangle y$ on the right, we can write

$$|\triangle y|\left(\left|\frac{\partial f}{\partial y}\right| - |\varepsilon|\right) \leq |\triangle x|\left(\left|\frac{\partial f}{\partial x}\right| + |\varepsilon|\right)$$

or, dividing by the term on the left,

$$|\triangle y| \leq |\triangle x|\left(\frac{|\partial f/\partial x|+|\varepsilon|}{|\partial f/\partial y|-|\varepsilon|}\right). \qquad (3.21)$$

Now, since $\varepsilon \to 0$, the ratio on the right converges to the ratio of the partials, and so is bounded by, say that ratio plus one, for $\triangle x$ sufficiently near zero:

$$\left|\frac{\triangle y}{\triangle x}\right| \leq \left(\left|\frac{\partial f/\partial x}{\partial f/\partial y}\right|+1\right).$$

This in turn says that the term multiplying $|\varepsilon|$ in Equation (3.20) is bounded, so $\varepsilon \to 0$ implies the desired equation

$$\phi'(x) = \lim_{\triangle x\to 0}\frac{\triangle y}{\triangle x} = -\frac{\partial f/\partial x}{\partial f/\partial y}.$$

This shows that ϕ is differentiable, with partials given by Equation (3.18), and since the right hand side is a continuous function of x, ϕ is *continuously* differentiable. □

We note some features of this theorem:

- The hypothesis that $f(x, y)$ is *continuously* differentiable is crucial; there are examples of differentiable (but not continuously differentiable) functions for which the conclusion is false: there is no function $\phi(x)$ as required by Equation (3.17) (Exercise 6).

- The statement that $\mathcal{L}(f, c) \cap R$ is the graph of $\phi(x)$ means that the function $\phi(x)$ is uniquely determined by Equation (3.17).

- Equation (3.18) is simply implicit differentiation: since

  $$y = \phi(x)$$

  we can write Equation (3.18) formally as

  $$\begin{aligned} \frac{dy}{dx} &= -\frac{\partial f}{\partial x} \div \frac{\partial f}{\partial y} \\ &= \frac{df}{dx}\frac{dy}{df} \\ &= \frac{dy}{df}\frac{df}{dx} \end{aligned}$$

  which is precisely what we get from implicitly differentiating

  $$f(x, y) = c.$$

- Equation (3.18) can also be interpreted as saying that a vector tangent to the level curve has slope

  $$\phi'(x) = -\left[\frac{\partial f}{\partial x}(x, \phi(x))\right] \Big/ \left[\frac{\partial f}{\partial y}(x, \phi(x))\right],$$

  which means that it is perpendicular to $\overrightarrow{\nabla} f(x, \phi(x))$. Of course, this could also be established using the Chain Rule (Exercise 4); the point of the proof above is that one can *take* a vector tangent to $\mathcal{L}(f, c)$, or equivalently that $\phi(x)$ is differentiable.

- In the statement of the theorem, the roles of x and y can be interchanged: if $\frac{\partial f}{\partial x}(x_0, y_0) \neq 0$, then the level set can be expressed as the graph of a function $x = \psi(y)$.

> At a regular point, at least one of these two situations occurs: some partial is nonzero. The theorem says that if the partial of f at $\overrightarrow{x}_0$ with respect to *one* of the variables is nonzero, then near $\overrightarrow{x}_0$ we can solve the equation
>
> $$f(x,y) = f(x_0, y_0)$$
>
> for *that* variable in terms of the *other*.

As an illustration of this last point, we again consider the function

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

The level set $\mathcal{L}(f,1)$ is the circle of radius 1 about the origin

$$x^2 + y^2 + 1.$$

We can solve this equation for y in terms of x on any open arc which does not include either of the points $(\pm 1, 0)$: if the point (x_0, y_0) with $|x_0| < 1$ has $y_0 > 0$, the solution near (x_0, y_0) is

$$\phi(x) = \sqrt{1 - x^2}$$

whereas if $y_0 < 0$ it is

$$\phi(x) = -\sqrt{1 - x^2}.$$

Since

$$\frac{\partial f}{\partial y} = 2y,$$

at the two points $(\pm 1, 0)$

$$\frac{\partial f}{\partial y}(\pm 1, 0) = 0$$

and the theorem does not guarantee the possibility of solving for y in terms of x; in fact, for x near ± 1 there are two values of y giving a point on the curve, given by the two formulas above. However, since at these points

$$\frac{\partial f}{\partial x}(\pm 1, 0) = \pm 2 \neq 0,$$

the theorem *does* guarantee a solution for x in terms of y; in fact, near any point other than $(0, \pm 1)$ (the “north pole” and “south pole”) we can write $x = \psi(y)$, where

$$\psi(y) = \sqrt{1 - y^2}$$

for points on the right semicircle and

$$\psi(y) = -\sqrt{1 - y^2}$$

on the left semicircle.

Slicing Surfaces

The level curves of a function $f(x, y)$ can be thought of as a “topographical map” of the graph of $f(x, y)$: a sketch of several level curves $\mathcal{L}(f, c)$, labeled with their corresponding c-values, allows us to formulate a rough idea of the shape of the graph: these are “slices” of the graph by horizontal planes at different heights. By studying the intersection of the graph with suitably chosen *vertical* planes, we can see how these horizontal pieces fit together to form the surface.

Consider for example the function

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

We know that the horizontal slice at height $c = a^2 > 0$ is the circle

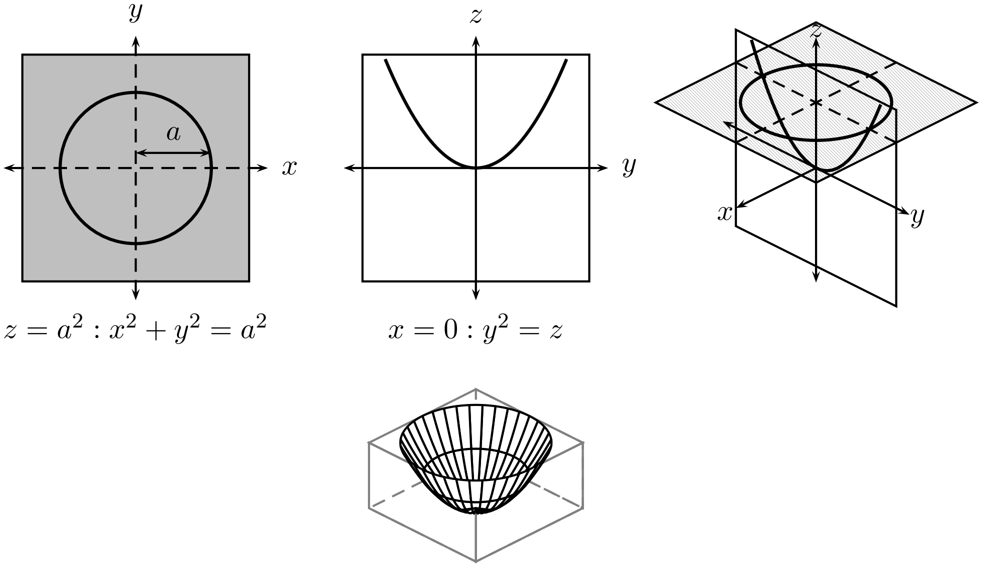
$$x^2 + y^2 = a^2$$

of radius $a = \sqrt{c}$ about the origin; in particular, $\mathcal{L}\big(f, a^2\big)$ crosses the y-axis at the pair of points $(0, \pm a)$. To see how these circles fit together to form the graph of $f(x, y)$, we consider the intersection of the graph

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

with the yz-plane

$$x = 0;$$



$z = a^2 : x^2 + y^2 = a^2$

$x = 0 : y^2 = z$

Figure 3.1: Slicing the Surface $x^2 + y^2 = z$

the intersection is found by substituting the second equation in the first to get

$$z = y^2$$

and we see that the "profile" of our surface is a parabola, with vertex at the origin, opening up. (See Figure 3.1)

If instead we consider the function

$$f(x, y) = 4x^2 + y^2,$$

the horizontal slice at height $c = a^2 > 0$ is the ellipse

$$\frac{x^2}{(a/2)^2} + \frac{y^2}{a^2} = 1$$

centered at the origin, with major axis along the y-axis and minor axis along the x-axis. $\mathcal{L}\big(f, a^2\big)$ again crosses the y-axis at the pair of points $(0, \pm a)$, and it crosses the x-axis at the pair of points $(\pm a/2, 0)$. To see how these ellipses fit together to form the graph of $f(x, y)$, we consider the intersection of the graph

$$z = 4x^2 + y^2$$

with the yz-plane

$$x = 0;$$

the intersection is found by substituting the second equation in the first to get the parabola

$$z = y^2.$$

Similarly, the intersection of the graph with the xz-plane

$$y = 0$$

is a different parabola

$$z = 4x^2.$$

One might say that the "shadow" of the graph on the xz-plane is a narrower parabola than the shadow on the yz-plane. (See Figure 3.2.) This surface is called an **elliptic paraboloid**.

A more interesting example is given by the function

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

The horizontal slice at height $c \neq 0$ is a hyperbola which opens along the x-axis if $c > 0$ and along the y-axis if $c < 0$; the level set $\mathcal{L}(f, 0)$ is the pair of diagonal lines

$$y = \pm x$$

which are the common asymptotes of each of these hyperbolas. (See Figure 3.3.)

To see how these fit together to form the graph, we again slice along the coordinate planes. The intersection of the graph

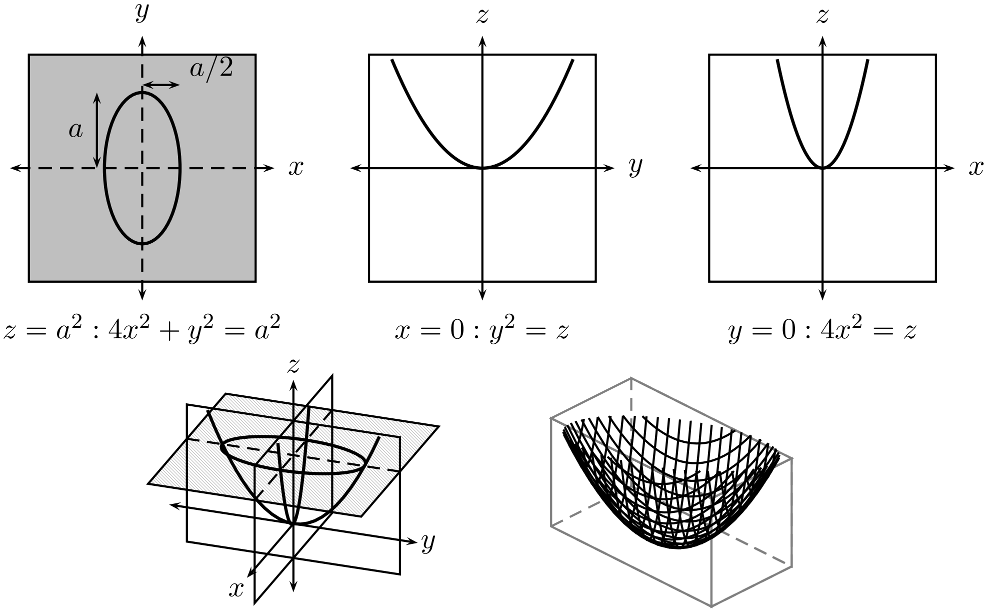
$$z = x^2 - y^2$$

with the xz-plane

$$y = 0$$

is a parabola opening *up*: these points are the "vertices" of the hyperbolas $\mathcal{L}(f, c)$ for *positive* c. The intersection with the yz-plane

$$x = 0$$



$z = a^2 : 4x^2 + y^2 = a^2$ $x = 0 : y^2 = z$ $y = 0 : 4x^2 = z$

Figure 3.2: Slicing the Surface $4x^2 + y^2 = z$

is a parabola opening *down*, going through the vertices of the hyperbolas $\mathcal{L}(f, c)$ for *negative c*.

Fitting these pictures together, we obtain a surface shaped like a saddle (imagine the horse's head facing parallel to the x-axis, and the rider's legs parallel to the yz-plane). It is often called the **saddle surface**, but its official name is the **hyperbolic paraboloid**. (See Figure 3.4.)

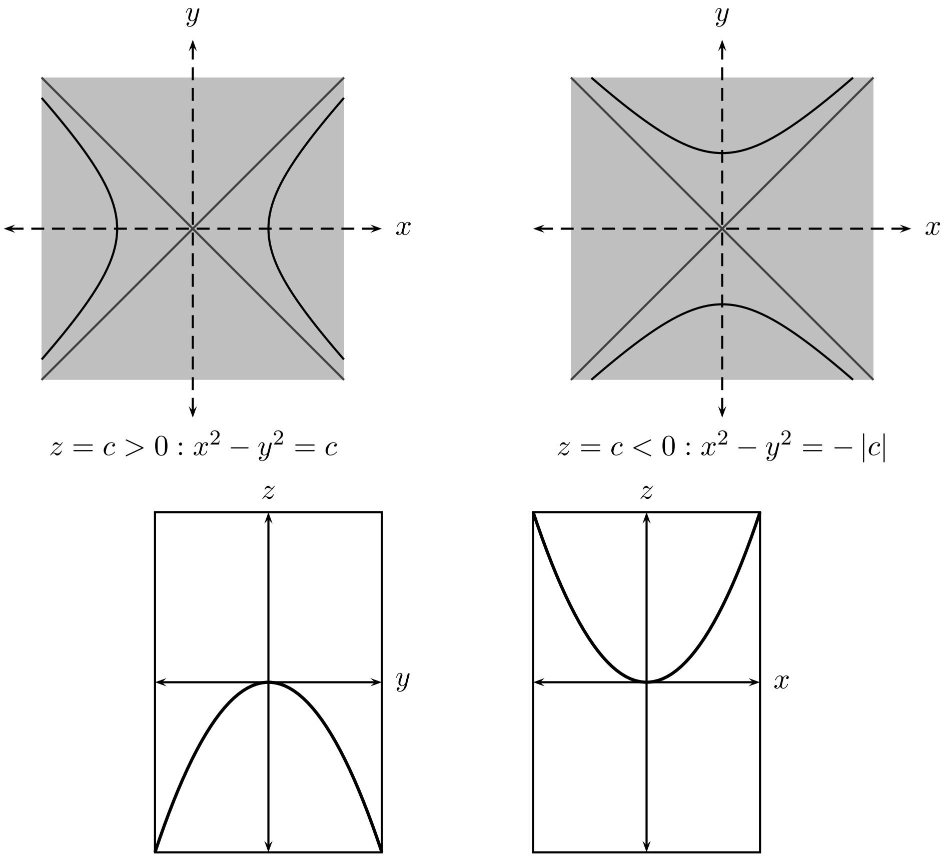
These slicing techniques can also be used to study surfaces given by equations in x, y and z which are not explicitly graphs of functions. We consider three examples.

The first is given by the equation

$$\frac{x^2}{4} + y^2 + z^2 = 1.$$

The intersection of this with the xy-plane $z = 0$ is the ellipse

$$\frac{x^2}{4} + y^2 = 1$$



$z = c > 0 : x^2 - y^2 = c$

$z = c < 0 : x^2 - y^2 = -|c|$

$x = 0 : z = -y^2$

$y = 0 : z = x^2$

Figure 3.3: Slicing the Surface $z = x^2 - y^2$

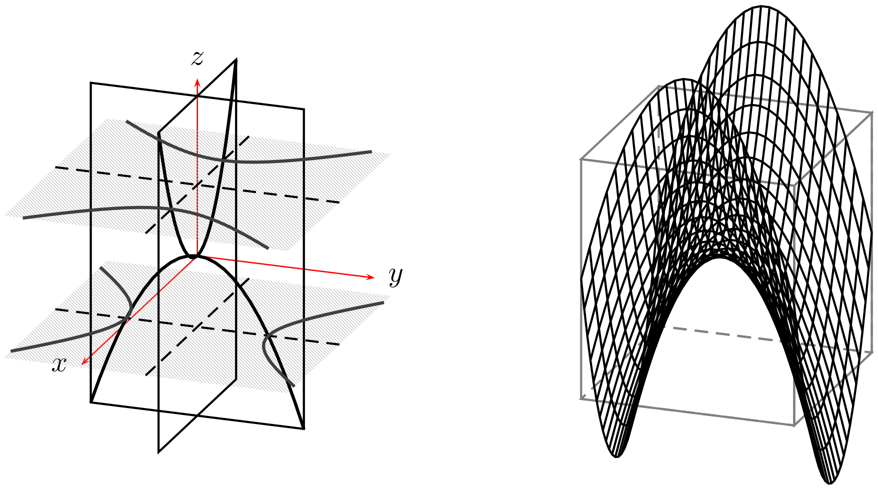


Figure 3.4: Combining Slices to Form the Surface $z = x^2 - y^2$

centered at the origin and with the ends of the axes at $(\pm 2, 0, 0)$ and $(0, \pm 1, 0)$; the intersection with any other horizontal plane $z = c$ for which $|c| < 1$ is an ellipse similar to this and with the same center, but scaled down:

$$\frac{x^2}{4} + y^2 = 1 - c^2$$

or

$$\frac{x^2}{4(1-c^2)} + \frac{y^2}{1-c^2} = 1.$$

There are no points on this surface with $|z| > 1$. Similarly, the intersection with a vertical plane parallel to the xz-plane, $y = c$ (again with $|c| < 1$) is a scaled version of the same ellipse, but in the xz-plane

$$\frac{x^2}{4} + z^2 = 1 - c^2$$

and again no points with $|y| > 1$. Finally, the intersection with a plane parallel to the yz-plane, $x = c$, is nonempty provided $\left|\frac{x}{2}\right| < 1$ or $|x| < 2$, and in that case is a circle centered at the origin in the yz-plane of radius $r = \sqrt{1 - \frac{c^2}{4}}$

$$y^2 + z^2 = 1 - \frac{c^2}{4}.$$

This surface is like a sphere, but “elongated” in the direction of the x-axis by a factor of 2 (see Figure 3.5); it is called an **ellipsoid**.

Our second example is the surface given by the equation

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

The intersection with any horizontal plane

$$z = c$$

is a circle

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

of radius $r = \sqrt{c^2 + 1}$ about the origin (actually, about the intersection of the plane $z = c$ with the z-axis). Note that always $r \geq 1$; the smallest circle is the intersection with the xy-plane.

If we slice along the xz-plane

$$y = 0$$

we get the hyperbola

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

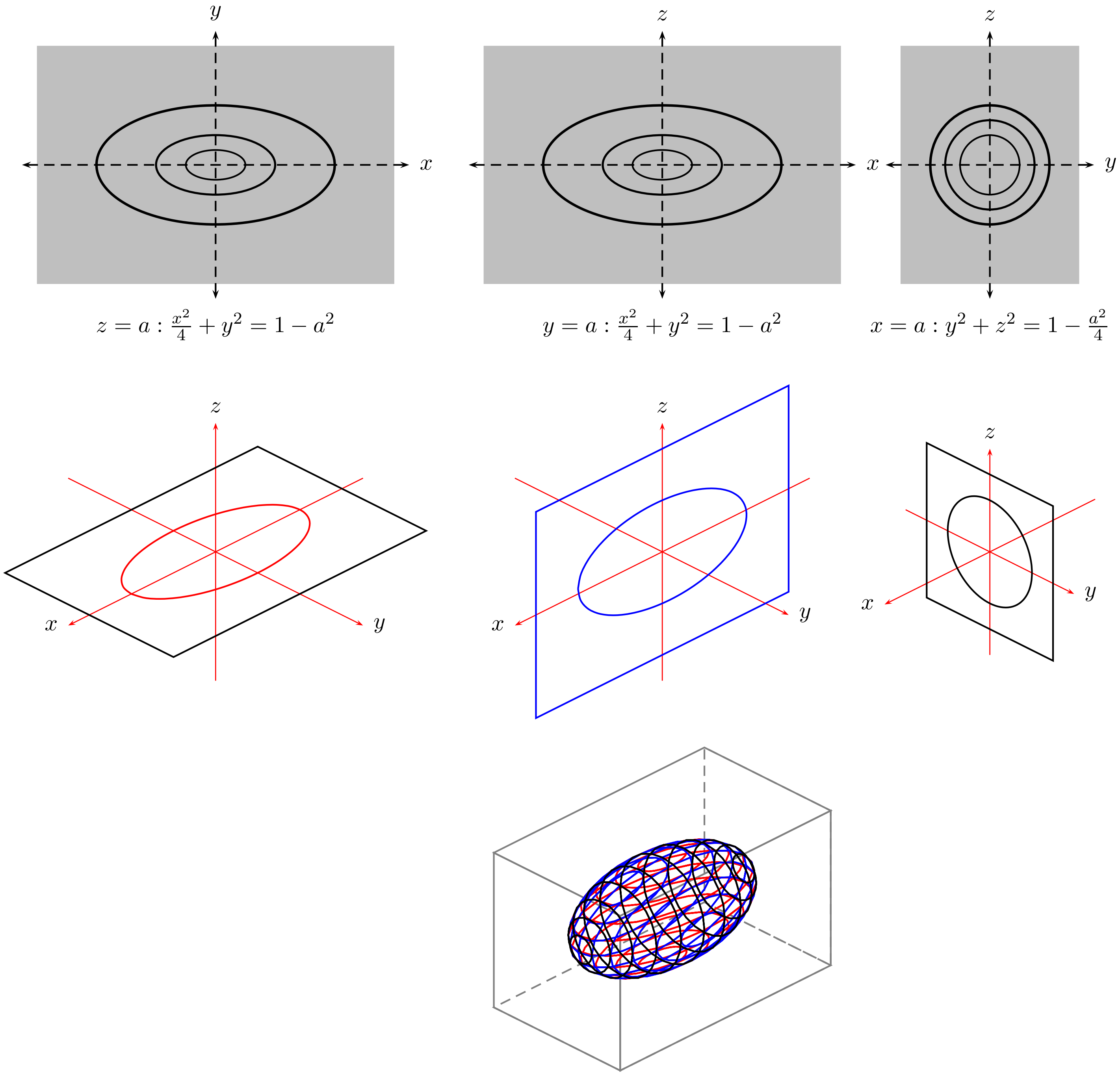
whose vertices lie on the small circle in the xy-plane. Slicing along the yz-plane we get a similar picture, since x and y play exactly the same role in the equation. The shape we get, like a cylinder that has been squeezed in the middle, is called a **hyperboloid of one sheet** (Figure 3.6).

Now, let us simply change the sign of the constant in the previous equation:

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

The intersection with the horizontal plane

$$z = c$$

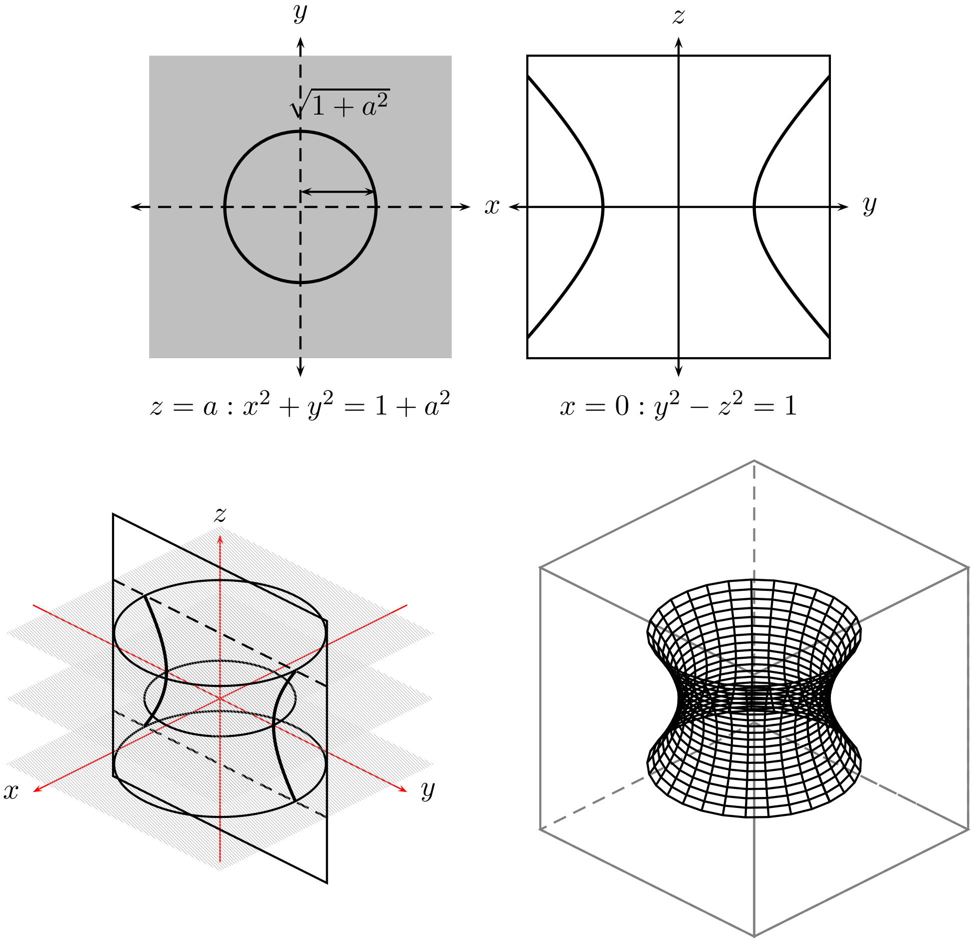


$z = a : \frac{x^2}{4} + y^2 = 1 - a^2$

$y = a : \frac{x^2}{4} + y^2 = 1 - a^2$

$x = a : y^2 + z^2 = 1 - \frac{a^2}{4}$

Figure 3.5: Slicing the Surface $\frac{x^2}{4} + y^2 + z^2 = 1$



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

$x = 0 : y^2 - z^2 = 1$

Figure 3.6: Slicing the Surface $x^2 + y^2 - z^2 = 1$

is a circle

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

of radius $r = \sqrt{c^2 + 1}$ about the "origin", *provided* $c^2 > 1$; for $c = \pm 1$ we get a single point, and for $|c| < 1$ we get the empty set. In particular, our surface consists of two pieces, one for $z \geq 1$ and another for $z \leq -1$.

If we slice along the xz-plane

$$y = 0$$

we get the hyperbola

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

or

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

which opens up and down; again, it is clear that the same thing happens along the yz-plane. Our surface consists of two "bowl"-like surfaces whose shadow on a vertical plane is a hyperbola. This is called a **hyperboloid of two sheets** (see Figure 3.7).

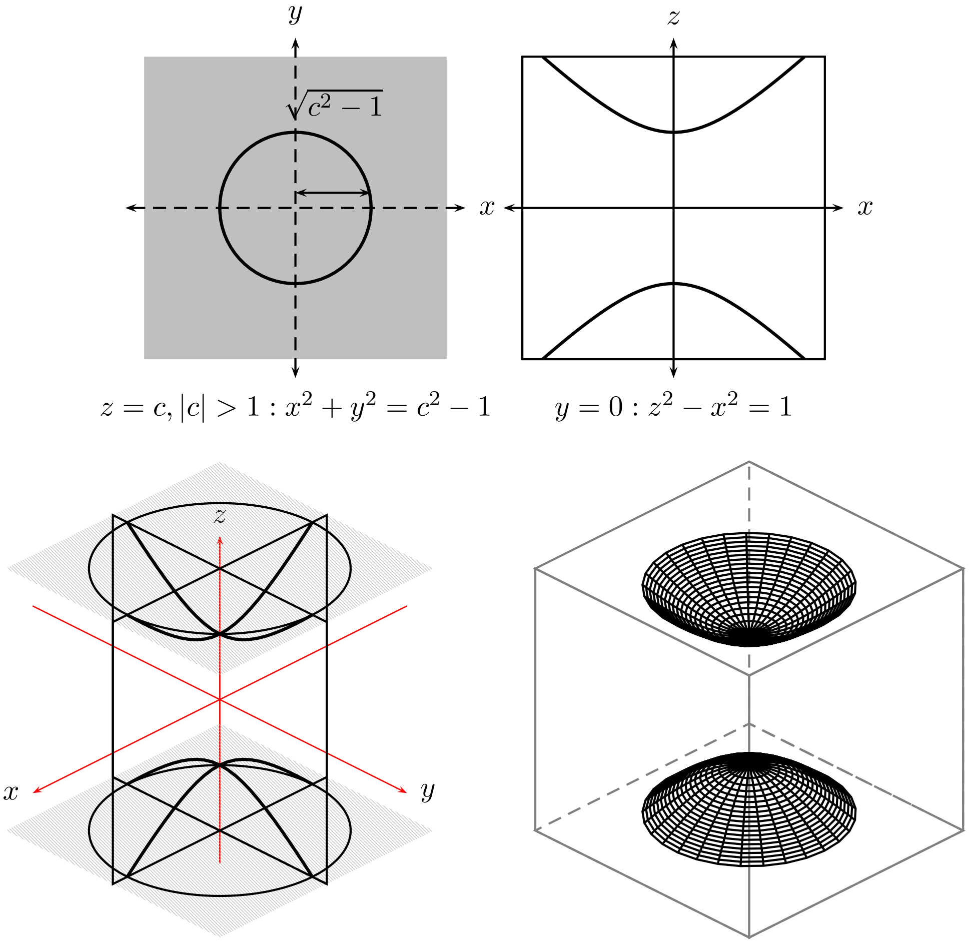
The reader may have noticed that the equations we have considered are the three-variable analogues of the model equations for parabolas, ellipses and hyperbolas, the quadratic curves; in fact, these are the basic models for equations given by quadratic polynomials in three coordinates, and are known collectively as the **quadric surfaces**.

Level Surfaces

For a real-valued function $f(x, y, z)$ of *three* variables, the level set $\mathcal{L}(f, c)$ is defined by an equation in *three* variables, and we expect it to be a *surface.*

For example, the level sets $\mathcal{L}(f, c)$ of the function

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



$z = c, |c| > 1 : x^2 + y^2 = c^2 - 1$ $y = 0 : z^2 - x^2 = 1$

Figure 3.7: Slicing the Surface $x^2 + y^2 - z^2 = -1$

are spheres (of radius $\sqrt{c}$) centered at the origin if $c > 0$; again for $c = 0$ we get a single point and for $c < 0$ the empty set: the origin is the one place where

$$\vec{\nabla} f(x, y, z) = 2x\vec{\imath} + 2y\vec{\jmath} + 2z\vec{k}$$

vanishes.

Similarly, the function

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

can be seen, following the analysis in § 3.4, to have as its level sets $\mathcal{L}(f, c)$ a family of hyperboloids[9]—of one sheet for $c > 0$ and two sheets for $c < 0$. For $c = 0$, the level set is given by the equation

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

which can be rewritten in polar coordinates

$$r^2 = z^2;$$

we recognize this as the cylindrical surface we used to study the conics in § 2.1. This is a reasonable surface, except at the origin, which again is the only place where the gradient grad f vanishes.

This might lead us to expect an analogue of Theorem 3.4.2 for functions of *three* variables. Before stating it, we introduce a useful bit of notation. By the **ε-ball** or **ball of radius ε** about $\vec{x}_0$, we mean the set of all points at distance at most $\varepsilon > 0$ from $\vec{x}_0$:

$$B_\varepsilon(\vec{x}_0) := \{\vec{x} \mid \operatorname{dist}(\vec{x}, \vec{x}_0) \leq \varepsilon\}.$$

For points on the line, this is the interval $[x_0 - \varepsilon, x_0 + \varepsilon]$; in the plane, it is the disc $\{(x, y) \mid (x - x_0)^2 + (y - y_0)^2 \leq \varepsilon\}$, and in space it is the actual ball $\{(x, y, z) \mid (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \leq \varepsilon\}$.

Theorem 3.4.3 (Implicit Function Theorem for $\mathbb{R}^3 \to \mathbb{R}$)**.** *The level set of a continuously differentiable function $f: \mathbb{R}^3 \to \mathbb{R}$ can be expressed near each of its regular points as the graph of a function.*

Specifically, suppose that at

$$\vec{x}_0 = (x_0, y_0, z_0)$$

[9] Our analysis in § 3.4 clearly carries through if 1 is replaced by any positive number $|c|$

we have

$$f(\overrightarrow{x}_0) = c$$

and

$$\frac{\partial f}{\partial z}(\overrightarrow{x}_0) \neq 0.$$

Then there exists a set of the form

$$R = B_\varepsilon((x_0, y_0)) \times [z_0 - \delta, z_0 + \delta]$$

(where $\varepsilon > 0$ *and* $\delta > 0$*), such that the intersection of* $\mathcal{L}(f, c)$ *with* R *is the graph of a* C^1 *function* $\phi(x, y)$*, defined on* $B_\varepsilon((x_0, y_0))$ *and taking values in* $[z_0 - \delta, z_0 + \delta]$*. In other words, if* $\overrightarrow{x} = (x, y, z) \in R$*, then*

$$f(x, y, z) = c \iff z = \phi(x, y)\,. \tag{3.22}$$

Furthermore, at any point $(x, y) \in B_\varepsilon(\overrightarrow{x}_0)$*, the partial derivatives of* ϕ *are*

$$\begin{aligned} \frac{\partial \phi}{\partial x} &= -\frac{\partial f/\partial x}{\partial f/\partial z} \\ \frac{\partial \phi}{\partial y} &= -\frac{\partial f/\partial y}{\partial f/\partial z} \end{aligned} \tag{3.23}$$

where the partial on the left is taken at $(x, y) \in B_\varepsilon \subset \mathbb{R}^2$ *and the partials on the right are taken at* $(x, y, \phi(x, y)) \in R \subset \mathbb{R}^3$*.*

Note that the statement of the general theorem says when we can solve for z in terms of x and y, but an easy argument (Exercise 5) shows that we can replace this with *any* variable whose partial is nonzero at $\overrightarrow{x} = \overrightarrow{x}_0$.

Proof sketch: This is a straightforward adaptation of the proof of Theorem 3.4.2 for functions of two variables.

Recall that the original proof had two parts. The first was to show simply that $\mathcal{L}(f, c) \cap R$ is the graph of a function on $B_\varepsilon(\overrightarrow{x}_0)$. The argument for this in the three-variable case is almost verbatim the argument in the original proof: assuming that

$$\frac{\partial f}{\partial z} > 0$$

for all $\overrightarrow{x}$ near $\overrightarrow{x}_0$, we see that F is strictly increasing along a short vertical line segment through any point (x', y', z') near $\overrightarrow{x}_0$,

$$I_{(x', y')} = \{(x', y', z) \mid z' - \delta \leq z \leq z' + \delta\}.$$

In particular, assuming $c = 0$ for convenience, we have at (x_0, y_0)

$$f(x_0, y_0, z_0 - \delta) < 0 < f(x_0, y_0, z_0 - \delta)$$

and so for $x = x_0 + \triangle x, y = y_0 + \triangle y$, $\triangle x$ and $\triangle y$ small ($\|(\triangle x, \triangle y)\| < \varepsilon$), we also have f positive at the top and negative at the bottom of the segment $I_{(x_0+\triangle x, y_0+\triangle y)}$:

$$f(x_0 + \triangle x, y_0 + \triangle y, z_0 - \delta) < 0 < f(x_0 + \triangle x, y_0 + \triangle y, z_0 + \delta)\,.$$

The Intermediate Value Theorem then guarantees that $f = 0$ for at least one point on each vertical segment in R, and the strict monotonicity of f along each segment also guarantees that there is *precisely* one such point along each segment. This analogue of the "vertical line test" proves that the function $\phi(x, y)$ is well-defined in $B_\varepsilon\,(x_0, y_0)$.

The second part of the original proof, showing that this function ϕ is continuously differentiable, could be reformulated in the three variable case, although it is perhaps less clear how the various ratios could be handled. But there is an easier way. The original proof that $\phi'(x)$ is the negative ratio of $\partial f/\partial x$ and $\partial f/\partial y$ in the two variable case is easily adapted to prove that the restriction of our new function $\phi(x, y)$ to a line parallel to either the x-axis or y-axis is differentiable, and that the derivative of the *restricition* (which is nothing other than a partial of $\phi(x, y)$) is the appropriate ratio of partials of f, as given in Equation (3.23). But then, rather than trying to prove *directly* that ϕ is differentiable as a function of two variables, we can appeal to Theorem 3.3.4 to conclude that, since its partials are continuous, the function is differentiable. This concludes the proof of the Implicit Function Theorem for real-valued functions of three variables. □

As an example, consider the level surface $\mathcal{L}(f, 1)$, where

$$f(x, y, z) = 4x^2 + y^2 - z^2 :$$

The partial derivatives of $f(x, y, z)$ are

$$\begin{aligned}
\frac{\partial f}{\partial x}(x, y, z) &= 8x \\
\frac{\partial f}{\partial y}(x, y, z) &= 2y \\
\frac{\partial f}{\partial z}(x, y, z) &= -2z;
\end{aligned}$$

at the point $(1,-1,2)$, these values are

$$\begin{aligned}\frac{\partial f}{\partial x}(1,-1,2) &= 8\\ \frac{\partial f}{\partial y}(1,-1,2) &= -2\\ \frac{\partial f}{\partial z}(1,-1,2) &= -2\end{aligned}$$

so we see from the Implicit Function Theorem that we can solve for any one of the variables in terms of the other two. For example, near this point we can write

$$z = \phi(x,y)$$

where

$$4x^2 + y^2 - \phi(x,y)^2 = 1$$

and

$$\phi(1,-1) = 2;$$

the theorem tells us that $\phi(x,y)$ is differentiable at $x=1$, $y=-1$, with

$$\begin{aligned}\frac{\partial \phi}{\partial x}(1,-1) &= -\frac{\partial f/\partial x}{\partial f/\partial z}\\ &= -\frac{8}{-2}\\ &= 4\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \phi}{\partial y}(1,-1) &= -\frac{\partial f/\partial y}{\partial f/\partial z}\\ &= -\frac{-2}{-2}\\ &= -1.\end{aligned}$$

Of course, in this case, we can verify the conclusion by solving explicitly:

$$\phi(x,y) = \sqrt{1-(4x^2+y^2)};$$

you should check that the properties of this function are as advertized. However, at $(0,1,0)$, the situation is different: since

$$\begin{aligned}\frac{\partial f}{\partial x}(0,1,0) &= 0\\ \frac{\partial f}{\partial y}(0,1,0) &= -2\\ \frac{\partial f}{\partial z}(0,1,0) &= 0\end{aligned}$$

we can only hope to solve for y in terms of x and z; the theorem tells us that in this case

$$\begin{aligned}\frac{\partial y}{\partial x}(0,0) &= 0\\ \frac{\partial y}{\partial z}(0,0) &= 0.\end{aligned}$$

We note in passing that Theorem 3.4.3 can be formulated for a function of any number of variables, and the passage from three variables to more is very much like the passage from two to three. However, some of the geometric setup to make this rigorous would take us too far afield. There is also a very slick proof of the most general version of this theorem based on the "contraction mapping theorem"; this is the version that you will probably encounter in higher math courses.

Exercises for § 3.4

Practice problems:

1. For each equation below, investigate several slices and use them to sketch the locus of the equation. For quadric surfaces, decide which kind it is (*e.g.*, hyperbolic paraboloid, ellipsoid, hyperboloid of one sheet, etc.)

   (a) $z = 9x^2 + 4y^2$ (b) $z = 1 - x^2 - y^2$ (c) $z = x^2 - 2x + y^2$

   (d) $x^2 + y^2 + z = 1$ (e) $9x^2 = y^2 + z$ (f) $x^2 - y^2 - z^2 = 1$

   (g) $x^2 - y^2 + z^2 = 1$ (h) $z^2 = x^2 + y^2$ (i) $x^2 + 4y^2 + 9z^2 = 36$

2. For each curve defined implicitly by the given equation, decide at each given point whether one can solve locally for (a) $y = \phi(x)$, (b) $x = \psi(y)$, and find the derivative of the function if it exists:

(a) $x^3 + 2xy + y^3 = -2$, at $(1, -1)$ and at $(2, -6)$.

(b) $(x - y)e^{xy} = 1$, at $(1, 0)$ and at $(0, -1)$.

(c) $x^2y + x^3y^2 = 0$, at $(1, -1)$ and at $(0, 1)$

3. For each surface defined implicitly, decide at each given point whether one can solve locally for (a) z in terms of x and y; (b) x in terms of y and z; y in terms of x and z. Find the partials of the function if it exists.

   (a) $x^3z^2 - z^3xy = 0$ at $(1, 1, 1)$ and at $(0, 0, 0)$.

   (b) $xy + z + 3xz^5 = 4$ at $(1, 0, 1)$

   (c) $x^3 + y^3 + z^3 = 10$ at $(1, 2, 1)$ and at $(\sqrt[3]{5}, 0, \sqrt[3]{5})$.

   (d) $\sin x \cos y - \cos x \sin z = 0$ at $(\pi, 0, \frac{\pi}{2})$.

Theory problems:

4. Prove that the gradient vector $\overrightarrow{\nabla}\mathrm{f}$ is perpendicular to the level surfaces of f, using the Chain Rule instead of Equation (3.18).

5. Mimic the argument for Theorem 3.4.3 to show that we can solve for *any* variable whose partial does not vanish at our point.

Challenge problem:

6. The following example (based on [32, pp. 58-9] or [49, p. 201]) shows that the hypothesis that f be *continuously* differentiable cannot be ignored in Theorem 3.4.2. Define $f(x, y)$ by

$$f(x, y) = \begin{cases} xy + y^2 \sin \frac{1}{y} & \text{if } y \neq 0, \\ 0 & \text{if } y = 0. \end{cases}$$

   (a) Show that for $y \neq 0$

$$\frac{\partial f}{\partial y}(x, y) = x + 2y \sin \frac{1}{y} - y \cos \frac{1}{y}.$$

   (b) Show that

$$\frac{\partial f}{\partial y}(x, 0) = x.$$

   (c) Show that $\partial f / \partial y$ is not continuous at $(1, 0)$.

(d) Show that if $f(x,y) = 0$ and $y \neq 0$, then

$$x = -y \sin \frac{1}{y}.$$

(e) Show that the only solutions of $f(x,y) = 0$ in the rectangle $\left[\frac{2}{3}, \frac{4}{3}\right] \times \left[-\frac{1}{3}, \frac{1}{3}\right]$ are on the y-axis.

(f) Conclude that there is no function ϕ defined on $I = \left[\frac{2}{3}, \frac{4}{3}\right]$ such that $f(x, \phi(x)) = 0$ for all $x \in I$.

3.5 Surfaces and their Tangent Planes

In this section, we study various ways of specifying a surface, and finding its tangent plane (when it exists) at a point. As a starting point, we deal first with surfaces defined as graphs of functions of two variables.

Graph of a Function

The graph of a real-valued function $f(x)$ of one real variable is the subset of the plane defined by the equation

$$y = f(x)\,,$$

which is of course a curve—in fact an arc (at least if $f(x)$ is continuous, and defined on an interval). Similarly, the graph of a function $f(x,y)$ of two real variables is the locus of the equation

$$z = f(x,y)\,,$$

which is a surface in $\mathbb{R}^3$, at least if $f(x,y)$ is continuous and defined on a reasonable region in the plane.

For a curve in the plane given as the graph of a differentiable function $f(x)$, the tangent to the graph at the point corresponding to $x = x_0$ is the line through that point, $P(x_0, f(x_0))$, with slope equal to the derivative $f'(x_0)$. Another way to look at this, though, is that *the tangent at $x = x_0$ to the graph of $f(x)$ is the graph of the linearization $T_{x_0}f(x)$ of $f(x)$ at $x = x_0$.* We can take this as the definition in the case of a general graph:

Definition 3.5.1. *The* ***tangent plane*** *at $\overrightarrow{x} = \overrightarrow{x}_0$ to the graph $z = f(\overrightarrow{x})$ of a differentiable function $f\colon \mathbb{R}^3 \to \mathbb{R}$ is the graph of the linearization of $f(\overrightarrow{x})$ at $\overrightarrow{x} = \overrightarrow{x}_0$; that is, it is the locus of the equation*

$$z = T_{\overrightarrow{x}_0} f(\overrightarrow{x}) = f(\overrightarrow{x}_0) + d_{\overrightarrow{x}_0} f(\triangle \overrightarrow{x})$$

where $\triangle \overrightarrow{x} = \overrightarrow{x} = \overrightarrow{x}_0$.

Note that in the definition above we are specifying where the tangent plane is being found by the value of the *input* $\overrightarrow{x}$; when we regard the graph as simply a surface in space, we should really think of the plane at $(x, y) = (x_0, y_0)$ as the tangent plane at the *point* $P(x_0, y_0, z_0)$ in space, where $z_0 = f(x_0, y_0)$.

For example, consider the function

$$f(x, y) = x^2 + 2y^2 :$$

the partials are

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2x \\ \frac{\partial f}{\partial y} &= 4y\end{aligned}$$

so taking

$$\overrightarrow{x}_0 = (1, -1),$$

we find

$$\begin{aligned}f(1, -1) &= 3 \\ \frac{\partial f}{\partial x}(1, -1) &= 2 \\ \frac{\partial f}{\partial y}(1, -1) &= -4\end{aligned}$$

and the linearization of $f(x, y)$ at $\overrightarrow{x} = (1, -1)$ is

$$T_{(1,-1)}f(x, y) = 3 + 2(x - 1) - 4(y + 1).$$

If we use the parameters

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

then the tangent plane is parametrized by

$$\begin{array}{rcrrr} x &=& 1 & +s & \\ y &=& -1 & & +t \\ z &=& 3 & +2s & -4t; \end{array} \tag{3.24}$$

the basepoint of this parametrization is $P(1,-1,3)$.

If we should want to express this tangent plane by an equation, we would need to find a normal vector. To this end, note that the parametrization above has the natural direction vectors

$$\begin{aligned}\vec{v}_1 &= \vec{\imath} + 2\vec{k}\\ \vec{v}_2 &= \vec{\jmath} - 4\vec{k}\,.\end{aligned}$$

Thus, we can find a normal vector by taking their cross product

$$\begin{aligned}\vec{N} &= \vec{v}_1 \times \vec{v}_2\\ &= \begin{vmatrix} \vec{\imath} & \vec{\jmath} & \vec{k}\\ 1 & 0 & 2\\ 0 & 1 & -4 \end{vmatrix}\\ &= -2\vec{\imath} + 4\vec{\jmath} + \vec{k}\,.\end{aligned}$$

It follows that the tangent plane has the equation

$$0 = -2(x-1) + 4(y+1) + (z-3)$$

which we recognize as a restatement of the equation Equation (3.24) identifying this plane as a graph:

$$\begin{aligned}z &= 3 + 2(x-1) - 4(y+1)\\ &= T_{(1,-1)}f(x,y)\,.\end{aligned}$$

These formulas have a geometric interpretation. The parameter $s = x-1$ represents a displacement of the input from the base input $(1,-1)$ parallel to the x-axis—that is, holding y constant (at the base value $y=-1$). The intersection of the graph $z = f(x,y)$ with this plane $y=-1$ is the curve

$$z = f(x,-1)$$

which is the graph of the function

$$z = x^2 + 2;$$

at $x=1$, the derivative of this function is

$$\begin{aligned}\left.\frac{dz}{dx}\right|_1 &= \frac{\partial f}{\partial x}(1,-1)\\ &= 2\end{aligned}$$

and the line through the point $x = 1$, $z = 3$ in this plane with slope 2 lies in the plane tangent to the graph of $f(x, y)$ at $(1, -1)$; the vector $\overrightarrow{v}_1 = \overrightarrow{\imath} + 2\overrightarrow{k}$ is a direction vector for this line: the line alone is parametrized by

$$\begin{array}{rcrl} x & = & 1 & +s \\ y & = & -1 & \\ z & = & 3 & +2s \end{array}$$

which can be obtained from the parametrization of the full tangent plane by fixing $t = 0$.

Similarly, the intersection of the graph $z = f(x, y)$ with the plane $x = 1$ is the curve

$$z = f(1, y)$$

which is the graph of the function

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

at $y = -1$, the derivative of this function is

$$\begin{aligned} \left.\frac{dz}{dy}\right|_{-1} &= \frac{\partial f}{\partial y}(1, -1) \\ &= -4 \end{aligned}$$

and $\overrightarrow{v}_2 = \overrightarrow{\jmath} - 4\overrightarrow{k}$ is the direction vector for the line of slope -4 through $y = -1$, $z = 3$ in this plane—a line which also lies in the tangent plane. This line is parametrized by

$$\begin{array}{rcrl} x & = & 1 & \\ y & = & -1 & +t \\ z & = & 3 & -4t \end{array}$$

which can be obtained from the parametrization of the full tangent plane by fixing $s = 0$.

The alert reader (this means *you*!) will have noticed that the whole discussion above could have been applied to the graph of *any* differentiable function of two variables. We summarize it below.

Remark 3.5.2. *If the function $f(x,y)$ is differentiable at $\overrightarrow{x}_0 = (x_0, y_0)$, then the plane tangent to the graph*

$$z = f(x,y)$$

at

$$\begin{aligned} x &= x_0 \\ y &= y_0, \end{aligned}$$

which is the graph of the linearization of $f(x,y)$

$$\begin{aligned} z &= T_{(x_0,y_0)} f(x,y) \\ &= f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial y}(y_0,y_0)(y-y_0), \end{aligned}$$

is the plane through the point

$$P(x_0,y_0,z_0),$$

where

$$z_0 = f(x_0,y_0),$$

with direction vectors

$$\overrightarrow{v}_1 = \overrightarrow{\imath} + \frac{\partial f}{\partial x}(x_0,y_0)\overrightarrow{k}$$

and

$$\overrightarrow{v}_2 = \overrightarrow{\jmath} + \frac{\partial f}{\partial y}(x_0,y_0)\overrightarrow{k}.$$

These represent the direction vectors of the lines tangent at $P(x_0,y_0,z_0)$ to the intersection of the planes

$$y = y_0$$

and

$$x = x_0,$$

respectively, with our graph.

A parametrization of the tangent plane is

$$\begin{aligned} x &= x_0 + s \\ y &= y_0 + t \\ z &= z_0 + \frac{\partial f}{\partial x}(x_0, y_0)\, s + \frac{\partial f}{\partial y}(x_0, y_0)\, t \end{aligned}$$

and the two lines are parametrized by setting t (resp. s) equal to zero.

A normal vector to the tangent plane is given by the cross product

$$\begin{aligned} \overrightarrow{n} &= \overrightarrow{v}_1 \times \overrightarrow{v}_2 \\ &= -\frac{\partial f}{\partial x}(x_0, y_0)\, \overrightarrow{\imath} - \frac{\partial f}{\partial y}(x_0, y_0)\, \overrightarrow{\jmath} + \overrightarrow{k}. \end{aligned}$$

The adventurous reader is invited to think about how this extends to graphs of functions of more than two variables.

Parametrized Surfaces

In § 2.2 we saw how to go beyond *graphs* of *real-valued* functions of a real variable to express more general curves as images of *vector-valued* functions of a real variable. In this subsection, we will explore the analogous representation of a surface in space as the image of a vector-valued function of *two* variables. Of course, we have already seen such a representation for planes.

Just as continuity and limits for functions of several variables present new subtleties compared to their single-variable cousins, an attempt to formulate the idea of a "surface" in $\mathbb{R}^3$ using only continuity notions will encounter a number of difficulties. We shall avoid these by starting out immediately with differentiable parametrizations.

Definition 3.5.3. *A vector-valued function*

$$\overrightarrow{p}(s,t) = (x_1(s,t)\,, x_2(s,t)\,, x_3(s,t))$$

of two real variables is ***differentiable*** *(resp.* ***continuously differentiable,*** *or* $\boldsymbol{C^1}$*) if each of the coordinate functions $x_j\colon \mathbb{R}^2 \to \mathbb{R}$ is differentiable (resp.*

continuously differentiable). We know from Theorem 3.3.4 that a C^1 function is automatically differentiable.

We define the ***partial derivatives*** *of a differentiable function $\overrightarrow{p}(s,t)$ to be the vectors*

$$\begin{aligned}\frac{\partial \overrightarrow{p}}{\partial s} &= \left(\frac{\partial x_1}{\partial s}, \frac{\partial x_2}{\partial s}, \frac{\partial x_3}{\partial s}\right)\\ \frac{\partial \overrightarrow{p}}{\partial t} &= \left(\frac{\partial x_1}{\partial t}, \frac{\partial x_2}{\partial t}, \frac{\partial x_3}{\partial t}\right).\end{aligned}$$

We will call $\overrightarrow{p}(s,t)$ ***regular*** *if it is C^1 and at every pair of parameter values (s,t) in the domain of $\overrightarrow{p}$ the partials are linearly independent—that is, neither is a scalar multiple of the other. The image of a regular parametrization*

$$\mathfrak{S} := \{\overrightarrow{p}(s,t) \mid (s,t) \in \mathrm{dom}(\overrightarrow{p})\}$$

is a surface in $\mathbb{R}^3$, and we will refer to $\overrightarrow{p}(s,t)$ as a ***regular parametrization*** *of $\mathfrak{S}$.*

As an example, you should verify (Exercise 9a) that the graph of a (continuously differentiable) function $f(x,y)$ is a surface parametrized by

$$\overrightarrow{p}(s,t) = (s,t,f(s,t)).$$

As another example, consider the function

$$\overrightarrow{p}(\theta,t) = (\cos\theta, \sin\theta, t);$$

this can also be written

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

The first two equations give a parametrization of the circle of radius one about the origin in the xy-plane, while the third moves such a circle vertically by t units: we see that this parametrizes a cylinder with axis the z-axis, of radius 1 (Figure 3.8).

The partials are

$$\begin{aligned}\frac{\partial \overrightarrow{p}}{\partial \theta}(\theta,t) &= -(\sin\theta)\overrightarrow{\imath} + (\cos\theta)\overrightarrow{\jmath}\\ \frac{\partial \overrightarrow{p}}{\partial t}(\theta,t) &= \overrightarrow{k}\end{aligned}$$

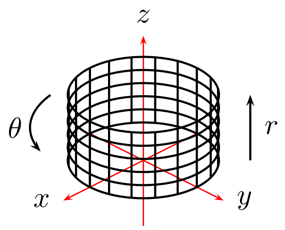


Figure 3.8: Parametrized Cylinder

Another function is

$$\overrightarrow{p}(r,\theta) = (r\cos\theta, r\sin\theta, 0)$$

or

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= 0 \end{aligned}$$

which describes the xy-plane in polar coordinates; the partials are

$$\begin{aligned} \frac{\partial\overrightarrow{p}}{\partial r}(r,\theta) &= (\cos\theta)\overrightarrow{\imath} + (\sin\theta)\overrightarrow{\jmath} \\ \frac{\partial\overrightarrow{p}}{\partial\theta}(r,\theta) &= -(r\sin\theta)\overrightarrow{\imath} + (r\cos\theta)\overrightarrow{\jmath}; \end{aligned}$$

these are independent unless $r = 0$, so we get a regular parametrization of the xy-plane provided we stay away from the origin.

We can similarly parametrize the sphere of radius R by using spherical coordinates:

$$\overrightarrow{p}(\theta,\phi) = (R\sin\phi\cos\theta, R\sin\phi\sin\theta, R\cos\phi) \tag{3.25}$$

or

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

the partials are

$$\frac{\partial \overrightarrow{p}}{\partial \phi}(\phi, \theta) = (R\cos\phi\cos\theta)\overrightarrow{\imath} + (R\cos\phi\sin\theta)\overrightarrow{\jmath} - (R\sin\phi)\overrightarrow{k}$$
$$\frac{\partial \overrightarrow{p}}{\partial \theta}(\phi, \theta) = -(R\sin\phi\sin\theta)\overrightarrow{\imath} + (R\sin\phi\cos\theta)\overrightarrow{\jmath}$$

which are independent provided $R \neq 0$ and ϕ is not a multiple of π; the latter is required because

$$\begin{aligned}\frac{\partial \overrightarrow{p}}{\partial \theta}(n\pi, \theta) &= -(R\sin(n\pi)\sin\theta)\overrightarrow{\imath} - (R\sin(n\pi)\cos\theta)\overrightarrow{\jmath} \\ &= \overrightarrow{0}.\end{aligned}$$

Regular parametrizations of surfaces share a pleasant property with regular parametrizations of curves:

Proposition 3.5.4. *A regular function $\overrightarrow{p}\colon \mathbb{R}^2 \to \mathbb{R}^3$ is* ***locally one-to-one****—that is, for every point (s_0, t_0) in the domain there exists $\delta > 0$ such that the restriction of $\overrightarrow{p}(s,t)$ to parameter values with*

$$\begin{aligned}|s - s_0| &< \delta \\ |t - t_0| &< \delta\end{aligned}$$

is one-to-one:

$$(s_1, t_1) \neq (s_2, t_2)$$

guarantees that

$$\overrightarrow{p}(s_1, t_1) \neq \overrightarrow{p}(s_2, t_2).$$

Note as before that the condition $(s_1, t_1) \neq (s_2, t_2)$ allows *one* pair of coordinates to be equal, provided the *other* pair is not; similarly, $\overrightarrow{p}(s_1, t_1) \neq \overrightarrow{p}(s_2, t_2)$ requires only that they differ in *at least one* coordinate.

Before proving Proposition 4.4.5, we establish a technical lemma.

Lemma 3.5.5. *Suppose $\overrightarrow{v}$ and $\overrightarrow{w}$ are linearly independent vectors. Then there exists a number $K(v, w) > 0$, depending continuously on $\overrightarrow{v}$ and $\overrightarrow{w}$, such that for any θ*

$$\|(\cos\theta)\overrightarrow{v} + (\sin\theta)\overrightarrow{w}\| \geq K(\overrightarrow{v}, \overrightarrow{w}).$$

The significance of this particular combination of $\overrightarrow{v}$ and $\overrightarrow{w}$ is that the coefficients, regarded as a vector $(\cos\theta, \sin\theta)$, form a *unit* vector. Any other combination of $\overrightarrow{v}$ and $\overrightarrow{w}$ is a scalar multiple of one of this type.

Proof of Lemma 3.5.5. For any θ,

$$\begin{aligned}\left\|(\cos\theta)\overrightarrow{v} + (\sin\theta)\overrightarrow{w}\right\|^2 &= ((\cos\theta)\overrightarrow{v} + (\sin\theta)\overrightarrow{w}) \cdot ((\cos\theta)\overrightarrow{v} + (\sin\theta)\overrightarrow{w})\\ &= \left\|\overrightarrow{v}\right\|^2 (\cos^2\theta) + 2\overrightarrow{v} \cdot \overrightarrow{w} \cos(\theta\sin\theta) + \left\|\overrightarrow{w}\right\|^2 (\sin^2\theta)\\ &= \frac{1}{2}\left\|\overrightarrow{v}\right\| (1+\cos 2\theta) + \overrightarrow{v} \cdot \overrightarrow{w} \sin 2\theta + \frac{1}{2}\left\|\overrightarrow{w}\right\| (1-\cos 2\theta);\end{aligned}$$

a standard calculation (Exercise 10) shows that the extreme values of this function of θ occur when

$$\tan 2\theta = \frac{2\overrightarrow{v} \cdot \overrightarrow{w}}{\left\|\overrightarrow{v}\right\|^2 - \left\|\overrightarrow{w}\right\|^2};$$

denote by θ_0 the value where the minimum occurs. It is clear that we can express θ_0 as a function of $\overrightarrow{v}$ and $\overrightarrow{w}$; let

$$K(\overrightarrow{v}, \overrightarrow{w}) = \left\|(\cos\theta_0)\overrightarrow{v} + (\sin\theta_0)\overrightarrow{w}\right\|.$$

Since $\overrightarrow{v}$ and $\overrightarrow{w}$are linearly independent, we automatically have

$$K(\overrightarrow{v}, \overrightarrow{w}) > 0.$$

□

Proof of Proposition 4.4.5. We apply Lemma 3.5.5 to the vectors

$$\begin{aligned}\overrightarrow{v} &= \frac{\partial\overrightarrow{p}}{\partial s}\\ \overrightarrow{w} &= \frac{\partial\overrightarrow{p}}{\partial t}\end{aligned}$$

to find a positive, continuous function $K(s,t)$ defined on the domain of $\overrightarrow{p}$ such that for every θ the vector

$$\overrightarrow{v}(s,t,\theta) = (\cos\theta)\frac{\partial\overrightarrow{p}}{\partial s}(s,t) + (\sin\theta)\frac{\partial\overrightarrow{p}}{\partial t}(s,t)$$

has

$$\left\|\overrightarrow{v}(s,t,\theta)\right\| \geq K(s,t).$$

In particular, given s, t, and an angle θ, we know that some *component* of the vector $\overrightarrow{v}(s,t,\theta)$ must have absolute value exceeding $K(s,t)/2$:

$$|v_j(s,t,\theta)| > \frac{K(s,t)}{2}.$$

(Do you see why this is necessary?)

Given (s_0,t_0) in the domain of $\overrightarrow{p}$, we can use the continuity of $K(s,t)$ to replace it with a positive constant K that works for all (s,t) sufficiently near (s_0,t_0). In particular, we can identify three (overlapping) sets of θ-values, say Θ_j $(j = 1,2,3)$ such that every θ belongs to at least one of them, and for every $\theta \in \Theta_j$ the estimate above works at (s_0,t_0) using the j^{th} coordinate:

$$|v_j(s_0,t_0,\theta)| > \frac{K}{2}.$$

But if this estimate works using the j^{th} component for $\overrightarrow{v}(s_0,t_0,\theta)$, then it also works for all $\overrightarrow{v}(s,t,\theta)$ with (s,t) sufficiently near (s_0,t_0). Thus by restricting to parameter values close to our "base" (s_0,t_0), we can pick an index $j(\theta)$ independent of (s,t) for which the j^{th} component of $\overrightarrow{v}(s,t,\theta)$ always has absolute value greater that $K/2$.

Now suppose (s_i,t_i), $i = 1,2$ are distinct pairs of parameter values near (s_0,t_0), and consider the straight line segment joining them in parameter space. This line segment can itself be parametrized as

$$(s(\tau),t(\tau)) = (s_1,t_1) + \tau(\triangle s, \triangle t), \quad 0 \leq \tau \leq 1$$

where

$$\begin{aligned}\triangle s &= s_2 - s_1\\ \triangle t &= t_2 - t_1.\end{aligned}$$

Choose θ satisfying

$$\begin{aligned}\triangle s &= \|\overrightarrow{v}\| \cos\theta\\ \triangle t &= \|\overrightarrow{v}\| \sin\theta\end{aligned}$$

and suppose

$$\theta \in \Theta_j$$

as above. Assume without loss of generality that $j = 1$, so the j^{th} component is x. Then writing x as a function of τ along our line segment, we have

$$x(\tau) = x(s_1 + \tau\triangle s, t_1 + \tau\triangle t)$$

so differentiating

$$\begin{aligned} x'(\tau) &= \triangle s \frac{\partial x}{\partial s}(s(\tau), t(\tau)) + \triangle t \frac{\partial x}{\partial t}(s(\tau), t(\tau)) \\ &= \left(\sqrt{\triangle s^2 + \triangle t^2}\right) v_j(s, t, \theta) \end{aligned}$$

which has absolute value at least $(K/2)\sqrt{\triangle s^2 + \triangle t^2}$, and in particular is nonzero. Thus the value of the x coordinate is a strictly monotonic function of τ along the curve $\overrightarrow{p}(s(\tau), t(\tau))$ joining $\overrightarrow{p}(s_1, t_1)$ to $\overrightarrow{p}(s_2, t_2)$, and hence the points are distinct. □

The parametrization of the sphere (Equation (3.25)) shows that the conclusion of Proposition 4.4.5 breaks down if the parametrization is not regular: when $\phi = 0$ we have

$$\overrightarrow{p}(\phi, \theta) = (0, 0, 1)$$

independent of θ; in fact, the curves corresponding to fixing ϕ at a value slightly above zero are circles of constant latitude around the North Pole, while the curves corresponding to fixing θ are great circles, all going through this pole. This is reminiscent of the breakdown of polar coordinates at the origin. A point at which a C^1 function $\overrightarrow{p}: \mathbb{R}^2 \to \mathbb{R}^3$ has dependent partials (including the possibility that at least one partial is the zero vector) is called a **singular point**; points at which the partials are independent are **regular points**. Proposition 4.4.5 can be rephrased as saying that $\overrightarrow{p}: \mathbb{R}^2 \to \mathbb{R}^3$ is locally one-to-one at each of its regular points. Of course, continuity says that every point sufficiently near a given regular point (that is, corresponding to nearby parameter values) is also regular; a region in the domain of $\overrightarrow{p}: \mathbb{R}^2 \to \mathbb{R}^3$ consisting of regular points, and on which $\overrightarrow{p}$ is one-to-one is sometimes called a **coordinate patch** for the surface it is parametrizing.

We consider one more example. Let us start with a circle in the xy-plane of radius $a > 0$, centered at the origin: this can be expressed in cylindrical coordinates as

$$r = a,$$

and the point on this circle which also lies in the vertical plane corresponding to a fixed value of θ has rectangular coordinates

$$(a\cos\theta, a\sin\theta, 0).$$

We are interested, however, not in this circle, but in the surface consisting of points in $\mathbb{R}^3$ at distance b from this circle, where $0 < b < a$; this is called a **torus**. It is reasonable to assume (and this will be verified later) that for any point P not on the circle, the nearest point to to P *on* the circle lies in the vertical plane given by fixing θ at its value for P, say $\theta = \alpha$. This means that if P has cylindrical coordinates (r, α, z) then the nearest point to P on the circle is the point $Q(a\cos\alpha, a\sin\alpha, 0)$ as given above. The vector $\overrightarrow{QP}$ lies in the plane $\theta = \alpha$; its length is, by assumption, b, and if we denote the angle it makes with the radial line $\mathcal{O}Q$ by β (Figure 3.9), then we have

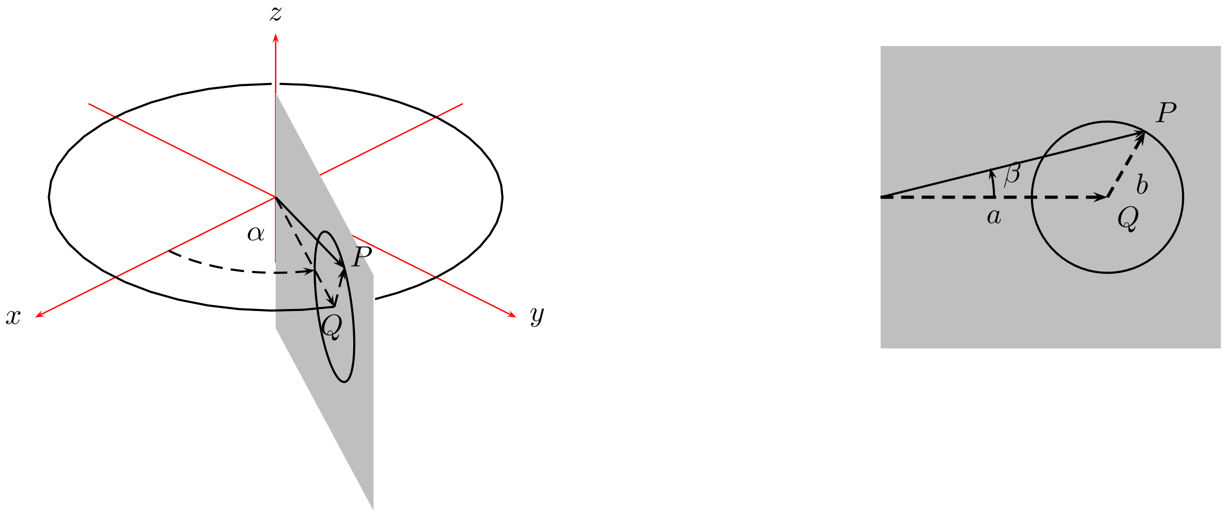


Figure 3.9: Parametrization of Torus

$$\overrightarrow{QP} = (b\cos\beta)\vec{v}_\alpha + (b\sin\beta)\vec{k}$$

where

$$\vec{v}_\alpha = (\cos\alpha)\vec{\imath} + (\sin\alpha)\vec{\jmath}$$

is the horizontal unit vector making angle α with the x-axis. Since

$$\begin{aligned}\overrightarrow{\mathcal{O}Q} &= a\vec{v}_\alpha \\ &= (a\cos\alpha)\vec{\imath} + (a\sin\alpha)\vec{\jmath}\end{aligned}$$

we see that the position vector of P is

$$\begin{aligned}\overrightarrow{\mathcal{O}P} &= \overrightarrow{\mathcal{O}Q} + \overrightarrow{QP}\\ &= [(a\cos\alpha)\overrightarrow{\imath} + (a\sin\alpha)\overrightarrow{\jmath}] + [(b\cos\beta)\overrightarrow{v}_{\alpha} + (b\sin\beta)\overrightarrow{k}]\\ &= [(a\cos\alpha)\overrightarrow{\imath} + (a\sin\alpha)\overrightarrow{\jmath}] + (b\cos\beta)[(\cos\alpha)\overrightarrow{\imath} + (\sin\alpha)\overrightarrow{\jmath}] + (b\sin\beta)\overrightarrow{k}\end{aligned}$$

so the torus (sketched in Figure 3.10) is parametrized by the vector-valued function

$$\overrightarrow{p}(\alpha,\beta) = (a + b\cos\beta)[(\cos\alpha)\overrightarrow{\imath} + (\sin\alpha)\overrightarrow{\jmath}] + (b\sin\beta)\overrightarrow{k} \tag{3.26}$$

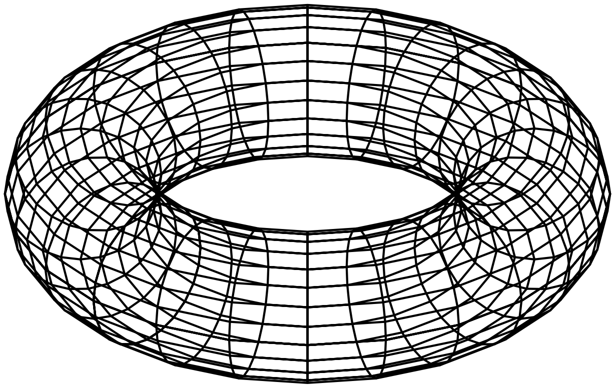


Figure 3.10: Torus

The partial derivatives of this function are

$$\begin{aligned}\frac{\partial\overrightarrow{p}}{\partial\alpha} &= (a + b\cos\beta)[(-\sin\alpha)\overrightarrow{\imath} + (\cos\alpha)\overrightarrow{\jmath}]\\ \frac{\partial\overrightarrow{p}}{\partial\beta} &= (a - b\sin\beta)[(\cos\alpha)\overrightarrow{\imath} + (\sin\alpha)\overrightarrow{\jmath}] + (b\cos\beta)\overrightarrow{k}.\end{aligned}$$

To see that these are independent, we note first that if $\cos\beta \neq 0$ this is obvious, since $\frac{\partial\overrightarrow{p}}{\partial\beta}$ has a nonzero vertical component while $\frac{\partial\overrightarrow{p}}{\partial\alpha}$ does not. If $\cos\beta = 0$, we simply note that the two partial derivative vectors are perpendicular to each other (in fact, in retrospect, this is true whatever value β has). Thus, every point is a regular point. of course, increasing either α or β by 2π will put us at the same position, so to get a coordinate patch we need to restrict each of our parameters to intervals of length $< 2\pi$.

To define the tangent plane to a regularly parametrized surface, we can think, as we did for the graph of a function, in terms of slicing the surface and finding lines tangent to the resulting curves. A more fruitful view, however, is to think in terms of arbitrary curves in the surface. Suppose $\overrightarrow{p}(r,s)$ is a C^1 function parametrizing the surface $\mathfrak{S}$ in $\mathbb{R}^3$ and $P = \overrightarrow{p}(r_0, s_0)$ is a regular point; by restricting the domain of $\overrightarrow{p}$ we can assume that we have a coordinate patch for $\mathfrak{S}$. Any curve in $\mathfrak{S}$ can be represented as

$$\overrightarrow{\gamma}(t) = \overrightarrow{p}(r(t)\,, s(t))$$

or

$$\begin{aligned} x &= x(r(t)\,, s(t)) \\ y &= y(r(t)\,, s(t)) \\ z &= z(r(t)\,, s(t)) \end{aligned}$$

—that is, we can "pull back" the curve on $\mathfrak{S}$ to a curve in the parameter space. If we want the curve to pass through P when $t = 0$, we need to require

$$\begin{aligned} r(0) &= r_0 \\ s(0) &= s_0. \end{aligned}$$

If $r(t)$ and $s(t)$ are differentiable, then by the Chain Rule $\gamma(t)$ is also differentiable, and its velocity vector can be found via

$$\begin{aligned} \overrightarrow{v}(t) &= \dot{\overrightarrow{\gamma}}(t) \\ &= \left(\frac{dx}{dt}, \frac{dx}{dt}, \frac{dx}{dt}\right) \end{aligned}$$

where

$$\begin{aligned} \frac{dx}{dt} &= \frac{\partial x}{\partial r}\frac{dr}{dt} + \frac{\partial x}{\partial s}\frac{ds}{dt} \\ \frac{dy}{dt} &= \frac{\partial y}{\partial r}\frac{dr}{dt} + \frac{\partial y}{\partial s}\frac{ds}{dt} \\ \frac{dz}{dt} &= \frac{\partial z}{\partial r}\frac{dr}{dt} + \frac{\partial z}{\partial s}\frac{ds}{dt}. \end{aligned}$$

We expect that for *any* such curve, $\overrightarrow{v}(0)$ will be parallel to the tangent plane to $\mathfrak{S}$ at P. In particular, the two curves obtained by holding one of the parameters constant will give a vector in this plane: holding s constant at $s = s_0$, we can take $r = r_0 + t$ to get

$$\overrightarrow{\gamma}(t) = \overrightarrow{p}(r_0 + t, s_0)$$

whose velocity at $t = t_0$ is

$$\overrightarrow{v}_r(0) = \frac{\partial \overrightarrow{p}}{\partial r}$$

and similarly, the velocity obtained by holding $r = r_0$ and letting $s = s_0 + t$ will be

$$\overrightarrow{v}_s(0) = \frac{\partial \overrightarrow{p}}{\partial s}.$$

Because P is a regular point, these are linearly independent and so form direction vectors for a parametrization of a plane

$$T_{(r_0,s_0)}\overrightarrow{p}(r_0 + \triangle r, s_0 + \triangle s) = \overrightarrow{p}(r_0, s_0) + \triangle r \frac{\partial \overrightarrow{p}}{\partial r} + \triangle s \frac{\partial \overrightarrow{p}}{\partial s}.$$

By looking at the components of this vector equation, we easily see that each component of $T_{(r_0,s_0)}\overrightarrow{p}(r_0 + \triangle r, s_0 + \triangle s)$ is the linearization of the corresponding component of $\overrightarrow{p}(r, s)$, and so has first order contact with it at $t = 0$. It follows, from arguments that are by now familiar, that for *any* curve in $\mathfrak{S}$

$$\begin{aligned}\overrightarrow{\gamma}(t) &= \overrightarrow{p}(r(t), s(t)) \\ &= (x(r(t), s(t)), y(r(t), s(t)), z(r(t), s(t)))\end{aligned}$$

the velocity vector

$$\overrightarrow{v}(0) = \frac{\partial \overrightarrow{p}}{\partial r}\frac{dr}{dt} + \frac{\partial \overrightarrow{p}}{\partial s}\frac{ds}{dt}$$

lies in the plane parametrized by $T\overrightarrow{p}$. It is also a straightforward argument to show that this parametrization of the tangent plane has **first order contact** with $\overrightarrow{p}(r, s)$ at $(r, s) = (r_0, s_0)$, in the sense that

$$\left\| \overrightarrow{p}(r_0 + \triangle r, s_0 + \triangle s) - T_{(r_0,s_0)}\overrightarrow{p}(r_0 + \triangle r, s_0 + \triangle s) \right\| = \mathfrak{o}(\|(\triangle r, \triangle s)\|) \text{ as } (\triangle r, \triangle s) \to \overrightarrow{0}.$$

The parametrization $T_{(r_0,s_0)}\overrightarrow{p}$ assigns to each vector $\overrightarrow{v} \in \mathbb{R}^2$ a vector $T_{(r_0,s_0)}\overrightarrow{p}(\overrightarrow{v})$ in the tangent plane at (r_0, s_0): namely if $\gamma(\tau)$ is a curve in the (s,t)-plane going through (r_0, s_0) with velocity $\overrightarrow{v}$ then the corresponding curve $\overrightarrow{p}(\gamma(\tau))$ in $\mathfrak{S}$ goes through $\overrightarrow{p}(r_0, s_0)$ with velocity $T_{(r_0,s_0)}\overrightarrow{p}(\overrightarrow{v})$. $T_{(r_0,s_0)}\overrightarrow{p}$ is sometimes called the **tangent map** at (r_0, s_0) of the parametrization $\overrightarrow{p}$.

We can also use the two partial derivative vectors $\frac{\partial\overrightarrow{p}}{\partial r}$ and $\frac{\partial\overrightarrow{p}}{\partial s}$ to find an equation for the tangent plane to $\mathfrak{S}$ at P. Since they are direction vectors for the plane, their cross product gives a normal to the plane:

$$\overrightarrow{N} = \frac{\partial\overrightarrow{p}}{\partial r} \times \frac{\partial\overrightarrow{p}}{\partial s}$$

and then the equation of the tangent plane is given by

$$\overrightarrow{N} \cdot [(x,y,z) - \overrightarrow{p}(r_0, s_0)] = 0.$$

You should check that in the special case when $\mathfrak{S}$ is the graph of a function $f(x,y)$, and $\overrightarrow{p}$ is the parametrization of $\mathfrak{S}$ as

$$\overrightarrow{p}(x,y) = (x, y, f(x,y))$$

then

$$\begin{aligned}\frac{\partial\overrightarrow{p}}{\partial x} &= \overrightarrow{\imath} + \frac{\partial f}{\partial x}\overrightarrow{k}\\ \frac{\partial\overrightarrow{p}}{\partial y} &= \overrightarrow{\jmath} + \frac{\partial f}{\partial y}\overrightarrow{k}\\ \overrightarrow{N} &= -\frac{\partial f}{\partial x}\overrightarrow{\imath} - \frac{\partial f}{\partial y}\overrightarrow{\jmath} + \overrightarrow{k}\end{aligned}$$

yielding the usual equation for the tangent plane.

We summarize these observations in the following

Remark 3.5.6. *If $\overrightarrow{p}:\mathbb{R}^2 \to \mathbb{R}^3$ is regular at (r_0, s_0), then*

1. *The linearization of $\overrightarrow{p}(r,s)$ at $r = r_0$, $s = s_0$*

$$T_{(r_0,s_0)}\overrightarrow{p}(r_0 + \triangle r, s_0 + \triangle s) = \overrightarrow{p}(r_0, s_0) + \triangle r\frac{\partial\overrightarrow{p}}{\partial r} + \triangle s\frac{\partial\overrightarrow{p}}{\partial s}$$

has first-order contact with $\overrightarrow{p}(r,s)$ at $r = r_0$, $s = s_0$;

2. *it parametrizes a plane through* $P = \overrightarrow{p}(r_0, s_0) = (x_0, y_0, z_0)$ *which contains the velocity vector of any curve passing through* P *in the surface* $\mathfrak{S}$ *parametrized by* $\overrightarrow{p}$*;*

3. *the equation of this plane is*

$$\overrightarrow{N} \cdot (x - x_0, y - y_0, z - z_0) = 0$$

where

$$\overrightarrow{N} = \frac{\partial \overrightarrow{p}}{\partial r} \times \frac{\partial \overrightarrow{p}}{\partial s}.$$

This plane is the ***tangent plane*** *to* $\mathfrak{S}$ *at* P*.*

Let us consider two quick examples.

First, we consider the sphere parametrized using spherical coordinates in Equation (3.25); using $R = 1$ we have

$$\overrightarrow{p}(\theta, \phi) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi).$$

Let us find the tangent plane at

$$P\left(\frac{1}{2}, -\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$$

which corresponds to

$$\begin{aligned}\phi &= \frac{\pi}{6}\\ \theta &= -\frac{\pi}{4}.\end{aligned}$$

The partials are

$$\begin{aligned}\frac{\partial \overrightarrow{p}}{\partial \phi}\left(\frac{\pi}{6}, -\frac{\pi}{4}\right) &= \left(\cos\frac{\pi}{6}\cos\left(-\frac{\pi}{4}\right)\right)\overrightarrow{\imath} + \left(\cos\frac{\pi}{6}\sin\left(-\frac{\pi}{4}\right)\right)\overrightarrow{\jmath} - \left(\sin\frac{\pi}{6}\right)\overrightarrow{k}\\ &= \frac{1}{2\sqrt{2}}\overrightarrow{\imath} - \frac{1}{2\sqrt{2}}\overrightarrow{\jmath} + \frac{\sqrt{3}}{2}\overrightarrow{k}\\ \frac{\partial \overrightarrow{p}}{\partial \theta}\left(\frac{\pi}{6}, -\frac{\pi}{4}\right) &= -\left(\sin\frac{\pi}{6}\sin\left(-\frac{\pi}{4}\right)\right)\overrightarrow{\imath} + \left(\sin\frac{\pi}{6}\cos\left(-\frac{\pi}{4}\right)\right)\overrightarrow{\jmath}\\ &= -\frac{\sqrt{3}}{2\sqrt{2}}\overrightarrow{\imath} + \frac{\sqrt{3}}{2\sqrt{2}}\overrightarrow{\jmath}\end{aligned}$$

so a parametrization of the tangent plane is given by

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

to find an equation for the tangent plane, we compute the normal

$$\begin{aligned}
\overrightarrow{N} &= \left(\frac{1}{2\sqrt{2}}\overrightarrow{\imath} - \frac{1}{2\sqrt{2}}\overrightarrow{\jmath} + \frac{\sqrt{3}}{2}\overrightarrow{k}\right) \times \left(-\frac{\sqrt{3}}{2\sqrt{2}}\overrightarrow{\imath} + \frac{\sqrt{3}}{2\sqrt{2}}\overrightarrow{\jmath}\right) \\
&= -\frac{3}{4\sqrt{2}}\overrightarrow{\imath} - \frac{3}{4\sqrt{2}}\overrightarrow{\jmath} + \frac{\sqrt{3}}{4}\overrightarrow{k}
\end{aligned}$$

so the equation of the tangent plane is

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

Next, we consider the torus with outer radius $a = 2$ and inner radius $b = 1$ parametrized by

$$\overrightarrow{p}(\alpha, \beta) = (2 - \sin\beta)[(\cos\alpha)\overrightarrow{\imath} + (\sin\alpha)\overrightarrow{\jmath}] + (\cos\beta)\overrightarrow{k}$$

at

$$P\left(\frac{3}{4}, \frac{3\sqrt{3}}{4}, \frac{\sqrt{3}}{2}\right)$$

which corresponds to

$$\begin{aligned}
\alpha &= \frac{\pi}{3} \\
\beta &= \frac{\pi}{6}.
\end{aligned}$$

The partials are

$$\begin{aligned}
\frac{\partial \overrightarrow{p}}{\partial \alpha} &= \left(2+\cos\frac{\pi}{6}\right)\left[\left(-\sin\frac{\pi}{3}\right)\overrightarrow{\imath}+\left(\cos\frac{\pi}{3}\right)\overrightarrow{\jmath}\right] \\
&= \left(2+\frac{\sqrt{3}}{2}\right)\left[-\frac{\sqrt{3}}{2}\overrightarrow{\imath}+\frac{1}{2}\overrightarrow{\jmath}\right] \\
&= -\left(\frac{4\sqrt{3}+3}{4}\right)\overrightarrow{\imath}+\left(\frac{4\sqrt{3}+3}{4}\right)\overrightarrow{\jmath} \\
\frac{\partial \overrightarrow{p}}{\partial \beta} &= \left(2-\sin\frac{\pi}{6}\right)\left[\left(\cos\frac{\pi}{3}\right)\overrightarrow{\imath}+\left(\sin\frac{\pi}{3}\right)\overrightarrow{\jmath}\right]+\left(\cos\frac{\pi}{6}\right)\overrightarrow{k} \\
&= \left(2-\frac{1}{2}\right)\left[\frac{1}{2}\overrightarrow{\imath}+\frac{\sqrt{3}}{2}\overrightarrow{\jmath}\right]+\frac{\sqrt{3}}{2}\overrightarrow{k} \\
&= \frac{3}{4}\overrightarrow{\imath}+\frac{3\sqrt{3}}{4}\overrightarrow{\jmath}+\frac{\sqrt{3}}{2}\overrightarrow{k}
\end{aligned}$$

so a parametrization of the tangent plane is

$$\begin{aligned}
x &= \frac{3}{4}-\left(\frac{4\sqrt{3}+3}{4}\right)\triangle r+\frac{3}{4}\triangle s \\
y &= \frac{3\sqrt{3}}{4}+\left(\frac{4\sqrt{3}+3}{4}\right)\triangle r+\frac{\sqrt{3}}{2}\triangle s \\
z &= \frac{\sqrt{3}}{2}+\frac{\sqrt{3}}{2}\triangle s.
\end{aligned}$$

The normal to the tangent plane is

$$\begin{aligned}
\overrightarrow{N} &= \left(-\left(\frac{4\sqrt{3}+3}{4}\right)\overrightarrow{\imath}+\left(\frac{4\sqrt{3}+3}{4}\right)\overrightarrow{\jmath}\right)\times\left(\frac{3}{4}\overrightarrow{\imath}+\frac{3\sqrt{3}}{4}\overrightarrow{\jmath}+\frac{\sqrt{3}}{2}\overrightarrow{k}\right) \\
&= \left(\frac{12+3\sqrt{3}}{8}\right)\overrightarrow{\imath}+\left(\frac{12+3\sqrt{3}}{8}\right)\overrightarrow{\jmath}-\left(\frac{45+21\sqrt{3}}{16}\right)\overrightarrow{k}
\end{aligned}$$

so an equation for the plane is

$$\left(\frac{12+3\sqrt{3}}{8}\right)\left(x-\frac{3}{4}\right)+\left(\frac{12+3\sqrt{3}}{8}\left(y-\frac{3\sqrt{3}}{4}\right)\right)-\left(\frac{45+21\sqrt{3}}{16}\right)\left(z-\frac{\sqrt{3}}{2}\right)=0.$$

Level Surfaces

When a surface is defined by an equation in x, y and z, it is being presented as a level surface of a function $f(x, y, z)$. The Implicit Function Theorem for $\mathbb{R}^3 \to \mathbb{R}$ (Theorem 3.4.3) tells us that in theory, we can express the locus of such an equation near a regular point of f as the graph of a function expressing one of the variables in terms of the other two, and tells us how to find the tangent plane to the level surface at this point. However, this can be done more directly, in terms of the gradient or linearization of f, and justified geometrically.

When we looked at the tangent plane to a parametrized surface, we saw that the tangent plane at a point $P = \overrightarrow{p}(s_0, t_0)$ is the image of the linearization of the parametrizing map at (s_0, t_0). Another way to think of this is to say that the tangent plane at P is made up of all the velocity vectors for curves in the surface as they pass through P. Of course, in this view, we no longer think of these vectors as "free": they are firmly attached at their base to P. This point of view is useful in identifying the tangent plane to a level surface of a function $f(x, y, z)$.

Recall that $\overrightarrow{a} \in \mathbb{R}^3$ is a **regular point** of the real-valued function $f(x, y, z)$ if the gradient $\overrightarrow{\nabla} f(\overrightarrow{a})$ at that point is not the zero vector (in other words, at least one partial does not vanish). Suppose $P(x_0, y_0, z_0)$ is a regular point of f, and suppose $\overrightarrow{p}(t)$ is a differentiable curve in the level surface $\mathcal{L}(f, c)$ through P (so $c = f(P)$), with $\overrightarrow{p}(0) = P$. On one hand, by the Chain Rule (3.3.6) we know that

$$\left.\frac{d}{dt}\right|_{t=0} [f(\overrightarrow{p}(t))] = \overrightarrow{\nabla} f(P) \cdot \vec{p}\,'(0)\,;$$

on the other hand, since $\overrightarrow{p}(t)$ lies in the level set $\mathcal{L}(f, c)$, $f(\overrightarrow{p}(t)) = c$ for all t, and in particular,

$$\left.\frac{d}{dt}\right|_{t=0} [f(\overrightarrow{p}(t))] = 0.$$

It follows that

$$\overrightarrow{\nabla} f(P) \cdot \vec{p}\,'(0) = 0$$

for every vector tangent to $\mathcal{L}(f, c)$ at P; in other words,[10]

[10] Strictly speaking, we have only shown that every tangent vector is perpendicular to

Remark 3.5.7. *If P is a regular point of $f(x,y,z)$, then the tangent plane to the level set $\mathcal{L}(f,c)$ through P is the plane through P perpendicular to the gradient vector $\overrightarrow{\nabla} f(P)$ of f at P.*

If we write this out in terms of coordinates, we find that a point $(x,y,z) = (x_0 + \triangle x, y_0 + \triangle y, z_0 + \triangle z)$ lies on the plane tangent at (x_0,y_0,z_0) to the surface $f(x,y,z) = c = f(x_0,y_0,z_0)$ if and only if

$$\left(\frac{\partial f}{\partial x}(x_0,y_0,z_0)\right)\triangle x + \left(\frac{\partial f}{\partial y}(x_0,y_0,z_0)\right)\triangle y + \left(\frac{\partial f}{\partial z}(x_0,y_0,z_0)\right)\triangle z = 0,$$

in other words, if

$$d_{(x_0,y_0,zso)}f(x - x_0, y - y_0, z - z_0) = 0.$$

Yet a third way to express this is to add $c = f(x_0,y_0,z_0)$ to both sides, noting that the left side then becomes the linearization of f at P:

$$T_{x_0,y_0,z_0}f(x,y,z) = f(x_0,y_0,z_0)\,.$$

We summarize all of this in

Proposition 3.5.8. *Suppose $P(x_0,y_0,z_0)$ is a regular point of the real-valued function $f(x,y,z)$ and $f(x_0,y_0,z_0) = c$. Then the level set of f through P*

$$\mathcal{L}(f,c) := \{(x,y,z) \mid f(x,y,z) = c\}$$

has a tangent plane $\mathcal{P}$ at P, which can be characterized in any of the following ways:

- *$\mathcal{P}$ is the plane through P with normal vector $\overrightarrow{\nabla} f(P)$;*
- *$\mathcal{P}$ is the set of all points $P + \overrightarrow{v}$ where*

$$d_P f(\overrightarrow{v}) = 0;$$

- *$\mathcal{P}$ is the level set $\mathcal{L}(T_P f, f(P))$ through P of the linearization of f at P.*

$\overrightarrow{\nabla} f$; we need to also show that every vector which is perpendicular to $\overrightarrow{\nabla} f$ is the velocity vector of some curve in $\mathcal{L}(f,c)$ as it goes through P. See Exercise 11.

Let us see how this works out in practice for a few examples.

First, let us find the plane tangent to the ellipsoid

$$x^2 + 3y^2 + 4z^2 = 20$$

at the point $P(2, -2, -1)$. This can be regarded as the level set $\mathcal{L}(f, 20)$ of the function

$$f(x, y, z) = x^2 + 3y^2 + 4z^2.$$

We calculate the partials

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2x \\ \frac{\partial f}{\partial y} &= 6y \\ \frac{\partial f}{\partial z} &= 8z \end{aligned}$$

which gives the gradient

$$\overrightarrow{\nabla} f(2, -2, -1) = 4\overrightarrow{\imath} - 12\overrightarrow{\jmath} - 8\overrightarrow{k}.$$

Thus the tangent plane is the plane through $P(2, -2, -1)$ perpendicular to $4\overrightarrow{\imath} - 12\overrightarrow{\jmath} - 8\overrightarrow{k}$, which has equation

$$4(x - 2) - 12(y + 2) - 8(z + 1) = 0$$

or

$$4x - 12y - 8z = 8 + 24 + 8 = 40.$$

We note that this is the same as

$$d_{(2,-2,-1)} f(\triangle x, \triangle y, \triangle z) = 0$$

with

$$\begin{aligned} \triangle x &= x - 2 \\ \triangle y &= y - (-2) \\ \triangle z &= z - (-1), \end{aligned}$$

or, calculating the linearization

$$\begin{aligned} T_{(2,-2,-1)}f(x,y,z) &= 20 + 4(x-2) - 12(y+2) - 8(z+1) \\ &= 4x - 12y - 8z - 20 \end{aligned}$$

the tangent plane is the level set of the linearization

$$\mathcal{L}\Big(T_{(2,-2,-1)}f, 20\Big) = \{(x,y,z) \,|\, T_{(2,-2,-1)}f(x,y,z) = 20\}.$$

We note in passing that in this case we could also have solved for z in terms of x and y:

$$\begin{aligned} 4z^2 &= 20 - x^2 - 3y^2 \\ z^2 &= 5 - \frac{x^2}{4} - \frac{3y^2}{4} \\ z &= \pm\sqrt{5 - \frac{x^2}{4} - \frac{3y^2}{4}} \end{aligned}$$

and since at our point z is negative, the nearby solutions are

$$z = -\sqrt{5 - \frac{x^2}{4} - \frac{3y^2}{4}}.$$

This would have given us an expression for the ellipsoid near $(2,-2,-1)$ as the graph $z = \phi(x,y)$ of the function of x and y

$$\phi(x,y) = -\sqrt{5 - \frac{x^2}{4} - \frac{3y^2}{4}}.$$

The partials of this function are

$$\begin{aligned} \frac{\partial\phi}{\partial x} &= -\frac{-x/4}{\sqrt{5 - \frac{x^2}{4} - \frac{3y^2}{4}}} \\ \frac{\partial\phi}{\partial y} &= -\frac{-3y/4}{\sqrt{5 - \frac{x^2}{4} - \frac{3y^2}{4}}}; \end{aligned}$$

at our point, these have values

$$\begin{aligned} \frac{\partial\phi}{\partial x}(2,-2) &= \frac{1}{2} \\ \frac{\partial\phi}{\partial y}(2,-2) &= -\frac{3}{2} \end{aligned}$$

so the parametric form of the tangent plane is

$$\begin{cases} x &= \quad 2 \quad +s \\ y &= -2 \qquad\qquad +t \\ z &= -1 \quad +\frac{s}{2} \quad -\frac{3t}{2} \end{cases}$$

while the equation of the tangent plane can be formulated in terms of the normal vector

$$\begin{aligned} \overrightarrow{n} &= \overrightarrow{v}_x \times \overrightarrow{v}_y \\ &= (\overrightarrow{\imath} + \frac{1}{2}\overrightarrow{k}) \times (\overrightarrow{\jmath} - \frac{3}{2}\overrightarrow{k}) \\ &= -\left(\frac{1}{2}\right)\overrightarrow{\imath} - \left(-\frac{3}{2}\right)\overrightarrow{\jmath} + \overrightarrow{k} \end{aligned}$$

as

$$-\frac{1}{2}(x-2) + \frac{3}{2}(y+2) + (z+1) = 0$$

or

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

which we recognize as our earlier equation, divided by -8.

As a second example, we consider the surface

$$x^3y^2z + x^2y^3z + xyz^3 = 30$$

near the point $P(-2,3,1)$. This time, it is not feasible to solve for any one of the variables in terms of the others; our only choice is to work directly with this as a level surface of the function

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

The partials of this function are

$$\begin{aligned} \frac{\partial f}{\partial x} &= 3x^2y^2z + 2zy^3z + yz^3 \\ \frac{\partial f}{\partial y} &= 2x^3yz + 3x^2y^2z + xz^3 \\ \frac{\partial f}{\partial z} &= x^3y^2 + x^2y^3 + 3xyz^2. \end{aligned}$$

The values of these at our point are

$$\frac{\partial f}{\partial x}(-2,3,1) = 3$$
$$\frac{\partial f}{\partial y}(-2,3,1) = 58$$
$$\frac{\partial f}{\partial z}(-2,3,1) = 28$$

giving as the equation of the tangent plane

$$3(x+2) + 58(y-3) + 28(z-1) = 0$$

or

$$3x + 58y + 28z = 196.$$

You should check that this is equivalent to any one of the forms of the equation given in Proposition 3.5.8.

Exercises for § 3.5

Practice Problems:

For each given surface, express the tangent plane (a) as an equation in x, y and z (b) in parametrized form:

1. $z = x^2 - y^2$, $(1,-2,-3)$, $(2,-1,3)$
2. $z^2 = x^2 + y^2$, $(1,1,\sqrt{2})$, $(2,-1,\sqrt{5})$
3. $x^2 + y^2 - z^2 = 1$, $(1,-1,1)$, $(\sqrt{3},0,\sqrt{2})$
4. $x^2 + y^2 + z^2 = 4$, $(1,1,\sqrt{2})$, $(\sqrt{3},1,0)$
5. $x^3 + 3xy + z^2 = 2$, $(1,\frac{1}{3},0)$, $(0,0,\sqrt{2})$
6. $$\begin{cases} x &= 2s & \\ y &= s^2 & +t \\ z &= & t^2 \end{cases} \text{ at } (0,1,1)$$
7. $$\begin{cases} x &= u^2 & -v^2 \\ y &= u & +v \\ z &= u^2 & +4v \end{cases} \text{ at } (-\frac{1}{4},\frac{1}{2},2).$$

8.
$$\begin{cases} x &= (2-\cos v) & \cos u \\ y &= (2-\cos v) & \sin u \\ z &= & \sin v \end{cases}$$

at any point (give in terms of u and v).

Theory problems:

9. (a) Verify that $\overrightarrow{p}(s,t) = (s,t,f(s,t))$ is a regular parametrization of the graph $z = f(x,y)$ of any $\mathcal{C}^1$ function $f(x,y)$ of two variables.

   (b) What is the appropriate generalization for $n > 2$ variables?

10. Show that the extreme values of the function $\left\|(\cos\theta)\overrightarrow{v} + (\sin\theta)\overrightarrow{w}\right\|^2$ occur when
$$\tan 2\theta = \frac{2\overrightarrow{v}\cdot\overrightarrow{w}}{\left\|\overrightarrow{v}\right\|^2 - \left\|\overrightarrow{w}\right\|^2}.$$

Challenge problem:

11. Suppose $P(x_0, y_0, z_0)$ is a regular point of the $\mathcal{C}^1$ function $f(x,y,z)$; for definiteness, assume $\frac{\partial f}{\partial z}(P) \neq 0$. Let $\overrightarrow{v}$ be a nonzero vector perpendicular to $\overrightarrow{\nabla} f(P)$.

    (a) Show that the projection $\overrightarrow{w} = (v_1, v_2)$ of $\overrightarrow{v}$ onto the xy-plane is a nonzero vector.

    (b) By the Implicit Function Theorem, the level set $\mathcal{L}(f,c)$ of f through P near P can be expressed as the graph $z = \phi(x,y)$ of some $\mathcal{C}^1$ function $\phi(x,y)$. Show that (at least for $|t| < \varepsilon$ for some $\varepsilon > 0$) the curve $\overrightarrow{p}(t) = (x_0 + v_1 t, y_0 + v_2 t, \phi(x_0 + v_1 t, y_0 + v_2 t))$ lies on $\mathcal{L}(f,c)$, and that $\vec{p}\,'(0) = \overrightarrow{v}$.

    (c) This shows that every vector in the plane perpendicular to the gradient is the velocity vector of some curve in $\mathcal{L}(f,c)$ as it goes through P, at least if $\overrightarrow{\nabla} f(P)$ has a nonzero z-component. What do you need to show this assuming only that $\overrightarrow{\nabla} f(P)$ is a nonzero vector?

3.6 Extrema

Bounded Functions

Recall the following definitions from single-variable calculus:

Definition 3.6.1. *Suppose S is a set of real numbers.*

1. *$\alpha \in \mathbb{R}$ is a **lower bound** for S if*

   $$\alpha \leq s \text{ for every } s \in S.$$

   *The set S is **bounded below** if there exists a lower bound for S.*

2. *$\beta \in \mathbb{R}$ is an **upper bound** for S if*

   $$s \leq \beta \text{ for every } s \in S.$$

   *The set S is **bounded above** if there exists an upper bound for S.*

3. *A set of real numbers is **bounded** if it is bounded below and bounded above.*

4. *If S is bounded below, there exists a unique lower bound A for S such that every lower bound α for S satisfies $\alpha \leq A$; it is called the **infimum** of S, and denoted* **inf** $\boldsymbol{S}$.

   A lower bound α for S equals $\inf S$ *precisely if there exists a sequence $\{s_i\}$ of elements of S with $s_i \to \alpha$.*

5. *If S is bounded above, there exists a unique upper bound B for S such that every upper bound β for S satisfies $\beta \geq B$; it is called the **supremum** of S, and denoted* **sup** $\boldsymbol{S}$.

   An upper bound β for S equals $\sup S$ *precisely if there exists a sequence $\{s_i\}$ of elements of S with $s_i \to \beta$.*

6. *A lower (resp. upper) bound for S is the **minimum** (resp. **maximum**) of S if it belongs to S. When it exists, the minimum (resp. maximum) of S is also its infimum (resp. supremum).*

These notions can be applied to the image, or set of values taken on by a real-valued function on a set of points in $\mathbb{R}^2$ or $\mathbb{R}^3$ (we shall state these for $\mathbb{R}^3$; the two-dimensional analogues are essentially the same):

Definition 3.6.2. *Suppose $f:\mathbb{R}^3 \to \mathbb{R}$ is a real-valued function with domain* $\operatorname{dom}(f) \subset \mathbb{R}^3$*, and let $S \subset \operatorname{dom}(f)$ be any subset of the domain of f. The* ***image*** *of S under f is the set of values taken on by f among the points of S:*

$$f(S) := \{f(s) \mid s \in S\}.$$

1. *f is* ***bounded*** *(resp.* ***bounded below***, ***bounded above****) on S if $f(S)$ is bounded (resp. bounded below, bounded above).*
2. *The* ***supremum*** *(resp.* ***infimum****) of f on S is defined by*

$$\sup_{x \in S} f(x) = \sup f(S)$$
$$\inf_{x \in S} f(x) = \inf f(S)\,.$$

3. *The function f* ***achieves its maximum*** *(resp.* ***achieves its minimum****) on S at $x \in S$ if*

$$f(x) \geq \ \textit{(resp. } \leq\textit{)}\ f(s) \ \textit{ for all } s \in S.$$

We shall say that x is an ***extreme point*** *of $f(x)$ on S if $f(x)$ achieves its maximum or minimum on S at x; the value $f(x)$ will be referred to as an* ***extreme value*** *of $f(x)$ on S.*

In all the statements above, when the set S is not mentioned explicitly, it is understood to be the whole domain of f.

The Extreme Value Theorem

A basic result in single-variable calculus is the Extreme Value Theorem, which says that a continuous function achieves its maximum and minimum on any closed, bounded interval $[a, b]$. We wish to extend this to result to real-valued functions defined on subsets of $\mathbb{R}^3$. First, we need to set up some terminology.

Definition 3.6.3. *A set $S \subset \mathbb{R}^3$ of points in $\mathbb{R}^3$ is* ***closed*** *if for any convergent sequence s_i of points in S, the limit also belongs to S:*

$$s_i \to L \textit{ and } s_i \in S \textit{ for all } i \ \Rightarrow L \in S.$$

It is an easy exercise (Exercise 9) to show that each of the following are examples of closed sets:

1. closed intervals $[a, b]$ in $\mathbb{R}$, as well as half-closed intervals of the form $[a, \infty)$ or $(-\infty, b]$;
2. level sets $\mathcal{L}(g, c)$ of a continuous function g, as well as sets defined by weak inequalities like $\{x \in \mathbb{R}^3 \mid g(x) \leq c\}$ or $\{x \in \mathbb{R}^3 \mid g(x) \geq c\}$;
3. any set consisting of a convergent sequence s_i together with its limit, or any set consisting of a sequence together with all of its accumulation points.

We also want to formulate the idea of a bounded set in $\mathbb{R}^3$. We cannot talk about such a set being "bounded above" or "bounded below"; the appropriate definition is

Definition 3.6.4. *A set $S \subset \mathbb{R}^3$ is **bounded** if the set of lengths of elements of S $\{\|s\| \mid s \in S\}$ is bounded—that is, if there exists $M \in \mathbb{R}$ such that*

$$\|s\| \leq M \text{ for all } s \in S.$$

(This is the same as saying that there exists some ball $B_\varepsilon(\mathcal{O})$—where $\varepsilon > 0$ is in general not assumed small—which contains S.)

A basic and important property of $\mathbb{R}^3$ is stated in the following.

Proposition 3.6.5. *For a subset $S \subset \mathbb{R}^3$, the following are equivalent:*

1. *S is closed and bounded;*
2. *S is **sequentially compact**: every sequence s_i of points in S has a subsequence which converges to a point of S.*

We shall abuse terminology and refer to such sets as **compact** sets.[11]

Proof. If S is bounded, then by the Bolzano-Weierstrass Theorem (Proposition 2.3.7) every sequence in S has a convergent subsequence, and if S is also closed, then the limit of this subsequence must also be a point of S.

Conversely, if S is *not bounded*, it cannot be sequentially compact since there must exist a sequence s_k of points in S with $\|s_k\| > k$; such a sequence has no convergent subsequence. Similarly, if S is *not closed*, there must exist a convergent sequence s_k of points in S whose limit L lies outside S; since every subsequence also converges to L, S cannot be sequentially compact. □

[11] The property of being *compact* has a specific definition in very general settings; however, in the context of $\mathbb{R}^3$, this is equivalent to either sequential compactness or being closed and bounded.

With these definitions, we can formulate and prove the following.

Theorem 3.6.6 (Extreme Value Theorem)**.** *If $S \subset \mathbb{R}^3$ is compact, then every real-valued function f which is continuous on S achieves its minimum and maximum on S.*

Note that this result includes the Extreme Value Theorem for functions of one variable, since closed intervals are compact, but even in the single variable setting, it applies to functions continuous on sets more general than intervals.

Proof. The strategy of this proof is: first, we show that f must be bounded on S, and second, we prove that there exists a point $s \in S$ where $f(s) = \sup_{x\in S} f(x)$ (*resp.* $f(s) = \inf_{x\in S} f(x)$).[12]

Step 1: *$f(x)$ is bounded on S:* Suppose $f(x)$ is *not* bounded on S: this means that there exist points in S at which $|f(x)|$ is arbitrarily high: thus we can pick a sequence $s_k \in S$ with $|f(s_k)| > k$. Since S is (sequentially) compact, we can find a subsequence—which without loss of generality can be assumed to be the whole sequence—which converges to a point of S: $s_k \to s_0 \in S$. Since $f(x)$ is continuous on S, we must have $f(s_k) \to f(s_0)$; but this contradicts the assumption that $|f(s_k)| > k$.

Step 2: *$f(x)$ achieves its maximum and minimum on S:* We will show that $f(x)$ achieves its maximum on S; the case of the minimum is entirely analogous. Since $f(x)$ is bounded on S, the set of values on S has a supremum, say $\sup_{x\in S} f(x) = A$; by the remarks in Definition 3.6.1, there exists a sequence $f(s_i)$ converging to A, where s_i all belong to S; pick a subsequence of s_i which converges to $s_0 \in S$; by continuity $f(s_0) = A$ and we are done. □

Local Extrema

How do we find the extreme values of a function on a set? For a function of one variable on an interval, we looked for local extrema interior to the interval and compared them to the values at the ends. Here we need to formulate the analogous items. The following is the natural higher-dimension analogue of local extrema for single-variable functions.

Definition 3.6.7. *The function $f(x)$ has a* ***local maximum*** *(resp.* ***local minimum****) at $\vec{x}_0 \in \mathbb{R}^3$ if there exists a ball $B_\varepsilon(\vec{x}_0)$, $\varepsilon > 0$, such that*

[12] A somewhat different proof, based on an idea of Daniel Reem, is worked out in Exercise 13.

1. *$f(x)$ is defined on all of $B_\varepsilon(\overrightarrow{x}_0)$; and*

2. *$f(x)$ achieves its maximum (resp. minimum) on $B_\varepsilon(\overrightarrow{x}_0)$ at $\overrightarrow{x} = \overrightarrow{x}_0$.*

*A **local extremum** of $f(x)$ is a local maximum or local minimum.*

To handle sets more complicated than intervals, we need to formulate the analogues of interior points and endponts.

Definition 3.6.8. *Let $S \subset \mathbb{R}^3$ be any set in $\mathbb{R}^3$.*

1. *A point $\overrightarrow{x} \in \mathbb{R}^3$ is an **interior point** of S if S contains some ball about $\overrightarrow{x}$:*

$$B_\varepsilon(\overrightarrow{x}) \subset S.$$

   *The set of all interior points of S is called the **interior** of S, denoted **int S**.*

   *A set S is **open** if every point is an interior point: $S = \text{int } S$.*

2. *A point $\overrightarrow{x} \in \mathbb{R}^3$ is a **boundary point** of S if every ball $B_\varepsilon(\overrightarrow{x})$, $\varepsilon > 0$ contains points* in *S as well as points* not in *S:*

$$B_\varepsilon(\overrightarrow{x}) \cap S \neq \emptyset, \text{ but } B_\varepsilon(\overrightarrow{x}) \not\subset S.$$

   *The set of boundary points of S is called the **boundary** and denoted $\boldsymbol{\partial S}$.*

The following are relatively easy observations (Exercise 10):

Remark 3.6.9. 1. *For any set $S \subset \mathbb{R}^3$,*

$$S \subseteq \text{int } S \cup \partial S.$$

2. *The boundary ∂S of any set is closed.*

3. *S is closed precisely if it contains its boundary points:*

$$S \text{ closed} \Leftrightarrow \partial S \subset S.$$

4. *$S \subset \mathbb{R}^3$ is* closed *precisely if its complement*

$$\mathbb{R}^3 \setminus S := \{x \in \mathbb{R}^3 \,|\, x \notin S\}$$

   is open.

The lynchpin of our strategy for finding extrema in the case of single-variable functions was that every local extremum is a critical point, and in most cases there are only finitely many of these. The analogue for our present situation is the following.

Theorem 3.6.10 (Critical Point Theorem)**.** *If $f:\mathbb{R}^3\to\mathbb{R}$ has a local extremum at $\overrightarrow{x}=\overrightarrow{x}_0$ and is differentiable there, then it is a critical point of $f(\overrightarrow{x})$:*

$$\overrightarrow{\nabla} f(\overrightarrow{x}_0)=\overrightarrow{0}.$$

Proof. If $\overrightarrow{\nabla} f(\overrightarrow{x}_0)$ is not the zero vector, then some partial derivative, say $\frac{\partial f}{\partial x_j}$, is nonzero. But this means that along the line through $\overrightarrow{x}_0$ parallel to the x_j-axis, the function is locally monotone:

$$\frac{d}{dt}\left[f(\overrightarrow{x}_0+t\overrightarrow{e}_j)\right]=\frac{\partial f}{\partial x_j}(\overrightarrow{x}_0)\neq 0$$

means that there are nearby points where the function exceeds, and others where it is less than, the value at $\overrightarrow{x}_0$; therefore $\overrightarrow{x}_0$ is *not* a local extreme point of $f(\overrightarrow{x})$. □

Finding Extrema

Putting all this together, we can formulate a strategy for finding the extreme values of a function on a subset of $\mathbb{R}^3$, analogous to the strategy used in single-variable calculus:

Given a function $f(\overrightarrow{x})$ defined on the set $S\subset\mathbb{R}^3$, search for extreme values as follows:

1. **Critical Points:** Locate all the critical points of $f(\overrightarrow{x})$ interior to S, and evaluate $f(\overrightarrow{x})$ at each.

2. **Boundary Behavior:** Find the maximum and minimum values of $f(\overrightarrow{x})$ on the boundary ∂S; if the set is unbounded, study the limiting values as $\|\overrightarrow{x}\|\to\infty$ in S.

3. **Comparison:** Compare these values: the lowest (*resp.* highest) of all the values is the infimum (*resp.* supremum), and if the point at which it is achieved lies in S, it is the minimum (*resp.* maximum) value of f on S.

In practice, this strategy is usually applied to sets of the form $S=\{\overrightarrow{x}\in\mathbb{R}^3\mid g(\overrightarrow{x})\leq c\}$. We consider a few examples.

First, let us find the maximum and minimum of the function

$$f(x,y)=x^2-2x+y^2$$

inside the disc of radius 2

$$x^2+y^2\leq 4.$$

Critical Points:

$$\vec{\nabla}f(x,y)=(2x-2)\vec{\imath}+2y\vec{\jmath}$$

this vanishes only at the point

$$\begin{aligned}x&=1\\ y&=0\end{aligned}$$

and the value of $f(x,y)$ at the critical point $(1,0)$, which lies inside the disc, is

$$\begin{aligned}f(1,0)&=1-2+0\\ &=-1.\end{aligned}$$

Boundary Behavior:

The boundary is the circle of radius 2

$$x^2+y^2=4$$

which we can parametrize as

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

so the function restricted to the boundary can be written

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

To find the extrema of this, we can either use common sense (how?) or take the derivative:

$$\frac{dg}{d\theta} = 4\sin\theta.$$

This vanishes when

$$\theta = 0, \pi.$$

The values at these places are

$$\begin{aligned} g(0) &= 4 - 4 \\ &= 0 \\ g(\pi) &= 4 + 4 \\ &= 8 \end{aligned}$$

and we see that

$$\begin{aligned} \max_{x^2+y^2\leq 4} x^2 - 2x + y^2 &= 8 \\ &= g(\pi) \\ &= f(-2,0) \\ \min_{x^2+y^2\leq 4} x^2 - 2x + y^2 &= -1 \\ &= f(1,0)\,. \end{aligned}$$

Next, let's find the extreme values of the same function on the *unbounded* set defined by

$$x \leq y:$$

here, the lone critical point $(1,0)$ lies *outside* the set, so all the extreme behavior is "at the boundary". There are two parts to this: first, we look at the behavior on the boundary points of S, which is the line

$$x = y.$$

Along this line we can write

$$\begin{aligned} g(x) &= f(x,x) \\ &= 2x^2 - 2x; \\ g'(x) &= 4x - 2 \end{aligned}$$

vanishes at

$$x = \frac{1}{2}$$

and the value there is

$$\begin{aligned} g\left(\frac{1}{2}\right) &= f\left(\frac{1}{2}, \frac{1}{2}\right) \\ &= -\frac{1}{2}. \end{aligned}$$

But we also need to consider what happens when $\|(x,y)\| \to \infty$ in our set. It is easy to see that for *any* point (x,y), $f(x,y) \geq x^2 - 2x \geq -1$, and also that $x^2 - 2x \to \infty$ if $|x| \to \infty$. For any sequence (x_j, y_j) with $\|(x_j, y_j)\| \to \infty$, either $|x| \to \infty$ (so $f(x,y) \geq x^2 - 2x \to \infty$) or $|y| \to \infty$ (so $f(x,y) \geq y^2 - 1 \to \infty$); in either case, $f(x_j, y_j) \to \infty$. Since there exist such sequences with $x_j \leq y_j$, the function is not bounded above. Now, if $\overrightarrow{s}_i = (x_i, y_i)$ is a sequence with $x_i \leq y_i$ and $f(\overrightarrow{s}_i) \to \inf_{x \leq y} f(x,y)$, either $\overrightarrow{s}_i$ have no convergent subsequence, and hence $\|\overrightarrow{s}_i\| \to \infty$, or some accumulation point of $\overrightarrow{s}_i$ is a local minimum for f. The first case is impossible, since we already know that then $f(\overrightarrow{s}_i) \to \infty$, while in the second case this accumulation point must be $\left(\frac{1}{2}, \frac{1}{2}\right)$, and then $f(\overrightarrow{s}_i) \to \frac{1}{2}$. From this it follows that

$$\min_{x \leq y}(x^2 - 2x + y^2) = -\frac{1}{2} = f\left(\frac{1}{2}, \frac{1}{2}\right).$$

Lagrange Multipliers

For problems in two variables, the boundary is a curve, which can often be parametrized, so that the problem of optimizing the function on the boundary is reduced to a one-variable problem. However, when three or more variables are involved, the boundary can be much harder to parametrize. Fortunately, there is an alternative approach, pioneered by Joseph Louis Lagrange (1736-1813) in connection with isoperimetric problems (for example, find the triangle of greatest area with a fixed perimeter)[13]

The method is applicable to problems of the form: find the extreme values of the function $f(\overrightarrow{x})$ on a level set $\mathcal{L}(g,c)$ of the differentiable function

[13] According to [48, pp. 169-170], when Lagrange communicated his method to Euler in 1755 (at the age of 18!), the older master was so impressed that he delayed publication of some of his own work on inequalities to give the younger mathematician the credit he was due for this elegant method.

$g(\overrightarrow{x})$ containing no critical points of g (we call c a **regular value** of $g(\overrightarrow{x})$ if $\overrightarrow{\nabla} g(\overrightarrow{x}) \neq \overrightarrow{0}$ whenever $g(\overrightarrow{x}) = c$). These are sometimes called **constrained extremum** problems.

The idea is this: suppose the function $f(\overrightarrow{x})$ *when restricted to the level set* $\mathcal{L}(g,c)$ has a local maximum at $\overrightarrow{x}_0$: this means that, while it might be possible to find nearby points where the function takes values higher than $f(\overrightarrow{x}_0)$, they cannot lie on the level set. Thus, we are interested in finding those points for which the function has a local maximum along any curve through the point *which lies in the level set.* Suppose that $\overrightarrow{p}(t)$ is such a curve; that is, we are assuming that

$$g(\overrightarrow{p}(t)) = c$$

for all t, and that

$$\overrightarrow{p}(0) = \overrightarrow{x}_0.$$

in order for $f(\overrightarrow{p}(t))$ to have a local maximum at $t = 0$, the derivative must vanish—that is,

$$\begin{aligned} 0 &= \left.\frac{d}{dt}\right|_{t=0} [f(\overrightarrow{p}(t))] \\ &= \overrightarrow{\nabla} f(\overrightarrow{x}_0) \cdot \overrightarrow{v} \end{aligned}$$

where

$$\overrightarrow{v} = \dot{\overrightarrow{p}}(0)$$

is the velocity vector of the curve as it passes $\overrightarrow{x}_0$: the velocity must be perpendicular to the gradient of f. This must be true for *any* curve in the level set as it passes through $\overrightarrow{x}_0$, which is the same as saying that it must be true for any vector in the plane tangent to the level set $\mathcal{L}(g,c)$ at $\overrightarrow{x}_0$: in other words, $\overrightarrow{\nabla} f(\overrightarrow{x}_0)$ must be normal to this tangent plane. But we already know that the gradient of g is normal to this tangent plane; thus the two gradient vectors must point along the same line—they must be linearly dependent! This proves

Proposition 3.6.11 (Lagrange Multipliers)**.** *If $\overrightarrow{x}_0$ is a local extreme point of the restriction of the function $f(\overrightarrow{x})$ to the level set $\mathcal{L}(g,c)$ of the function*

$g(\overrightarrow{x})$, and c is a regular value of g. Then $\overrightarrow{\nabla} f(\overrightarrow{x}_0)$ and $\overrightarrow{\nabla} g(\overrightarrow{x}_0)$ must be linearly dependent:

$$\overrightarrow{\nabla} f(\overrightarrow{x}_0) = \lambda \overrightarrow{\nabla} g(\overrightarrow{x}_0) \tag{3.27}$$

for some real number λ.

The number λ is called a **Lagrange multiplier**. We have formulated the linear dependence of the gradients as $\overrightarrow{\nabla} f$ being a multiple of $\overrightarrow{\nabla} g$, rather than the other way around, because we assume that $\overrightarrow{\nabla} g$ is nonvanishing, while this formulation allows $\overrightarrow{\nabla} f$ to vanish—that is, this equation holds automatically if $\overrightarrow{x}_0$ is a genuine critical point of f. We will refer to this weaker situation by saying $\overrightarrow{x}_0$ is a **relative critical point** of $f(\overrightarrow{x})$—that is, it is critical relative to the constraint $g(\overrightarrow{x}) = c$.

To see this method in practice, we consider a few examples.

First, let us find the extreme values of

$$f(x, y, z) = x - y + z$$

on the sphere

$$x^2 + y^2 + z^2 = 4 :$$

we have

$$\overrightarrow{\nabla} f(x, y, z) = \overrightarrow{\imath} - \overrightarrow{\jmath} + \overrightarrow{k}$$

and $g(x, y, z) = x^2 + y^2 + z^2$, so

$$\overrightarrow{\nabla} g(x, y, z) = 2x\overrightarrow{\imath} + 2y\overrightarrow{\jmath} + 2z\overrightarrow{k}.$$

The Lagrange Multiplier equation

$$\overrightarrow{\nabla} f(\overrightarrow{x}_0) = \lambda \overrightarrow{\nabla} g(\overrightarrow{x}_0)$$

amounts to the three scalar equations

$$\begin{aligned} 1 &= 2\lambda x \\ -1 &= 2\lambda y \\ 1 &= 2\lambda z \end{aligned}$$

which constitute 3 equations in 4 unknowns; a fourth equation is the specification that we are on $\mathcal{L}(g,4)$:

$$x^2 + y^2 + z^2 = 4.$$

Note that none of the four variables can equal zero (why?), so we can rewrite the three Lagrange equations in the form

$$\begin{aligned} x &= \frac{1}{2\lambda} \\ y &= -\frac{1}{2\lambda} \\ z &= \frac{1}{2\lambda}. \end{aligned}$$

Substituting this into the fourth equation, we obtain

$$\frac{1}{4\lambda^2} + \frac{1}{4\lambda^2} + \frac{1}{4\lambda^2} = 4$$

or

$$\begin{aligned} 3 &= 16\lambda^2 \\ \lambda &= \pm\frac{\sqrt{3}}{4}. \end{aligned}$$

This yields two relative critical points:

$$\lambda = \frac{\sqrt{3}}{4}$$

gives the point

$$\left(\frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}, \frac{2}{\sqrt{3}}\right)$$

where

$$f\left(\frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}, \frac{2}{\sqrt{3}}\right) = 2\sqrt{3}$$

while

$$\lambda = -\frac{\sqrt{3}}{4}$$

gives the point

$$\left(-\frac{2}{\sqrt{3}}, \frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}\right)$$

where

$$f\left(-\frac{2}{\sqrt{3}}, \frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}\right) = -2\sqrt{3}.$$

Thus,

$$\begin{aligned}\max_{x^2+y^2+z^2} f(x,y,z) &= f\left(\frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}, \frac{2}{\sqrt{3}}\right)\\ &= 2\sqrt{3}\\ \max_{x^2+y^2+z^2} f(x,y,z) &= f\left(-\frac{2}{\sqrt{3}}, \frac{2}{\sqrt{3}}, -\frac{2}{\sqrt{3}}\right)\\ &= -2\sqrt{3}.\end{aligned}$$

As a second example, let us find the point on the surface

$$xyz = 1$$

closest to the origin. We characterize the surface as $\mathcal{L}(g,1)$, where

$$\begin{aligned}g(x,y,z) &= xyz\\ \overrightarrow{\nabla} g(x,y,z) &= (yz, xz, xy).\end{aligned}$$

As is usual in distance-optimizing problems, it is easier to work with the square of the distance; this is minimized at the same place(s) as the distance, so we take

$$\begin{aligned}f(x,y,z) &= \operatorname{dist}((x,y,z),(0,0,0))^2\\ &= x^2+y^2+z^2\\ \overrightarrow{\nabla} f(x,y,z) &= (2x, 2y, 2z).\end{aligned}$$

The Lagrange Multiplier Equation

$$\overrightarrow{\nabla} f = \lambda \overrightarrow{\nabla} g$$

reads

$$\begin{aligned}2x &= \lambda yz\\ 2y &= \lambda xz\\ 2z &= \lambda xy.\end{aligned}$$

Note first that if $xyz = 1$, all three coordinates must be nonzero. Thus, we can solve each of these equations for λ:

$$\begin{aligned}\lambda &= \frac{2x}{yz}\\ \lambda &= \frac{2y}{xz}\\ \lambda &= \frac{2z}{xy}.\end{aligned}$$

Thus, we can eliminate λ—whose value is of no direct importance to us—by setting the three right-hand sides equal:

$$\frac{2x}{yz} = \frac{2y}{xz} = \frac{2z}{xy}.$$

Cross-multiplying the first equation yields

$$2x^2z = 2y^2z$$

and since $z \neq 0$ (why?)

$$x^2 = y^2;$$

similarly, we cross-multiply the second equation to get

$$y^2 = z^2.$$

In particular, all three have the same absolute value, so

$$|x|^3 = 1$$

implies

$$|x| = |y| = |z| = 1$$

and an even number of the variables can be negative. This yields four relative critical points, at all of which $f(x, y, z) = 3$:

$$\begin{aligned}
&(1, 1, 1),\\
&(1, -1, -1),\\
&(-1, -1, 1),\\
&(-1, 1, -1).
\end{aligned}$$

To see that they are the closest (not the furthest) from the origin, simply note that there are points on this surface arbitrarily far from the origin, so the distance to the origin is not bounded above.

Finally, let us consider a "full" optimization problem: to find the extreme values of

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

inside the unit ball

$$x^2 + y^2 + z^2 \leq 1.$$

We begin by looking for the **Critical Points** of f:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= 4x\\
\frac{\partial f}{\partial y} &= 2y\\
\frac{\partial f}{\partial z} &= -2y
\end{aligned}$$

all vanish only at the origin, and

$$f(0, 0, 0) = 0.$$

as to **Boundary Behavior**:

$$\begin{aligned}
\overrightarrow{\nabla} f(x, y, z) &= 4x\overrightarrow{\imath} + 2y\overrightarrow{\jmath} - 2z\overrightarrow{k}\\
\overrightarrow{\nabla} g(x, y, z) &= 2x\overrightarrow{\imath} + 2y\overrightarrow{\jmath} + 2z\overrightarrow{k}
\end{aligned}$$

and the Lagrange Multiplier Equations read

$$\begin{aligned} 4x &= 2\lambda x \\ 2y &= 2\lambda y \\ -2z &= 2\lambda z. \end{aligned}$$

The first equation tells us that *either*

$$\lambda = 2$$

or

$$x = 0;$$

the second says that *either*

$$\lambda = 1$$

or

$$y = 0$$

while the third says that *either*

$$\lambda = -1$$

or

$$z = 0.$$

Since only one of the three named λ-values can hold, two of the coordinates must be zero, which means in terms of the constraint that the third is ± 1. Thus we have six relative critical points, with respective f-values

$$\begin{aligned} f(\pm 1, 0, 0) &= 2 \\ f(0, \pm 1, 0) &= 1 \\ f(0, 0, \pm 1) &= -1. \end{aligned}$$

Combining this with the critical value 0 at the origin, we have

$$\begin{aligned} \min_{x^2+y^2+z^2 \leq 1} (2x^2 + y^2 - z^2) &= f(0, 0, \pm 1) = -1 \\ \max_{x^2+y^2+z^2 \leq 1} (2x^2 + y^2 - z^2) &= f(\pm 1, 0, 0) = 2. \end{aligned}$$

Multiple Constraints

The method of Lagrange Multipliers can be extended to problems in which there is more than one constraint present. We illustrate this with a single example, involving two constraints.

The intersection of the cylinder

$$x^2 + y^2 = 4$$

with the plane

$$x + y + z = 1$$

is an ellipse; we wish to find the points on this ellipse nearest and farthest from the origin. Again, we will work with the square of the distance from the origin:

$$\begin{aligned} f(x,y,z) &= x^2 + y^2 + z^2 \\ \overrightarrow{\nabla} f(x,y,z) &= (2x, 2y, 2z). \end{aligned}$$

We are looking for the extreme values of this function on the curve of intersection of two level surfaces. In principle, we could parametrize the ellipse, but instead we will work directly with the constraints and their gradients:

$$\begin{aligned} g_1(x,y,z) &= x^2 + y^2 \\ \overrightarrow{\nabla} g_1 &= (2x, 2y, 0) \\ g_2(x,y,z) &= x + y + z \\ \overrightarrow{\nabla} g_2 &= (1, 1, 1). \end{aligned}$$

Since our curve lies in the intersection of the two level surfaces $\mathcal{L}(g_1, 4)$ and $\mathcal{L}(g_2, 1)$, its velocity vector must be perpendicular to both gradients:

$$\begin{aligned} \vec{v} \cdot \overrightarrow{\nabla} g_1 &= 0 \\ \vec{v} \cdot \overrightarrow{\nabla} g_2 &= 0. \end{aligned}$$

At a place where the restriction of f to this curve achieves a local (relative) extremum, the velocity must also be perpendicular to the gradient of f:

$$\vec{v} \cdot \overrightarrow{\nabla} f = 0.$$

But the two gradient vectors $\overrightarrow{\nabla} g_1$ and $\overrightarrow{\nabla} g_2$ are linearly independent, and hence span the plane perpendicular to $\overrightarrow{v}$. It follows that $\overrightarrow{\nabla} f$ must lie in this plane, or stated differently, it must be a linear combination of the $\overrightarrow{\nabla} g$'s:

$$\overrightarrow{\nabla} f = \lambda_1 \overrightarrow{\nabla} g_1 + \lambda_2 \overrightarrow{\nabla} g_2.$$

Written out, this gives us three equations in the five unknowns x, y, z, λ_1 and λ_2:

$$\begin{aligned} 2x &= 2\lambda_1 x + \lambda_2 \\ 2y &= 2\lambda_1 y + \lambda_2 \\ 2z &= \lambda_2. \end{aligned}$$

The other two equations are the constraints:

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

We can solve the first three equations for λ_2 and eliminate it:

$$\begin{aligned} (1 - \lambda_1)x &= (1 - \lambda_1)y \\ &= 2z. \end{aligned}$$

The first of these equations says that either

$$\lambda_1 = 1$$

or

$$x = y.$$

If $\lambda_1 = 1$, then the second equality says that $z = 0$, so $y = 1 - x$. In this case the first constraint gives us

$$\begin{aligned} x^2 + (1 - x)^2 &= 4 \\ 2x^2 - 2x - 3 &= 0 \\ x &= \frac{1}{2}(1 \pm \sqrt{7}) \\ y &= \frac{1}{2}(1 \mp \sqrt{7}) \end{aligned}$$

yielding two relative critical points, at which the function f has value

$$f\left(\frac{1}{2}(1\pm\sqrt{7}),\frac{1}{2}(1\mp\sqrt{7}),0\right)=\frac{9}{4}.$$

If $x=y$, then the first constraint tells us

$$\begin{aligned} x^2+x^2&=4\\ x=y&=\pm\sqrt{2}\end{aligned}$$

and then the second constraint says

$$\begin{aligned} z&=1-2x\\ &=1\mp2\sqrt{2}\end{aligned}$$

yielding another pair of relative critical points, with respective values for f

$$\begin{aligned} f\left(\sqrt{2},\sqrt{2},1-2\sqrt{2}\right)&=13-4\sqrt{2}\\ f\left(-\sqrt{2},-\sqrt{2},1+2\sqrt{2}\right)&=13+4\sqrt{2}.\end{aligned}$$

Comparing these various values, we see that the point farthest from the origin is $(-\sqrt{2},-\sqrt{2},1+2\sqrt{2})$ and the closest are the two points $\left(\frac{1}{2}(1\pm\sqrt{7}),\frac{1}{2}(1\mp\sqrt{7}),0\right)$.

Exercises for § 3.6

Practice problems:

1. Find the minimum and maximum values of

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

inside the unit disc

$$x^2+y^2\leq1.$$

2. Find the minimum and maximum values of

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

inside the disc

$$x^2 + y^2 \le 4.$$

3. Find the minimum and maximum values of

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

inside the elliptic disc

$$x^2 + 4y^2 \le 4.$$

4. Find the minimum and maximum values of

$$f(x,y) = \sin x \sin y \sin(x+y)$$

inside the square

$$\begin{aligned} 0 &\le x \le \pi \\ 0 &\le y \le \pi \end{aligned}$$

5. Find the minimum and maximum values of

$$f(x,y) = (x^2 + 2y^2)e^{-(x^2+y^2)}$$

in the plane.

6. Find the minimum and maximum values of

$$f(x,y,z) = xyz$$

on the sphere

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

7. Find the point on the sphere

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

which is farthest from the point $(1, 2, 3)$.

8. Find the rectangle of greatest perimeter inscribed in the ellipse

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

Theory problems:

9. Show that each of the following is a closed set, according to Definition 3.6.3:

   (a) Any close interval $[a, b]$ in $\mathbb{R}$;

   (b) any half-closed interval of the form $[a, \infty)$ or $(-\infty, b]$;

   (c) any level set $\mathcal{L}(g, c)$ of a continuous function g;

   (d) any set defined by weak inequalities like $\{x \in \mathbb{R}^3 \,|\, g(x) \leq c\}$ or $\{x \in \mathbb{R}^3 \,|\, g(x) \geq c\}$;

10. Prove Remark 3.6.9:

   (a) For any set $S \subset \mathbb{R}^3$,
   $$S \subseteq \text{int } S \cup \partial S.$$

   (b) The boundary ∂S of any set is closed.

   (c) S is closed precisely if it contains its boundary points:
   $$S \text{ closed} \Leftrightarrow \partial S \subset S.$$

   (d) $S \subset \mathbb{R}^3$ is *closed* precisely if its complement
   $$\mathbb{R}^3 \setminus S := \{x \in \mathbb{R}^3 \,|\, x \notin S\}$$
   is *open*.

Challenge problems:

11. (a) Show that any set consisting of a *convergent* sequence s_i together with its limit is a closed set;

    (b) Show that any set consisting of a (not necessarily convergent) sequence together with all of its accumulation points is a closed set.

12. Prove that if $\alpha, \beta > 0$ satisfy

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1$$

then for all $x, y \geq 0$

$$xy \leq \frac{1}{\alpha} x^\alpha + \frac{1}{\beta} y^\beta$$

as follows:

   (a) The inequality is clear for $xy = 0$, so we can assume $xy \neq 0$.

   (b) If it is true (given α and β) for a given pair (x, y), then it is also true for the pair $(t^{1/\alpha} x, t^{1/\beta} y)$ (verify this!), and so we can assume without loss of generality that $xy = 1$

   (c) Prove the inequality in this case by minimizing

$$f(x, y) = \frac{1}{\alpha} x^\alpha + \frac{1}{\beta} y^\beta$$

   over the hyperbola

$$xy = 1.$$

13. Here is a somewhat different proof of Theorem 3.6.6, based on an idea of Daniel Reem [44]. Suppose $S \subset \mathbb{R}^3$ is compact.

   (a) Show that for every integer $k = 1, 2, \ldots$ there is a *finite* subset $S_k \subset S$ such that for every point $x \in S$ there is at least one point in S_k whose coordinates differ from those of x by at most 10^{-k}. In particular, for every $x \in S$ there is a sequence of points $\{x_k\}_{k=1}^{\infty}$ such that $x_k \in S_k$ for $k = 1, \ldots$ and $x = \lim x_k$.

   (b) Show that these sets can be picked to be nested: $S_k \subset S_{k+1}$ for all k.

   (c) Now, each of the sets S_k is finite, so f has a minimum $\min_{s \in S_k} f(s) = f(m_k)$ and a maximum $\max_{s \in S_k} f(s) = f(M_k)$. Show that

$$\begin{aligned} f(m_k) &\geq f(m_{k+1}) \\ f(M_k) &\leq f(M_{k+1}). \end{aligned}$$

(d) Also, by the Bolzano-Weierstrass Theorem, each of the sequences $\{m_k\}_{k=1}^{\infty}$ and $\{M_k\}_{k=1}^{\infty}$ has a convergent subsequence. Let m (*resp.* M) be the limit of such a subsequence. Show that $m, M \in S$ and

$$f(m) = \inf f(m_k) = \lim f(m_k)$$
$$f(M) = \sup f(M_k) = \lim f(M_k)\,.$$

(e) Finally, show that

$$f(m) \leq f(x) \leq f(M)$$

for every $x \in S$, as follows: given $x \in S$, by part (a), there is a sequence $x_k \to x$ with $x_k \in S_k$. Thus,

$$f(m_k) \leq f(x_k) \leq f(M_k)$$

and so by properties of limits (which?) the desired conclusion follows.

3.7 Higher Derivatives

For a function of one variable, the higher-order derivatives give more subtle information about the function near a point: while the first derivative specifies the "tilt" of the graph, the second derivative tells us about the way the graph curves, and so on. Specifically, the second derivative can help us decide whether a given critical point is a local maximum, local minimum, or neither.

In this section we develop the basic theory of higher-order derivatives for functions of several variables, which can be a bit more complicated than the single-variable version. Most of our energy will be devoted to second-order derivatives.

Higher-order Partial Derivatives

The partial derivatives of a function of several variables are themselves functions of several variables, and we can try to find *their* partial derivatives. Thus, if $f(x, y)$ is differentiable, it has two first-order partials

$$\frac{\partial f}{\partial x}, \quad \frac{\partial f}{\partial y}$$

and, if they are also differentiable, each has two partial derivatives, which are the **second-order partials** of f:

$$\begin{aligned}\frac{\partial^2 f}{\partial^2 x} &= \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial x}\right]\\ \frac{\partial^2 f}{\partial y \partial x} &= \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial x}\right]\\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial y}\right]\\ \frac{\partial^2 f}{\partial^2 y} &= \frac{\partial}{\partial y}\left[\frac{\partial f}{\partial y}\right].\end{aligned}$$

In subscript notation, the above would be written

$$\begin{aligned}f_{xx} &= (f_x)_x\\ f_{xy} &= (f_x)_y\\ f_{yx} &= (f_y)_x\\ f_{yy} &= (f_y)_y\,.\end{aligned}$$

Notice that in the "partial" notation, the order of differentiation is right-to-left, while in the subscript version it is left-to-right. (We shall see shortly that for $\mathcal{C}^2$ functions, this is not an issue.)

For example, the function

$$f(x,y) = x^2 + 2xy + y - 1 + xy^3$$

has first-order partials

$$\begin{aligned}f_x &= \frac{\partial f}{\partial x} = 2x + 2y + y^3\\ f_y &= \frac{\partial f}{\partial y} = 2x + 1 + 3xy^2\end{aligned}$$

and second-order partials

$$\begin{aligned} f_{xx} &= \frac{\partial^2 f}{\partial^2 x} = 2 \\ f_{xy} &= \frac{\partial^2 f}{\partial y \partial x} = 2 + 3y^2 \\ f_{yx} &= \frac{\partial^2 f}{\partial x \partial y} = 2 + 3y^2 \\ f_{yy} &= \frac{\partial^2 f}{\partial^2 y} = 6xy. \end{aligned}$$

It is clear that the game of successive differentiation can be taken further; in general a sufficiently smooth function of two (*resp.* three) variables will have 2^r (*resp.* 3^r) partial derivatives of order r. Recall that a function is called *continuously differentiable*, or C^1, if its (first-order) partials exist and are continuous; Theorem 3.3.4 tells us that such functions are automatically differentiable. We shall extend this terminology to higher derivatives: a function is **r times continuously differentiable** or **C^r** if all of its partial derivatives of order $1, 2, ..., r$ exist and are continuous. In practice, we shall seldom venture beyond the second-order partials.

The alert reader will have noticed that the two *mixed partials* of the function above are equal. This is no accident; the phenomenon was first noted around 1718 by Nicolaus I Bernoulli (1687-1759);[14] in 1734 Leonard Euler (1707-1783) and Alexis-Claude Clairaut (1713-1765) published proofs of the following result.

Theorem 3.7.1 (Equality of Mixed Partials)**.** *If a real-valued function f of two or three variables is twice continuously differentiable (C^2), then for any pair of indices i, j*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

While it is formulated for second-order partials, this theorem automatically extends to partials of higher order (Exercise 6): if f is C^r, then the

[14] There were at least six Bernoullis active in mathematics in the late seventeenth and early eighteenth century: the brothers Jacob Bernoulli (1654-1705) and Johann Bernoulli (1667-1748)—who was the tutor to L'Hôpital—their nephew, son of the painter Nicolaus and also named Nicolaus—who is denoted Nicolaus I—and Johann's three sons, Nicolaus II Bernoulli (1695-1726), Daniel Bernoulli (1700-1782) and Johann II Bernoulli (1710-1790). I am following the numeration given by [12, pp. 92-94], which has a brief biographical account of Nicolaus I in addition to a detailed study of his contributions to partial differentiation.

order of differentiation in any mixed partial derivative of order up to r does not affect its value. This reduces the number of *different* partial derivatives of a given order tremendously.

Proof. We shall give the proof for a function of two variables; after finishing the proof, we shall note how this actually gives the same conclusion for three variables.

The proof is based on looking at **second-order differences**: given two points (x_0, y_0) and $(x_1, y_1) = (x_0 + \triangle x, y_0 + \triangle y)$, we can go from the first to the second in two steps: increase one of the variables, holding the other fixed, then increase the other variable. This can be done in two ways, depending on which variable we change first; the two paths form the sides of a rectangle with (x_i, y_i), $i = 1, 2$ at opposite corners (Figure 3.11). Let

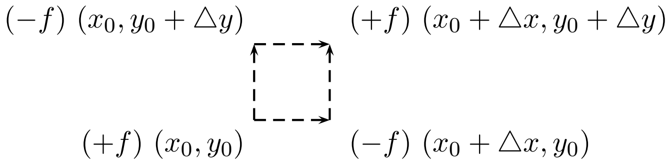

$(-f)\ (x_0, y_0 + \triangle y)$
$(+f)\ (x_0 + \triangle x, y_0 + \triangle y)$
$(+f)\ (x_0, y_0)$
$(-f)\ (x_0 + \triangle x, y_0)$

Figure 3.11: Second order differences

us now consider the difference between the values of $f(x, y)$ at the ends of one of the horizontal edges of the rectangle: the difference along the bottom edge

$$\triangle_x f\,(y_0) = f(x_0 + \triangle x, y_0) - f(x_0, y_0)$$

represents the change in $f(x, y)$ when y is held at $y = y_0$ and x increases by $\triangle x$ from $x = x_0$, while the difference along the *top* edge

$$\triangle_x f\,(y_0 + \triangle y) = f(x_0 + \triangle x, y_0 + \triangle y) - f(x_0, y_0 + \triangle y)$$

represents the change in $f(x, y)$ when y is held at $y = y_0 + \triangle y$ and x increases by $\triangle x$ from $x = x_0$. We wish to compare these two changes, by subtracting the first from the second:

$$\begin{aligned}
\triangle_y \triangle_x f &= \triangle_x f\,(y_0 + \triangle y) - \triangle_x f\,(y_0) \\
&= [f(x_0 + \triangle x, y_0 + \triangle y) - f(x_0, y_0 + \triangle y)] \\
&\quad - [f(x_0 + \triangle x, y_0) - f(x_0, y_0)] \\
&= f(x_0 + \triangle x, y_0 + \triangle y) - f(x_0, y_0 + \triangle y) \\
&\quad - f(x_0 + \triangle x, y_0) + f(x_0, y_0)\,.
\end{aligned}$$

(Note that the signs attached to the four values of $f(x,y)$ correspond to the signs in Figure 3.11.) Each of the first-order differences $\triangle_x f\,(y_0)$ (*resp.* $\triangle_x f\,(y_0+\triangle y)$) is an approximation to $\frac{\partial f}{\partial x}$ at (x_0,y_0) (*resp.* $(x_0,y_0+\triangle y)$), multiplied by $\triangle x$; their difference is then an approximation to $\frac{\partial^2 f}{\partial y \partial x}$ at (x_0,y_0), multiplied by $\triangle y \triangle x$; we shall use the Mean Value Theorem to make this claim precisely.

But first consider the *other* way of going: the differences along the two *vertical* edges

$$\begin{aligned}\triangle_y f\,(x_0) &= f(x_0,y_0+\triangle y) - f(x_0,y_0)\\ \triangle_y f\,(x_0+\triangle x) &= f(x_0+\triangle x,y_0+\triangle y) - f(x_0+\triangle x,y_0)\end{aligned}$$

represent the change in $f(x,y)$ as x is held constant at one of the two values $x=x_0$ (*resp.* $x=x_0+\triangle x$) and y increases by $\triangle y$ from $y=y_0$; this roughly approximates $\frac{\partial f}{\partial y}$ at (x_0,y_0) (*resp.* $(x_0+\triangle x,y_0)$), multiplied by $\triangle y$, and so the difference of *these* two differences

$$\begin{aligned}\triangle_x \triangle_y f &= \triangle_y f\,(x_0+\triangle x) - \triangle_y f\,(x_0)\\ &= [f(x_0+\triangle x,y_0+\triangle y) - f(x_0+\triangle x,y_0)]\\ &\quad - [f(x_0,y_0+\triangle y) - f(x_0,y_0)]\\ &= f(x_0+\triangle x,y_0+\triangle y) - f(x_0+\triangle x,y_0)\\ &\quad - f(x_0,y_0+\triangle y) + f(x_0,y_0)\end{aligned}$$

approximates $\frac{\partial^2 f}{\partial x \partial y}$ at (x_0,y_0), multiplied by $\triangle x \triangle y$. But a close perusal shows that these two second-order differences are the same—and this will be the punch line of our proof.

Actually, for technical reasons, we don't follow the strategy suggested above precisely. Let's concentrate on the first (second-order) difference: counterintuitively, our goal is to show that

$$\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0) = \lim_{(\triangle x,\triangle y)\to(0,0)} \frac{\triangle_y \triangle_x f}{\triangle y \triangle x}.$$

To this end, momentarily fix $\triangle x$ and $\triangle y$ and define

$$\begin{aligned}g(t) &= \triangle_x f\,(y_0+t\triangle y)\\ &= f(x_0+\triangle x,y_0+t\triangle y) - f(x_0,y_0+t\triangle y)\,;\end{aligned}$$

then

$$g'(t) = \left[\frac{\partial f}{\partial y}\,(x_0+\triangle x,y_0+t\triangle y) - \frac{\partial f}{\partial y}\,(x_0,y_0+t\triangle y)\right]\triangle y.$$

Now,

$$\triangle_y \triangle_x f = g(1) - g(0)$$

and the Mean Value Theorem applied to $g(t)$ tells us that for some $\tilde{t} \in (0,1)$, this difference

$$\begin{aligned} &= g'\left(\tilde{t}\right) \\ &= \left[\frac{\partial f}{\partial y}\left(x_0 + \triangle x, y_0 + \tilde{t}\triangle y\right) - \frac{\partial f}{\partial y}\left(x_0, y_0 + \tilde{t}\triangle y\right)\right]\triangle y \end{aligned}$$

or, writing $\tilde{y} = y_0 + \tilde{t}\triangle y$, and noting that $\tilde{y}$ lies between y_0 and $y_0 + \triangle y$, we can say that

$$\triangle_y \triangle_x f = \left[\frac{\partial f}{\partial y}\left(x_0 + \triangle x, \tilde{y}\right) - \frac{\partial f}{\partial y}\left(x_0, \tilde{y}\right)\right]\triangle y$$

where $\tilde{y}$ is some value between y_0 and $y_0 + \triangle y$.

But now apply the Mean Value Theorem to

$$h(t) = \frac{\partial f}{\partial y}\left(x_0 + t\triangle x, \tilde{y}\right)$$

with derivative

$$h'(t) = \frac{\partial^2 f}{\partial x \partial y}(x_0 + t\triangle x, \tilde{y})\triangle x$$

so for some $t' \in (0,1)$

$$\begin{aligned} \left[\frac{\partial f}{\partial y}\left(x_0 + \triangle x, \tilde{y}\right) - \frac{\partial f}{\partial y}\left(x_0, \tilde{y}\right)\right] &= h(1) - h(0) \\ &= h'\left(t'\right) \\ &= \frac{\partial^2 f}{\partial x \partial y}(x_0 + t'\triangle x, \tilde{y})\triangle x \end{aligned}$$

and we can say that

$$\begin{aligned} \triangle_y \triangle_x f &= \left[\frac{\partial f}{\partial y}\left(x_0 + \triangle x, \tilde{y}\right) - \frac{\partial f}{\partial y}\left(x_0, \tilde{y}\right)\right]\triangle y \\ &= \frac{\partial^2 f}{\partial x \partial y}(\tilde{x}, \tilde{y})\triangle x \triangle y \end{aligned}$$

where $\tilde{x} = x_0 + t'\triangle x$ is between x_0 and $x_0 + \triangle x$, and $\tilde{y} = y_0 + \tilde{t}\triangle y$ lies between y_0 and $y_0 + \triangle y$. Now, if we divide both sides of the equation above by $\triangle x \triangle y$, and take limits, we get the desired result:

$$\begin{aligned}\lim_{(\triangle x, \triangle y)\to(0,0)} \frac{\triangle_y \triangle_x f}{\triangle x \triangle y} &= \lim_{(\triangle x, \triangle y)\to(0,0)} \frac{\partial^2 f}{\partial x \partial y}(\tilde{x}, \tilde{y}) \\ &= \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)\end{aligned}$$

because $(\tilde{x}, \tilde{y}) \to (x_0, y_0)$ as $(\triangle x, \triangle y) \to (0, 0)$ and the partial is assumed to be continuous at (x_0, y_0).

But now it is clear that by reversing the roles of x and y we get, in the same way,

$$\lim_{(\triangle x, \triangle y)\to(0,0)} \frac{\triangle_x \triangle_y f}{\triangle y \triangle x} = \frac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$$

which, together with our earlier observation that

$$\triangle_y \triangle_x f = \triangle_x \triangle_y f$$

completes the proof. □

At first glance, it might seem that a proof for functions of more than two variables might need some work over the one given above. However, when we are looking at the equality of two specific mixed partials, say $\frac{\partial^2 f}{\partial x_i \partial x_j}$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}$, we are holding *all other variables* constant, so the proof above goes over verbatim, once we replace x with x_i and y with x_j (Exercise 5).

Taylor Polynomials

The higher derivatives of a function of one variable can be used to construct a polynomial that has high-order contact with the function at a point, and hence is a better local approximation to the function. An analogous construction is possible for functions of several variables, however more work is needed to combine the various partial derivatives of a given order into the appropriate polynomial.

A polynomial in several variables consists of monomial terms, each involving powers of the different variables; the degree of the term is the **exponent sum**: the sum of the exponents of all the variables appearing in that term.[15] Thus, each of the monomial terms $3x^2yz^3$, $2xyz^4$ and $5x^6$ has

[15] The variables that don't appear have exponent zero

exponent sum 6. We group the terms of a polynomial according to their exponent sums: the group with exponent sum k on its own is a **homogeneous** function of degree k. This means that inputs scale via the k^{th} power of the scalar. We already saw that homogeneity of degree one is exhibited by linear functions:

$$\ell(c\overrightarrow{x}) = c\ell(\overrightarrow{x})\,.$$

The degree k analogue is

$$\varphi(c\overrightarrow{x}) = c^k\varphi(\overrightarrow{x})\,;$$

for example,

$$\varphi(x,y,z) = 3x^2yz^3 + 2xyz^4 + 5x^6$$

satisfies

$$\begin{aligned}\varphi(cx,cy,cz) &= 3(cx)^2(cy)(cz)^3 + 2(cx)(cy)(cz)^4 + 5(cx)^6\\ &= c^6(3x^2yz^3 + 2xyz^4 + 5x^6)\end{aligned}$$

so this function is homogeneous of degree 6. In general, it is easy to see that a polynomial (in any number of variables) is homogeneous precisely if the exponent sum of each term appearing in it is the same, and this sum equals the degree of homogeneity.

For functions of one variable, the k^{th} derivative determines the term of degree k in the Taylor polynomial, and similarly for a function of several variables the partial derivatives of order k determine the part of the Taylor polynomial which is homogeneous of degree k. Here, we will concentrate on degree two.

For a $\mathcal{C}^2$ function $f(x)$ of one variable, the Taylor polynomial of degree two

$$T_2f\left(\overrightarrow{a}\right)\overrightarrow{x} := f(a) + f'(a)\,(x-a) + \frac{1}{2}f''(a)\,(x-a)^2$$

has contact of order two with $f(x)$ at $x = a$, and hence is a closer approximation to $f(x)$ (for x near a) than the linearization (or degree one Taylor polynomial). To obtain the analogous polynomial for a function f of two or three variables, given $\overrightarrow{a}$ and a nearby point $\overrightarrow{x}$, we consider the restriction of f to the line segment from $\overrightarrow{a}$ to $\overrightarrow{x}$, parametrized as

$$g(t) = f(\overrightarrow{a} + t\triangle\overrightarrow{x})\,,\quad 0 \le t \le 1$$

where $\triangle\vec{x} = \vec{x} - \vec{a}$. Taylor's Theorem with Lagrange Remainder for functions of one variable ((*Calculus Deconstructed*, Theorem 6.1.7)) tells us that

$$g(t) = g(0) + tg'(0) + \frac{t^2}{2}g''(s) \tag{3.28}$$

for some $0 \leq s \leq t$. By the Chain Rule (Proposition 3.3.6)

$$g'(t) = \vec{\nabla} f(\vec{a} + t\triangle\vec{x}) \cdot \triangle\vec{x} = \sum_j \frac{\partial f}{\partial x_j}(\vec{a} + t\triangle\vec{x})\,\triangle_j\vec{x}$$

and so

$$g''(s) = \sum_i \sum_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a} + s\triangle\vec{x})\,\triangle_i\vec{x}\,\triangle_j\vec{x}.$$

This is a homogeneous polynomial of degree two, or **quadratic form**, in the components of $\triangle\vec{x}$. By analogy with our notation for the total differential, we denote it by

$$d^2_{\vec{a}} f(\triangle\vec{x}) = \sum_i \sum_j \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a})\,\triangle_i\vec{x}\,\triangle_j\vec{x}.$$

We shall refer to this particular quadratic form—the analogue of the second derivative—as the **Hessian form** of f, after Ludwig Otto Hesse (1811-1874), who introduced it in 1857 [28].

Again by analogy with the single-variable setting, we define the degree two **Taylor polynomial** of f at $\vec{a}$ as the sum of the function with its (total) differential and half the quadratic form at $\vec{a}$, both applied to $\triangle\vec{x} = \vec{x} - \vec{a}$. Note that in the quadratic part, equality of cross-partials allows us to combine any pair of terms involving distinct indices into one term, whose coefficient is precisely the relevant partial derivative; we use this in writing the last expression below. (We write the version for a function of three variables; for a function of two variables, we simply omit any terms that are supposed to involve x_3.)

$$\begin{aligned} T_2 f(\vec{a})\,\vec{x} &= f(a) + d_{\vec{a}} f(\triangle\vec{x}) + \frac{1}{2} d^2_{\vec{a}} f(\triangle\vec{x}) \\ &= f(a) + \sum_{j=1}^{3} \frac{\partial f}{\partial x_j}(\vec{a})\,\triangle x_j + \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a})\,\triangle_i\vec{x}\,\triangle_j\vec{x} \\ &= f(a) + \sum_{j=1}^{3} \frac{\partial f}{\partial x_j}(\vec{a})\,\triangle_j\vec{x} + \frac{1}{2}\sum_{i=1}^{3} \frac{\partial^2 f}{\partial^2 x_i}(\vec{a})\,\triangle x_i^2 + \sum_{1\leq i<j\leq 3} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{a})\,\triangle x_i \triangle x_j. \end{aligned}$$

We consider two examples.

The function

$$f(x,y) = e^{2x}\cos y$$

has

$$\begin{aligned}
\frac{\partial f}{\partial x}(x_0,y_0) &= 2e^{2x}\cos y\\
\frac{\partial f}{\partial y}(x_0,y_0) &= -e^{2x}\sin y\\
\frac{\partial^2 f}{\partial^2 x}(x_0,y_0) &= 4e^{2x}\cos y\\
\frac{\partial^2 f}{\partial x\partial y}(x_0,y_0) &= -2e^{2x}\sin y\\
\frac{\partial^2 f}{\partial^2 y}(x_0,y_0) &= -e^{2x}\cos y.
\end{aligned}$$

At $\overrightarrow{a} = \left(0, \frac{\pi}{3}\right)$, these values are

$$\begin{aligned}
f\left(0,\frac{\pi}{3}\right) &= e^0\cos\frac{\pi}{3} &&= \frac{1}{2}\\
\frac{\partial f}{\partial x}\left(0,\frac{\pi}{3}\right) &= 2e^0\cos\frac{\pi}{3} &&= 1\\
\frac{\partial f}{\partial y}\left(0,\frac{\pi}{3}\right) &= -e^0\sin\frac{\pi}{3} &&= -\frac{\sqrt{3}}{2}\\
\frac{\partial^2 f}{\partial^2 x}\left(0,\frac{\pi}{3}\right) &= 4e^0\cos\frac{\pi}{3} &&= 2\\
\frac{\partial^2 f}{\partial x\partial y}\left(0,\frac{\pi}{3}\right) &= -2e^0\sin\frac{\pi}{3} &&= -\sqrt{3}\\
\frac{\partial^2 f}{\partial^2 y}\left(0,\frac{\pi}{3}\right) &= -e^0\cos\frac{\pi}{3} &&= -\frac{1}{2}
\end{aligned}$$

so the degree two Taylor polynomial at $\overrightarrow{a} = \left(0, \frac{\pi}{3}\right)$ is

$$T_2 f\left(\left(0,\frac{\pi}{3}\right)\right)\triangle x, \triangle y = \frac{1}{2} + \triangle x - \left(\frac{\sqrt{3}}{2}\right)\triangle y + \triangle x^2 - \frac{1}{4}\triangle y^2 - \sqrt{3}\triangle x\triangle y.$$

Let us compare the value $f\left(0.1, \frac{\pi}{2}\right)$ with $f\left(0, \frac{\pi}{3}\right) = 0.5$:

- The exact value is

$$f\left(0.1, \frac{\pi}{2}\right) = e^{0.2} \cos \frac{\pi}{2} = 0.$$

- The linearization (degree one Taylor polynomial) gives an estimate of

$$T_{\left(0, \frac{\pi}{3}\right)} f\left(0.1, \frac{\pi}{6}\right) = \frac{1}{2} + 0.1 - \left(\frac{\sqrt{3}}{2}\right) \frac{\pi}{6} \approx 0.14655.$$

- The quadratic approximation (degree two Taylor polynomial) gives

$$\begin{aligned} T_2 f\left(\left(0, \frac{\pi}{3}\right)\right) 0.1, \frac{\pi}{6} &= \frac{1}{2} + 0.1 - \left(\frac{\sqrt{3}}{2}\right) \frac{\pi}{6} \\ &\quad + (0.1)^2 - \frac{1}{4}\left(\frac{\pi}{6}\right)^2 - \sqrt{3}(0.1)\left(\frac{\pi}{6}\right) \\ &\approx -.00268 \end{aligned}$$

  a much better approximation.

As a second example, consider the function

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

which has

$$\begin{aligned} f_x = 2xy^3 z,\ f_y &= 3x^2 y^2 z,\ f_z = x^2 y^3 \\ f_{xx} = 2y^3 z,\ f_{xy} &= 6xy^2 z,\ f_{xz} = 2xy^3 \\ f_{yy} &= 6x^2 yz,\ f_{yz} = 3x^2 y^2 \\ &\qquad\qquad f_{zz} = 0. \end{aligned}$$

Evaluating these at $\overrightarrow{a} = \left(1, \frac{1}{2}, 2\right)$ yields

$$\begin{aligned}
f\left(1, \frac{1}{2}, 2\right) &= (1)^2 \left(\frac{1}{2}\right)^3 (2) &&= \frac{1}{4} \\
f_x\left(1, \frac{1}{2}, 2\right) &= 2(1) \left(\frac{1}{2}\right)^3 (2) &&= \frac{1}{2} \\
f_y\left(1, \frac{1}{2}, 2\right) &= 3(1)^2 \left(\frac{1}{2}\right)^2 (2) &&= \frac{3}{2} \\
f_z\left(1, \frac{1}{2}, 2\right) &= (1)^2 \left(\frac{1}{2}\right)^3 &&= \frac{1}{8} \\
f_{xx}\left(1, \frac{1}{2}, 2\right) &= 2 \left(\frac{1}{2}\right)^3 (2) &&= \frac{1}{2} \\
f_{xy}\left(1, \frac{1}{2}, 2\right) &= 6(1)^2 \left(\frac{1}{2}\right)^2 (2) &&= 3 \\
f_{xz}\left(1, \frac{1}{2}, 2\right) &= 2(1) \left(\frac{1}{2}\right)^3 &&= \frac{1}{4} \\
f_{yy}\left(1, \frac{1}{2}, 2\right) &= 6(1)^2 \left(\frac{1}{2}\right) (2) &&= 6 \\
f_{yz}\left(1, \frac{1}{2}, 2\right) &= 3(1)^2 \left(\frac{1}{2}\right)^2 &&= \frac{3}{4} \\
f_{zz}\left(1, \frac{1}{2}, 2\right) & &&= 0.
\end{aligned}$$

The degree two Taylor polynomial is

$$\begin{aligned}
T_2 f\left(\overrightarrow{a}\right) \triangle x, \triangle y, \triangle z &= \frac{1}{4} + \left(\frac{1}{2}\triangle x + \frac{3}{2}\triangle y + \frac{1}{8}\triangle z\right) \\
&\quad + \frac{1}{2}\left(2\triangle x^2 + 6\triangle y^2 + 0\triangle z^2\right) + \left(6\triangle x\triangle y + 2\triangle x\triangle z + 3\triangle y\triangle z\right) \\
&= 0.25 + 0.5\triangle x + 1.5\triangle y + 0.125\triangle z \\
&\quad + 0.25\triangle x^2 + 0.25\triangle x\triangle z + 3\triangle x\triangle y + 3\triangle y^2 + 0.75\triangle y\triangle z.
\end{aligned}$$

Let us compare the value $f(\overrightarrow{a}) = f(1, 0.5, 2.0) = .25$ with $f(1.1, 0.4, 1.8)$:

- The exact value is

$$f(1.1, 0.4, 1.8) = (1.1)^2(0.4)^3(1.8) = 0.139392.$$

- The linear approximation, with $\overrightarrow{a} = (1.0, 0.5, 2.0)$, $\triangle x = 0.1$, $\triangle y = -0.1$ and $\triangle z = -0.2$ is

$$\begin{aligned}
&T_{\overrightarrow{a}} f(0.1, -0.1, -0.2) \\
&\qquad = 0.25 + (0.5)(0.1) + (1.5)(-0.1) + (0.125)(-0.2) \\
&\qquad = 0.125.
\end{aligned}$$

- The quadratic approximation is

$$\begin{aligned}
&T_2 f\,(\overrightarrow{a})\, 0.1, -0.1, -0.2 \\
&\qquad = 0.25 + (0.5)(0.1) + (1.5)(-0.1) + (0.125)(-0.2) \\
&\qquad\qquad + (0.25)(0.1)^2 + (3)(-0.1)^2 \\
&\qquad + (3)(0.1)(-0.1) + (0.25)(0.1)(-0.2) + (0.75)(-0.1)(-0.2) \\
&\qquad = 0.1375
\end{aligned}$$

  again a better approximation.

These examples illustrate that the quadratic approximation, or degree two Taylor polynomial $T_2 f\,(\overrightarrow{a})\,\overrightarrow{x}$, provides a better approximation than the linearization $T_{\overrightarrow{a}} f(\overrightarrow{x})$. This was the expected effect, as we designed $T_2 f\,(\overrightarrow{a})\,\overrightarrow{x}$ to have contact of order two with $f(x)$ at $\overrightarrow{x} = \overrightarrow{a}$. Let us confirm that this is the case.

Proposition 3.7.2 (Taylor's Theorem for $f\colon \mathbb{R}^3 \to \mathbb{R}$ (degree 2))**.** *If $f\colon \mathbb{R}^3 \to \mathbb{R}$ is $\mathcal{C}^2$ (f has continuous second-order partials), then $T_2 f\,(\overrightarrow{a})\,\overrightarrow{x}$ and $f(\overrightarrow{x})$ have contact of order two at $\overrightarrow{x} = \overrightarrow{a}$:*

$$\lim_{\overrightarrow{x} \to \overrightarrow{a}} \frac{|f(\overrightarrow{x}) - T_2 f\,(\overrightarrow{a})\,\overrightarrow{x}|}{\|\overrightarrow{x} - \overrightarrow{a}\|^2} = 0.$$

Proof. Equation (3.28), evaluated at $t = 1$ and interpreted in terms of f, says that, fixing $\overrightarrow{a} \in \mathbb{R}^3$, for any $\overrightarrow{x}$ in the domain of f,

$$f(\overrightarrow{x}) = f(\overrightarrow{a}) + d_{\overrightarrow{a}} f(\triangle \overrightarrow{x}) + \frac{1}{2} d^2_{\vec{s}} f(\triangle \overrightarrow{x})$$

where $\vec{s}$ lies on the line segment from $\overrightarrow{a}$ to $\overrightarrow{x}$. Thus,

$$\begin{aligned}
f(\overrightarrow{x}) - T_2 f\,(\overrightarrow{a})\,\overrightarrow{x} &= \frac{1}{2}\left(d^2_{\overrightarrow{a}} f(\triangle \overrightarrow{x}) - d^2_{\vec{s}} f(\triangle \overrightarrow{x})\right) \\
&= \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\overrightarrow{a}) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{s}) \right) \triangle x_i \triangle x_j
\end{aligned}$$

and so

$$\begin{aligned}\frac{|f(\overrightarrow{x}) - T_2 f(\overrightarrow{a})\overrightarrow{x}|}{\|\overrightarrow{x} - \overrightarrow{a}\|^2} \\ &\leq \frac{1}{2}\left(\sum_{i=1}^{3}\sum_{j=1}^{3}\left|\frac{\partial^2 f}{\partial x_i \partial x_j}(\overrightarrow{a}) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{s})\right|\right)\frac{|\triangle x_i \triangle x_j|}{\triangle \overrightarrow{x}^2} \\ &\leq \frac{n^2}{2}\max_{i,j}\left|\frac{\partial^2 f}{\partial x_i \partial x_j}(\overrightarrow{a}) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{s})\right|\max_{i,j}\frac{|\triangle x_i \triangle x_j|}{\triangle \overrightarrow{x}^2}. \qquad (3.29)\end{aligned}$$

By an argument analogous to that giving Equation (2.20) on p. 154 (Exercise 7), we can say that

$$\max_{i,j}\frac{|\triangle x_i \triangle x_j|}{\|\triangle \overrightarrow{x}\|^2} \leq 1$$

and by continuity of the second-order partials, for each i and j

$$\lim_{\overrightarrow{x}\to\overrightarrow{a}}\frac{\partial^2 f}{\partial x_i \partial x_j}(\overrightarrow{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\overrightarrow{a}).$$

Together, these arguments show that the right-hand side of Equation (3.29) goes to zero as $\overrightarrow{x} \to \overrightarrow{a}$ (since also $\vec{s} \to \overrightarrow{a}$), proving our claim. □

We note in passing that higher-order “total” derivatives, and the corresponding higher-degree Taylor polynomials, can also be defined and shown to satisfy higher-order contact conditions. However, the formulation of these quantities involves more complicated multi-index formulas, and since we shall not use derivatives beyond order two in our theory, we leave these constructions and proofs to your imagination.

Exercises for § 3.7

Practice problems:

1. Find $\frac{\partial^2 f}{\partial^2 x}$, $\frac{\partial^2 f}{\partial^2 y}$, and $\frac{\partial^2 f}{\partial y \partial x}$ for each function below:

   (a) $f(x,y) = x^2 y$ (b) $f(x,y) = \sin x + \cos y$ (c) $f(x,y) = x^3 y + 3xy^2$

   (d) $f(x,y) = \sin(x^2 y)$ (e) $f(x,y) = \sin(x^2 + 2y)$ (f) $f(x,y) = \ln(x^2 y)$

   (g) $f(x,y) = \ln(x^2 y + xy^2)$ (h) $f(x,y) = \dfrac{x+y}{x^2+y^2}$ (i) $f(x,y) = \dfrac{xy}{x^2+y^2}$

2. Find all second-order derivatives of each function below:

(a) $f(x,y,z) = x^2 + y^2 + z^2$ (b) $f(x,y,z) = xyz$ (c) $f(x,y,z) = \sqrt{xyz}$

(d) $f(x,y,z) = \dfrac{1}{x^2+y^2+z^2}$ (e) $f(x,y,z) = e^{x^2+y^2+z^2}$ (f) $f(x,y,z) = \dfrac{xyz}{x^2+y^2+z^2}$

3. Find the degree two Taylor polynomial of the given function at the given point:

   (a) $f(x,y) = x^3y^2$ at $(-1,-2)$.

   (b) $f(x,y) = \frac{x}{y}$ at $(2,3)$.

   (c) $f(x,y,z) = \frac{xy}{z}$ at $(2,-3,5)$.

4. Let

$$f(x,y) = \frac{xy}{x+y}.$$

   (a) Calculate an approximation to $f(0.8,1.9)$ using the degree one Taylor polynomial at $(1,2)$, $T_{(1,2)}f(\triangle x, \triangle y)$.

   (b) Calculate an approximation to $f(0.8,1.9)$ using the degree two Taylor polynomial at $(1,2)$, $T_2 f\left((1,2)\right)\triangle x, \triangle y$.

   (c) Compare the two to the calculator value of $f(0.8,1.9)$; does the second approximation improve the accuracy, and by how much?

Theory problems:

5. (a) Show that for any $\mathcal{C}^2$ function $f(x,y,z)$,

$$\frac{\partial^2 f}{\partial x \partial z}(x_0,y_0,z_0) = \frac{\partial^2 f}{\partial z \partial x}(x_0,y_0,z_0)\,.$$

   (b) How many *distinct* partial derivatives of order 2 can a $\mathcal{C}^2$ function of n variables have?

6. (a) Show that for any $\mathcal{C}^3$ function $f(x,y)$,

$$\frac{\partial^3 f}{\partial x \partial y \partial x}(x_0,y_0) = \frac{\partial^3 f}{\partial y \partial^2 x}(x_0,y_0).$$

   (b) How many *distinct* partial derivatives of order n can a $\mathcal{C}^n$ function of two variables have?

   (c) How many *distinct* partial derivatives of order n can a $\mathcal{C}^n$ function of three variables have?

7. Prove the comment in the proof of Proposition 3.7.2, that

$$\max_{(i,j)} \frac{|\triangle x_i \triangle x_j|}{\|\triangle \overrightarrow{x}\|^2} \leq 1.$$

Challenge problem:

8. Consider the function of two variables

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{if } x^2+y^2 \neq 0, \\ 0 & \text{at } (x,y) = (0,0). \end{cases}$$

   (a) Calculate $\frac{\partial f}{\partial x}(x,y)$ and $\frac{\partial f}{\partial y}(x,y)$ for $(x,y) \neq (0,0)$.

   (b) Calculate $\frac{\partial f}{\partial x}(0,0)$ and $\frac{\partial f}{\partial y}(0,0)$.

   (c) Calculate the second-order partial derivatives of $f(x,y)$ at $(x,y) \neq (0,0)$.

   (d) Calculate the second-order partial derivatives of $f(x,y)$ at the origin. Note that

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) \neq \frac{\partial^2 f}{\partial y \partial x}(0,0).$$

   Explain.

3.8 Local Extrema

The Critical Point Theorem (Theorem 3.6.10) tells us that a local extremum must be a critical point: if a differentiable function $f: \mathbb{R}^3 \to \mathbb{R}$ has a local maximum (or local minimum) at $\overrightarrow{a}$, then $d_{\overrightarrow{a}} f(\overrightarrow{v}) = 0$ for all $\overrightarrow{v} \in \mathbb{R}^3$. The converse is not true: for example, the function $f(x) = x^3$ has a critical point at the origin but is strictly increasing on the whole real line. Other phenomena are possible for multivariate functions: for example, the restriction of $f(xy) = x^2 - y^2$ to the x-axis has a *minimum* at the origin, while its restriction to the y-axis has a *maximum* there. So to determine whether a critical point $\overrightarrow{a}$ is a local extremum, we need to study the (local) behavior in all directions—in particular, we need to study the Hessian $d^2_{\overrightarrow{a}} f$.

Definite Quadratic Forms

Since it is homogeneous, every quadratic form is zero at the origin. We call the quadratic form Q **definite** if it is nonzero everywhere else: $Q(\overrightarrow{x}) \neq 0$ for $\overrightarrow{x} \neq \overrightarrow{0}$. For example, the forms $Q(x,y) = x^2 + y^2$ and $Q(x,y) = -x^2 - 2y^2$ are definite, while $Q(x,y) = x^2 - y^2$ and $Q(x,y) = xy$ are not. We shall see that the form

$$Q(x,y) = 2(x+y)^2 + y(y-6x) = 2x^2 - 2xy + 3y^2$$

is definite, but *a priori* this is not entirely obvious.

A definite quadratic form cannot switch sign, since along any path where the endpoint values of Q have opposite sign there would be a point where $Q = 0$, and such a path could be picked to avoid the origin, giving a point other than the origin where $Q = 0$:

Remark 3.8.1. *If $Q(\overrightarrow{x})$ is a definite quadratic form, then one of the following inequalities holds:*

- *$Q(\overrightarrow{x}) > 0$ for all $\overrightarrow{x} \neq \overrightarrow{0}$ (Q is **positive definite**), or*
- *$Q(\overrightarrow{x}) < 0$ for all $\overrightarrow{x} \neq \overrightarrow{0}$ (Q is **negative definite**)*

Actually, in this case we can say more:

Lemma 3.8.2. *If $Q(\overrightarrow{x})$ is a positive definite (resp. negative definite) quadratic form, then there exists $K > 0$ such that*

$$Q(\overrightarrow{x}) \geq K \|\overrightarrow{x}\|^2 \text{ (resp. } Q(\overrightarrow{x}) \leq -K \|\overrightarrow{x}\|^2\text{) for all } x.$$

Proof. The inequality is trivial for $\overrightarrow{x} = \overrightarrow{0}$. If $\overrightarrow{x} \neq \overrightarrow{0}$, let $\overrightarrow{u} = \overrightarrow{x} / \|\overrightarrow{x}\|$ be the unit vector parallel to $\overrightarrow{x}$; then

$$Q(\overrightarrow{x}) = Q(\overrightarrow{u}) \|\overrightarrow{x}\|^2$$

and so we need only show that $|Q(\overrightarrow{x})|$ is bounded away from zero on the **unit sphere**

$$\mathcal{S} = \{\overrightarrow{u} \mid \|\overrightarrow{u}\| = 1\}.$$

In the plane, $\mathcal{S}$ is the unit circle $x^2 + y^2 = 1$, while in space it is the unit sphere $x^2 + y^2 + z^2 = 1$. Since $\mathcal{S}$ is closed and bounded (Exercise 3), it is sequentially compact, so $|Q(\overrightarrow{x})|$ achieves its minimum on $\mathcal{S}$, which is not zero, since Q is definite. It is easy to see that

$$K = \min_{\|\overrightarrow{u}\|=1} |Q(\overrightarrow{u})|$$

has the required property. □

Using Lemma 3.8.2 and Taylor's theorem (Proposition 3.7.2), we can show that a critical point with definite Hessian is a local extremum.

Proposition 3.8.3. *Suppose f is a $\mathcal{C}^2$ function and $\overrightarrow{a}$ is a critical point for f where the Hessian form $d^2_{\overrightarrow{a}}f$ is definite.*

Then f has a local extremum at $\overrightarrow{a}$:

- *If $d^2_{\overrightarrow{a}}f$ is positive definite, then f has a local minimum at $\overrightarrow{a}$;*
- *If $d^2_{\overrightarrow{a}}f$ is negative definite, then f has a local maximum at $\overrightarrow{a}$.*

Proof. The fact that the quadratic approximation $T_2 f\left(\overrightarrow{a}\right)\overrightarrow{x}$ has second order contact with $f(\overrightarrow{x})$ at $\overrightarrow{x} = \overrightarrow{a}$ can be written in the form

$$f(\overrightarrow{x}) = T_2 f\left(\overrightarrow{a}\right)\overrightarrow{x} + \varepsilon(\overrightarrow{x})\left\|\overrightarrow{x} - \overrightarrow{a}\right\|^2, \quad \text{where } \lim_{\overrightarrow{x}\to\overrightarrow{a}} \varepsilon(\overrightarrow{x}) = 0.$$

Since $\overrightarrow{a}$ is a critical point, $d_{\overrightarrow{a}}f(\triangle\overrightarrow{x}) = 0$, so

$$T_2 f\left(\overrightarrow{a}\right)\overrightarrow{x} = f(\overrightarrow{a}) + \frac{1}{2}d^2_{\overrightarrow{a}}f(\triangle\overrightarrow{x}),$$

or

$$f(\overrightarrow{x}) - f(\overrightarrow{a}) = \frac{1}{2}d^2_{\overrightarrow{a}}f(\triangle\overrightarrow{x}) + \varepsilon(\overrightarrow{x})\left\|\triangle\overrightarrow{x}\right\|^2.$$

Suppose $d^2_{\overrightarrow{a}}f$ is positive definite, and let $K > 0$ be the constant given in Lemma 3.8.2, such that

$$d^2_{\overrightarrow{a}}f(\triangle\overrightarrow{x}) \geq K\left\|\triangle\overrightarrow{x}\right\|^2.$$

Since $\varepsilon(\overrightarrow{x}) \to 0$ as $\overrightarrow{x} \to \overrightarrow{a}$, for $\|\triangle\overrightarrow{x}\|$ sufficiently small, we have

$$|\varepsilon(\overrightarrow{x})| < \frac{K}{4}$$

and hence

$$f(\overrightarrow{x}) - f(\overrightarrow{a}) \geq \{\frac{K}{2} - \varepsilon(\overrightarrow{x})\}\left\|\triangle\overrightarrow{x}\right\|^2 > \frac{K}{4}\left\|\triangle\overrightarrow{x}\right\|^2 > 0$$

or

$$f(\overrightarrow{x}) > f(\overrightarrow{a}) \text{ for } \overrightarrow{x} \neq \overrightarrow{a} \quad (\|\triangle\overrightarrow{x}\| \text{ sufficiently small}).$$

The argument when $d^2_{\overrightarrow{a}}f$ is negative definite is analogous (Exercise 4a). □

An analogous argument (Exercise 4b) gives

Lemma 3.8.4. *If $d^2_{\overrightarrow{a}}f$ takes both positive and negative values at the critical point $\overrightarrow{x} = \overrightarrow{a}$ of f, then f does* not *have a local extremum at $\overrightarrow{x} = \overrightarrow{a}$.*

Quadratic Forms in $\mathbb{R}^2$

To take advantage of Proposition 3.8.3 we need a way to decide whether or not a given quadratic form Q is positive definite, negative definite, or neither.

In the planar case, there is an easy and direct way to decide this.

If we write $Q(x_1, x_2)$ in the form

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2$$

then we can factor out "a" from the first two terms and complete the square:

$$\begin{aligned} Q(x_1, x_2) &= a\left((x_1 + \frac{b}{a}x_2)^2 - \frac{b^2}{a^2}\right) + cx_2^2 \\ &= a\left(x_1 + \frac{b}{a}x_2\right)^2 + \left(c - \frac{b^2}{a}\right)\left(x_2\right)^2. \end{aligned}$$

Thus, Q is definite provided the two coefficients in the last line have the same sign, or equivalently, if their product is positive:[16]

$$ac - b^2 > 0.$$

The quantity in this inequality will be denoted Δ_2; it can be written as the determinant of the matrix

$$[Q] = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

which is the **matrix representative** of Q.

If $\Delta_2 > 0$, then Q is *definite*, which is to say the two coefficients in the expression for $Q(x_1, x_2)$ have the same sign; to tell whether it is *positive* definite or *negative* definite, we need to decide if this sign is positive or negative, and this is most easily seen by looking at the sign of a, which we will denote Δ_1. The significance of these observations will become clear later.

With this notation, we have

Proposition 3.8.5. *A quadratic form*

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2$$

[16] Note that if either coefficient is zero, then there is a whole line along which $Q = 0$, so it is not definite.

is definite only if

$$\Delta_2 := ac - b^2 > 0;$$

it is positive *definite if in addition*

$$\Delta_1 := a > 0$$

and negative *definite if*

$$\Delta_1 < 0.$$

If $\Delta_2 < 0$, then $Q(\overrightarrow{x})$ takes both (strictly) positive and (strictly) negative values.

Let us see what this tells us about the forms we introduced at the beginning of this section:

$$Q(x,y) = x^2 + y^2$$

has

$$A = [Q] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so

$$\begin{aligned} \Delta_1 &= 1 > 0 \\ \Delta_2 &= 1 > 0 \end{aligned}$$

and Q is positive definite.

$$Q(x,y) = -x^2 - 2y^2$$

has

$$A = [Q] = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}$$

so

$$\begin{aligned} \Delta_1 &= -1 < 0 \\ \Delta_2 &= 2 > 0 \end{aligned}$$

and Q is negative definite.

$$Q(x,y) = x^2 - y^2$$

has

$$A = [Q] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

so

$$\Delta_2 = -1 < 0$$

and Q is not definite.

$$Q(x,y) = xy$$

has

$$A = [Q] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

so

$$\Delta_2 = -1 < 0$$

and Q is not definite.

Finally, for the one we couldn't decide in an obvious way:

$$Q(x,y) = 2x^2 - 2xy + 3y^2$$

has

$$A = [Q] = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}$$

so

$$\begin{aligned} \Delta_1 &= 2 > 0 \\ \Delta_2 &= 5 > 0 \end{aligned}$$

and Q is positive definite.

When applied to the Hessian of $f\colon\mathbb{R}^2\to\mathbb{R}$, the matrix representative of the Hessian form is the matrix of partials of f, sometimes called the **Hessian matrix** of f:

$$Hf(\overrightarrow{a}) = \begin{bmatrix} f_{xx}(\overrightarrow{a}) & f_{xy}(\overrightarrow{a}) \\ f_{xy}(\overrightarrow{a}) & f_{yy}(\overrightarrow{a}) \end{bmatrix}.$$

this gives us [17]

Theorem 3.8.6 (Second Derivative Test, Two Variables)**.** *If $f\colon\mathbb{R}^2\to\mathbb{R}$ is $\mathcal{C}^2$ and has a critical point at $\overrightarrow{x}=\overrightarrow{a}$, consider the determinant of the Hessian matrix* [18]

$$\Delta = \Delta_2(\overrightarrow{a}) = f_{xx}(\overrightarrow{a})\, f_{yy}(\overrightarrow{a}) - f_{xy}(\overrightarrow{a})^2\,,$$

and its upper left entry

$$\Delta_1(\overrightarrow{a}) = f_{xx}.$$

Then:

1. *if $\Delta > 0$, then $\overrightarrow{a}$ is a local extremum of f:*
   - *(a) it is a local minimum if $\Delta_1(\overrightarrow{a}) = f_{xx} > 0$*
   - *(b) it is a local maximum if $\Delta_1(\overrightarrow{a}) = f_{xx} < 0$;*
2. *if $\Delta < 0$, then $\overrightarrow{a}$ is not a local extremum of f;*
3. *$\Delta = 0$ does not give enough information to distinguish the possibilities.*

Proof.
1. We know that $d^2_{\overrightarrow{a}}f$ is positive (*resp.* negative) definite by Proposition 3.8.5, and then apply Proposition 3.8.3.
2. Apply Proposition 3.8.5 and then Lemma 3.8.4 in the same way.
3. Consider the following three functions:

$$\begin{aligned} f(x,y) &= (x+y)^2 = x^2+2xy+y^2 \\ g(x,y) &= f(x,y)+y^4 = x^2+2xy+y^2+y^4 \\ h(x,y) &= f(x,y)-y^4 = x^2+2xy+y^2-y^4. \end{aligned}$$

[17] The Second Derivative Test was published by Joseph Louis Lagrange (1736-1813) in his very first mathematical paper [33] ([20, p. 323]).

[18] sometimes called the **discriminant** of f

They all have second order contact at the origin, which is a critical point, and all have Hessian matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

so all have $\Delta = 0$. However:

- f has a weak local minimum at the origin: the function is non-negative everywhere, but equals zero along the whole line $y = -x$;
- g has a strict minimum at the origin: $g(\overrightarrow{x}) > 0$ for all $\overrightarrow{x} \neq \overrightarrow{0}$, and
- h has saddle behavior: its restriction to the x-axis has a minimum at the origin, while its restriction to the line $y = -x$ has a maximum at the origin.

□

As an example, consider the function

$$f(x,y) = 5x^2 + 6xy + 5y^2 - 8x - 8y.$$

We calculate the first partials

$$\begin{aligned} f_x(x,y) &= 10x + 6y - 8 \\ f_y(x,y) &= 6x + 10y - 8 \end{aligned}$$

and set both equal to zero to find the critical points:

$$\begin{aligned} 10x + 6y &= 8 \\ 6x + 10y &= 8 \end{aligned}$$

has the unique solution

$$(x,y) = \left(\frac{1}{2}, \frac{1}{2}\right).$$

Now we calculate the second partials

$$\begin{aligned} f_{xx}(x,y) &= 10 \\ f_{xy}(x,y) &= 6 \\ f_{yy}(x,y) &= 10. \end{aligned}$$

Thus, the discriminant is

$$\Delta_2(x,y) := f_{xx}f_{yy} - (f_{xy})^2 = (10)\cdot(10) - (6)^2 > 0$$

and since also

$$\Delta_1(x,y) = f_x(x,y) = 6 > 0$$

the function has a local minimum at $\left(\frac{1}{2},\frac{1}{2}\right)$.

As another example,

$$f(x,y) = 5x^2 + 26xy + 5y^2 - 36x - 36y + 12$$

has

$$\begin{aligned} f_x(x,y) &= 10x + 26y - 36 \\ f_y(x,y) &= 26x + 10y - 36 \end{aligned}$$

so the sole critical point is $(1,1)$; the second partials are

$$\begin{aligned} f_{xx}(x,y) &= 10 \\ f_{xy}(x,y) &= 26 \\ f_{yy}(x,y) &= 10 \end{aligned}$$

so the discriminant is

$$\Delta_2(1,1) = (10)\cdot(10) - (26)^2 < 0$$

and the function has a saddle point at $(1,1)$.

Finally, consider

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

We have

$$\begin{aligned} f_x(x,y) &= 3x^2 + 6x = 3x(x+2) \\ f_y(x,y) &= -3y^2 + 3 \end{aligned}$$

and these both vanish when $x = 0$ or -2 and $y = \pm 1$, yielding four critical points. The second partials are

$$\begin{aligned} f_{xx}(x,y) &= 6x + 6 \\ f_{xy}(x,y) &= 0 \\ f_{yy}(x,y) &= -6y \end{aligned}$$

so the discriminant is

$$\begin{aligned} \Delta_2(x,y) &= (6x+6)(-6y) - 0 \\ &= -36(x+1)y. \end{aligned}$$

The respective values at the four critical points are

$$\begin{gathered} \Delta(2)\, 0, -1 = 36 > 0 \\ \Delta(2)\, 0, 1 = -36 < 0 \\ \Delta(2)\, -2, -1 = -36 < 0 \\ \Delta(2)\, -2, 1 = 36 > 0 \end{gathered}$$

so $(0,1)$ and $(-2,-1)$ are saddle points, while $(0,-1)$ and $(-2,1)$ are local extrema; for further information about the extrema, we consider the first partials there:

$$\Delta_1(0,1) = f_x(0,1) = 6 > 0$$

so $f(x,y)$ has a local minimum there, while

$$\Delta_1(-2,1) = f_x(-2,1) = -12 < 0$$

so $f(x,y)$ has a local maximum there.

The situation for three or more variables is more complicated. In the next section, we establish the *Principal Axis Theorem* which gives us more detailed information about the behavior of quadratic forms. This will help us understand the calculations in this section, and also the more subtle considerations at play in $\mathbb{R}^3$.

Exercises for § 3.8

Practice problems:

1. For each quadratic form below, find its matrix representative, and use Proposition 3.8.5 to decide whether it is *positive definite*, *negative* definite, or not definite.

   (a) $Q(x,y) = x^2 - 2xy + y^2$
   (b) $Q(x,y) = x^2 + 4xy + y^2$
   (c) $Q(x,y) = 2x^2 + 2xy + y^2$
   (d) $Q(x,y) = x^2 - 2xy + 2y^2$
   (e) $Q(x,y) = 2xy$
   (f) $Q(x,y) = 4x^2 + 4xy$
   (g) $Q(x,y) = 4x^2 - 2xy$
   (h) $Q(x,y) = -2x^2 + 2xy - 2y^2$

2. For each function below, locate all critical points and classify each as a local maximum, local minimum, or saddle point.

   (a) $f(x,y) = 5x^2 - 2xy + 10y^2 + 1$
   (b) $f(x,y) = 3x^2 + 10xy - 8y^2 + 2$
   (c) $f(x,y) = x^2 - xy + y^2 + 3x - 2y + 1$
   (d) $f(x,y) = x^2 + 3xy + y^2 + x - y + 5$
   (e) $f(x,y) = 5x^2 - 2xy + y^2 - 2x - 2y + 25$
   (f) $f(x,y) = 5y^2 + 2xy - 2x - 4y + 1$
   (g) $f(x,y) = (x^3 - 3x)(y^2 - 1)$
   (h) $f(x,y) = x + y \sin x$

Theory problems:

3. Show that the unit sphere $\mathcal{S}$ is a closed and bounded set.

4. (a) Mimic the proof given in the *positive* definite case of Proposition 3.8.3 to prove the *negative* definite case.

   (b) Prove Lemma 3.8.4.

3.9 The Principal Axis Theorem

In this section, we extend the analysis of quadratic forms from two to three variables, which requires some new ideas.

First, we need to clarify the mysterious “matrix representative” that appeared, for a quadratic form in two variables, in § 3.8.

Matrix Representative of a Quadratic Form

We saw in § 3.2 that a linear real-valued function $\ell(\overrightarrow{x})$ can be expressed as multiplication of the coordinate column $[\overrightarrow{x}]$ of the input vector by a row of coefficients; for $\mathbb{R}^2$, this reads

$$\ell(\overrightarrow{x}) = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_1 \cdot x_1 + a_2 x_2 = a_1 x + a_2 y$$

while for $\mathbb{R}^3$ it reads

$$\ell(\overrightarrow{x}) = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = a_1 \cdot x_1 + a_2 x_2 + a_3 x_3. = a_1 x + a_2 y + a_3 z.$$

Analogously, we can express any quadratic form as a three-factor product, using the basic matrix arithmetic which is reviewed in Appendix E. For example, there are four kinds of quadratic terms in the two variables x and y: x^2, y^2, xy and yx (for the moment, let us ignore the fact that we can combine the last two). Then in the expression

$$Q(x, y) = \alpha x^2 + \beta xy + \gamma yx + \delta y^2$$

we can factor out the initial x factor from the first two terms and the initial y factor from the last two to write

$$Q(x, y) = x(\alpha x + \beta y) + y(\gamma x + \delta y)$$

which can be written as the product of a row with a column

$$= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \alpha x + \beta y \\ \gamma x + \delta y \end{bmatrix}.$$

The column on the right can be expressed in turn as the product of a 2×2 matrix with a column, leading to the three-factor product

$$= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The two outside factors are clearly the coordinate column of (x, y) and its transpose. The 2×2 matrix in the middle could be regarded as a matrix representing Q, but note that there is an ambiguity here: the two "mixed product" terms βxy and γyx can be rewritten in many other ways without changing their total value; all we need to do is to make sure that the sum $\beta+\gamma$ is unchanged. Thus, any other matrix with the same diagonal entries α and δ, and whose off-diagonal entries add up to $\beta+\gamma$, leads to the same function $Q(x, y)$. To standardize things, we require that the matrix be symmetric. This amounts to "balancing" the two mixed product terms: each is equal to half will have some useful consequences down the road. Thus the **matrix representative** of a quadratic form $Q(x, y)$ in two variables is the *symmetric* 2×2 matrix $[Q]$ satisfying

$$Q(\overrightarrow{x}) = [\overrightarrow{x}]^T\,[Q]\,[\overrightarrow{x}]\,. \tag{3.30}$$

You should confirm that this is the same as the matrix representative we used in § 3.8.

When we apply Equation (3.30) to a quadratic form in three variables $Q(x_1, x_2, x_3)$, we get a symmetric 3×3 matrix. The diagonal entries of $[Q]$ are the coefficients a_{ii} of the "square" terms x_i^2, and each off-diagonal entry is *half* of the coefficient b_{ij} of a "mixed product" term x_ix_j: if

$$\begin{aligned} Q(x_1, x_2, x_3) = a_{11}x_1^2 &+ b_{12}x_1x_2 + b_{13}x_1x_3\\ &+ a_{22}x_2^2 + b_{23}x_2x_3 + a_{33}x_3^2 \end{aligned}$$

then we rewrite it in "balanced" form

$$\begin{aligned} Q(x_1, x_2, x_3) = a_{11}x_1^2 &+ a_{12}x_1x_2 + a_{13}x_1x_3\\ &+ a_{21}x_2x_1 + a_{22}x_2^2 + a_{23}x_2x_3\\ &+ a_{31}x_3x_1 + a_{32}x_3x_2 + a_{33}x_3^2 \end{aligned}$$

where

$$\begin{aligned} a_{12} = a_{21} &= \frac{1}{2}b_{12}\\ a_{13} = a_{31} &= \frac{1}{2}b_{13}\\ a_{23} = a_{32} &= \frac{1}{2}b_{23} \end{aligned}$$

and its matrix representative is

$$[Q] = \begin{bmatrix} a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & \frac{1}{2}b_{12} & \frac{1}{2}b_{13}\\ \frac{1}{2}b_{12} & a_{22} & \frac{1}{2}b_{23}\\ \frac{1}{2}b_{13} & \frac{1}{2}b_{23} & a_{33} \end{bmatrix}.$$

The Principal Axis Theorem

Using the language of matrices, Proposition 3.8.5 can be rephrased as: Q is positive (*resp.* negative) definite if the determinant of its matrix representative is positive and its upper-lefthand entry a_{11} is positive (*resp.* negative). This does not carry over to forms in three or more variables. For example, the quadratic form

$$Q(x,y,z) = x^2 - y^2 - z^2$$

which is clearly *not* definite, has $a_{11} = 1 > 0$ and

$$[Q] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

with determinant $1 > 0$. It turns out that we have to look at a minor of the determinant, as well.

To understand this, we approach our problem differently, taking a clue from the proof of Lemma 3.8.2: to know that Q is positive definite[19] we need to establish that the minimum value of its restriction to the unit sphere

$$\mathcal{S}^2 = \{\overrightarrow{u} \in \mathbb{R}^3 \mid \|\overrightarrow{u}\| = 1\}$$

is positive. This means we need to consider the constrained optimization problem: find the minimum of

$$f(\overrightarrow{x}) = Q(\overrightarrow{x})$$

subject to the constraint

$$g(\overrightarrow{x}) = \overrightarrow{x} \cdot \overrightarrow{x} = 1.$$

This can be attacked using Lagrange multipliers: we know that the point $\overrightarrow{u} \in \mathcal{S}^2$ where the minimum occurs satisfies the condition

$$\overrightarrow{\nabla} f(\overrightarrow{u}) = \lambda \overrightarrow{\nabla} g(\overrightarrow{u}) \text{ for some } \lambda \in \mathbb{R}.$$

We already know that

$$\overrightarrow{\nabla} g(\overrightarrow{u}) = 2\overrightarrow{u};$$

we need to calculate $\overrightarrow{\nabla} f(\overrightarrow{u})$.

[19] we return to the negative definite case at the end of this subsection

To this end, we write $f(x_1, x_2, x_3) = Q(x_1, x_2, x_3)$ in the matrix form

$$f(x_1, x_2, x_3) = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \sum_{i=1}^{3} \sum_{j=1}^{3} a_{ij} x_i x_j.$$

To find $\frac{\partial f}{\partial x_1}$, we locate all the terms involving x_1: they are

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{21}x_2x_1 + a_{31}x_3x_1;$$

using the symmetry of A we can combine some of these terms to get

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3.$$

Differentiating with respect to x_1, this gives

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 2a_{11}x_1 + 2a_{12}x_2 + 2a_{13}x_3 \\ &= 2(a_{11}x_1 + a_{12}x_2 + a_{13}x_3). \end{aligned}$$

Note that the quantity in parentheses is exactly the product of the first row of $A = [Q]$ with $[\overrightarrow{x}]$, or equivalently the first entry of $A[\overrightarrow{x}]$. For convenience, we will abuse notation, and write simply $A\overrightarrow{x}$ for the vector whose coordinate column is A times the coordinate column of $\overrightarrow{x}$:

$$[A\overrightarrow{x}] = A[\overrightarrow{x}].$$

You should check that the other two partials of f are the other coordinates of $A\overrightarrow{x}$, so

$$\overrightarrow{\nabla} f(\overrightarrow{u}) = 2A\overrightarrow{u}.$$

If we also recall that the dot product of two vectors can be written in terms of their coordinate columns as

$$\overrightarrow{x} \cdot \overrightarrow{y} = [\overrightarrow{x}]^T [\overrightarrow{y}]$$

then the matrix form of $f(\overrightarrow{x}) = Q(\overrightarrow{x})$ becomes

$$f(\overrightarrow{x}) = \overrightarrow{x} \cdot A\overrightarrow{x};$$

we separate out this calculation as

Remark 3.9.1. *The gradient of a function of the form*

$$f(\overrightarrow{x}) = \overrightarrow{x} \cdot A\overrightarrow{x}$$

is

$$\overrightarrow{\nabla} f(\overrightarrow{x}) = 2A\overrightarrow{x}.$$

Note that, while our calculation was for a 3×3 matrix, the analogous result holds for a 2×2 matrix as well.

Now, the Lagrange multiplier condition for extrema of f on $\mathcal{S}^2$ becomes

$$A\overrightarrow{u} = \lambda\overrightarrow{u}. \tag{3.31}$$

Geometrically, this means that $A\overrightarrow{u}$ and $\overrightarrow{u}$ have the same direction (up to reversal, or possibly squashing to zero). Such situations come up often in problems involving matrices; we call a nonzero vector $\overrightarrow{u}$ which satisfies Equation (3.31) an **eigenvector** of A; the associated scalar λ is called the **eigenvalue** of A associated to $\overrightarrow{u}$.[20]

Our discussion has shown that every symmetric matrix A has an eigenvector, corresponding to the minimum of the associated quadratic form $Q(\overrightarrow{x}) = \overrightarrow{x} \cdot A\overrightarrow{x}$ on the unit sphere. In fact, we can say more:

Proposition 3.9.2 (Principal Axis Theorem for $\mathbb{R}^3$)**.** *If A is a symmetric 3×3 matrix, then there exist three mutually perpendicular unit eigenvectors for A: $\overrightarrow{u}_i$, $i = 1, 2, 3$ satisfying*

$$A\overrightarrow{u}_i = \lambda_i \overrightarrow{u}_i \textit{ for some scalars } \lambda_i \in \mathbb{R}, \quad i = 1, 2, 3,$$
$$\overrightarrow{u}_i \cdot \overrightarrow{u}_j = \begin{cases} 0 & \textit{if } i \neq j, \\ 1 & \textit{if } i = j. \end{cases}$$

Proof. Since the unit sphere $\mathcal{S}^2$ is sequentially compact, the restriction to $\mathcal{S}^2$ of the quadratic form $Q(\overrightarrow{x}) = \overrightarrow{x} \cdot A\overrightarrow{x}$ achieves its minimum somewhere, say $\overrightarrow{u}_1$, and this is a solution of the equations

$$\begin{aligned} A\overrightarrow{u}_1 &= \lambda_1 \overrightarrow{u}_1 \quad (\text{some } \lambda_1) \\ \overrightarrow{u}_1 \cdot \overrightarrow{u}_1 &= 1. \end{aligned}$$

Now, consider the plane $\mathcal{P}$ through the origin perpendicular to $\overrightarrow{u}_1$

$$\mathcal{P} = \overrightarrow{u}_1^{\perp} = \{\overrightarrow{x} \in \mathbb{R}^3 \,|\, \overrightarrow{x} \cdot \overrightarrow{u} = 0\}$$

[20] Another terminology calls an eigenvector a **characteristic vector** and an eigenvalue a **characteristic value** of A.

and look at the restriction of Q to the circle

$$\mathcal{P} \cap \mathcal{S}^2 = \{\overrightarrow{u} \in \mathbb{R}^3 \mid \overrightarrow{u} \cdot \overrightarrow{u}_1 = 0 \text{ and } \overrightarrow{u} \cdot \overrightarrow{u} = 1\}.$$

This has a minimum at $\overrightarrow{u}_2$ and a maximum at $\overrightarrow{u}_3$, each of which is a solution of the Lagrange multiplier equations

$$\begin{aligned}\overrightarrow{\nabla} f(\overrightarrow{u}) &= \lambda \overrightarrow{\nabla} g_1(\overrightarrow{u}) + \mu \overrightarrow{\nabla} g_2(\overrightarrow{u})\\ g_1(\overrightarrow{u}) = \overrightarrow{u} \cdot \overrightarrow{u} &= 1\\ g_2(\overrightarrow{u}) = \overrightarrow{u} \cdot \overrightarrow{u}_1 &= 0.\end{aligned}$$

Again, we have, for $i = 1, 2, 3$,

$$\begin{aligned}\overrightarrow{\nabla} f(\overrightarrow{u}) &= 2A\overrightarrow{u}\\ \overrightarrow{\nabla} g_1(\overrightarrow{u}) &= 2\overrightarrow{u}\end{aligned}$$

and clearly

$$\overrightarrow{\nabla} g_2(\overrightarrow{u}) = \overrightarrow{u}_1,$$

so the first equation reads

$$2A\overrightarrow{u} = 2\lambda\overrightarrow{u} + \mu\overrightarrow{u}_1.$$

If we take the dot product of both sides of this with $\overrightarrow{u}_1$, using the fact that $\overrightarrow{u} \cdot \overrightarrow{u}_1 = 0$ and $\overrightarrow{u}_1 \cdot \overrightarrow{u}_1 = 1$, we obtain

$$0 = \mu$$

so for $\overrightarrow{u} = \overrightarrow{u}_i$, $i = 2, 3$, the first equation, as before, is the eigenvector condition

$$A\overrightarrow{u}_i = \lambda_i \overrightarrow{u}_i.$$

We already know that $\overrightarrow{u}_1 \cdot \overrightarrow{u}_2 = \overrightarrow{u}_1 \cdot \overrightarrow{u}_3 = 0$ and $\overrightarrow{u}_2 \cdot \overrightarrow{u}_2 = \overrightarrow{u}_3 \cdot \overrightarrow{u}_3 = 1$, but what about $\overrightarrow{u}_2 \cdot \overrightarrow{u}_3$? If Q is constant on $\mathcal{P} \cap \mathcal{S}^2$, then every vector in $\mathcal{P} \cap \mathcal{S}^2$ qualifies as $\overrightarrow{u}_2$ and/or $\overrightarrow{u}_3$, so we simply *pick* these to be mutually perpendicular. If not, then

$$\begin{aligned}\lambda_2 = \overrightarrow{u}_2 \cdot (\lambda_2 \overrightarrow{u}_2) = Q(\overrightarrow{u}_2) &= \min_{\overrightarrow{u} \in \mathcal{P}\cap\mathcal{S}^2} Q(\overrightarrow{u})\\ &< \max_{\overrightarrow{u} \in \mathcal{P}\cap\mathcal{S}^2} Q(\overrightarrow{u}) = Q(\overrightarrow{u}_3) = \lambda_3\end{aligned}$$

so $\lambda_2 \neq \lambda_3$. So far, we haven't used the symmetry of A, which gives us

$$\left([\overrightarrow{u}_2]^T A\,[\overrightarrow{u}_3]\right)^T = [\overrightarrow{u}_3]^T A\,[\overrightarrow{u}_2]\,.$$

This can be reinterpreted as saying that

$$\overrightarrow{u}_2 \cdot (A\overrightarrow{u}_3) = \overrightarrow{u}_3 \cdot (A\overrightarrow{u}_2).$$

This lets us say that

$$\begin{aligned}(\lambda_2\overrightarrow{u}_2) \cdot \overrightarrow{u}_3 &= (A\overrightarrow{u}_2) \cdot \overrightarrow{u}_3 \\ &= \overrightarrow{u}_2 \cdot (A\overrightarrow{u}_3) \\ &= \overrightarrow{u}_2 \cdot (\lambda_3\overrightarrow{u}_3) \\ &= \lambda_3\overrightarrow{u}_2 \cdot \overrightarrow{u}_3.\end{aligned}$$

Since $\lambda_2 \neq \lambda_3$, we must have $\overrightarrow{u}_2 \cdot \overrightarrow{u}_3 = 0$, completing the proof of the proposition. □

What does this result tell us about quadratic forms?

We start with some consequences of the second property of the eigenvectors $\overrightarrow{u}_i$:

$$\overrightarrow{u}_i \cdot \overrightarrow{u}_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

Geometrically, this says two things: first (using $i = j$) they are *unit* vectors ($|\overrightarrow{u}_i|^2 - \overrightarrow{u}_i \cdot \overrightarrow{u}_i = 1$), and second, they are mutually perpendicular. A collection of vectors with both properties is called an **orthonormal** set of vectors. Since they define three mutually perpendicular directions in space, a set of three orthonormal vectors in $\mathbb{R}^3$ can be used to set up a new rectangular coordinate system: any vector $\overrightarrow{x} \in \mathbb{R}^3$ can be located by its projections onto the directions of these vectors, which are given as the dot products of $\overrightarrow{x}$ with each of the $\overrightarrow{u}_i$. You should check that using these coordinates, we can express any vector $\overrightarrow{x} \in \mathbb{R}^3$ as a linear combination of the $\overrightarrow{u}_i$:

$$\overrightarrow{x} = (\overrightarrow{x} \cdot \overrightarrow{u}_1)\overrightarrow{u}_1 + (\overrightarrow{x} \cdot \overrightarrow{u}_2)\overrightarrow{u}_2 + (\overrightarrow{x} \cdot \overrightarrow{u}_3)\overrightarrow{u}_3.$$

Any collection $\mathcal{B} = \{\overrightarrow{v}_1, \overrightarrow{v}_2\overrightarrow{v}_3\}$ of three vectors in $\mathbb{R}^3$ with the property that each 3-vector is a linear combination of them is called a **basis** for $\mathbb{R}^3$, and the **coordinates** of $\overrightarrow{x}$ with respect to $\mathcal{B}$ are the coefficients in this combination[21]; we arrange them in a column to form the **coordinate column** of $\overrightarrow{x}$ with respect to $\mathcal{B}$:

$$_{\mathcal{B}}[\overrightarrow{x}] = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}$$

where

$$\overrightarrow{x} = \xi_1 \overrightarrow{v}_1 + \xi_2 \overrightarrow{v}_2 + \xi_3 \overrightarrow{v}_3.$$

The **standard basis** for $\mathbb{R}^3$ is $\mathcal{E} = \{\overrightarrow{\imath}, \overrightarrow{\jmath}, \overrightarrow{k}\}$, and in the coordinate column of any vector $\overrightarrow{x} = (x, y, z)$ with respect to the standard basis is the one we have been using all along:

$$_{\mathcal{E}}[\overrightarrow{x}] = [\overrightarrow{x}] = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Now, using both properties given by Proposition 3.9.2 we can use the coordinates with respect to the eigenvectors of $[Q]$ to obtain a particularly simple and informative expression for the quadratic form Q:

Corollary 3.9.3. *If*

$$\mathcal{B} = \{\overrightarrow{u}_1, \overrightarrow{u}_2, \overrightarrow{u}_3\}$$

is the basis of unit eigenvectors for the matrix representative $A = [Q]$ of a quadratic form, with respective eigenvalues λ_i, $i = 1, 2, 3$, then the value $Q(\overrightarrow{x})$ of Q at any vector can be expressed in terms of its coordinates with respect to $\mathcal{B}$ as

$$Q(\overrightarrow{x}) = \lambda_1 \xi_1^2 + \lambda_2 \xi_2^2 + \lambda_3 \xi_3^2 \tag{3.32}$$

where

$$_{\mathcal{B}}[\overrightarrow{x}] = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}$$

or in other words

$$\xi_i = \overrightarrow{u}_i \cdot \overrightarrow{x}.$$

[21] which can be shown to be unique

The expression for $Q(\overrightarrow{x})$ given by Equation (3.32) is called the **weighted squares** expression for Q.

We note in passing that the analogous statements hold for two instead of three variables. [22] The proof is a simplified version of the proofs above (Exercise 3):

Remark 3.9.4 (Principal Axis Theorem for $\mathbb{R}^2$)**.** *Suppose A is a symmetric 2×2 matrix. Then:*

1. *A has a pair of mutually perpendicular unit eigenvectors $\overrightarrow{u}_i$, $i = 1, 2$, with corresponding eigenvalues λ_i, $i = 1, 2$.*

2. *The quadratic form $Q(x, y)$ with matrix representative $[Q] = A$ has a weighted squares expression*

$$Q(\overrightarrow{x}) = \lambda_1 \xi_1^2 + \lambda_2 \xi_2^2$$

where

$$\begin{aligned} \xi_1 &= \overrightarrow{u}_1 \cdot \overrightarrow{x} \\ \xi_2 &= \overrightarrow{u}_2 \cdot \overrightarrow{x}. \end{aligned}$$

Finding Eigenvectors

How do we find the eigenvectors of a matrix A? If we knew the eigen*values*, then for each eigenvalue λ we would need to simply solve the system of linear equations

$$A\overrightarrow{u} = \lambda\overrightarrow{u}$$

for $\overrightarrow{u}$. To find the eigenvalues, we use our observations about singularity and determinants in Appendix E. We rewrite the eigenvector equation in the form

$$A\overrightarrow{u} - \lambda\overrightarrow{u} = \overrightarrow{0}$$

or, using the distributive law for matrix multiplication,

$$(A - \lambda I)\overrightarrow{u} = \overrightarrow{0}. \tag{3.33}$$

You should verify that the matrix $A - \lambda I$ is obtained from A by subtracting λ from each diagonal entry, and leaving the rest alone. Now $\overrightarrow{u}$ is by assumption a *nonzero* vector, and this means that the matrix $A - \lambda I$ is *singular*

[22] Indeed, there is a version for an arbitrary finite number of variables.

(since it sends a nonzero vector to the zero vector). From Appendix E, we see that this forces

$$\det(A - \lambda I) = 0. \tag{3.34}$$

Now given a 3×3 matrix A, Equation (3.34) is an equation in the unknown λ; you should verify (Exercise 4) that the left side of this equation is a polynomial of degree three in λ; it is called the **characteristic polynomial** of A, and Equation (3.34) is called the **characteristic equation** of A. As a corollary of Proposition 3.9.2 we have

Remark 3.9.5. *Every eigenvalue of A is a zero of the characteristic polynomial*

$$p(\lambda) = \det(A - \lambda I).$$

If A is symmetric and 3×3, then $p(\lambda)$ is a cubic polynomial with three real zeroes, and these are the eigenvalues of A.

Thus, we find the eigen*values* of A first, by finding the zeroes of the characteristic polynomial (*i.e.*, solving the characteristic equation (3.34)); then for each eigenvalue λ we solve the system of equations (3.33) to find the corresponding eigen*vectors*.

Let us apply this to two examples.

The quadratic form

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

has matrix representative

$$A = [Q] = \left| \begin{array}{ccc} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{array} \right]$$

with characteristic polynomial

$$\begin{aligned} p(\lambda) &= \det \left| \begin{array}{ccc} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{array} \right] \\ &= (-\lambda)\det \left| \begin{array}{cc} -\lambda & 1 \\ 1 & -\lambda \end{array} \right] - (1)\det \left| \begin{array}{cc} 1 & 1 \\ 1 & -\lambda \end{array} \right] + (1)\det \left| \begin{array}{cc} 1 & -\lambda \\ 1 & 1 \end{array} \right] \\ &= -\lambda(\lambda^2 - 1) - (-\lambda - 1) + (1 + \lambda) \\ &= (\lambda + 1)\{-\lambda(\lambda - 1) + 1 + 1\} \\ &= (\lambda + 1)\{-\lambda^2 + \lambda + 2\} \\ &= -(\lambda + 1)^2(\lambda - 2). \end{aligned}$$

So the eigenvalues are

$$\lambda_1 = 2, \quad \lambda_2 = -1.$$

The eigenvectors for $\lambda_1 = 2$ satisfy

$$\begin{array}{rcl} v_2 + v_3 &=& 2v_1 \\ v_1 \qquad + v_3 &=& 2v_2 \\ v_1 + v_2 \qquad &=& 2v_3 \end{array}$$

from which we conclude that $v_1 = v_2 = v_3$: so

$$\overrightarrow{u}_1 = \frac{1}{\sqrt{3}}(1,1,1).$$

For $\lambda_2 = -1$, we need

$$\begin{array}{rcl} v_2 + v_3 &=& -v_1 \\ v_1 \qquad + v_3 &=& -v_2 \\ v_1 + v_2 \qquad &=& -v_3 \end{array}$$

which amounts to $v_1 + v_2 + v_3 = 0$. This defines a plane of solutions. If we set $v_3 = 0$, we get

$$\overrightarrow{u}_2 = \frac{1}{\sqrt{2}}(1,-1,0).$$

This is automatically perpendicular to $\overrightarrow{u}_1$ (but check that it is!). We need a third eigenvector perpendicular to both $\overrightarrow{u}_1$ and $\overrightarrow{u}_2$. We can take their cross-product, which is

$$\begin{aligned} \overrightarrow{u}_3 &= \overrightarrow{u}_1 \times \overrightarrow{u}_2 \\ &= \frac{1}{\sqrt{6}} \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ 1 & 1 & 1 \\ 1 & -1 & 0 \end{vmatrix} \\ &= \frac{1}{\sqrt{6}}(\overrightarrow{\imath} + \overrightarrow{\jmath} - 2\overrightarrow{k}) \\ &= \frac{1}{\sqrt{6}}(1,1,-2). \end{aligned}$$

You should check that $\overrightarrow{u}_3$ is an eigenvector with $\lambda_3 = \lambda_2 = -1$.

The weighted-sums expression for Q, then, is

$$\begin{aligned} Q(x,y,z) &= 2\left(\frac{x+y+z}{\sqrt{3}}\right)^2 - \left(\frac{x-y}{\sqrt{2}}\right)^2 - \left(\frac{x+y-2z}{\sqrt{6}}\right)^2 \\ &= \frac{2}{3}(x+y+z)^2 - \frac{1}{2}(x-y)^2 - \frac{1}{6}(x+y-2z)^2. \end{aligned}$$

As another example the form

$$Q(x,y,z) = 4x^2 - y^2 - z^2 - 4xy + 4xz - 6yz$$

has matrix representative

$$[Q] = \left| \begin{matrix} 4 & -2 & 2 \\ -2 & -1 & -3 \\ 2 & -3 & -1 \end{matrix} \right]$$

with characteristic polynomial

$$\begin{aligned}
p(\lambda) &= \det \left| \begin{matrix} 4-\lambda & -2 & 2 \\ -2 & -1-\lambda & -3 \\ 2 & -3 & -1-\lambda \end{matrix} \right] \\
&= (4-\lambda)\det \left| \begin{matrix} -1-\lambda & -3 \\ -3 & -1-\lambda \end{matrix} \right] - (-2)\det \left| \begin{matrix} -2 & -3 \\ 2 & -1-\lambda \end{matrix} \right] \\
&\quad + (2)\det \left| \begin{matrix} -2-\lambda & -1-\lambda \\ 2 & -3 \end{matrix} \right] \\
&= (4-\lambda)\{(\lambda+1)^2 - 9\} + 2\{2(1+\lambda)+6\} + 2\{6 + 2(1+\lambda)\} \\
&= (4-\lambda)\{(\lambda+4)(\lambda-2)\} + 2\{2\lambda+8\} + 2\{2\lambda+8\} \\
&= (4-\lambda)(\lambda+4)(\lambda-2) + 8(\lambda+4) \\
&= (\lambda+4)\{(4-\lambda)(\lambda-2)+8\} \\
&= (\lambda+4)\{-\lambda^2+6\lambda\} \\
&= -\lambda(\lambda+4)(\lambda-6).
\end{aligned}$$

The eigenvalues of $[Q]$ are

$$\lambda_1 = 0, \quad \lambda_2 = -4, \quad \lambda_3 = 6.$$

To find the eigenvectors for $\lambda = 0$, we need

$$\begin{aligned}
4v_1 - 2v_2 + 2v_3 &= 0 \\
-2v_1 - v_2 - 3v_3 &= 0 \\
2v_1 - 3v_2 - v_3 &= 0.
\end{aligned}$$

The sum of the fist and second equations is the third, so we drop the last equation; dividing the first by 2, we have

$$\begin{aligned}
2v_1 - v_2 + v_3 &= 0 \\
-2v_1 - v_2 - 3v_3 &= 0.
\end{aligned}$$

The first of these two equations gives

$$v_2 = 2v_1 + v_3$$

and substituting this into the second gives

$$-4v_1 - 4v_2 = 0$$

so

$$v_1 + v_3 = 0$$

or

$$v_3 = -v_1$$

and then

$$\begin{aligned} v_2 &= 2v_1 + v_3 \\ &= 2v_1 - v_1 \\ &= v_1. \end{aligned}$$

Setting $v_1 = 1$ leads to

$$\overrightarrow{v} = (1, 1, -1)$$

and the unit eigenvector

$$\overrightarrow{u}_1 = \frac{1}{\sqrt{3}}(1, 1, -1).$$

The eigenvectors for $\lambda = -4$ must satisfy

$$\begin{aligned} 4v_1 - 2v_2 + 2v_3 &= -4v_1 \\ -2v_1 \quad - v_2 - 3v_3 &= -4v_2 \\ 2v_1 - 3v_2 \quad - v_3 &= -4v_3 \end{aligned}$$

or

$$\begin{aligned} 8v_1 - 2v_2 + 2v_3 &= 0 \\ -2v_1 + 3v_2 - 3v_3 &= 0 \\ 2v_1 - 3v_2 + 3v_3 &= 0. \end{aligned}$$

The second and third equations are (essentially) the same; the first divided by 2 is

$$4v_1 - v_2 + v_3 = 0$$

so

$$v_2 = 4v_1 + v_3.$$

Substituting this into the second equation we have

$$-2v_1 + 3(4v_1 + v_3) - 3v_3 = 0$$

or

$$10v_1 = 0.$$

Thus $v_1 = 0$ and $v_2 = v_3$. We find

$$\vec{u}_2 = \frac{1}{\sqrt{2}}(0, 1, 1).$$

Finally, $\lambda = 6$ leads to

$$\begin{aligned} 4v_1 - 2v_2 + 2v_3 &= 6v_1 \\ -2v_1 - v_2 - 3v_3 &= 6v_2 \\ 2v_1 - 3v_2 - v_3 &= 6v_3 \end{aligned}$$

or

$$\begin{aligned} -2v_1 - 2v_2 + 2v_3 &= 0 \\ -2v_1 - 7v_2 - 3v_3 &= 0 \\ 2v_1 - 3v_2 - 7v_3 &= 0. \end{aligned}$$

The first equation says

$$v_1 = -v_2 + v_3$$

and substituting this into the other two yields two copies of

$$-5v_2 - 5v_3 = 0$$

or

$$v_2 = -v_3$$

so

$$v_1 = 2v_3$$

and

$$\vec{u}_3 = \frac{1}{\sqrt{6}}(2, -1, 1).$$

The weighted squares expression for Q is

$$\begin{aligned} Q(x,y,z) &= 0\left(\frac{x+y+z}{\sqrt{3}}\right)^2 - 4\left(\frac{y+z}{\sqrt{2}}\right)^2 + 6\left(\frac{2x-y+z}{\sqrt{6}}\right)^2 \\ &= -2(y+z)^2 + (2x-y+z)^2. \end{aligned}$$

The Determinant Test for Three Variables

The weighted squares expression shows that the character of a quadratic form as positive or negative definite (or neither) can be determined from the signs of the eigenvalues of its matrix representative: *Q is positive (resp. negative) definite precisely if all three eigenvalues of $A = [Q]$ are strictly positive (resp. negative).* We would like to see how this can be decided using determinants, without solving the characteristic equation. We will at first concentrate on deciding whether a form is positive definite, and then at the end we shall see how to modify the test to decide when it is negative definite.

The first step is to analyze what the determinant of $A = [Q]$ tells us. Recall from § 1.7 that a 3×3 determinant can be interpreted as the signed volume of the parallelipiped formed by its rows. In our case, since A is symmetric, the rows are the same as the columns, and it is easy to check (see Appendix E) that the columns of any 3×3 matrix A are the positions of the standard basis vectors after they have been multiplied by A. The parallelipiped formed by the standard basis vectors is simply a unit cube, and so we can interpret the determinant of A as the signed volume of the parallelipiped created by the action of A on the unit cube with sides parallel to the coordinate planes. If we slice along the directions of the standard basis vectors to calculate volumes,, we see that we can interpret the determinant of A more broadly as the factor by which *any* signed volume is multiplied when we apply A—that is, for any region $E \subset \mathbb{R}^3$, the signed volume $\vec{\mathcal{V}}(E)$ of E and that of its image AE under multiplication by A are related by

$$\vec{\mathcal{V}}(AE) = \det A \vec{\mathcal{V}}(E)\,.$$

Now let us apply this to the parallelipiped E formed by the three eigenvectors of A: since they are orthonormal, E is a unit cube (but its sides need not be parallel to the coordinate planes), and (changing the numbering of the eigenvectors if necessary) we can assume that it is *positively* oriented, so the signed volume of this cube is $\overrightarrow{\mathcal{V}}(E) = 1$. It follows that $\overrightarrow{\mathcal{V}}(AE) = \det A$. But the sides of AE are the vectors $\lambda_i \overrightarrow{u}_i$, so this is again a cube, whose signed volume is *the product of the eigenvalues*: $\overrightarrow{\mathcal{V}}(AE) = \lambda_1\lambda_2\lambda_3$. Thus

$$\det A = \lambda_1\lambda_2\lambda_3$$

and we have the following observation:

> If Q is positive-definite, then $\Delta_3 = \det A > 0$.

Of course, the *product* of the eigenvalues (*i.e.*, $\det A$) can *also* be positive if we have *one* positive and *two* negative eigenvalues, so we need to know more to determine whether or not Q is positive-definite.

If Q is positive-definite, we know that its restriction to any plane in $\mathbb{R}^3$ is also positive-definite. In particular, we can consider its restriction to the xy-plane, that is, to all vectors of the form $\overrightarrow{x} = (x, y, 0)$. It is easy to check that for any such vector,

$$\begin{aligned} Q(\overrightarrow{x}) &= \overrightarrow{x}^T A \overrightarrow{x} \\ &= \begin{bmatrix} x & y & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} x & y & 0 \end{bmatrix} \begin{bmatrix} a_{11}x + a_{12}y \\ a_{21}x + a_{22}y \\ a_{31}x + a_{32}y \end{bmatrix} \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \end{aligned}$$

This shows that the restriction of Q to the xy-plane can be regarded as the quadratic form in two variables whose matrix representative is obtained from A by deleting the last row and last column—that is, the upper-left 2×2 minor submatrix. But for a quadratic form in two variables, we know that it is positive-definite precisely if the determinant of its matrix representative as well as its upper-left entry are both positive. Thus if we set Δ_2 to be the upper left (2×2) minor of $\det A$ and Δ_1 to be the upper-left entry, we have the necessity of the conditions in the following:

Proposition 3.9.6 (Determinant Test for Positive Definite Forms in $\mathbb{R}^3$). *The quadratic form Q on $\mathbb{R}^3$ is positive-definite if and only if its matrix representative*

$$[Q] = A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

satisfies

$$\Delta_3 > 0, \quad \Delta_2 > 0, \text{ and } \Delta_1 > 0$$

where

$$\Delta_3 = \det A = \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\Delta_2 = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and

$$\Delta_1 = a_{11}.$$

Proof. We have seen that the conditions are necessary. To see that they are sufficient, suppose all three determinants are positive. Then we know that the eigenvalues of A satisfy

$$\lambda_1\lambda_2\lambda_3 > 0.$$

Assuming $\lambda_1 \geq \lambda_2 \geq \lambda_3$, this means $\lambda_1 > 0$ and the other two eigenvalues are either both positive or both negative. Suppose they were both *negative*: then the restriction of Q to the plane $\vec{u}_1^{\perp}$ containing $\vec{u}_2$ and $\vec{u}_3$ would be *negative* definite. Now, this plane intersects the xy-plane in (at least) a line, so the restriction of Q to the xy-plane couldn't possibly be *positive* definite, contradicting the fact that $\Delta_1 > 0$ and $\Delta_2 > 0$. Thus λ_2 and λ_3 are both *positive*, and hence Q is positive definite on all of $\mathbb{R}^3$. □

What about deciding if Q is *negative* definite? The easiest way to get at this is to note that Q is *negative* definite precisely if its negative $(\bar{Q})(\vec{x}) := -Q(\vec{x})$ is *positive* definite, and that $\left[\bar{Q}\right] = -[Q]$. Now, the determinant of a $k \times k$ matrix M is related to the determinant of its negative by

$$\det(-M) = (-1)^k \det M$$

so we see that for $k = 1, 2, 3$

$$\Delta_k(\bar{Q}) = (-1)^k \Delta_k(Q)$$

from which we easily get the following test for a quadratic form in three variables to be *negative* definite:

Corollary 3.9.7 (Determinant Test for Negative Definite Forms in $\mathbb{R}^3$)**.** *The quadratic form Q in three variables is negative definite precisely if its matrix representative $A = [Q]$ satisfies*

$$(-1)^k \Delta_k > 0 \text{ for } k = 1, 2, 3$$

where Δ_k are the determinants given in Proposition 3.9.6.

Let us see how this test works on the examples studied in detail earlier in this section.

The form

$$Q(x, y, z) = x^2 - y^2 - z^2$$

has matrix representative

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

and it is easy to calculate that

$$\begin{aligned} \Delta_1 &= \det [Q] \\ &= (1)(-1)(-1) \\ &= 1 > 0 \end{aligned}$$

...which, so far, tells us that the form is not *negative* definite...

$$\begin{aligned} \Delta_2 &= (1)(-1) \\ &= -1 < 0 \end{aligned}$$

so that Q is also not *positive* definite. There is no further information to be gained from calculating

$$\Delta_1 = 1.$$

The form

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

has matrix representative

$$[Q] = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

with determinant

$$\begin{aligned} \Delta_1 &= \det{[Q]} \\ &= 0 - (1)(0-1) + 1(1-0) \\ &= 2 > 0 \end{aligned}$$

which again rules out the possibility that the form is not *negative* definite,

$$\begin{aligned} \Delta_2 &= (0)(0) - (1)(1) \\ &= -1 < 0 \end{aligned}$$

so that Q is also not *positive* definite. For completeness, we also note that

$$\Delta_1 = 0.$$

Finaly, the form

$$Q(x,y,z) = 4x^2 - y^2 - z^2 - 4xy + 4xz - 6yz$$

has matrix representative

$$[Q] = \begin{pmatrix} 4 & -2 & 2 \\ -2 & -1 & -3 \\ 2 & -3 & -1 \end{pmatrix}$$

with determinant

$$\begin{aligned} \Delta_1 &= \det{[Q]} \\ &= 4[(-1)(-1) - (-3)(-3)] - (-2)[(-2)(-1) - (2)(-3)] + (2)[(-2)(-3) - (2)(-1)] \\ &= 4[1-9] + 2[2+6] + 2[6+2] \\ &= -32 + 16 + 16 \\ &= 0. \end{aligned}$$

This already guarantees that Q is not definite (neither positive nor negative definite). In fact, this says that the product of the eigenvalues is zero, which forces at least one of the eigenvalues to be zero, something we saw earlier in a more direct way.

None of these three forms is definite. As a final example, we consider the form

$$Q(x,y,z) = 2xy + 8xz + 4yz - 3x^2 - 3y^2 - 10z^2$$

with matrix representative

$$[Q] = \begin{pmatrix} -3 & 1 & 4 \\ 1 & -3 & 2 \\ 4 & 2 & -10 \end{pmatrix}.$$

The determinant of this matrix is

$$\begin{aligned} \Delta_3 &= \det [Q] \\ &= (-3)[(-3)(-10) - (2)(2)] - (1)[(1)(-10) - (4)(2)] + (4)[(1)(2) - (4)(-3)] \\ &= (-3)[26] - [-18] + (4)[2 + 12] \\ &= -78 + 18 + 56 \\ &= -4 < 0 \end{aligned}$$

so the form is not *positive* definite;

$$\begin{aligned} \Delta_2 &= (-3)(-3) - (1)(1) \\ &= 8 > 0 \end{aligned}$$

which is still consistent with being negative definite, and finally

$$\Delta_1 = -3 < 0;$$

we see that Q satisfied the conditions of Corollary 3.9.7, and so it is *negative definite*. Note that the characteristic polynomial of $[Q]$ is

$$\det \begin{pmatrix} -3-\lambda & 1 & 4 \\ 1 & -3-\lambda & 2 \\ 4 & 2 & -10-\lambda \end{pmatrix} = -(\lambda^3 + 16\lambda^2 + 48\lambda + 4)$$

which has no obvious factorization (in fact, it has no integer zeroes). Thus we can determine that the form is negative definite far more easily than we can calculate its weighted squares expression.

Combining the analysis in Proposition 3.9.6 and Corollary 3.9.7 with Proposition 3.8.3 and Lemma 3.8.4, we can get the three-variable analogue of the Second Derivative Test which we obtained for two variables in Theorem 3.8.6:

Theorem 3.9.8 (Second Derivative Test, Three Variables)**.** *Suppose the $\mathcal{C}^2$ function $f(\overrightarrow{x}) = f(x, y, z)$ has a critical point at $\overrightarrow{x} = \overrightarrow{a}$. Consider the following three quantities:*

$$\begin{aligned}\Delta_1 &= f_{xx}(\overrightarrow{a})\\ \Delta_2 &= f_{xx}(\overrightarrow{a})\, f_{yy}(\overrightarrow{a}) - f_{xy}(\overrightarrow{a})^2\\ \Delta_3 &= \det Hf(\overrightarrow{a}).\end{aligned}$$

1. *If $\Delta_k > 0$ for $k = 1, 2, 3$, then $f(\overrightarrow{x})$ has a local minimum at $\overrightarrow{x} = \overrightarrow{a}$.*

2. *If $(-1)^k \Delta_k > 0$ for $k = 1, 2, 3$ (i.e., , $\Delta_2 > 0$ while $\Delta_1 < 0$ and $\Delta_3 < 0$), then $f(\overrightarrow{x})$ has a local maximum at $\overrightarrow{x} = \overrightarrow{a}$.*

3. *If all three quantities are nonzero but neither of the preceding conditions holds, then $f(\overrightarrow{x})$ does* not *have a local extremum at $\overrightarrow{x} = \overrightarrow{a}$.*

A word of warning: *when one of these quantities equals zero, this test gives no information.*

As an example of the use of Theorem 3.9.8, consider the function

$$f(x, y, z) = 2x^2 + 2y^2 + 2z^2 + 2xy + 2xz + 2yz - 6x + 2y + 4z;$$

its partial derivatives are

$$\begin{aligned}f_x(x, y, z) &= 4x + 2y + 2z - 6\\ f_y(x, y, z) &= 4y + 2x + 2z + 2\\ f_z(x, y, z) &= 4z + 2y + 2x + 4.\end{aligned}$$

Setting all of these equal to zero, we have the system of equations

$$\left\{\begin{array}{cccrr} 4x & +2y & +2z & = & 6\\ 2x & +4y & +2z & = & -2\\ 2x & +2y & +4z & = & -4\end{array}\right.$$

whose only solution is

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

The second partials are

$$\begin{array}{lll} f_{xx} = 4 & f_{xy} = 2 & f_{xz} = 2 \\ & f_{yy} = 4 & f_{yz} = 2 \\ & & f_{zz} = 2 \end{array}$$

so

$$\begin{aligned} \Delta_3 &= \det \begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 2 \end{pmatrix} \\ &= 4(8-4) - 2(4-4) + 2(4-8) \\ &= 16 - 0 - 8 \\ &= 8 > 0 \\ \Delta_2 &= (4)(4) - (2)(2) \\ &= 12 \\ \Delta_1 &= 4 > 0 \end{aligned}$$

so the Hessian is *positive definite*, and f has a *local minimum* at $(3, -1, -2)$.

Exercises for § 3.9

Practice problems:

1. For each quadratic form Q below, (i) write down its matrix representative $[Q]$; (ii) find all eigenvalues of $[Q]$; (iii) find corresponding unit eigenvectors; (iv) write down the weighted squares representative of Q.

   (a) $Q(x, y) = 17x^2 + 12xy + 8y^2$

   (b) $Q(x, y) = 11x^2 + 6xy + 19y^2$

   (c) $Q(x, y) = 3x^2 + 4xy$

   (d) $Q(x, y) = 19x^2 + 24xy + y^2$

   (e) $Q(x, y, z) = 6x^2 - 4xy + 6y^2 + z^2$

   (f) $Q(x, y, z) = 2x^2 - 2y^2 + 2z^2 + 8xy + 8xz$

2. For each function below, find all critical points and classify each as local minimum, local maximum, or neither.

   (a) $f(x, y, z) = 5x^2 + 3y^2 + z^2 - 2xy + 2yz - 6x - 8y - 2z$

(b) $f(x, y, z) = x^2 + y^2 + z^2 + xy + yz + xz - 2x$

(c) $f(x, y, z) = x^2 + y^2 + z^2 + xy + yz + xz - 3y - z$

(d) $f(x, y, z) = x^2 + y^2 + z^2 + xy + yz + xz - 2x - 3y - z$

(e) $f(x, y, z) = 2x^2 + 5y^2 - 6xy + 2xz - 4yz - 2x - 2z$

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

(g) $f(x, y) = x^3 + 2x^2 - 12x + y^2 + z^2 - 2xy - 2xz$

Theory problems:

3. (a) Adapt the proof of Proposition 3.9.2 to show that if

$$M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

is a symmetric 2×2 matrix, then there exist two unit vectors $\overrightarrow{u}_1$ and $\overrightarrow{u}_2$ and two scalars λ_1 and λ_2 satisfying

$$M\overrightarrow{u}_i = \lambda_i \overrightarrow{u}_i \quad \text{for } i = 1, 2.$$

(b) Show that if $\overrightarrow{u}_i$, $i = 1, 2, 3$ are orthonormal vectors, then an arbitrary vector $\overrightarrow{v} \in \mathbb{R}^3$ can be expressed as

$$\overrightarrow{x} = \xi_1 \overrightarrow{u}_1 + \xi_2 \overrightarrow{u}_2$$

where

$$\xi_i = \overrightarrow{x} \cdot \overrightarrow{u}_i \quad \text{for } i = 1, 2.$$

(c) Show that if $M = [Q]$ and $\overrightarrow{x} = (x, y)$ then Q has the weighted squares decomposition

$$Q(x, y) = \lambda_1 \xi_1^2 + \lambda_2 \xi_2^2.$$

4. Let

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

be any 2×2 matrix.

(a) Show that the characteristic polynomial

$$\det(A - \lambda I)$$

is a polynomial of degree 2 in the variable λ.

(b) Show that if A is 3×3, the same polynomial is of degree 3.

3.10 Quadratic Curves and Surfaces

In this section, we will use the Principal Axis Theorem to classify the curves (*resp.* surfaces) which arise as the locus of an equation of degree two in two (*resp.* three) variables.

Quadratic Curves

The general quadratic equation in x and y is

$$Ax^2 + Bxy + Cy^2 + Dx + Ey = F. \tag{3.35}$$

If all three of the leading terms vanish, then this is the equation of a line. We will assume henceforth that at least one of A, B and C is nonzero.

In § 2.1 we identified a number of equations of this form as "model equations" for the conic sections:

Parabolas: the equations

$$y = ax^2 \tag{3.36}$$

and its sister

$$x = ay^2 \tag{3.37}$$

are model equations for a parabola with vertex at the origin, focus on the y-axis (*resp.* x-axis) and horizontal (*resp.* vertical) directrix. These correspond to Equation (3.35) with $B = 0$, exactly one of A and C nonzero, and the linear (degree one) term corresponding to the "other" variable nonzero; you should check that moving the vertex from the origin results from allowing both D and E, and/or F, to be nonzero.

Ellipses and Circles: the model equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{3.38}$$

for a circle (if $a = b$) or an ellipse with axes parallel to the coordinate axes and center at the origin corresponds to $B = 0$, A, C and F of the same sign, and $D = E = 0$. Again you should check that moving the vertex results in the introduction of nonzero values for D and/or E and simultaneously raises the absolute value of F. However, when the linear terms are present, one needs to complete the square(s) to determine whether the given equation came from one of the type above, or one with zero or negative right-hand side.

Hyperbolas: the model equations

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \pm 1 \tag{3.39}$$

for a hyperbola centered at the origin and symmetry about both coordinate axes corresponds to $B = 0$, A and C of *opposite* signs, $F \neq 0$, and $D = 0 = E$. When $F = 0$ but the other conditions remain, we have the equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0 \tag{3.40}$$

which determines a pair of lines, the asymptotes of the hyperbolas with $F \neq 0$. As before, moving the center introduces linear terms, but completing the square is needed to decide whether an equation with either D or E (or both) nonzero corresponds to a hyperbola or a pair of asymptotes.

In effect, the list above (with some obvious additional degenerate cases) takes care of all versions of Equation (3.35) in which $B = 0$. Unfortunately, when $B \neq 0$ there is no quick and easy way to determine which, if any, of the conic sections is the locus. However, if it is, it must arise from rotation of one of the model versions above.

We will see that the locus of every instance of Equation (3.35) with not all leading terms zero has a locus fitting one of these descriptions (with different centers, foci and directrices), or a degenerate locus (line, point or empty set). To this end, we shall start from Equation (3.35) and show that in an appropriate coordinate system the equation fits one of the molds above.

Let us denote the polynomial on the left side of Equation (3.35) by $p(x, y)$:

$$p(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey.$$

Assuming they don't all vanish, the leading terms define a quadratic form

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

with matrix representative

$$\mathbf{A} = [Q] = \begin{bmatrix} A & B/2 \\ B/2 & C \end{bmatrix}.$$

By Remark 3.9.4, there is an orthonormal basis for $\mathbb{R}^2$ consisting of two unit eigenvectors $\overrightarrow{u}_1$ and $\overrightarrow{u}_2$ (with eigenvalues λ_1, λ_2) for $\mathbf{A}$. Note that the negative of an eigenvector is also an eigenvector, so we can assume that $\overrightarrow{u}_2$ is the result of rotating $\overrightarrow{u}_1$ counterclockwise by a right angle. Thus we can write

$$\begin{aligned} \overrightarrow{u}_1 &= (c, s) \\ \overrightarrow{u}_2 &= (-s, c) \end{aligned}$$

where

$$\begin{aligned} c &= \cos\theta \\ s &= \sin\theta \end{aligned}$$

(θ is the angle between $\overrightarrow{u}_1$ and the positive x-axis). These vectors define a rectangular coordinate system (with coordinate axes rotated counterclockwise from the standard axes) in which the point with standard coordinates (x, y) has coordinates in the new system

$$\begin{array}{lclcl} \xi_1 & = \overrightarrow{u}_1 \cdot \overrightarrow{x} & = & cx + sy \\ \xi_2 & = \overrightarrow{u}_2 \cdot \overrightarrow{x} & = & -sx + cy \end{array}.$$

You should check that these equations can by solved for x and y in terms of ξ_1 and ξ_2:

$$\begin{aligned} x &= c\xi_1 - s\xi_2 \\ y &= s\xi_1 + c\xi_2 \end{aligned}$$

so that $p(x, y)$ can be rewritten as

$$\begin{aligned} p(x, y) &= Q(x, y) + Dx + Ey \\ &= \lambda_1\xi_1^2 + \lambda_2^2\xi_2^2 + \alpha\xi_1 + \beta\xi_2 \end{aligned}$$

where

$$\begin{aligned} \alpha &= cD + sE \\ \beta &= -sD + cE. \end{aligned}$$

To finish our analysis, we distinguish two cases. By assumption, at least one of the eigenvalues is nonzero. For notational convenience, assume (renumbering if necessary) that $|\lambda_1| \geq |\lambda_2|$.

If only one of the eigenvalues is nonzero, then $\lambda_2 = 0$; we can complete the square in the terms involving ξ_1 to write the equation in the form

$$\lambda_1 \left(\xi_1 + \frac{\alpha}{2\lambda_1} \right)^2 + \beta\xi_2 = F + \frac{\alpha^2}{4\lambda_1};$$

The locus of this is a parabola as in Equation (3.36), but in the new coordinate system, displaced so the vertex is at $\xi_1 = -\alpha/2\lambda_1, \xi_2 = (4\lambda_1 F + \alpha^2)/4\lambda_1$.

If both eigenvalues are nonzero, then we complete the square in the terms involving ξ_2 as well as in those involving ξ_1 to obtain

$$\lambda_1 \left(\xi_1 + \frac{\alpha}{2\lambda_1} \right)^2 + \lambda_2 \left(\xi_2 + \frac{\beta}{2\lambda_2} \right)^2 = F + \frac{\alpha^2}{4\lambda_1} + \frac{\beta^2}{4\lambda_2}.$$

This is Equation (3.38), Equation (3.39), or Equation (3.40), with x (*resp.* y) replaced by $\xi_1 + \frac{\alpha}{2\lambda_1}$ (*resp.* $\xi_2 + \frac{\beta}{2\lambda_2}$), and so its locus is one of the other loci described above, in the new coordinate system, displaced so the origin moves to $\xi_1 = -\alpha/2\lambda_1$, $\xi_2 = -\beta/2\lambda_2$.

Quadric Surfaces

The most general equation of degree two in x, y and z consists of three "square" terms, three "mixed product" turns, three degree one terms (multiples of a single variable), and a constant term. A procedure similar to the one we used for two variables can be applied here: combining the six quadratic terms (the three squares and the three mixed products) into a quadratic form $Q(x,y,z)$, we can express the general quadratic equation in three variables as

$$Q(x,y,z) + Ax + By + Cz = D. \tag{3.41}$$

Using the Principal Axis Theorem (Proposition 3.9.2) we can create a new coordinate system, a rotation of the standard one, in which the quadratic form can be written

$$Q(\overrightarrow{x}) = \lambda_1\xi_1^2 + \lambda_2\xi_2^2 + \lambda_3\xi_3^2 \tag{3.42}$$

where λ_i, $i = 1,2,3$ are the eigenvalues of $[Q]$, with corresponding unit eigenvectors $\overrightarrow{u}_i$, and $\xi_i = \overrightarrow{u}_i \cdot \overrightarrow{x}$ are the coordinates of $\overrightarrow{x}$ with respect to our rotated system. We can also solve the equations defining these coordinates for the standard coordinates x_i in terms of the rotated ones ξ_i, and substitute these expressions in to the linear terms, to rewrite Equation (3.41) as

$$\lambda_1\xi_1^2 + \lambda_2\xi_2^2 + \lambda_3\xi_3^2 + \alpha_1\xi_1 + \alpha_2\xi_2 + \alpha_3\xi_3 = D;$$

by completing the square in each variable ξ_i for which $\lambda_i \neq 0$ we get an equation in which each variable appears either in the form $\lambda_i(\xi_i - \xi_{i0})^2$ (if $\lambda_i \neq 0$) or $\alpha_i(\xi_i - \xi_{i0})$ (if $\lambda_i = 0$). We shall not attempt an exhaustive catalogue of the possible cases, but will consider five "model equations" which cover all the important possibilities. In all of these, we will assume that $\xi_{10} = \xi_{20} = \xi_{30} = 0$ (which amounts to displacing the origin); in many cases we will also assume that the coefficient of each term is ± 1. The latter amounts to changing the scale of each coordinate, but not the general shape-classification of the surface.

1. The easiest scenario to analyze is when *z appears only to the first power*: we can then move everything except the "z" term to the right side of the equation, and divide by the coefficient of z, to write our equation as the expression for the graph of a function of x and y

   $$z = f(x, y)\,.$$

   In this scenario, the intersection of our surface with the horizontal plane $z = k$ is just the level set $\mathcal{L}(f, k)$ consisting of those points at which $f(x, y)$ takes the value k. For a quadratic equation, $f(x, y)$ takes one of three forms, corresponding to the three kinds of conic sections:

   (a) If another variable, say y, also appears to the first power, $f(x, y)$ has the form $ax^2 + y$, so our our level set is given by $ax^2 + by = k$, which defines a parabola. In Figure 3.12 we sketch the surface given by the "model equation"

   $$x^2 - y + z = 0 \tag{3.43}$$

   which corresponds to $a = b = -1$; the level sets are the parabolas $y = x^2 + k$. Note that these are all horizontal copies of the "standard" parabola $y = x^2$, but with their vertices lying along the line $z = y$ in the yz-plane.

   (b) If both x and y appear squared, and their coefficients have the *same* sign, then our level sets are ellipses (or circles) centered at

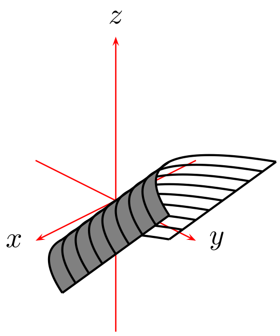


Figure 3.12: $x^2 - y + z = 0$

the origin, with axes (or radius) depending on z. For example, the surface given by the "model equation"

$$4x^2 + y^2 - z = 0 \tag{3.44}$$

intersects each horizontal plane $z = k$ with $k > 0$ in an ellipse of the form

$$x^2 + \frac{y^2}{4} = \frac{k}{4};$$

it is immediate that the major axis is parallel to the y-axis, while the minor axis is half as long, and is parallel to the x-axis (Figure 3.13).

To see how these ellipses fit together, we consider the intersection of the surface with the two vertical coordinate planes. The intersection with the xz-plane ($y = 0$) has equation

$$z = 4x^2$$

which is a parabola opening up from the origin, while the intersection with the yz-plane ($x = 0$) has equation

$$z = y^2$$

which also opens up from the origin, but is twice as "broad"(see Figure 3.14).

Fitting these pictures together, we see that the surface is a kind of bowl shape, known as an **elliptic paraboloid** (Figure 3.15).

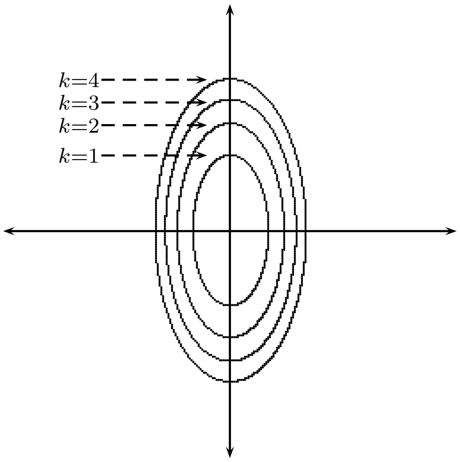


Figure 3.13: Level sets: $x^2 + \frac{y^2}{4} = \frac{k}{4}$

(c) If both x and y appear squared, and their coefficients have *opposite* signs, then our level sets are hyperbolas centered at the origin. For example, the surface given by the "model equation"

$$x^2 - y^2 = z \tag{3.45}$$

intersects the horizontal plane $z = k$ in a hyperbola opening in the direction of the x-axis (*resp.* y-axis) for $z > 0$ (*resp.* $z < 0$), and in the common asymptotes of all these hyperbolas when $z = 0$ (Figure 3.16).

To see how these fit together, we note that the vertices of the two branches lie on the curve $z = x^2$ in the xz-plane for $z > 0$ and on the curve $z = -y^2$ in the yz-plane for $z < 0$ (Figure 3.17).

The official name of this surface is a **hyperbolic paraboloid**, but it is colloquially referred to as a **saddle surface** (Figure 3.18).

2. If the form $Q(x, y, z)$ is definite, then all of the eigenvalues λ_i have the same sign, and we can model the locus (up to rotation and displacement of the origin) by

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

if $k = 0$ this is just the origin, and if $k < 0$ this gives an empty locus; if $k > 0$, then we can divide by k and modify the divisors on the left

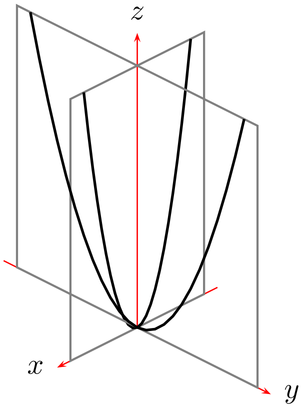


Figure 3.14: Intersection with vertical planes

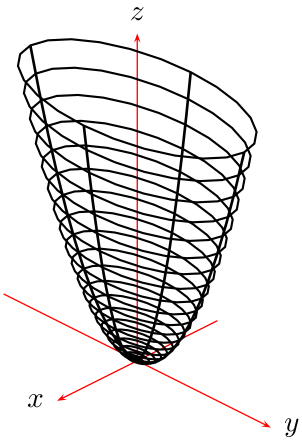


Figure 3.15: Elliptic Paraboloid $4x^2 + y^2 - z = 0$

to get an equation of the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{3.46}$$

We study the locus of this equation by **slicing**: that is, by looking at how it intersects various planes parallel to the coordinate planes. This is an elaboration of the idea of looking at level curves of a function.

The xy-plane ($z = 0$) intersects the surface in the ellipse

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

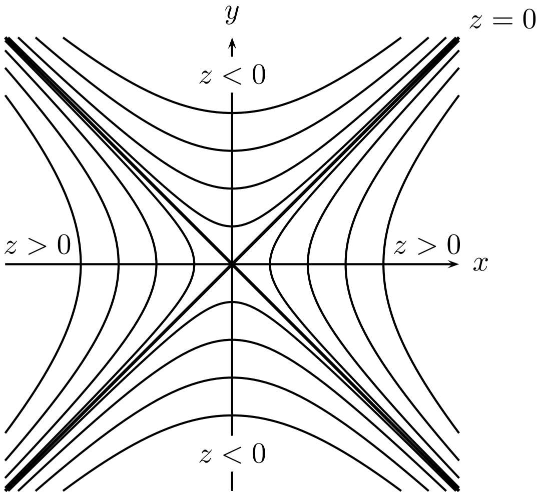


Figure 3.16: Level sets of $z = x^2 - y^2$

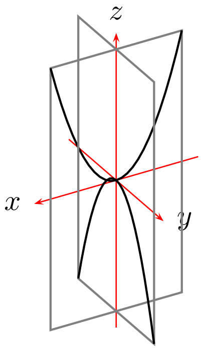


Figure 3.17: Locus of vertices

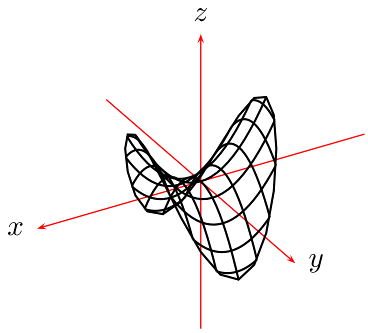


Figure 3.18: Hyperbolic Paraboloid (Saddle Surface) $x^2 - y^2 = z$

To find the intersection with another horizontal plane, $z = k$, we substitute this into the equation of the surface, getting

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 - \frac{k^2}{c^2};$$

to get a nonempty locus, we must have the right side nonnegative, or

$$|k| \leq c.$$

When we have equality, the intersection is a single point, and otherwise is an ellipse similar to that in Figure 3.19, but scaled down: we have superimposed a few of these "sections" of the surface in Figure 3.19. To see how these fit together, we can look at where the "vertices", or

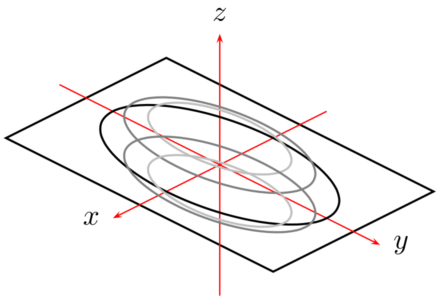


Figure 3.19: Horizontal sections

ends of the major and minor axes lie. These are the intersections of

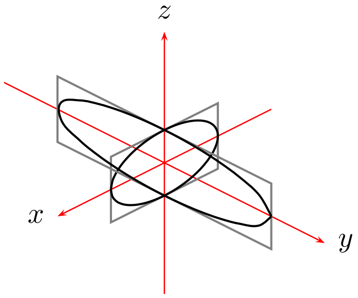


Figure 3.20: Vertical sections

our surface with the two vertical coordinate planes (Figure 3.20). In Figure 3.21 we sketch the surface

$$4x^2 + y^2 + 16z^2 = 4$$

which can be expressed more informatively as

$$\frac{x^2}{1^2} + \frac{y^2}{2^2} + \frac{z^2}{(1/2)^2} = 1;$$

this is called an **ellipsoid**; note that the three "axes" of our figure (along the coordinate axes) are precisely the square roots of the denominators of the second expression: $1, 2, \frac{1}{2}$ respectively.

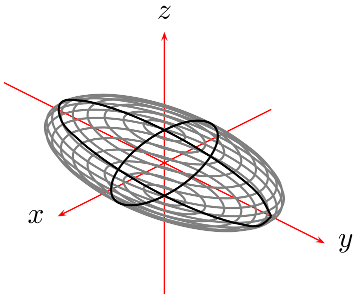


Figure 3.21: Ellipsoid $4x^2 + y^2 + 16z^2 = 4$

3. When all three variables appear to the second power but the quadratic form is not definite, there are three basic shapes that occur. These are

illustrated by "model equations" below. In each, we assume that x^2 occurs with coefficient 1 and z^2 with coefficient -1.

(a) If there is no constant term, then we have an equation of the form

$$x^2 \pm y^2 - z^2 = 0$$

which boils down to one of the two equations

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

or

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

Since the second of these equations results from the first by interchanging x with z, we will concentrate on the first.

The intersection of the surface

$$x^2 + y^2 = z^2 \tag{3.47}$$

with the horizontal plane $z = k$ is a circle, centered at the origin, of radius $|k|$ (Figure 3.22).

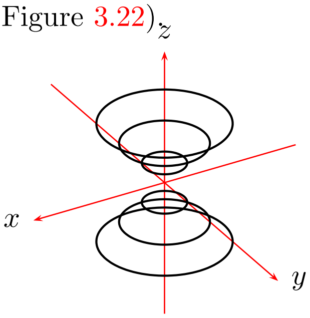


Figure 3.22: Horizontal Sections

The intersection of this surface with each of the vertical coordinate planes is a pair of lines (Figure 3.23)

and fitting this together, we see that this is just the conical surface $\mathcal{K}$ of Archimedes (§ 2.1) which we refer to in this context as simply a **cone** (Figure 3.24).

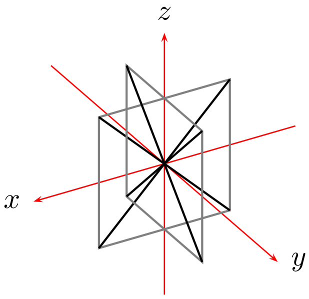


Figure 3.23: Vertical Sections

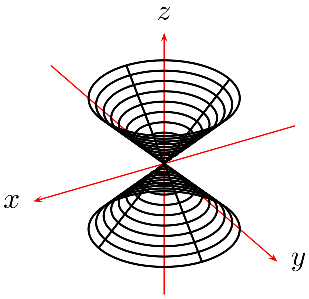


Figure 3.24: The cone $x^2 + y^2 = z^2$

(b) When there is a nonzero constant term, we will take it to be ± 1, and this leads to the possible equations

$$x^2 \pm y^2 = z^2 \pm 1.$$

Again, up to interchange of variables, there are two possible shapes, which can be modelled by the equation above with the coefficient of y^2 positive.

The surface given by

$$x^2 + y^2 = z^2 + 1 \tag{3.48}$$

intersects the horizontal plane $z = k$ in the circle centered on the z-axis of radius $\sqrt{1 + k^2}$ (Figure 3.25)

To see how they fit together, we consider the intersection of the surface with the two vertical coordinate planes, which are both hyperbolas opening horizontally (Figure 3.26).

The resulting surface (Figure 3.27) is called a **hyperboloid of one sheet** (Figure 3.27).

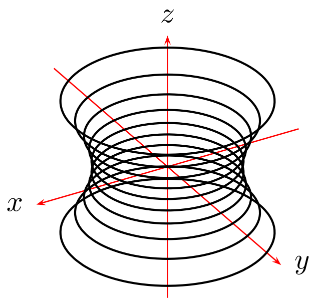


Figure 3.25: Horizontal Sections

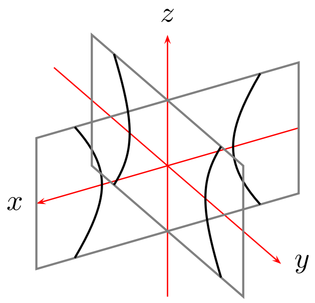


Figure 3.26: Vertical Sections

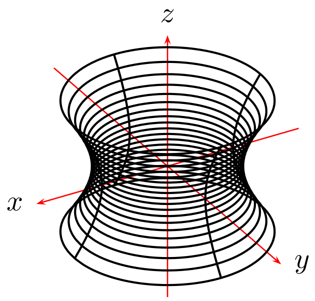


Figure 3.27: Hyperboloid of One Sheet $x^2 + y^2 = z^2 + 1$

(c) The surface given by

$$x^2 + y^2 = z^2 - 1 \tag{3.49}$$

intersects the horizontal plane $z = k$ in the locus of the equation $x^2 + y^2 = k^2 - 1$; for $|k| < 1$ the right side is negative, so there is no intersection; for $|k| > 1$ we again get a circle centered on the z-axis, with radius $\sqrt{k^2 - 1}$ (Figure 3.28).

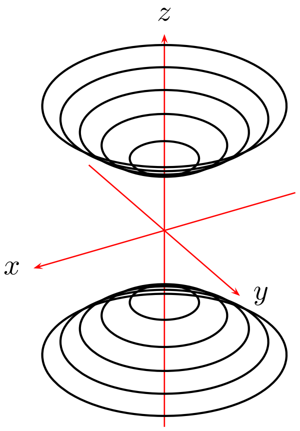


Figure 3.28: Horizontal Sections

To see how these fit together, we intersect the surface with the two vertical coordinate planes. The intersection with the xz-plane has equation $x^2 = z^2 - 1$ or $z^2 - x^2 = 1$, which is a hyperbola opening vertically; the intersection with the yz-plane is essentially identical (Figure 3.29)

The resulting surface consists of two bowl-like parts, and is called a **hyperboloid of two sheets** (Figure 3.30).

Exercises for § 3.10

Practice problems:

1.

2.

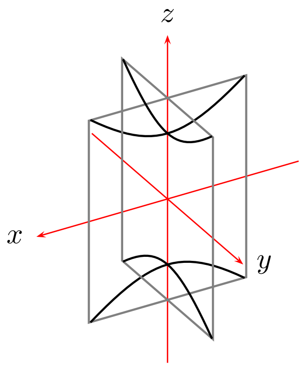


Figure 3.29: Vertical Sections

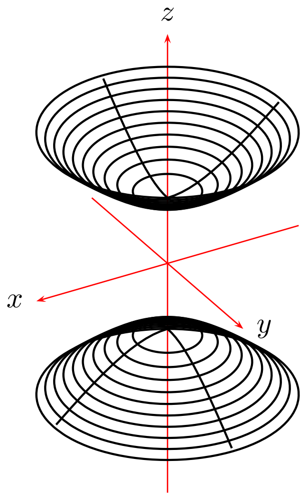


Figure 3.30: Hyperboloid of Two Sheets $x^2 + y^2 = z^2 - 1$

4

Mappings and Transformations: Vector-Valued Functions of Several Variables

In this chapter we extend differential calculus to vector-valued functions of a vector variable. We shall refer to a rule (call it F) which assigns to every vector (or point) $\overrightarrow{x}$ in its domain an unambiguous vector (or point) $\overrightarrow{y} = F(\overrightarrow{x})$ as a **mapping** from the domain to the target (the plane or space). This is of course a restatement of the definition of a function, except that the input and output are both vectors instead of real numbers.[1] We shall use the arrow notation first adopted in § 2.2 to indicate the domain and target of a mapping: using the notation $\mathbb{R}^2$ for the plane and $\mathbb{R}^3$ for space, we will write

$$F\colon \mathbb{R}^n \to \mathbb{R}^m$$

to indicate that the mapping F takes inputs from $\mathbb{R}^n$ ($n \leq 3$) and yields values in in $\mathbb{R}^m$ ($m \leq 3$). If we want to specify the domain $D \subset \mathbb{R}^n$, we write

$$F\colon D \to \mathbb{R}^m.$$

[1] More generally, the notion of a mapping from any set of objects to any (other) set is defined analogously, but this will not concern us.

If we expand the superscript notation by thinking of numbers as "1-vectors" ($\mathbb{R}^1 = \mathbb{R}$), then this definition and notation embrace all of the kinds of functions we have considered earlier. The term **transformation** is sometimes used when the domain and target live in the same dimension ($m = n$).

In Chapter 3 we identified the input to a function of several variables as a vector, while in Chapter 2 we identified the output of a vector-valued function F as a list of functions f_i, giving the coordinates of the output. In the present context, when we express the output as a list, we write down the coordinate column of the output vector: for example, a mapping $F\colon \mathbb{R}^2 \to \mathbb{R}^3$ from the plane to space could be expressed (using vector notation for the input) as

$$[F(\overrightarrow{x})] = \begin{bmatrix} f_1(\overrightarrow{x}) \\ f_2(\overrightarrow{x}) \\ f_3(\overrightarrow{x}) \end{bmatrix}$$

or, writing the input as a list of numerical variables,

$$[F(x_1, x_2, x_3)] = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \\ f_3(x_1, x_2) \end{bmatrix}.$$

Often we shall be sloppy and simply write

$$F(\overrightarrow{x}) = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \\ f_3(x_1, x_2) \end{bmatrix}.$$

We cannot draw (or imagine drawing) the "graph" of a mapping $F\colon \mathbb{R}^n \to \mathbb{R}^m$ if $m + n > 3$, but we can try to picture its action by looking at the images of various sets. For example, one can view a change of coordinates in the plane or space as a mapping: specifically, the calculation that gives the rectangular coordinates of a point in terms of its polar coordinates is the map $P\colon \mathbb{R}^2 \to \mathbb{R}^2$ from the (r, θ)-plane to the (x, y)-plane given by

$$P(r, \theta) = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}.$$

We get a picture of how it acts by noting that it takes horizontal (*resp.* vertical) lines to rays from (*resp.* circles around) the origin (Figure 4.1).

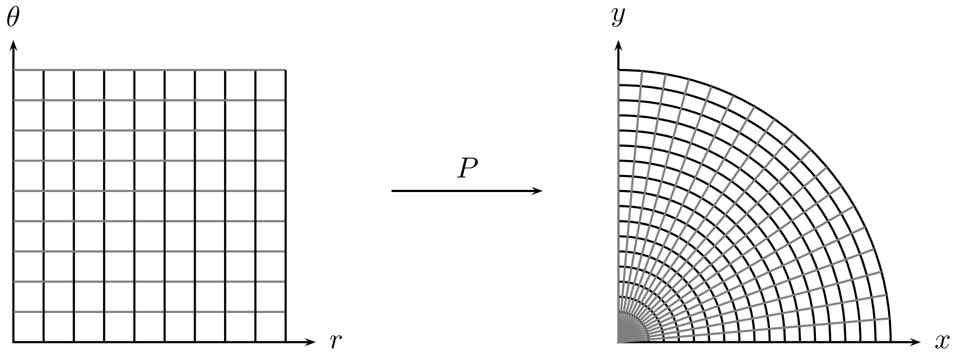


Figure 4.1: Polar Coordinates as a mapping

It is also possible (and, as we shall see, useful) to think of a system of one or more equations in several variables as a single equation involving a mapping: for example, the system of two equations in three unknowns

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

which geometrically represents the intersection of the unit sphere with the plane $x+y=z$ can also be thought of as finding a "level set" for the mapping $F\colon \mathbb{R}^3 \to \mathbb{R}^2$

$$F(x,y,z) = \begin{bmatrix} x^2+y^2+z^2 \\ x+y-z \end{bmatrix}$$

corresponding to the value $(1,0)$.

4.1 Linear Mappings

Recall that a linear function L on 3-space is just a homogeneous polynomial of degree one

$$L(x,y,z) = a_1x + a_2y + a_3z;$$

it is naturally defined on all of $\mathbb{R}^3$.These functions are the simplest to calculate, and play the role of derivatives for more general functions of several variables. By analogy, we call a mapping $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ **linear** if each of its component functions is linear:[2]

[2] To avoid tortuous constructions or notations, we will work here with mappings of space to space; the analogues when the domain or target (or both) lives in the plane or

$$L(x,y,z) = \begin{bmatrix} L_1(x,y,z) \\ L_2(x,y,z) \\ \vdots \\ L_m(x,y,z) \end{bmatrix} = \begin{bmatrix} a_{11}x + a_{12}y + a_{13}z \\ a_{21}x + a_{22}y + a_{23}z \\ a_{31}x + a_{32}y + a_{33}z \end{bmatrix}.$$

A more efficient way of writing this is via **matrix multiplication**: if we form the 3×3 matrix $[L]$, called the **matrix representative** of L, whose entries are the coefficients of the component polynomials

$$[L] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

then the coordinate column of the image $L(\overrightarrow{x})$ is the product of $[L]$ with the coordinate column of $\overrightarrow{x}$:

$$[L(\overrightarrow{x})] = [L] \cdot [\overrightarrow{x}]$$

or

$$\begin{aligned} L(\overrightarrow{x}) &= \begin{bmatrix} a_{11}x + a_{12}y + a_{13}z \\ a_{21}x + a_{22}y + a_{23}z \\ a_{31}x + a_{32}y + a_{33}z \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \end{aligned}$$

The last equation can be taken as the definition of the matrix product; if you are not familiar with matrix multiplication, see Appendix E for more details.

When a linear mapping is defined in some way other than giving the coordinate polynomials, there is an easy way to find its matrix representative. The proof of the following is outlined in Exercise 3:

Remark 4.1.1. *The j^{th} column of the matrix representative $[L]$ of a linear mapping $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ is the coordinate column of $L(\overrightarrow{e}_j)$, where $\{\overrightarrow{e}_1, \overrightarrow{e}_2, \overrightarrow{e}_3\}$ are the standard basis vectors for $\mathbb{R}^3$.*

There is a more geometric characterization of linear mappings which is often useful:

on the line are straightforward.

Remark 4.1.2. *A mapping $L: \mathbb{R}^3 \to \mathbb{R}^3$ is linear if and only if it preserves linear combinations: that is, for any two vectors $\overrightarrow{v}$ and $\overrightarrow{v}'$ and any two scalars α and β,*

$$L\big(\alpha\overrightarrow{x} + \beta\overrightarrow{v}'\big) = \alpha L(\overrightarrow{v}) + \beta L\big(\overrightarrow{v}'\big)\,.$$

Geometrically, this means that *the image under L of any triangle (with vertex at the origin) is again a triangle* (with vertex at the origin).

As an example, consider the mapping $P: \mathbb{R}^3 \to \mathbb{R}^3$ that takes each vector $\overrightarrow{x}$ to its perpendicular projection onto the plane

$$x + y + z = 0$$

through the origin with normal vector $\overrightarrow{n} = \overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}$ (Figure 4.2). It is geometrically clear that this takes triangles through the origin to triangles through the origin, and hence is linear. Since any vector is the sum of its projection on the plane and its projection on the normal line, we know that P can be calculated from the formula $P(\overrightarrow{x}) = \overrightarrow{x} - \text{proj}_{\overrightarrow{u}}\,\overrightarrow{x} = \overrightarrow{x} - (\overrightarrow{x} \cdot \overrightarrow{u})\overrightarrow{u}$, where $\overrightarrow{u} = \overrightarrow{n}/\sqrt{3}$ is the unit vector in the direction of $\overrightarrow{n}$. Applying this

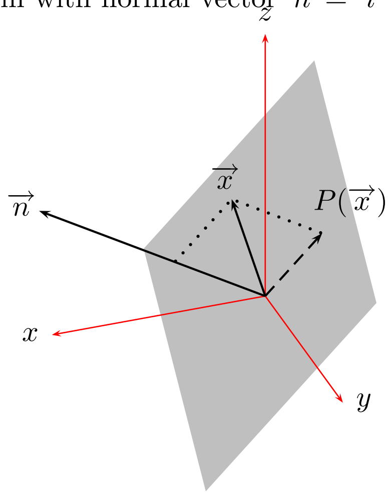


Figure 4.2: Projection onto a Plane

to the three standard basis vectors

$$\begin{aligned} P(\overrightarrow{\imath}) &= \overrightarrow{\imath} - (\overrightarrow{\imath} \cdot \overrightarrow{u})\overrightarrow{u} \\ &= \overrightarrow{\imath} - \frac{1}{3}(\overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}) \\ &= \begin{bmatrix} 1 - \frac{1}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} \\ P(\overrightarrow{\jmath}) &= \overrightarrow{\jmath} - (\overrightarrow{\jmath} \cdot \overrightarrow{u})\overrightarrow{u} \\ &= \begin{bmatrix} -\frac{1}{3} \\ 1 - \frac{1}{3} \\ -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ \frac{2}{3} \\ -\frac{1}{3} \end{bmatrix} \\ P\left(\overrightarrow{k}\right) &= \overrightarrow{k} - (\overrightarrow{k} \cdot \overrightarrow{u})\overrightarrow{u} \\ &= \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{3} \\ 1 - \frac{1}{3} \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} \\ -\frac{1}{3} \\ \frac{2}{3} \end{bmatrix} \end{aligned}$$

so

$$[P] = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & - -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

An example of a linear mapping $L: \mathbb{R}^2 \to \mathbb{R}^2$ is rotation by α radians counterclockwise; to find its matrix representative, we use Remark 4.1.1: from the geometric definition of L, the images of $\overrightarrow{\imath}$ and $\overrightarrow{\jmath}$ are easy to calculate (Figure 4.3):

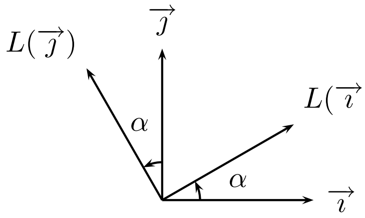

$L(\overrightarrow{\jmath})$
$\overrightarrow{\jmath}$
$L(\overrightarrow{\imath})$
α
α
$\overrightarrow{\imath}$

Figure 4.3: Rotating the Standard Basis

$$L(\overrightarrow{\imath}) = \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}$$
$$L(\overrightarrow{\jmath}) = \begin{bmatrix} -\sin\alpha \\ \cos\alpha \end{bmatrix}$$

so

$$[L] = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}.$$

Composition of Linear Mappings

Recall that the composition of two real-valued functions, say f and g, is the function obtained by applying one of the functions to the output of the other: $(f \circ g)(x) = f(g(x))$ and $(g \circ f)(x) = g(f(x))$. For the first of these to make sense, of course, x must belong to the domain of g, but also its image $g(x)$ must belong to the domain of f (in the other composition, the two switch roles). The same definition can be applied to mappings: in particular, suppose $L: \mathbb{R}^n \to \mathbb{R}^m$ and $L': \mathbb{R}^{n'} \to \mathbb{R}^{m'}$ are linear maps (so the natural domain of L (*resp.* L') is all of $\mathbb{R}^n$ (*resp.* $\mathbb{R}^{n'}$)); then the composition $L \circ L'$ is defined precisely if $n = m'$. It is easy to see that this composition is linear as well:

$$\begin{aligned}(L \circ L')\big(\alpha\overrightarrow{x} + \beta\overrightarrow{x}'\big) &= L\big(L'\big(\alpha\overrightarrow{x} + \beta\overrightarrow{x}'\big)\big) \\ &= L\big(\alpha L'(\overrightarrow{x}) + \beta L'\big(\overrightarrow{x}'\big)\big) \\ &= \alpha L\big(L'(\overrightarrow{x})\big) + \beta L\big(L'\big(\overrightarrow{x}'\big)\big) \\ &= \alpha(L \circ L')(\overrightarrow{x}) + \beta(L \circ L')\big(\overrightarrow{x}'\big)\,.\end{aligned}$$

It is equally easy to see that the matrix representative $[L \circ L']$ of a composition is the matrix product of the matrix representatives of the two maps:

$$\begin{aligned}\big[(L \circ L')(\overrightarrow{x})\big] &= \big[L\big(L'(\overrightarrow{x})\big)\big] \\ &= [L] \cdot \big[L'(\overrightarrow{x})\big] \\ &= [L] \cdot \big(\big[L'\big] \cdot [\overrightarrow{x}]\big) \\ &= \big([L] \cdot \big[L'\big]\big)\,[\overrightarrow{x}]\,.\end{aligned}$$

We formalize this in

Remark 4.1.3. *The composition of linear maps is linear, and the matrix representative of the composition is the product of their matrix representatives.*

For example, suppose $L'\colon \mathbb{R}^3 \to \mathbb{R}^2$ is defined by

$$L'\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y+z \end{bmatrix}$$

and $L\colon \mathbb{R}^2 \to \mathbb{R}^3$ is defined by

$$L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x-y \\ x+y \\ 2x-y \end{bmatrix};$$

then $L \circ L'\colon \mathbb{R}^3 \to \mathbb{R}^3$ is defined by

$$\begin{aligned}(L \circ L')\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) &= L\left(\begin{bmatrix} x+y \\ y+z \end{bmatrix}\right) \\ &= \begin{bmatrix} (x+y)-(y+z) \\ (x+y)+(y+z) \\ 2(x+y)-(y+z) \end{bmatrix} \\ &= \begin{bmatrix} x-z \\ x+2y+z \\ 2x+y-z \end{bmatrix}\end{aligned}$$

and the composition in the other order, $L' \circ L\colon \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$\begin{aligned}(L' \circ L)\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= L'\left(\begin{bmatrix} x-y \\ x+y \\ 2x-y \end{bmatrix}\right) \\ &= \begin{bmatrix} (x-y)+(x+y) \\ (x+y)+(2x-y) \end{bmatrix} \\ &= \begin{bmatrix} 2x \\ 3x+2y \end{bmatrix}.\end{aligned}$$

The respective matrix representatives are

$$[L'] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \qquad [L] = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 2 & -1 \end{bmatrix}$$

so

$$[L] \cdot [L'] = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 2 & 1 \\ 2 & 2 & -1 \end{bmatrix}$$

and

$$[L'] \cdot [L] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 3 & 2 \end{bmatrix}.$$

You should verify that these last two matrices are, in fact the matrix representatives of $L \circ L'$ and $L' \circ L$, respectively.

Exercises for § 4.1

Practice problems:

1. Which of the following maps are linear? Give the matrix representative for those which are linear.

   (a) $f(x,y) = (y,x)$

   (b) $f(x,y) = (x,x)$

   (c) $f(x,y) = (e^x \cos y, e^x \sin y)$

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

   (e) $f(x,y) = (x+y, x-y)$

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

   (g) $f(x,y) = (x+y, 2x-y, x+3y)$

   (h) $f(x,y) = (x-2y, x+y-1, 3x+5y)$

   (i) $f(x,y) = (x, y, x^2 - y^2)$

   (j) $f(x,y) = (x, y, xy)$

   (k) $f(x,y,z) = (2x+3y+4z, x+z, y+z)$

   (l) $f(x,y,z) = (y+z, x+z, x+y)$

   (m) $f(x,y,z) = (x-2y+1, y-z+2, x-y-z)$

   (n) $f(x,y,z) = (x+2y, z-y+1, x)$

   (o) $f(x,y,z) = (x+y, 2x-y, 3x+2y)$

   (p) $f(x,y,z) = (x+y+z, x-2y+3z)$

(q) Projection of $\mathbb{R}^3$ onto the plane $2x - y + 3z = 0$

(r) Projection of $\mathbb{R}^3$ onto the plane $3x + 2y + z = 1$

(s) Rotation of $\mathbb{R}^3$ around the z-axis by θ radians counterclockwise, seen from above.

(t) Projection of $\mathbb{R}^3$ onto the line $x = y = z$.

2. Express each affine map T below as $T(\overrightarrow{x}) = T(\overrightarrow{x}_0) + L(\triangle\overrightarrow{x})$ with the given $\overrightarrow{x}_0$ and linear map L.

(a) $T(x, y) = (x + y - 1, x - y + 2)$, $\overrightarrow{x}_0 = (1, 2)$

(b) $T(x, y) = (3x - 2y + 2, x - y)$, $\overrightarrow{x}_0 = (-2, -1)$

(c) $T(x, y, z) = (3x - 2y + z, z + 2)$, $\overrightarrow{x}_0 = (1, 1, -1)$

(d) $T(x, y) = (2x - y + 1, x - 2y, 2)$, $\overrightarrow{x}_0 = (1, 1)$

(e) $T(x, y, z) = (x + 2y, z - y + 1, x)$, $\overrightarrow{x}_0 = (2, -1, 1)$

(f) $T(x, y, z) = (x - 2y + 1, y - z + 2, x - y - z)$, $\overrightarrow{x}_0 = (1, -1, 2)$

(g) $T(x, y, z) = (x + 2y - z - 2, 2x - y + 1, z - 2)$, $\overrightarrow{x}_0 = (1, 1, 2)$

(h) T is projection onto the line $\overrightarrow{p}(t) = (t, t+1, t-1)$, $\overrightarrow{x}_0 = (1, -1, 2)$ (*Hint:* Find where the given line intersects the plane through $\overrightarrow{x}$ perpendicular to the line.)

(i) T is projection onto the plane $x + y + z = 3$, $\overrightarrow{x}_0 = (1, -1, 2)$ (*Hint:* first project onto the parallel plane through the origin, then translate by a suitable normal vector.)

Theory problems:

3. Prove Remark 4.1.1. (*Hint:* What is the coordinate column of the standard basis vector $\overrightarrow{e}_j$?)

4. Show that the composition of two affine maps is again affine.

5. Find the matrix representative for each kind of linear map $L: \mathbb{R}^2 \to \mathbb{R}^2$ described below:

(a) Horizontal Scaling: horizontal component gets scaled (multiplied) by $\lambda > 0$, vertical component is unchanged.

(b) Vertical Scaling: vertical component gets scaled (multiplied) by $\lambda > 0$, horizontal component is unchanged.

(c) Horizontal Shear: Each horizontal line $y = c$ is translated (horizontally) by an amount proportional to c.

(d) Vertical Shear: Each vertical line $x = c$ is translated (vertically) by an amount proportional to c.

(e) Reflection about the Diagonal: x and y are interchanged.

(f) Rotation: Each vector is rotated θ radians counterclockwise.

Challenge problems:

6. Suppose $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ is linear.

   (a) **Show** that the determinant of $[L]$ is nonzero iff the image vectors $L(\overrightarrow{\imath})$ and $L(\overrightarrow{\jmath})$ are independent.

   (b) **Show** that if $L(\overrightarrow{\imath})$ and $L(\overrightarrow{\jmath})$ are linearly independent, then L is an onto map.

   (c) **Show** that if $L(\overrightarrow{\imath})$ and $L(\overrightarrow{\jmath})$ are linearly *dependent*, then L maps $\mathbb{R}^2$ into a line, and so is *not* onto.

   (d) **Show** that if L is *not* one-to-one, then there is a nonzero vector $\overrightarrow{x}$ with $L(\overrightarrow{x}) = \overrightarrow{0}$.

   (e) **Show** that if L is not one-to-one, then $L(\overrightarrow{\imath})$ and $L(\overrightarrow{\jmath})$ are linearly dependent.

   (f) **Show** that if $L(\overrightarrow{\imath})$ and $L(\overrightarrow{\jmath})$ are dependent, then there is some nonzero vector sent to $\overrightarrow{0}$ by L.

   (g) Use this to prove that the following are equivalent:

      i. the determinant of $[L]$ is nonzero;

      ii. $L(\overrightarrow{\imath})$ and $L(\overrightarrow{\jmath})$ are linearly independent;

      iii. L is onto;

      iv. L is one-to-one.

   (h) L is **invertible** if there exists another map $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $L(F(x, y)) = (x, y) = F(L(x, y))$. Show that if F exists it must be linear.

7. **Show** that every invertible linear map $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ can be expressed as a composition of the kinds of mappings described in Exercise 5. (*Hint:* Given the desired images $L(\overrightarrow{\imath})$ and $L(\overrightarrow{\jmath})$, first adjust the angle, then get the lengths right, and finally rotate into position.)

4.2 Differentiable Mappings

We have seen several versions of the notion of a derivative in previous sections: for a real-valued function f of one real variable, the derivative is a number, which gives the slope of the line tangent to the graph $y = f(x)$ at the given point; for a *vector*-valued function of one *real* variable, the derivative is a vector, giving the velocity of the motion described by the function, or equivalently giving the coefficients of the "time" variable in the natural parametrization of the tangent line; for a *real*-valued function of a *vector* variable, the derivative is the linear part of an affine function making first-order contact with the function at the given point. We can combine these last two interpretations to formulate the derivative of a *vector*-valued function of a *vector* variable. Extending our terminology from real-valued functions (as in § 3.2) to (vector-valued) mappings, we define an **affine mapping** to be a mapping of the form $T(\overrightarrow{x}) = \overrightarrow{c} + \phi(\overrightarrow{x})$, where ϕ is linear and $\overrightarrow{c}$ is a constant vector. If we pick any point $\overrightarrow{x}_0$ in the domain, then we can write T in the form

$$T(\overrightarrow{x}) = T(\overrightarrow{x}_0) + \phi(\triangle\overrightarrow{x})$$

where

$$\triangle\overrightarrow{x} = \overrightarrow{x} - \overrightarrow{x}_0.$$

Definition 4.2.1. *A mapping*[3] *F is* ***differentiable*** *at a point $\overrightarrow{x}_0$ interior to its domain if there exists an affine mapping T which has first-order contact with F at $\overrightarrow{x} = \overrightarrow{x}_0$:*

$$\|F(\overrightarrow{x}) - T(\overrightarrow{x})\| = \mathfrak{o}\|\overrightarrow{x} - \overrightarrow{x}_0\|$$

as $\overrightarrow{x} \to \overrightarrow{x}_0$; in other words

$$\lim_{\overrightarrow{x}\to\overrightarrow{x}_0} \frac{\|F(\overrightarrow{x}) - T(\overrightarrow{x})\|}{\|\overrightarrow{x} - \overrightarrow{x}_0\|} = 0.$$

Arguments analogous to those for a real-valued map of several variables (§ 3.3) show that at most one affine function T can satisfy the requirements of this definition at a given point $\overrightarrow{x}_0$: we can write it in the form

$$T_{\overrightarrow{x}_0}F(\overrightarrow{x}) = F(\overrightarrow{x}_0) + \phi(\triangle\overrightarrow{x}) := F(\overrightarrow{x}_0) + \phi(\overrightarrow{x} - \overrightarrow{x}_0)\,.$$

[3] Of course, a mapping can be given either an upper- or lower-case name. We are adopting an upper-case notation to stress that our mapping is vector-valued.

The "linear part" ϕ is called the **derivative** or **differential**[4] of F at $\overrightarrow{x}_0$ and denoted either $d_{\overrightarrow{x}_0}F$ or $DF_{\overrightarrow{x}_0}$; we shall use the "derivative" terminology and the "D" notation. The "full" affine map will be denoted $T_{\overrightarrow{x}_0}F$, in keeping with the notation for Taylor polynomials: this is sometimes called the **linearization** of F at $\overrightarrow{x}_0$. Thus, the linearization of the differentiable mapping $F\colon \mathbb{R}^n \to \mathbb{R}^m$ at $\overrightarrow{x}_0$ is

$$T_{\overrightarrow{x}_0}F(\overrightarrow{x}) = F(\overrightarrow{x}_0) + DF_{\overrightarrow{x}_0}(\overrightarrow{x} - \overrightarrow{x}_0) = F(\overrightarrow{x}_0) + DF_{\overrightarrow{x}_0}(\triangle\overrightarrow{x}) .$$

To calculate the derivative, let us fix a point $\overrightarrow{x}_0$ and a velocity vector $\overrightarrow{v}$. If we write a mapping F with values in space as a column of functions

$$F(\overrightarrow{x}) = \begin{bmatrix} f_1(\overrightarrow{x}) \\ f_2(\overrightarrow{x}) \\ f_3(\overrightarrow{x}) \end{bmatrix}$$

then we can consider the action of the differentials at $\overrightarrow{x}_0$ of the various component functions f_i on $\overrightarrow{v}$: recall from § 3.3 that this can be interpreted as the derivative

$$d_{\overrightarrow{x}_0}(f_i)(\overrightarrow{v}) = \left.\frac{d}{dt}\right|_{t=0} [f_i(\overrightarrow{x}_0 + t\overrightarrow{v})] .$$

We can consider the full function $\overrightarrow{p}(t) = F(\overrightarrow{x}_0 + t\overrightarrow{v})$ as a parametrized curve—that is, as t varies, the input into F is a point moving steadily along the line in the domain of F which goes through $\overrightarrow{x}_0$ with velocity $\overrightarrow{v}$; the curve $\overrightarrow{p}(t)$ is the image of this curve under the mapping, and *its* velocity at $t = 0$ is the column consisting of the differentials above. If we add the initial vector $\overrightarrow{p}(0) = F(\overrightarrow{x}_0)$, we obtain an affine map from $\mathbb{R}$ to $\mathbb{R}^3$ which has first-order contact with $\overrightarrow{p}(t)$ at $t = 0$. From this we have

Remark 4.2.2. *The derivative of a mapping F at $\overrightarrow{x}_0$ can be evaluated on a vector $\overrightarrow{v}$ as the velocity of the image under F of the constant-velocity curve $\overrightarrow{p}(t) = F(\overrightarrow{x}_0 + t\overrightarrow{v})$ through $\overrightarrow{x}_0$ in $\mathbb{R}^3$ with velocity $\overrightarrow{v}$:*

$$DF_{\overrightarrow{x}_0}(\overrightarrow{v}) = \left.\frac{d}{dt}\right|_{t=0} [F(\overrightarrow{x}_0 + t\overrightarrow{v})] = \begin{bmatrix} d_{\overrightarrow{x}_0}f_1(\overrightarrow{v}) \\ d_{\overrightarrow{x}_0}f_2(\overrightarrow{v}) \\ \vdots \\ d_{\overrightarrow{x}_0}f_m(\overrightarrow{v}) \end{bmatrix} . \tag{4.1}$$

[4] It is also called the **tangent mapping** of F at $\overrightarrow{x}_0$.

In particular, when $\overrightarrow{v}$ is the j^{th} element of the standard basis for $\mathbb{R}^3$, this gives us the velocity of the image of the j^{th} coordinate axis, and as a column this consists of the j^{th} partial derivatives of the component functions. But this column is the j^{th} column of the matrix representative of $DF_{\overrightarrow{x}_0}$, giving us[5]

Remark 4.2.3. *The matrix representative of the derivative $DF_{\overrightarrow{x}_0}$ of $F\colon\mathbb{R}^3\to\mathbb{R}^3$ is the matrix of partial derivatives of the component functions of F:*

$$[DF] = \begin{bmatrix} \partial f_1/\partial x & \partial f_1/\partial y & \partial f_1/\partial z \\ \partial f_2/\partial x & \partial f_2/\partial y & \partial f_2/\partial z \\ \partial f_3/\partial x & \partial f_3/\partial y & \partial f_3/\partial z \end{bmatrix}.$$

The matrix above is called the **Jacobian matrix** of F, and denoted [6] JF.

As a special case, we note the following, whose (easy) proof is left to you (Exercise 3):

Remark 4.2.4. *If $F\colon\mathbb{R}^3\to\mathbb{R}^3$ is linear, then it is differentiable and*

$$DF_{\overrightarrow{x}_0} = F$$

for every $\overrightarrow{x}_0 \in \mathbb{R}^3$. In particular, the linearization (at any point) of an affine map is the map itself.

The Chain Rule

We have seen several versions of the Chain Rule before. The setting of mappings allows us to formulate a single unified version which includes the others as special cases. In the statements below, we assume the dimensions m, n and p are each 1, 2 or 3.

Theorem 4.2.5 (General Chain Rule)**.** *If $F\colon\mathbb{R}^n\to\mathbb{R}^m$ is differentiable at $\overrightarrow{y}_0 \in \mathbb{R}^n$ and $G\colon\mathbb{R}^p\to\mathbb{R}^n$ is differentiable at $\overrightarrow{x}_0 \in \mathbb{R}^p$ where $\overrightarrow{y}_0 = G(\overrightarrow{x}_0)$, then the composition $F\circ G\colon\mathbb{R}^p\to\mathbb{R}^m$ is differentiable at $\overrightarrow{x}_0$, and its derivative is the composition of the derivatives of G (at $\overrightarrow{x}_0$) and F (at $\overrightarrow{y}_0 = G(\overrightarrow{x}_0)$):*

$$D(F\circ G)_{\overrightarrow{x}_0} = (DF_{\overrightarrow{y}_0})\circ(DG_{\overrightarrow{x}_0});$$

[5] Again, the analogue when the domain or the target or both are the plane instead of space is straightforward.

[6] An older, but sometimes useful notation, based on viewing F as an m-tuple of functions, is $\frac{\partial(f_1,f_2,f_3)}{\partial(x,y,z)}$.

in matrix language, the Jacobian matrix of the composition is the product of the Jacobian matrices:

$$J(F \circ G)(\overrightarrow{x}_0) = JF(\overrightarrow{y}_0) \cdot JG(\overrightarrow{x}_0)\,.$$

Proof. We need to show that the "affine approximation" we get by assuming that the derivative of $F \circ G$ is the composition of the derivatives, say

$$T(\overrightarrow{x}_0 + \triangle\overrightarrow{x}) = (F \circ G)(\overrightarrow{x}_0) + (DF_{\overrightarrow{y}_0} \circ DG_{\overrightarrow{x}_0})(\triangle\overrightarrow{x})$$

has first-order contact at $\triangle\overrightarrow{x} = \overrightarrow{0}$ with $(F \circ G)(\overrightarrow{x}_0 + \triangle\overrightarrow{x})$. The easiest form of this to work with is to show that for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$(F \circ G)(\overrightarrow{x}_0 + \triangle\overrightarrow{x}) = (F \circ G)(\overrightarrow{x}_0) + (DF_{\overrightarrow{y}_0} \circ DG_{\overrightarrow{x}_0})(\triangle\overrightarrow{x}) + \mathcal{E}(\triangle\overrightarrow{x}) \quad (4.2)$$

such that $\|\mathcal{E}(\triangle\overrightarrow{x})\| < \varepsilon \|\triangle\overrightarrow{x}\|$ whenever $\|\triangle\overrightarrow{x}\| < \delta$.

To carry this out, we need first to establish an estimate on how much the length of a vector can be increased when we apply a linear mapping.

> **Claim:** *If $L: \mathbb{R}^n \to \mathbb{R}^m$ is linear, then there exists a number*[7] *M such that*
>
> $$\|L(\overrightarrow{x})\| \leq M \|\overrightarrow{x}\|$$
>
> *for every $n \in \mathbb{R}$. This number can be chosen to satisfy the estimate*
>
> $$M \leq mn\, a_{max}$$
>
> *where a_{max} is the maximum absolute value of entries in the matrix representative $[L]$.*

This is an easy application of the triangle inequality. Given a vector $\overrightarrow{x} = (v_1, \ldots, v_n)$, let v_{max} be the maximum absolute value of the components of $\overrightarrow{x}$, and let a_{ij} be the entry in row i, column j of $[L]$, The i^{th} component of $L(\overrightarrow{v})$ is

$$(L(\overrightarrow{v}))_i = a_{i1}v_1 + \cdots + a_{in}v_n$$

so we can write

$$\begin{aligned} |(L(\overrightarrow{v}))_i| &\leq |a_{1i}|\,|v_1| + \cdots + |a_{in}|\,|v_n| \\ &\leq n a_{max} v_{max}. \end{aligned}$$

[7] The *least* such number is called the **operator norm** of the mapping, and is denoted $\|\, L \,\|$

Now we know that the length of a vector is less than the sum of (the absolute values of) its components, so

$$\begin{aligned}\|L(\overrightarrow{x})\| &\leq |(L(\overrightarrow{v}))_1| + \cdots + |(L(\overrightarrow{v}))_m| \\ &\leq m(n\, a_{max} v_{max}) \\ &\leq mn\, a_{max}\, \|\overrightarrow{x}\|\end{aligned}$$

since the length of a vector is at least as large as any of its components.[8] This proves the claim.

Now, to prove the theorem, set

$$\begin{aligned}\overrightarrow{y} &= G(\overrightarrow{x}) \\ \overrightarrow{y}_0 &= G(\overrightarrow{x}_0)\end{aligned}$$

and

$$\triangle\overrightarrow{y} = G(\overrightarrow{x}_0 + \triangle\overrightarrow{x}) - G(\overrightarrow{x}_0)\,,$$

that is,

$$G(\overrightarrow{x}_0 + \triangle\overrightarrow{x}) = \overrightarrow{y}_0 + \triangle\overrightarrow{y}.$$

Then the differentiability of F at $\overrightarrow{y}_0$ says that, given $\varepsilon_1 > 0$, we can find $\delta_1 > 0$ such that $\|\triangle\overrightarrow{y}\| < \delta_1$ guarantees

$$F(\overrightarrow{y}) = F(\overrightarrow{y}_0) + DF_{\overrightarrow{y}_0}(\triangle\overrightarrow{y}) + \mathcal{E}_1(\triangle\overrightarrow{y})$$

where

$$\|\mathcal{E}_1(\triangle\overrightarrow{y})\| \leq \varepsilon_1\, \|\triangle\overrightarrow{y}\|\,.$$

Similarly, the differentiability of G at $\overrightarrow{x}_0$ says that, given $\varepsilon_2 > 0$, for $\|\triangle\overrightarrow{x}\| < \delta_2$ we can write

$$G(\overrightarrow{x}_0 + \triangle\overrightarrow{x}) = \overrightarrow{y}_0 + DG_{\overrightarrow{x}_0}(\triangle\overrightarrow{x}) + \mathcal{E}_2(\triangle\overrightarrow{x})$$

[8] Another way to get at the existence of such a number (without necessarily getting the estimate in terms of entries of $[L]$) is to note that the function $f(\overrightarrow{x}) = \|L(\overrightarrow{x})\|$ is continuous, and so takes its maximum on the (compact) unit sphere in $\mathbb{R}^n$. We leave you to work out the details (Exercise 4).

where

$$\|\mathcal{E}_2(\triangle\overrightarrow{x})\| \leq \varepsilon_2 \|\triangle\overrightarrow{x}\|.$$

Note that our expression for $G(\overrightarrow{x}_0 + \triangle\overrightarrow{x})$ lets us express $\triangle\overrightarrow{y}$ in the form

$$\triangle\overrightarrow{y} = DG_{\overrightarrow{x}_0}(\triangle\overrightarrow{x}) + \mathcal{E}_2(\triangle\overrightarrow{x}).$$

Applying the claim to $DG_{\overrightarrow{x}_0}$ we can say that for some $M_1 > 0$

$$\left\|DG_{\overrightarrow{x}_0}(\triangle\overrightarrow{x})\right\| \leq M_1 \|\triangle\overrightarrow{x}\|$$

so

$$\begin{aligned}\|\triangle\overrightarrow{y}\| &= \left\|DG_{\overrightarrow{x}_0}(\triangle\overrightarrow{x}) + \mathcal{E}_2(\triangle\overrightarrow{x})\right\| \\ &\leq (M_1 + \varepsilon_2)\|\triangle\overrightarrow{x}\|.\end{aligned}$$

Thus for

$$\|\triangle\overrightarrow{x}\| \leq \max\left(\delta_2, \frac{\delta_1}{M_1 + \varepsilon_2}\right)$$

we have

$$\|\triangle\overrightarrow{y}\| \leq \delta_1$$

so

$$F(\overrightarrow{y}) - F(\overrightarrow{y}_0) = DF_{\overrightarrow{y}_0}(\triangle\overrightarrow{y}) + \mathcal{E}_1(\triangle\overrightarrow{y})$$

with $\|\mathcal{E}_1(\triangle\overrightarrow{y})\| < \varepsilon_1$. Substituting our expression for $\triangle\overrightarrow{y}$ into this, and using the linearity of $DF_{\overrightarrow{y}_0}$, we have

$$\begin{aligned}F(\overrightarrow{y}) - F(\overrightarrow{y}_0) &= DF_{\overrightarrow{y}_0}\big(DG_{\overrightarrow{x}_0}(\triangle\overrightarrow{x}) + \mathcal{E}_2(\triangle\overrightarrow{x})\big) + \mathcal{E}_1(\triangle\overrightarrow{y}) \\ &= (DF_{\overrightarrow{y}_0} \circ DG_{\overrightarrow{x}_0})(\triangle\overrightarrow{x}) + DF_{\overrightarrow{y}_0}(\mathcal{E}_2(\triangle\overrightarrow{x})) + \mathcal{E}_1(\triangle\overrightarrow{y})\end{aligned}$$

so in Equation (4.2), we can write

$$\mathcal{E}(\triangle\overrightarrow{x}) = DF_{\overrightarrow{y}_0}(\mathcal{E}_2(\triangle\overrightarrow{x})) + \mathcal{E}_1(\triangle\overrightarrow{y});$$

we need to estimate this in terms of $\|\triangle\overrightarrow{x}\|$. Now we know from the claim that there exists $M_2 > 0$ such that

$$\left\|DF_{\overrightarrow{y}_0}(\mathcal{E}_2(\triangle\overrightarrow{x}))\right\| \leq M_2 \left\|\mathcal{E}_2(\triangle\overrightarrow{x})\right\|;$$

Using the triangle inequality as well as our previous estimates, we see that for

$$\|\triangle\overrightarrow{x}\| \leq \max\left(\delta_2, \frac{\delta_1}{M_1+\varepsilon_2}\right)$$

we have

$$\begin{aligned}
\|\mathcal{E}(\triangle\overrightarrow{x})\| &\leq M_2 \|\mathcal{E}_2(\triangle\overrightarrow{x})\| + \|\mathcal{E}_1(\triangle\overrightarrow{y})\| \\
&\leq M_2\varepsilon_2 \|\triangle\overrightarrow{x}\| + M_2\varepsilon_1 \|\triangle\overrightarrow{y}\| \\
&= [M_2\varepsilon_2 + M_2\varepsilon_1(M_1+\varepsilon_2)] \|\triangle\overrightarrow{x}\|.
\end{aligned}$$

Thus, if we pick

$$\varepsilon_2 < \frac{\varepsilon}{2M_2}$$

and

$$\varepsilon_1 < \frac{\varepsilon}{2M_2M_1}$$

then

$$\begin{aligned}
\|\mathcal{E}(\triangle\overrightarrow{x})\| &\leq [M_2\varepsilon_2 + M_2\varepsilon_1(M_1+\varepsilon_2)] \|\triangle\overrightarrow{x}\| \\
&< [M_2\varepsilon_2 + M_1M_2\varepsilon_1] \|\triangle\overrightarrow{x}\| \\
&< \frac{\varepsilon}{2} \|\triangle\overrightarrow{x}\| + \frac{\varepsilon}{2} \|\triangle\overrightarrow{x}\| \\
&= \varepsilon \|\triangle\overrightarrow{x}\|,
\end{aligned}$$

as required. □

Let us consider a few special cases, to illustrate how this chain rule subsumes the earlier ones.

First, a totally trivial example: if f and g are both real-valued functions of one real variable, and $y_0 = g(x_0)$, then the Jacobian matrix of each is a 1×1 matrix

$$\begin{aligned}
Jf(y_0) &= [f'(y_0)] \\
Jg(x_0) &= [g'(x_0)]
\end{aligned}$$

and the Jacobian of their composition is

$$\begin{aligned}[(f\circ g)'(x_0)] &= J(f\circ g)(x_0)\\ &= Jf(y_0)\cdot Jg(x_0)\\ &= [f'(y_0)\,g'(x_0)].\end{aligned}$$

Second, if $\overrightarrow{p}\colon\mathbb{R}\to\mathbb{R}^3$ is a parametrization $\overrightarrow{p}(t)$ of the curve $\mathcal{C}$ and $f\colon\mathbb{R}^3\to\mathbb{R}$ is a function defined on $\mathcal{C}$, then $(f\circ\overrightarrow{p})(t) = f(\overrightarrow{p}(t))$ gives f as a function of the parameter t: letting $\overrightarrow{x} = \overrightarrow{p}(t)$,

$$\begin{aligned}Jf(\overrightarrow{x}) &= [d_{\overrightarrow{x}}f]\\ &= \begin{bmatrix}\frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z}\end{bmatrix}\\ J\overrightarrow{p}(t) &= \begin{bmatrix}x'\\ y'\\ z'\end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}\left[\frac{d}{dt}\left[f(\overrightarrow{p}(t))\right]\right] &= J(f\circ\overrightarrow{p})(t)\\ &= Jf(\overrightarrow{x})\cdot J\overrightarrow{p}(t)\\ &= \left[\frac{\partial f}{\partial x}x' + \frac{\partial f}{\partial y}y' + \frac{\partial f}{\partial z}z'\right].\end{aligned}$$

Third, if again $\overrightarrow{p}\colon\mathbb{R}\to\mathbb{R}^3$ and $\overrightarrow{q}\colon\mathbb{R}\to\mathbb{R}^3$ are two parametrizations ($\overrightarrow{p}(t)$ and $\overrightarrow{q}(s)$) of the curve $\mathcal{C}$, and $\tau\colon\mathbb{R}\to\mathbb{R}$ is the change-of-parameter function $t = \tau(s)$ (*i.e.*, $\overrightarrow{q}(s) = \overrightarrow{p}(\tau(s))$, or $\overrightarrow{q} = \overrightarrow{p}\circ\tau$), then

$$\begin{aligned}J\overrightarrow{p}(t) &= \begin{bmatrix}\frac{dx}{dt} & \frac{dy}{dt} & \frac{dz}{dt}\end{bmatrix}\\ J\overrightarrow{q}(s) &= \begin{bmatrix}\frac{dx}{ds} & \frac{dy}{ds} & \frac{dz}{ds}\end{bmatrix}\\ J\tau(s) &= \begin{bmatrix}\tau'(s)\end{bmatrix} = \begin{bmatrix}\frac{dt}{ds}\end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}J\overrightarrow{q}(s) &= J(\overrightarrow{p}\circ\tau)(s)\\ &= J\overrightarrow{p}(t)\cdot J\tau(s)\\ &= \begin{bmatrix}\frac{dx}{dt}\frac{dt}{ds} & \frac{dy}{dt}\frac{dt}{ds} & \frac{dz}{dt}\frac{dt}{ds}\end{bmatrix}\end{aligned}$$

in other words,

$$\vec{q}'(s) = \vec{p}'(t)\,\frac{dt}{ds}.$$

The second and third examples above have further generalizations in light of Theorem 4.2.5.

If $f\colon \mathbb{R}^3 \to \mathbb{R}$ is a function defined on the surface $\mathfrak{S}$ and $\overrightarrow{p}(s,t)$ is a parametrization of $\mathfrak{S}$ ($\overrightarrow{p}\colon \mathbb{R}^2 \to \mathbb{R}^3$), then $f \circ \overrightarrow{p}$ expresses f as a function of the two parameters s and t, and the Chain Rule gives the partials of f with respect to them: setting $\overrightarrow{x} = \overrightarrow{p}(s,t)$,

$$\begin{aligned} Jf(\overrightarrow{x}) &= \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} & \dfrac{\partial f}{\partial y} \end{bmatrix} \\ J\overrightarrow{p}(s,t) &= \begin{bmatrix} \partial x/\partial s & \partial x/\partial t \\ \partial y/\partial s & \partial y/\partial t \\ \partial z/\partial s & \partial z/\partial t \end{bmatrix} \end{aligned}$$

so

$$\begin{aligned} J(f \circ \overrightarrow{p})(s,t) &= Jf(\overrightarrow{x}) \cdot J\overrightarrow{p}(s,t) \\ &= \begin{bmatrix} \dfrac{\partial f}{\partial x}\dfrac{\partial x}{\partial s} + \dfrac{\partial f}{\partial y}\dfrac{\partial y}{\partial s} + \dfrac{\partial f}{\partial z}\dfrac{\partial z}{\partial s} & \dfrac{\partial f}{\partial x}\dfrac{\partial x}{\partial t} + \dfrac{\partial f}{\partial y}\dfrac{\partial y}{\partial t} + \dfrac{\partial f}{\partial z}\dfrac{\partial z}{\partial t} \end{bmatrix}; \end{aligned}$$

the first entry says

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial s}$$

while the second says

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial t}.$$

We can think of changing coordinates for a function of two variables as the analogue of this when $\mathfrak{S}$ is the xy-plane. In this case we drop the third variable.

In particular, if a measurement is expressed as a function $m = f(x,y)$ of the rectangular coordinates, then its expression in terms of polar coordinates is $(f \circ (Pol))(r,\theta)$, where $Pol\colon \mathbb{R}^2 \to \mathbb{R}^2$ is the change-of-coordinates map

$$Pol(r,\theta) = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}$$

with Jacobian

$$J(Pol)(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} \tag{4.3}$$

and the Chain Rule tells us that

$$\begin{aligned}
\begin{bmatrix} \dfrac{\partial m}{\partial r} & \dfrac{\partial m}{\partial \theta} \end{bmatrix} &= J(f\circ Pol)(r,\theta) \\
&= Jf(x,y)\cdot J(Pol)(r,\theta) \\
&= \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\partial f}{\partial x}\cos\theta + \dfrac{\partial f}{\partial y}\sin\theta & -\dfrac{\partial f}{\partial x} r\sin\theta + \dfrac{\partial f}{\partial y} r\cos\theta \end{bmatrix}
\end{aligned}$$

in other words,

$$\begin{aligned}
\frac{\partial m}{\partial r} &= \frac{\partial m}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial m}{\partial y}\frac{\partial y}{\partial r} \\
&= (\cos\theta)\frac{\partial f}{\partial x} + (\sin\theta)\frac{\partial f}{\partial y}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial m}{\partial \theta} &= \frac{\partial m}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial m}{\partial y}\frac{\partial y}{\partial \theta} \\
&= (-r\sin\theta)\frac{\partial f}{\partial x} + (r\cos\theta)\frac{\partial f}{\partial y}.
\end{aligned}$$

For example, if

$$m = f(x,y) = \frac{y}{x}$$

then

$$m = \frac{r\sin\theta}{r\cos\theta} = \tan\theta;$$

using the Chain Rule, we have

$$\begin{aligned}
\frac{\partial m}{\partial r} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} \\
&= \left(-\frac{y}{x^2}\right)(\cos\theta) + \left(\frac{1}{x}\right)(\sin\theta) \\
&= \frac{-r\sin\theta\cos\theta}{r^2\cos^2\theta} + \frac{\sin\theta}{r\cos\theta} \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}\frac{\partial m}{\partial \theta} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} \\ &= \left(-\frac{y}{x^2}\right)(-r\sin\theta) + \left(\frac{1}{x}\right)(r\cos\theta) \\ &= \frac{r^2\sin^2\theta}{r^2\cos^2\theta} + \frac{r\sin\theta}{r\cos\theta} \\ &= \tan^2\theta + 1 \\ &= \sec^2\theta.\end{aligned}$$

While this may seem a long-winded way to go about performing the differentiation (why not just differentiate $\tan\theta$?), this point of view has some very useful theoretical consequences, which we shall see later.

Similarly, change-of-coordinate transformations in three variables can be handled via their Jacobians.

The transformation $Cyl: \mathbb{R}^3 \to \mathbb{R}^3$ going from cylindrical to rectangular coordinates is just Pol together with keeping z unchanged:

$$Cyl(r,\theta,z) = \begin{bmatrix} r\cos\theta \\ r\sin\theta \\ z \end{bmatrix}$$

with Jacobian

$$J(Cyl)(r,\theta,z) = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.4}$$

(Note that the upper-left 2×2 part of this is just $J(Pol)$.)

The transformation $Sph: \mathbb{R}^3 \to \mathbb{R}^3$ from spherical to rectangular coordinates is most easily understood as the composition of the transformation $SC: \mathbb{R}^3 \to \mathbb{R}^3$ from spherical to *cylindrical* coordinates

$$SC(\rho,\phi,\theta) = \begin{bmatrix} \rho\sin\phi \\ \theta \\ \rho\cos\phi \end{bmatrix}$$

with Jacobian

$$J(SC)(\rho,\phi,\theta) = \begin{bmatrix} \sin\phi & \rho\cos\phi & 0 \\ 0 & 0 & 1 \\ \cos\phi & -\rho\sin\phi & 0 \end{bmatrix} \tag{4.5}$$

and the transformation Cyl from cylindrical to rectangular coordinates, which we studied above. Then $Sph = (Cyl) \circ (SC)$, and its Jacobian is

$$\begin{aligned} J(Sph)(\rho,\phi,\theta) &= J((Cyl)\circ(SC))(\rho,\phi,\theta) \\ &= J(Cyl)(r,\theta,z)\cdot J(SC)(\rho,\phi,\theta) \\ &= \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sin\phi & \rho\cos\phi & 0 \\ 0 & 0 & 1 \\ \cos\phi & -\rho\sin\phi & 0 \end{bmatrix} \\ &= \begin{bmatrix} \sin\phi\cos\theta & \rho\cos\phi\cos\theta & -r\sin\theta \\ \sin\phi\sin\theta & \rho\cos\phi\sin\theta & r\cos\theta \\ \cos\phi & -\rho\sin\phi & 0 \end{bmatrix} \end{aligned}$$

and substituting $r = \rho\sin\phi$,

$$J(Sph)(\rho,\phi,\theta) = \begin{bmatrix} \sin\phi\cos\theta & \rho\cos\phi\cos\theta & -\rho\sin\phi\sin\theta \\ \sin\phi\sin\theta & \rho\cos\phi\sin\theta & \rho\sin\phi\cos\theta \\ \cos\phi & -\rho\sin\phi & 0 \end{bmatrix}. \tag{4.6}$$

This can be used to study motion which is most easily expressed in spherical coordinates. For example, suppose $\overrightarrow{p}(t)$, $0 < t < \pi$ is the curve on the unit sphere consisting of latitude decreasing and longitude increasing at a steady rate, given in spherical coordinates by

$$\begin{aligned} \rho &= 1 \\ \phi &= t \\ \theta &= 4t \end{aligned}$$

(Figure 4.4). To find the velocity (in rectangular coordinates) of this moving point as it crosses the equator at $t = \frac{\pi}{2}$, we note that since

$$\begin{aligned} \frac{d\rho}{dt} &= 0 \\ \frac{d\phi}{dt} &= 1 \\ \frac{d\theta}{dt} &= 4 \end{aligned}$$

and when $t = \pi/2$ (so $\phi = \pi/2$ and $\theta = 2\pi$),

$$\begin{aligned} \sin\phi &= 1 \\ \sin\theta &= 0 \\ \cos\phi &= 0, \\ \cos\theta &= 1, \end{aligned}$$

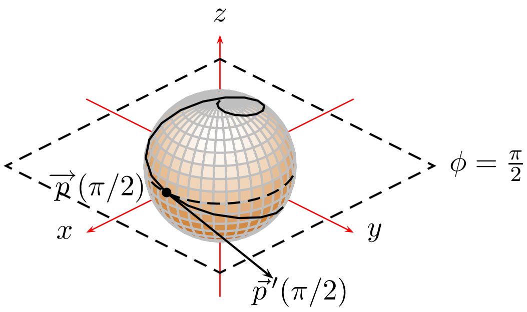


Figure 4.4: Curve on the Sphere

so

$$J(Sph)(1, \pi/2, 2\pi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

and we have

$$\begin{aligned} \left[\vec{p}'(\pi/2)\right] &= \begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix}. \end{aligned}$$

Exercises for § 4.2

Practice problems:

1. For each mapping below, find the Jacobian matrix $JF(\overrightarrow{x}_0)$ and the linearization $T_{\overrightarrow{x}_0}F$ at the given point $\overrightarrow{x}_0$:

   (a) $F(x,y) = (y,x)$, $\overrightarrow{x}_0 = (1,2)$.

   (b) $F(x,y) = (e^x \cos y, e^x \sin y)$, $\overrightarrow{x}_0 = (0, \frac{\pi}{3})$.

(c) $F(x,y) = (x^2+y^2, 2xy)$, $\overrightarrow{x}_0 = (-1,1)$.

(d) $F(x,y) = (x+y, x-y)$, $\overrightarrow{x}_0 = (-1,2)$.

(e) $F(x,y) = (x, y, x^2-y^2)$, $\overrightarrow{x}_0 = (2,-1)$.

(f) $F(x,y) = (x, y, xy)$, $\overrightarrow{x}_0 = (2,-1)$.

(g) $F(x,y) = (x-2y, x+y-1, 3x+5y)$, $\overrightarrow{x}_0 = (2,-1)$.

(h) $F(x,y) = (x^2, 2xy, y^2)$, $\overrightarrow{x}_0 = (1,-3)$.

(i) $F(x,y,z) = (y+z, xy+z, xz+y)$, $\overrightarrow{x}_0 = (2,-1,3)$.

(j) $F(x,y,z) = (xyz, x-y+z^2)$, $\overrightarrow{x}_0 = (2,1,-1)$.

2. In each part below, you are given a mapping described in terms of rectangular coordinates. Use the Chain Rule together with one of the equations (4.3), (4.4), (4.5), or (4.6) to find the indicated partial derivative when the input is given in one of the other coordinated systems.

   (a) $F(x,y) = (x^2-y^2, 2xy)$; find $\frac{\partial F_1}{\partial r}$ and $\frac{\partial F_2}{\partial \theta}$ at the point with polar coordinates $r = 2$ and $\theta = \frac{\pi}{3}$.

   (b) $F(x,y,z) = (x^2+y^2+z^2, xyz)$; find $\frac{\partial F_1}{\partial r}$ and $\frac{\partial F_2}{\partial \theta}$ at the point with cylindrical coordinates $r = 2$, $\theta = \frac{2\pi}{3}$, and $z = 1$.

   (c) $F(x,y,z) = (\frac{x}{1-z}, \frac{y}{1-z})$; find $\frac{\partial F_1}{\partial \rho}$, $\frac{\partial F_2}{\partial \phi}$ and $\frac{\partial F_2}{\partial \theta}$ at the point with spherical coordinates $\rho = 1$, $\phi = \frac{\pi}{2}$, and $\theta = \frac{\pi}{3}$.

   (d) $F(x,y,z) = (x^2+y^2+z^2, xy+yz, xyz)$; find $\frac{\partial F_1}{\partial \rho}$, $\frac{\partial F_2}{\partial \phi}$ and $\frac{\partial F_3}{\partial \theta}$ at the point with spherical coordinates $\rho = 4$, $\phi = \frac{\pi}{3}$, and $\theta = \frac{2\pi}{3}$.

Theory problems:

3. Prove Remark 4.2.4.

4. Show that for any linear map $L: \mathbb{R}^n \to \mathbb{R}^m$,

   (a) the number

$$\| L \| = \sup\{\|L(\overrightarrow{u})\| \mid \overrightarrow{u} \in \mathbb{R}^n \text{ and } \|\overrightarrow{u}\| = 1\} \tag{4.7}$$

   satisfies

$$\|L(\overrightarrow{x})\| \leq \| L \| \, \|\overrightarrow{x}\| \tag{4.8}$$

   for every vector $\overrightarrow{x} \in \mathbb{R}^n$;

   (b) $\| L \|$ is actually a maximum in Equation (4.7);

(c) $\| L \|$ is the least number satisfying Equation (4.8).

5. Find the operator norm $\| L \|$ for each linear map L below:

   (a) $L(x, y) = (y, x)$.

   (b) $L(x, y) = (x + y, x - y)$.

   (c) $L(x, y) = (x + y\sqrt{2}, x)$.

   (d) $L: \mathbb{R}^2 \to \mathbb{R}^2$ is reflection across the diagonal $x = y$.

   (e) $L: \mathbb{R}^2 \to \mathbb{R}^3$ defined by $L(x, y) = (x, x - y, x + y)$.

   (f) $L: \mathbb{R}^3 \to \mathbb{R}^3$ defined by $L(x, y, z) = (x, x - y, x + y)$.

4.3 Linear Systems of Equations

A system of equations can be viewed as a single equation involving a mapping. For example, the system of two equations in three unknowns

$$\left\{ \begin{array}{cccccc} x & +2y & +5z & = & 5 \\ 2x & + & y & +7z & = & 4 \end{array} \right. \tag{4.9}$$

can be viewed as the vector equation

$$L(\overrightarrow{x}) = \overrightarrow{y}$$

where $L: \mathbb{R}^3 \to \mathbb{R}^2$ is the mapping given by

$$[L(x, y, z)] = \begin{bmatrix} x + 2y + 5z \\ 2x + y + 7z \end{bmatrix}$$

and the right-hand side is the vector with

$$[\overrightarrow{y}] = \begin{bmatrix} 5 \\ 4 \end{bmatrix};$$

we want to solve for the unknown vector with

$$[\overrightarrow{x}] = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

We can think of the solution(s) of this system as the "level set" of the mapping corresponding to the output value $\overrightarrow{y}$.

Solving Linear Systems: Row Reduction

When the mapping is linear, as above, we can use a process of **elimination** to solve for some variables in terms of others, and ultimately to exhibit all solutions of the system in a convenient form. The process can be streamlined using matrix notation, and in this setting it is called **row reduction**. We review and illustrate this process briefly below; if you are not familiar with it, see § E.2 for a more detailed and motivated discussion.

The relevant data in a linear system consists of the coefficients of the different variables together with the numbers to the right of the "equals" signs. We display this in an array, called the **augmented matrix**, often separating the coefficients from the right-hand sides using a vertical bar. The augmented matrix of the system (4.9) is

$$[A|\overrightarrow{y}] = \left[\begin{array}{ccc|c} 1 & 2 & 5 & 5 \\ 2 & 1 & 7 & 4 \end{array}\right].$$

The **matrix of coefficients**—the subarray to the left of the bar—is the matrix representative of the linear mapping L

$$A = [L] = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 1 & 7 \end{bmatrix},$$

while the array to the right is the coordinate column of the vector $\overrightarrow{y}$; we will often refer to the augmented matrix using the notation $[A|\overrightarrow{y}]$.

The process of row reduction uses three **row operations** on a matrix:

1. Multiply all entries of one row by the same nonzero number.

2. Add (or subtract) to a *given* row a multiple of *another* row (the latter remains unchanged).

3. Occasionally, we need to rearrange the order in which the rows occur: the basic operation is an interchange of two rows.

Our goal is to end up with a matrix in **reduced row-echelon form** (or, informally, a **reduced matrix**), characterized by the following conditions:

- The leading entry in any (nonzero) row is a 1.

- The leading entries move right as one goes down the rows, with any rows consisting entirely of zeroes appearing at the bottom.

- A leading entry is the only nonzero entry in its *column*.

When this process is applied to the augmented matrix of a system of linear equations, the successive augmented matrices represent systems with exactly the same set of solutions (Exercise 5). A system whose augmented matrix is in reduced row-echelon form exhibits its solutions explicitly. We illustrate with a few examples.

A reduction of the augmented matrix for Equation (4.9) can be summarized as follows:

$$\left[\begin{array}{ccc|c} 1 & 2 & 5 & 5 \\ 2 & 1 & 7 & 4 \end{array}\right] \mapsto \left[\begin{array}{ccc|c} 1 & 2 & 5 & 5 \\ 0 & -3 & -3 & -6 \end{array}\right] \mapsto \left[\begin{array}{ccc|c} 1 & 2 & 5 & 5 \\ 0 & 1 & 1 & 2 \end{array}\right] \mapsto \left[\begin{array}{ccc|c} 1 & 0 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{array}\right].$$

The steps here are the following:

1. Subtract twice row 1 from row 2 (leaving row 1 unchanged) to get 0 below the leading 1 in the first row.

2. Divide the second row by its leading entry (-3) to make its leading entry 1.

3. Subtract twice row 2 from row 1 to get 0 above the leading 1 in row 2.

The system whose augmented matrix is the reduced one is

$$\left\{\begin{array}{cccc} x & & +3z & = & 1 \\ & y & +z & = & 2 \end{array}\right. .$$

This can be rewritten as expressing the **leading variables** x and y in terms of z:

$$\left\{\begin{array}{cccc} x & = & 1 & -3z \\ y & = & 2 & -z \end{array}\right. .$$

The value of z is not constrained by any equation in the system: it is a **free variable**. As we pick different values for the free variable, we run through all the possible solutions of the system. In other words, choosing a parameter t and setting z equal to it we can exhibit the solutions of the system (4.9) (*i.e.,* the level set $\mathcal{L}(L, \overrightarrow{y})$) as the *line* in space parametrized by

$$\begin{cases} x &= 1 & -3t \\ y &= 2 & -t \\ z &= & t \end{cases}$$

or

$$\overrightarrow{p}(t) = (1 - 3t, 2 - t, t).$$

As another example, consider the system

$$\begin{cases} & 2y & +z & = & 1 \\ 2x & -y & -z & = & 3 \\ x & +y & +z & = & 0 \end{cases}.$$

The reduction of its augmented matrix is summarized below. The first step is a row interchange, and in subsequent steps we have indicated in bold face the leading entries being used to clear the various columns.

$$\left[\begin{array}{rrr|r} 0 & 2 & 1 & 1 \\ 2 & -1 & -1 & 3 \\ 1 & 1 & 1 & 0 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 1 & 1 & 0 \\ 2 & -1 & -1 & 3 \\ 0 & 2 & 1 & 1 \end{array}\right]$$

$$\mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 1 & 1 & 0 \\ 0 & -3 & -3 & 3 \\ 0 & 2 & 1 & 1 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 1 & 1 & 0 \\ 0 & \mathbf{1} & 1 & -1 \\ 0 & 2 & 1 & 1 \end{array}\right]$$

$$\mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 0 & 0 & 1 \\ 0 & \mathbf{1} & 1 & -1 \\ 0 & 0 & -1 & 3 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 0 & 0 & 1 \\ 0 & \mathbf{1} & 0 & 2 \\ 0 & 0 & \mathbf{1} & -3 \end{array}\right].$$

The last matrix represents the system

$$\begin{array}{ccccc} x & & & = & 1 \\ & y & & = & 2 \\ & & z & = & -3 \end{array}$$

which clearly exhibits the *unique* solution

$$(x, y, z) = (1, 2, -3)$$

of the system.

The system

$$\begin{array}{ccccccc} x & + & y & -2z & = & 1 \\ x & & -2y & +z & = & 7 \\ x & & +7y & -8z & = & 4 \end{array}$$

has an augmented matrix which reduces according to

$$\begin{aligned}
&\left[\begin{array}{rrr|r} \mathbf{1} & 1 & -2 & 1 \\ 1 & -2 & 1 & 7 \\ 1 & 7 & -8 & 4 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 1 & -2 & 1 \\ 0 & -3 & 3 & 6 \\ 0 & 6 & -6 & 3 \end{array}\right] \mapsto \\
&\qquad \mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 1 & -2 & 1 \\ 0 & \mathbf{1} & -1 & -2 \\ 0 & 6 & -6 & 3 \end{array}\right] \mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 0 & -1 & 3 \\ 0 & \mathbf{1} & -1 & -2 \\ 0 & 0 & 0 & 15 \end{array}\right] \mapsto \\
&\qquad\qquad \mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 0 & -1 & 0 \\ 0 & \mathbf{1} & -1 & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{array}\right].
\end{aligned}$$

The last matrix represents the system

$$\begin{array}{rrcl} x & -z & = & 0 \\ y & -z & = & 0 \\ & 0 & = & 1. \end{array}$$

The first two equations look fine—as in the first example, they express their leading variables in terms of the free variable z. However, the third equation, $0 = 1$, has *no* solutions. Thus the *full* system of three equations has no solutions, implying that the same is true of the original system: it is **inconsistent**. You should check that this occurs precisely if the last *column* contains a leading entry of some row in the reduced matrix.

You undoubtedly noted in this last example that the leading entries skipped a column. This does not necessarily imply inconsistency of the system: for example, you should check that if the right side of the third equation in the original system of the last example had been -11 instead of 4, the reduction would have led to

$$\left[\begin{array}{rrr|r} \mathbf{1} & 1 & -2 & 1 \\ 1 & -2 & 1 & 7 \\ 1 & 7 & -8 & -11 \end{array}\right] \mapsto \cdots \mapsto \left[\begin{array}{rrr|r} \mathbf{1} & 0 & -1 & 3 \\ 0 & \mathbf{1} & -1 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The system corresponding to this matrix

$$\begin{array}{rrcr} x & -z & = & 3 \\ y & -z & = & -2 \\ & 0 & = & 0 \end{array}$$

has a *line* of solutions, determined by the first and second equations; the third equation is *always* satisfied.

The scenarios we have seen in these examples reflect the different ways that planes can intersect in space:

- In general, a linear equation in three variables determines a *plane* in $\mathbb{R}^3$; thus *two* linear equations in three variables represent the intersection of two planes, which we expect (with rare exceptions) to be a *line*, as in our first example, and *three* equations are expected to determine a single *point*, as in the second.

- However, if two of the equations represent the same plane, then geometrically the solutions are really just the intersection of *two* planes; in this case reduction will eventually lead to a row of zeroes. If *all three* equations represent the same plane, reduction will lead to *two* rows of zeroes.

- Even if the three planes are distinct, they can intersect along a common line; in this case reduction will still result in a row of zeroes. Algebraically, this means that one of the equations can be deduced directly from the other two, so the situation is the same as if there were only two equations present (this occurs in the last of our four examples).

- If at least two of the planes are parallel, then algebraically we have two equations which can be written with the same coefficients, but different right-hand sides: this will result in a row which is zero, except for a nonzero entry in the last column (so there will be a leading entry in the last column, as in our third example). Geometrically, no point belongs to all three planes, so there are no solutions. Even if no two of the planes are parallel, their pairwise intersection lines might include two parallel lines. This will also yield a leading entry in the last column, indicating that again there are no solutions to the system.

You should work out the possibilities for systems of equations in two unknowns, in terms of the arrangements of lines in the plane (Exercise 6).

Our intuition—that each equation of a system "eliminates" one variable, in the sense that for three variables, a *single* equation should lead to a plane of solutions, a *pair* leads to a line of solutions, and *three* equations have a unique solution—needs to be modified, using the **rank** of a matrix. This can be defined as the number of independent rows in the matrix, or equivalently as the number of *nonzero* rows in the equivalent reduced matrix. Interpreted in terms of a system of equations, the rank of the augmented matrix is the same as the number of algebraically independent equations in the system; if the number of rows (equations) exceeds the rank, then some equations in the system can be algebraically obtained from the others, and therefore are redundant, as far as solving the system is concerned; the process of row reduction replaces these rows with rows of zeroes.

The rank of a matrix is clearly no more than the total number of rows, or "height" of the matrix. Thus, the solution set of a system of two equations in three unknowns is *at least* a line (so uniqueness of solutions is impossible), but it may happen that the two equations are multiples of each other, so the solutions form a plane. There is also the possibility that the *left* sides of the two equations are multiples of each other, but the *right* sides are inconsistent with this, so that there are *no* solutions. One way to codify this is to compare the rank of the *coefficient* matrix A with that of the *augmented* matrix $[A|\overrightarrow{y}]$. Since row reduction proceeds from left to right, a reduction of the augmented matrix $[A|\overrightarrow{y}]$ includes a reduction of the coefficient matrix A, so the rank of A equals the number of leading entries occurring to the left of the vertical line in the reduced matrix equivalent to $[A|\overrightarrow{y}]$. We see, then, that the rank of A either equals that of $[A|\overrightarrow{y}]$—in which case the system is consistent, and the solution set is nonempty—or it is one less—in which case the system is inconsistent, and there are no solutions. When there *are* solutions, the geometric nature of the solution set is determined by the rank of the coefficient matrix: the rank of A equals the number of leading variables, which are determined as affine functions of the remaining, *free* variables. So for a system of linear equations in three variables whose coefficient matrix has rank r, there are $3 - r$ free variables: when $r = 3$, solutions are unique, when $r = 2$, any nonempty solution set is a line, and when $r = 1$, it is a plane in space.[9]

Linear Systems and Linear Mappings

When we think of a system of equations in terms of the level set of a linear mapping, we use the fact that the coefficient matrix of the system $L(\overrightarrow{x}) = \overrightarrow{y}$ is the matrix representative of the related mapping L, and refer to the rank of the matrix representative $A = [L]$ as the **rank** of L. While this is by definition the number of independent *rows* in A, it can be shown (Exercise 7) that the rank also gives the number of independent *columns* in A. A quick calculation (Exercise 8) shows that the columns of A are the coordinate columns of the vectors $L(\overrightarrow{e}_j)$, the images under L of the standard basis vectors $\overrightarrow{e}_j$, and the full **image** of L consists of all linear combinations of these vectors. Looked at differently, the full image of L, the set of all outputs of L, is the set of all possible right-hand sides $\overrightarrow{y}$ in the output

[9] Can the rank of a matrix equal zero? When? What happens then?

space for which the system $L(\overrightarrow{x}) = \overrightarrow{y}$ has at least one solution, and this has dimension equal to the rank of L. On the input side, for each such $\overrightarrow{y}$ the solution set of the system $L(\overrightarrow{x}) = \overrightarrow{y}$ is the **preimage** of $\overrightarrow{y}$ under L,

$$(L)^{-1}(\overrightarrow{y}) := \{\overrightarrow{x} \mid L(\overrightarrow{x}) = \overrightarrow{y}\}.$$

This set is parametrized by the number of free variables, which is the difference between the "width" of A (the dimension of the input space) and its rank: this is sometimes called the **nullity** of A; stated differently, the rank of L is the difference between the dimension of (nonempty) preimages under L and the dimension of the input space where they live—this is sometimes called the **codimension** of the preimages. To summarize:

Remark 4.3.1. *If $A = [L]$ is the matrix representative of the linear mapping L, then the rank of A is the dimension of the image of L, and for every element $\overrightarrow{y}$ of the image, its preimage has codimension equal to this rank.*

In particular, the rank of A cannot exceed either the height of A (the output dimension of L) or its width (the input dimension). When the rank of L equals the output dimension, the mapping L is **onto**: every point in the target is hit by at least one input via L—or equivalently, every system of the form $L(\overrightarrow{x}) = \overrightarrow{y}$ has at least one solution. By contrast, when the rank of A equals the input dimension, the solution to the system $L(\overrightarrow{x}) = \overrightarrow{y}$, if nonempty, is a single point: the mapping is **one-to-one**.

We see that the only way that both conditions can possibly hold is if the matrix A (and, by abuse of language, the system) is **square**: its width equals its height. Of course, the rank of a square matrix need not equal its size–it could be less. But in a sense the "typical" matrix has rank as high as its dimensions allow. When a square matrix *has* the maximal possible rank, the corresponding map is both one-to-one and onto; using the French-derived term, it is **bijective**. A bijective map L automatically has an **inverse** L^{-1}, defined by

$$\overrightarrow{x} = (L)^{-1}(\overrightarrow{y}) \Leftrightarrow L(\overrightarrow{x}) = \overrightarrow{y}.$$

In terms of equations, this is the map assigning to each $\overrightarrow{y}$ in the full range of L the unique $\overrightarrow{x}$ satisfying $L(\overrightarrow{x}) = \overrightarrow{y}$. It is fairly straightforward to see that the inverse of a bijective linear mapping is itself also linear (Exercise 9). The definition of the inverse can be formulated in terms of two different equations:

$$(L \circ L^{-1})(\overrightarrow{y}) = \overrightarrow{y}$$

and

$$(L^{-1} \circ L)(\overrightarrow{x}) = \overrightarrow{x}.$$

If we define the **identity map** id to be the "trivial" map which uses its input directly as output, we can express the defining equations of L^{-1} as

$$L \circ L^{-1} = \mathrm{id} = L^{-1} \circ L. \tag{4.10}$$

In matrix terms, we can define the **identity matrix** I, which has every diagonal entry equal to 1 and every off-diagonal entry zero: the 3×3 version is

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The identity matrix has no effect on any vector: $I\overrightarrow{x} = \overrightarrow{x}$, so I is the matrix representative of id, and Equation (4.10) can be written in terms of matrix representatives: if $A = [L]$ and $B = \left[L^{-1}\right]$:

$$AB = I = BA. \tag{4.11}$$

Given a 3×3 matrix A of rank 3, there is a unique 3×3 matrix B satisfying Equation (4.11); it is called the **inverse** of A, and denoted A^{-1}; in this case A is called **invertible** or **nonsingular**.

One way to decide whether a matrix is invertible and, if it is, to find its inverse, is to take advantage of Remark 4.1.1, which says that the columns of the matrix representative $[L]$ of a linear mapping are the coordinate columns of the images $L(\overrightarrow{e}_j)$ of the standard basis. If $A = [L]$, then the columns of $A^{-1} = \left[L^{-1}\right]$ are the coordinate columns of the *pre*-images of the standard basis elements $\overrightarrow{e}_j$; that is, they are solutions of the systems of equations $L(\overrightarrow{x}) = \overrightarrow{e}_j$. The coefficient matrix of all these systems (for different j) is A, so we can attempt to simultaneously solve all of them using the **super-augmented matrix** $[A|I]$ consisting of A augmented by the coordinate columns of the standard basis, which together form the identity matrix. If we try to row-reduce this matrix, one of two things will happen: either we will end up with a leading entry to the right of the vertical bar, indicating that at least one of the systems has no solution—that is, one of the standard basis elements has no preimage under the corresponding mapping L (and hence L is not onto), so the matrix is not invertible, or the leading entries all appear to the left of the vertical bar. In this case, the submatrix appearing to the left of the bar is the identity matrix, and the submatrix to the right of the bar is A^{-1}:

Remark 4.3.2. *If A is invertible, then row-reduction of the super-augmented matrix $[A|I]$ results in the matrix $[I|A^{-1}]$.*

Let us consider an example. To check whether the matrix

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

is invertible, we form the super-augmented matrix

$$[A|I] = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 0 & 1 \end{array}\right]$$

and row-reduce:

$$\left[\begin{array}{ccc|ccc} \mathbf{1} & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \mapsto \left[\begin{array}{ccc|ccc} \mathbf{1} & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & -1 & -1 & -2 & 0 & 1 \end{array}\right] \mapsto$$

$$\mapsto \left[\begin{array}{ccc|ccc} \mathbf{1} & 1 & 1 & 1 & 0 & 0 \\ 0 & \mathbf{1} & 1 & 2 & 0 & -1 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right] \mapsto \left[\begin{array}{ccc|ccc} \mathbf{1} & 0 & 0 & -1 & 0 & 1 \\ 0 & \mathbf{1} & 1 & 2 & 0 & -1 \\ 0 & 0 & \mathbf{1} & -1 & 1 & 0 \end{array}\right] \mapsto$$

$$\mapsto \left[\begin{array}{ccc|ccc} \mathbf{1} & 0 & 0 & -1 & 0 & 1 \\ 0 & \mathbf{1} & 0 & 3 & -1 & -1 \\ 0 & 0 & \mathbf{1} & -1 & 1 & 0 \end{array}\right].$$

Thus, we can read off that A is invertible, with inverse

$$A^{-1} = \begin{bmatrix} -1 & 0 & 1 \\ 3 & -1 & -1 \\ -1 & 1 & 0 \end{bmatrix}.$$

Another useful criterion for invertibility of a matrix is via determinants. Recall that the determinant of a 3×3 matrix A is the signed volume of the parallelepiped whose sides are the rows of A, regarded as vectors. A is invertible precisely if it has rank 3, which is to say its rows are linearly independent—and this in turn means that they span a parallelepiped of nonzero volume. Thus

Remark 4.3.3. *A 3×3 matrix A is invertible if and only if its determinant is nonzero.*

An alternative argument for this is outlined in § F.3.

Exercises for § 4.3

Practice problems:

1. Which of the matrices below are in reduced row-echelon form?

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

(c) $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$

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

2. For each system of equations below, (i) write down the augmented matrix; (ii) row-reduce it; (iii) write down the corresponding system of equations; (iv) give the solution set in parametrized form, or explain why there are no solutions.

(a) $\left\{\begin{array}{rrcr} 2x & -y & = & 3 \\ x & +3y & = & -2 \end{array}\right.$ (b) $\left\{\begin{array}{rrcr} 4x & +2y & = & 2 \\ 2x & +y & = & 1 \end{array}\right.$

(c) $\left\{\begin{array}{rrrcr} x & +2y & +8z & = & 5 \\ 2x & +y & +7z & = & 4 \end{array}\right.$ (d) $\left\{\begin{array}{rrrcr} x & +y & +3z & = & 3 \\ x & +2y & +5z & = & 4 \\ 2x & +y & +4z & = & 5 \end{array}\right.$

(e) $\left\{\begin{array}{rrrcr} x & +2y & +5z & = & 4 \\ 2x & +y & 4z & = & 5 \\ 3x & +2y & +3z & = & 2 \end{array}\right.$ (f) $\left\{\begin{array}{rrrcr} x & +2y & +3z & = & 6 \\ 2x & -y & z & = & 2 \\ x & +y & +2z & = & 5 \end{array}\right.$.

3. Find the rank and nullity of each matrix below:

(a) $\begin{bmatrix} 1 & 2 & 4 \\ 2 & 1 & 5 \\ 1 & 1 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -3 & -6 & -9 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 3 \\ 1 & 2 & 3 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$

4. Find the inverse of each matrix below, or show that none exists.

(a) $\begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 6 \\ 1 & -1 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 6 \\ 2 & 3 & 4 \end{bmatrix}$

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

(e) $\begin{bmatrix} 2 & 3 & 1 \\ 4 & 4 & 3 \\ 3 & 3 & 2 \end{bmatrix}$ (f) $\begin{bmatrix} 3 & -1 & 1 \\ 1 & 1 & 3 \\ 1 & 2 & 5 \end{bmatrix}$

Theory problems:

5. Show that row equivalent matrices represent systems with the same set of solutions, as follows:
   (a) Clearly, interchanging rows and multiplying a row by a nonzero number corresponds to operations on the equations which don't change the solutions.
   (b) Show that, if we replace a single equation in a system with the difference between it and another equation (and leave all other equations unchanged) then any solution of the old system is also a solution of the new system. That is, we don't *lose* any solutions via this operation.
   (c) Show that we can get from the new system in the previous item to the old one by a similar operation. That is, every solution of the *new* system also solves the *old* system—the original operation did not *create* any new solutions.

6. In general, a single linear equation in two variables determines a line in the plane. Two equations then determine a pair of lines; describe what the solution set of the resulting system is when the lines are the same, parallel, or non-parallel. What are the possible configurations for *three* equations in two variables?

7. (a) Show that, in a reduced matrix, the columns which contain the leading entries of the rows are linearly independent.
   (b) Show that, in a reduced matrix, a column which does not contain a leading entry of some row is a linear combination of the columns containing leading entries to its left.

(c) Conclude that the number of independent rows in a reduced matrix equals the number of independent columns.

(d) Show that, if one column of a matrix is a linear combination of other columns, then this relation is unchanged by row operations.

(e) Conclude that the number of independent rows in any matrix equals the number of independent columns in that matrix.

8. Show that the columns of the matrix representative $A = [L]$ of a linear mapping $L: \mathbb{R}^3 \to \mathbb{R}^3$ are the coordinate columns of $L(\overrightarrow{\imath})$, $L(\overrightarrow{\jmath})$ and $L\left(\overrightarrow{k}\right)$. (*Hint:* What are the coordinate columns of $\overrightarrow{\imath}$, $\overrightarrow{\jmath}$, and $\overrightarrow{k}$?)

Challenge problem:

9. Suppose $L: \mathbb{R}^3 \to \mathbb{R}^3$ is linear and bijective. Show that for any two vectors $\overrightarrow{x}_1$ and $\overrightarrow{x}_2$ and scalars α_1, α_2,

$$(L)^{-1}(\alpha_1 \overrightarrow{x}_1 + \alpha_2 \overrightarrow{x}_2) = \alpha_1 \, (L)^{-1}(\overrightarrow{x}_1) + \alpha_2 \, (L)^{-1}(\overrightarrow{x}_2)\,.$$

4.4 Nonlinear Systems of Equations: The Implicit and Inverse Mapping Theorems

How does the discussion of linear systems in the preceding section extend to *non*-linear systems of equations? We cannot expect to always solve such equations by some systematic method analogous to row reduction. However, we saw in § 3.4 that for a real-valued function of 2 or 3 variables, the level set corresponding to a regular value is *locally* the graph of a function (Theorems 3.4.2 and 3.4.3, the Implicit Function Theorem).

This result can be generalized to arbitrary differentiable mappings: the general philosophy is that the linearization of the mapping gives us local information about the nonlinear mapping.

Two Equations in Two Unknowns: The Inverse Mapping Theorem

A system of two equations in two unknowns

$$\begin{cases} f_1(x,y) &= a \\ f_2(x,y) &= b \end{cases}$$

can be interpreted as the vector equation $F(\overrightarrow{x}) = \overrightarrow{y}$, where

$$\overrightarrow{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \overrightarrow{y} = \begin{bmatrix} a \\ b \end{bmatrix},$$

and $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$F(\overrightarrow{x}) = \begin{bmatrix} f_1(\overrightarrow{x}) \\ f_2(\overrightarrow{x}) \end{bmatrix}.$$

The analogous situation for one equation in one unknown is that if the real-valued function f of one real variable (*i.e.*, $f\colon \mathbb{R}^1 \to \mathbb{R}^1$) has nonvanishing derivative $f'(x_0)$ at x_0 then it has an inverse $g = f^{-1}$ defined (at least) on a neighborhood $(x_0 - \varepsilon, x_0 + \varepsilon)$ of x_0, and the derivative of the inverse is the reciprocal of the derivative: writing $f(x_0) = y_0$,

$$g'(y_0) = 1/f'(x_0).$$

In other words, if x_0 is a regular point of f then f is locally invertible there, and the derivative of the inverse is the inverse of the derivative.

To extend this to the situation of $F\colon \mathbb{R}^2 \to \mathbb{R}^2$, we need first to define what we mean by a regular point. Motivated by the linear situation, we take it to mean that the rank of the derivative is as high as possible, given the dimensions of input and output.

Definition 4.4.1. *Suppose $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a differentiable mapping with domain an open set $\mathcal{D} \subset \mathbb{R}^2$.*

1. *A point $\overrightarrow{x}_0 \in \mathcal{D}$ is a **regular point** of F if the rank of $DF_{\overrightarrow{x}_0}$ is 2.*

2. *It is a **critical point** of F otherwise—that is, if the rank of $DF_{\overrightarrow{x}_0}$ is 1 or 0.*

Recall that a point in the domain of a real-valued function $f(\overrightarrow{x})$ on $\mathbb{R}^2$ is *regular* if the gradient is nonvanishing there, which is the same as saying that the rank of the derivative is 1, while it is *critical* if the gradient is the zero vector (all partials are zero)—that is, the rank of the derivative is zero. If the function is *continuously* differentiable, then all points near a regular point are also regular.

For a mapping

$$F(\overrightarrow{x}) = \begin{bmatrix} f_1(\overrightarrow{x}) \\ f_2(\overrightarrow{x}) \end{bmatrix}$$

a point is regular if the two gradients $\overrightarrow{\nabla} f_1$ and $\overrightarrow{\nabla} f_2$ are linearly independent. Thus a point can be critical in two ways: if it is a critical point of one of the component functions, or if it is regular for both, but their gradients at the point are parallel (this is equivalent to saying that the two component functions have first-order contact at the point). Again, if the mapping is *continuously* differentiable (*i.e.*, both component functions are $\mathcal{C}^1$), then every regular point has a neighborhood consisting of regular points.

The two gradients $\overrightarrow{\nabla} f_1$ and $\overrightarrow{\nabla} f_2$ are linearly independent if and only if the triangle they form has nonzero area , that is, if the determinant of partials $\left|\frac{\partial(f_1,f_2)}{\partial(x,y)}\right|$ is nonzero:

$$\begin{aligned}\left|\frac{\partial\left(f_1,f_2\right)}{\partial\left(x,y\right)}\right| &:= \det\begin{bmatrix}\partial f_1/\partial x & \partial f_1/\partial y\\ \partial f_2/\partial x & \partial f_2/\partial y\end{bmatrix}\\ &= \frac{\partial f_1}{\partial x}\frac{\partial f_2}{\partial y} - \frac{\partial f_1}{\partial y}\frac{\partial f_2}{\partial x}\\ &\neq 0.\end{aligned}$$

To generalize the differentiation formula from one to two variables, we should reinterpret the derivative $f'(x_0)$ of a function of one variable—which is a number—as the (1×1) matrix representative of the derivative "mapping" $Df_{x_0}\colon\mathbb{R}^1\to\mathbb{R}^1$, which multiplies every input by $f'(x_0)$. The inverse of *multiplying* by a number is *dividing* by it, which is *multiplying* by its *reciprocal*. This point of view leads naturally to the following formulation for plane mappings analogous to the situation for mappings of the real line:

Theorem 4.4.2 (Inverse Mapping Theorem for $\mathbb{R}^2$)**.** *Suppose*

$$F(\overrightarrow{x}) = \begin{bmatrix} f_1(\overrightarrow{x})\\ f_2(\overrightarrow{x})\end{bmatrix}$$

is a $\mathcal{C}^1$ mapping of the plane to itself, and $\overrightarrow{x}_0$ is a regular point for F—that is, its Jacobian determinant at $\overrightarrow{x}_0$ is nonzero:

$$\begin{aligned}\left|\frac{\partial\left(f_1,f_2\right)}{\partial\left(x,y\right)}\right|(\overrightarrow{x}_0) &:= \det\begin{bmatrix}\partial f_1/\partial x(\overrightarrow{x}_0) & \partial f_1/\partial y(\overrightarrow{x}_0)\\ \partial f_2/\partial x(\overrightarrow{x}_0) & \partial f_2/\partial y(\overrightarrow{x}_0)\end{bmatrix}\\ &= \frac{\partial f_1}{\partial x}(\overrightarrow{x}_0)\frac{\partial f_2}{\partial y}(\overrightarrow{x}_0) - \frac{\partial f_1}{\partial y}(\overrightarrow{x}_0)\frac{\partial f_2}{\partial x}(\overrightarrow{x}_0)\\ &\neq 0.\end{aligned}$$

Then F is locally invertible at $\overrightarrow{x}_0$: there exist neighborhoods V of $\overrightarrow{x}_0$ and W of $\overrightarrow{y}_0 = F(\overrightarrow{x}_0) = (c,d)$ such that $F(V) = W$, together with a $\mathcal{C}^1$

mapping $G = F^{-1}\colon W \to V$ which is the inverse of F (restricted to V):

$$G(\overrightarrow{y}) = \overrightarrow{x} \Leftrightarrow \overrightarrow{y} = F(\overrightarrow{x}_0)\,.$$

Furthermore, the derivative of G at y_0 is the inverse of the derivative of F at $\overrightarrow{x}_0$:

$$DF^{-1}_{\overrightarrow{y}_0} = \left(DF_{\overrightarrow{x}_0}\right)^{-1} \tag{4.12}$$

(equivalently, the linearization of the inverse is the inverse of the linearization.)

The proof of Theorem 4.4.2 has three parts:

Claim 1: *There exist neighborhoods V of $\overrightarrow{x}_0$ and W of $\overrightarrow{y}_0$ such that F maps V one-to-one onto W.*

Claim 2: *The inverse mapping $G\colon W \to V$ is continuous.*

Claim 3: *G is $\mathcal{C}^1$ and Equation (4.12) holds.*

Claim 1 requires a proof specific to our situation in $\mathbb{R}^2$, which we give below. However, Claims 2 and 3 will be useful in our later contexts, and their proof does not rely on the dimension of the ambient space, so we will formulate and prove them as independent lemmas:

Lemma 4.4.3. *If $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuous on $V \subset \mathbb{R}^n$ and maps V onto $W \subset \mathbb{R}^n$ in a one-to-one way, then the inverse map $G\colon W \to V$ defined by*

$$G(y) = x \Leftrightarrow y = F(x)$$

is continuous on W.

Lemma 4.4.4. *If $F\colon \mathbb{R}^n \to \mathbb{R}^n$ maps $V \subset \mathbb{R}^n$ onto $W \subset \mathbb{R}^n$ in a one-to-one way, and is differentiable at $\overrightarrow{x}_0$ with invertible derivative map $DF_{\overrightarrow{x}_0}$, then the inverse map $G\colon W \to V$ is differentiable at $\overrightarrow{y}_0 = F(\overrightarrow{x}_0)$, and*

$$DG_{\overrightarrow{y}_0} = \left(DF_{\overrightarrow{x}_0}\right)^{-1}. \tag{4.13}$$

Proof of Claim 1:

In order to have

$$\left|\frac{\partial\,(f_1, f_2)}{\partial\,(x, y)}\right| = \frac{\partial f_1}{\partial x}\frac{\partial f_2}{\partial y} - \frac{\partial f_1}{\partial y}\frac{\partial f_2}{\partial x} \neq 0,$$

at least one of the two products is nonzero; we will assume the first one is nonzero at $\overrightarrow{x}_0$, so

$$\frac{\partial f_1}{\partial x}(\overrightarrow{x}_0) \neq 0$$

and

$$\frac{\partial f_2}{\partial y}(\overrightarrow{x}_0) \neq 0.$$

Applying the Implicit Function Theorem (Theorem 3.4.2) to f_2, we see that the level set $\mathcal{L}(f_2, d)$ of f_2 through $\overrightarrow{x}_0$ is locally the graph of a $\mathcal{C}^1$ function $y = \varphi(x)$, and similarly the level set $\mathcal{L}(f_1, c)$ of f_1 through $\overrightarrow{x}_0$ is locally the graph of a $\mathcal{C}^1$ function $x = \psi(y)$. Let R be the rectangle $[x_0 - \varepsilon_1, x_0 + \varepsilon_1] \times [y_0 - \varepsilon_2, y_0 + \varepsilon_2]$ where the function φ is defined on $[x_0 - \varepsilon_1, x_0 + \varepsilon_1]$ and ψ is defined on $[y_0 - \varepsilon_2, y_0 + \varepsilon_2]$. By picking ε_1 and ε_2 sufficiently small, we can assume that the graphs of φ and ψ are contained in R (Figure 4.5).

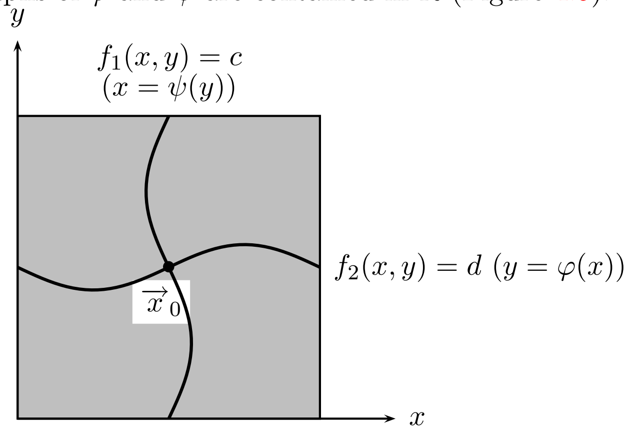


Figure 4.5: The level sets of f_1 and f_2 through $\vec{x}_0$ in R.

Now every point near $\overrightarrow{x}_0$ also has $\frac{\partial f_2}{\partial y}$ and $\frac{\partial f_1}{\partial x}$ nonzero, so picking ε_1 and ε_2 sufficiently small, we can assume that this holds at every point of R. Again the Implicit Function Theorem guarantees that the level set of f_2 (*resp.* of f_1) through any point of R is the graph of y as a $\mathcal{C}^1$ function of x (*resp.* of x as a function of y). However, *a priori* there is no guarantee that these new functions are defined over a whole edge of R. But suppose some level set of f_2 in R can *not* be given as the graph of a function defined on all of $[x_0 - \varepsilon_1, x_0 + \varepsilon_1]$. Then the endpoint of this curve cannot be interior to R, and hence must lie on the bottom or top edge of R. We claim that this does not happen for level sets which come close to $\overrightarrow{x}_0$.

Assuming for simplicity that $\frac{\partial f_2}{\partial y}$ is positive at all points of R, we can find a positive lower bound m for $\frac{\partial f_2}{\partial y}$. This means that for any point $\overrightarrow{x}$ on the level set $\mathcal{L}(f_2, d)$ through $\overrightarrow{x}_0$, the value of f_2 at the point on the top (*resp.* bottom) edge of R directly above (*resp.* below) $\overrightarrow{x}$ is at least $d + m\delta_+$ (*resp.* at most $d - m\delta_-$) where δ_+ (*resp.* δ_-) is the distance from $\overrightarrow{x}$ to the top (*resp.* bottom) edge of R. But these numbers are bounded away from zero, so the level sets $\mathcal{L}(f_2, d + \triangle d)$, for $|\triangle d|$ sufficiently small, are curves going from the left to the right edge of R—that is, they are the graphs of functions defined along the whole bottom edge of R. Similarly, all level sets of f_1 close to the one through $\overrightarrow{x}_0$ are graphs of $\mathcal{C}^1$ functions of x as a function of y defined along the whole left edge of R. Form a curvilinear quadrilateral from two pairs of such graphs, one pair $\mathcal{L}(f_2, d \pm \triangle d)$ bracketing $\mathcal{L}(f_2, d)$, and the other pair $\mathcal{L}(f1, c \pm \triangle c)$ bracketing $\mathcal{L}(f_1, c)$; call it V (Figure 4.6).

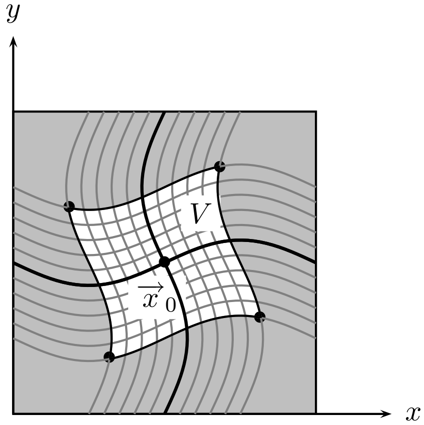


Figure 4.6: The quadrilateral V

The four corners of V map under F to the four points $(c \pm \triangle c, d \pm \triangle d)$; let W be the rectangle with these corners (Figure 4.7). We claim V maps one-to-one onto W.

Given a point $(a, b) \in W$, we know that a point (x, y) with $F(x, y) = (a, b)$, if it exists, must be a point of intersection of $\mathcal{L}(f_1, a)$ and $\mathcal{L}(f_2, b)$. Start at the lower-left corner of V, which maps to $(c - \triangle c, d - \triangle d)$. The value of f_2 there is *at most* b, while at the upper-left corner it is *at least* b; thus the Intermediate Value Theorem guarantees that somewhere along the left edge of V we have $f_2 = b$; moreover, since the coordinate y increases

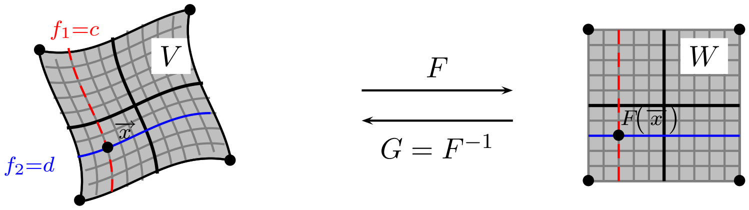


Figure 4.7: The rectangle W

monotonically along this curve and $\frac{\partial f_2}{\partial y} \neq 0$, f_2 increases monotonically along this edge of V, so there is exactly one point on this edge where $f_2 = b$. By construction, $f_1 = c - \triangle c \leq a$ at this point. Traveling along the level set of f_2 through this point, we know that by the time we reach the right edge of V we have $f_1 \geq a$. Again, a combination of the Intermediate Value Theorem and monotonicity of f_1 guarantee that there is a unique point along this curve where also $f_1 = a$.

This simultaneously proves two things: that every point in W gets hit by some point of V (F maps V *onto* W), and this point is unique (F is *one-to-one* on V), proving Claim 1. □

Proof of Lemma 4.4.3:

Given a sequence $\overrightarrow{y}_k \to \overrightarrow{y}_\infty \in W$, let $\overrightarrow{x}_k = G(\overrightarrow{y}_k)$ and $\overrightarrow{x}_\infty = G(\overrightarrow{y}_\infty)$; we need to show that $\overrightarrow{x}_k \to \overrightarrow{x}_\infty$.

The two relations can be rewritten

$$\begin{aligned} F(\overrightarrow{x}_k) &= \overrightarrow{y}_k \\ F(\overrightarrow{x}_\infty) &= \overrightarrow{y}_\infty . \end{aligned}$$

Since V is closed and bounded, the Bolzano-Weierstrass Theorem says that the sequence $\{\overrightarrow{x}_k\}$ in V has at least one accumulation point; we will show that *every* such accumulation point must be $\overrightarrow{x}_\infty$.

Suppose $\{\overrightarrow{x}_{k_i}\}$ is a convergent subsequence of $\{\overrightarrow{x}_k\}$. Then, since F is continuous, $\overrightarrow{y}_{k_i} = F(\overrightarrow{x}_{k_i})$ converge to $F(\lim \overrightarrow{x}_{k_i})$. But this is a subsequence of the *convergent* sequence $\{\overrightarrow{y}_k\}$, so its limit must be $\overrightarrow{y}_\infty$. In other words,

$$F(\lim \overrightarrow{x}_{k_i}) = \overrightarrow{y}_\infty = F(\overrightarrow{x}_\infty) ,$$

or, since F is one-to-one,

$$\lim \overrightarrow{x}_{k_i} = \overrightarrow{x}_{\infty}.$$

We have shown that the bounded sequence $\overrightarrow{x}_k$ has the unique accumulation point $\overrightarrow{x}_{\infty}$; this implies that the sequence converges to it. This proves Claim 2. □

Proof of Lemma 4.4.4:

Since we have a candidate for the derivative, we need only show that this candidate has first-order contact with G at $\overrightarrow{y}_0$, that is, we need to show that

$$\lim_{\overrightarrow{y} \to \overrightarrow{y}_0} \frac{G(\overrightarrow{y}) - G(\overrightarrow{y}_0) - (L)^{-1}(\triangle \overrightarrow{y})}{\|\triangle \overrightarrow{y}\|} = 0$$

where

$$L = DF_{\overrightarrow{x}_0}$$

and

$$\triangle \overrightarrow{y} = \overrightarrow{y} - \overrightarrow{y}_0.$$

In other words, if we set

$$\mathcal{E}_1(\overrightarrow{y}) = G(\overrightarrow{y}) - G(\overrightarrow{y}_0) - (L)^{-1}(\triangle \overrightarrow{y}) \tag{4.14}$$

then given $\varepsilon > 0$, we need to find $\delta > 0$ such that $\|\triangle \overrightarrow{y}\| = |\overrightarrow{y} - \overrightarrow{y}_0| \leq \delta$ guarantees

$$\|\mathcal{E}_1(\overrightarrow{y})\| \leq \varepsilon \|\triangle \overrightarrow{y}\|. \tag{4.15}$$

We will do this via some estimates.

Given $\overrightarrow{y}$, let

$$\begin{aligned} \overrightarrow{x} &= G(\overrightarrow{y}) \\ \triangle \overrightarrow{x} &= \overrightarrow{x} - \overrightarrow{x}_0. \end{aligned}$$

Then we can rewrite Equation (4.14) as

$$\mathcal{E}_1(\overrightarrow{y}) = \triangle \overrightarrow{x} - (L)^{-1}(\triangle \overrightarrow{y}) \tag{4.16}$$

or

$$\triangle \overrightarrow{x} = (L)^{-1}(\triangle \overrightarrow{y}) + \mathcal{E}_1(\overrightarrow{y}). \tag{4.17}$$

We will also make use of the relation

$$\triangle\overrightarrow{y} = F(\overrightarrow{x}) - F(\overrightarrow{x}_0) = L(\triangle\overrightarrow{x}) + \mathcal{E}_2(\overrightarrow{x}) \tag{4.18}$$

where

$$\lim_{\overrightarrow{x}\to\overrightarrow{x}_0} \frac{\mathcal{E}_2(\overrightarrow{x})}{\|\triangle\overrightarrow{x}\|} = 0.$$

This means that, given $\varepsilon_2 > 0$, we can pick $\delta_2 > 0$ such that $\|\triangle\overrightarrow{x}\| \leq \delta_2$ guarantees

$$\|\mathcal{E}_2(\overrightarrow{x})\| \leq \varepsilon_2 \|\triangle\overrightarrow{x}\|. \tag{4.19}$$

(We will pick a specific value for ε_2 below.)

Applying the linear mapping L^{-1} to both sides of Equation (4.18), we have

$$(L)^{-1}(\triangle\overrightarrow{y}) = \triangle\overrightarrow{x} + (L)^{-1}(\mathcal{E}_2(\overrightarrow{x}))$$

and substituting this into Equation (4.16), we get

$$\mathcal{E}_1(\overrightarrow{y}) = -(L)^{-1}(\mathcal{E}_2(\overrightarrow{x}))$$

Now, we know that there is a constant $m > 0$ such that for every $\overrightarrow{v} \in \mathbb{R}^2$,

$$\left\|(L)^{-1}(\overrightarrow{v})\right\| \leq m\|\overrightarrow{v}\|;$$

applying this to the previous equation we have

$$\|\mathcal{E}_1(\overrightarrow{y})\| \leq m\|\mathcal{E}_2(\overrightarrow{x})\| \tag{4.20}$$

and applying it to Equation (4.17), and substituting Equation (4.20), we have

$$\|\triangle\overrightarrow{x}\| \leq m\|\triangle\overrightarrow{y}\| + m\|\mathcal{E}_2(\overrightarrow{x})\|. \tag{4.21}$$

Finally, since G is continuous, we can find $\delta > 0$ such that $\|\triangle\overrightarrow{y}\| < \delta$ guarantees $\|\triangle\overrightarrow{x}\| < \delta_2$, so Equation (4.19) holds.

Now observe that

1. If $\varepsilon_2 < \frac{1}{2m}$, then for $\|\triangle\overrightarrow{y}\| < \delta$, substituting Equation (4.19) into Equation (4.21) gives

$$\begin{aligned}\|\triangle\overrightarrow{x}\| &\leq m\|\triangle\overrightarrow{y}\| + m\varepsilon_2\|\triangle\overrightarrow{x}\| \\ &\leq m\|\triangle\overrightarrow{y}\| + m\left(\frac{1}{2m}\right)\|\triangle\overrightarrow{x}\| \\ &\leq m\|\triangle\overrightarrow{y}\| + \frac{1}{2}\|\triangle\overrightarrow{x}\|\end{aligned}$$

from which it follows that

$$\|\triangle\overrightarrow{x}\| < 2m\,\|\triangle\overrightarrow{y}\|$$

2. If $\varepsilon_2 < \frac{\varepsilon}{2m^2}$, then for $\|\triangle\overrightarrow{y}\| < \delta$ Equation (4.20) and Equation (4.19) tell us that

$$\begin{aligned}\|\mathcal{E}_1(\overrightarrow{y})\| &\leq m\,\|\mathcal{E}_2(\overrightarrow{x})\| \\ &\leq m\varepsilon_2\,\|\triangle\overrightarrow{x}\| \\ &\leq m\left(\frac{\varepsilon}{2m^2}\right)\|\triangle\overrightarrow{x}\| \\ &= \left(\frac{\varepsilon}{2m}\right)\|\triangle\overrightarrow{x}\|\,.\end{aligned}$$

So, given $\varepsilon > 0$, if we

- pick
$$0 < \varepsilon_2 < \min\left\{\frac{1}{2m}, \frac{\varepsilon}{2m^2}\right\}$$
and then
- pick $\delta_2 > 0$ so that $\|\triangle\overrightarrow{x}\| \leq \delta_2$ guarantees Equation (4.19) holds, and finally
- pick $\delta > 0$ (by continuity of G, as above) so that $\|\triangle\overrightarrow{y}\| < \delta$ guarantees $\|\triangle\overrightarrow{x}\| < \delta_2$,

then both conditions above hold, and we can assert that

$$\begin{aligned}\|\mathcal{E}_1(\overrightarrow{y})\| &\leq \left(\frac{\varepsilon}{2m}\right)\|\triangle\overrightarrow{x}\| \\ &\leq \left(\frac{\varepsilon}{2m}\right)(2m\,\|\triangle\overrightarrow{y}\|) \\ &= \varepsilon\,\|\triangle\overrightarrow{y}\|\,.\end{aligned}$$

as required by Equation (4.15). □

Together, these claims establish Theorem 4.4.2.

Our prime example in § 3.4 of a regular (parametrized) surface was the graph of a function of two variables. As an application of Theorem 4.4.2, we see that *every* regular surface can be viewed locally as the graph of a function.

Proposition 4.4.5. *Suppose $\mathfrak{S}$ is a regular surface in $\mathbb{R}^3$, and $\overrightarrow{x}_0 \in \mathfrak{S}$ is a point on $\mathfrak{S}$. Let P be the plane tangent to $\mathfrak{S}$ at $\overrightarrow{x}_0$.*

Then there is a neighborhood $V \subset \mathbb{R}^3$ of $\overrightarrow{x}_0$ such that the following hold:

1. *If P is not vertical (i.e., P is not perpendicular to the xy-plane), then $\mathfrak{S} \cap V$ can be expressed as the graph $z = \varphi(x, y)$ of a $\mathcal{C}^1$ function defined on a neighborhood of (x_0, y_0), the projection of $\overrightarrow{x}_0$ on the xy-plane. Analogously, if P is not perpendicularl to the xz-plane (resp. yz-plane), then locally $\mathfrak{S}$ is the graph of y (resp. x) as a function of the other two variables. (Figure 4.8)*

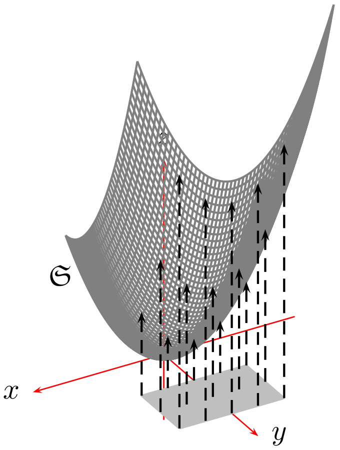


Figure 4.8: $\mathfrak{S}$ parametrized by projection on the xy-plane

2. *$\mathfrak{S} \cap V$ can be parametrized via its projection on P: there is a real-valued function f defined on $P \cap V$ such that*

$$\mathfrak{S} \cap V = \{\overrightarrow{x} + f(\overrightarrow{x})\,\overrightarrow{n} \mid \overrightarrow{x} \in V \cap P\}$$

where $\overrightarrow{n}$ is a vector normal to P (Figure 4.9).

Proof. Let $\overrightarrow{p}\colon \mathbb{R}^2 \to \mathbb{R}^3$

$$\overrightarrow{p}(s,t) = \begin{bmatrix} x(s,t) \\ y(s,t) \\ z(s,t) \end{bmatrix}$$

be a regular parametrization of $\mathfrak{S}$ near $\overrightarrow{x}_0$. Then we can take

$$\overrightarrow{n} = \frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}.$$

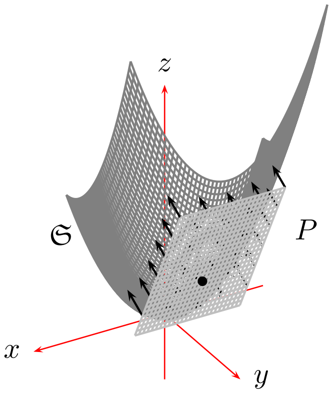


Figure 4.9: $\mathfrak{S}$ parametrized by projection on the tangent plane P

1. The condition that P is not vertical means that the third component of $\overrightarrow{n}$ is nonzero, or
$$\frac{\partial\,(x,y)}{\partial\,(s,t)} \neq 0.$$
Let
$$F(s,t) = \left[\begin{array}{c} x(s,t) \\ y(s,t) \end{array}\right]$$
be $\overrightarrow{p}$ followed by projection on the xy-plane. Then $\frac{\partial(x,y)}{\partial(s,t)}$ is the Jacobian of F, which is nonzero at the parameter values corresponding to $\overrightarrow{x}_0$. It follows from Theorem 4.4.2 that locally, F is invertible: in other words, there is a mapping
$$G(x,y) = \left[\begin{array}{c} s(x,y) \\ t(x,y) \end{array}\right]$$
from a neighborhood of (x_0, y_0) to the s,t-plane such that $F(G(x,y)) = (x,y)$. But then for each (x,y) in the domain of G,
$$\overrightarrow{p}(G(x,y)) = \left[\begin{array}{c} x = x(s(x,y)) \\ y = y(t(x,y)) \\ z(G(x,y)) \end{array}\right]$$
is the point on $\mathfrak{S}$ which projects to (x,y), and
$$\varphi(x,y) = z(G(x,y))$$
is the required function.

2. Let R be a rigid rotation of $\mathbb{R}^3$ about $\overrightarrow{x}_0$ which takes $\overrightarrow{n}$ to the vertical direction, and hence takes P to the horizontal plane P' through $\overrightarrow{x}_0$. Then $R \circ \overrightarrow{p}$ parametrizes $R(\mathfrak{S})$, which by the previous argument is locally parametrized by its projection on the xy-plane—that is, the mapping $(x, y) \mapsto (x, y, \varphi(x, y))$ parametrizes $R(\mathfrak{S})$. Note that

   $$(x, y) \mapsto (x, y, \varphi(x, y)) = (x, y, z_0) + f(x, y)\, \|\overrightarrow{n}\|\, \overrightarrow{k}.$$

   Now rotate the whole picture back, using the inverse rotation R^{-1}: then

   $$\overrightarrow{p}\left(R^{-1}((x, y, z_0))\right) = R^{-1}((x, y, \varphi(x, y)))$$

   is a local parametrization of $\mathfrak{S}$ which assigns to each point of $R^{-1}(P') = P$ the point on $\mathfrak{S}$ which projects to it, and has the form

   $$\overrightarrow{x} \mapsto \overrightarrow{x} + f(R(x, y, z_0))\, \overrightarrow{n}.$$

□

As a corollary of Proposition 4.4.5, we can establish an analogue for parametrized surfaces of Proposition 2.4.6. Recall that a **coordinate patch** for a parametrization $\overrightarrow{p}: \mathbb{R}^2 \to \mathbb{R}^3$ of a surface is a region in the domain of $\overrightarrow{p}$ consisting of regular points, on which the mapping is one-to-one. By abuse of terminology, we will also use this term to refer to the image of such a region: that is, a (sub)surface such that every point is a regular value, and such that no point corresponds to two different pairs of coordinates. This is, of course, the two-dimensional analogue of an arc (but with further conditions on the derivative).

Corollary 4.4.6. *Suppose $\mathfrak{S}$ is simultaneously a coordinate patch for two regular parametrizations, $\overrightarrow{p}$ and $\overrightarrow{q}$. Then there exists a differentiable mapping $T: \mathbb{R}^2 \to \mathbb{R}^2$ which has no critical points, is one-to-one, and such that*

$$\overrightarrow{q} = \overrightarrow{p} \circ T. \tag{4.22}$$

We will refer to T as a **reparametrization** of $\mathfrak{S}$.

Proof. Let us first assume that $\overrightarrow{p}$ is a parametrization by projection on the xy-plane, as in the first part of Proposition 4.4.5, so $\overrightarrow{p}$ has the form

$$\overrightarrow{p}(x, y) = (x, y, \varphi(x, y)).$$

If $\overrightarrow{q}$ has the form

$$\overrightarrow{q}(s,t) = (x(s,t)\,,y(s,t)\,,z(s,t)),$$

then in particular

$$z(s,t) = \varphi(x(s,t)\,,y(s,t))\,.$$

Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T(s,t) = (x(s,t)\,,y(s,t));$$

then clearly T is $\mathcal{C}^1$, with Jacobian matrix

$$JT = \begin{pmatrix} \partial x/\partial s & \partial x/\partial t \\ \partial y/\partial s & \partial y/\partial t \end{pmatrix}.$$

Furthermore, since $\mathfrak{S}$ can be parametrized as the graph of a function, the tangent plane to $\mathfrak{S}$ at each point is not vertical; in other words, its normal has a component in the z direction. But in terms of $\overrightarrow{q}$, this component is given by the determinant of the matrix above, so that JT is invertible at each point. It follows that every point is a regular point for T; to see that it is one-to-one, we need only note that each point of $\mathfrak{S}$ corresponds to a unique pair of coordinates for either parametrization.

Note that F has a differentiable inverse by the Inverse Mapping Theorem.

Now for the general case, we first define $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ by Equation (4.22): to see that this is well-defined, we note that since $\overrightarrow{q}$ is a coordinate patch, the mapping that assigns to $q(s,t)$ the point $(s,t) \in \mathbb{R}^2$ is well-defined, and T is by definition the composition of $\overrightarrow{p}$ with this mapping. Since both factors are one-to-one, so is T. Finally, to see that T is differentiable at each point, we find a neighborhood of $\overrightarrow{p}(u,v) = \overrightarrow{q}(s,t)$ which can be reparametrized by projection on one of the coordinate planes (say the x,y-plane); call this new parametrization ρ: applying the first case to each of $\overrightarrow{p}$ and $\overrightarrow{q}$, we can write

$$\begin{aligned} \overrightarrow{p} &= \rho \circ T_1 \\ \overrightarrow{q} &= \rho \circ T_2. \end{aligned}$$

Then clearly,

$$T = T_1^{-1} \circ T_2$$

is $\mathcal{C}^1$ and satisfies

$$\overrightarrow{q} = \rho \circ T_2 = (\rho \circ T_1) \circ (T_1^{-1} \circ T_2) = \overrightarrow{p} \circ T.$$

□

Two Equations in Three Unknowns: The Implicit Mapping Theorem

We know that the solutions of a linear system of two equations in three unknowns cannot be unique; the maximal possible rank for the matrix of coefficients is 2, and in that case the solution sets are lines in $\mathbb{R}^3$. Thus, we expect the solutions of a typical system of two nonlinear equations in three unknowns—in other words, the level sets of a typical mapping $F\colon \mathbb{R}^3 \to \mathbb{R}^2$—to be composed of curves in 3-space.

The definitions of critical and regular points $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ are essentially the same as for mappings of the plane to itself:

Definition 4.4.7. *Suppose $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ is a differentiable mapping with domain an open set $\mathcal{D} \subset \mathbb{R}^3$.*

1. *A point $\overrightarrow{x}_0 \in \mathcal{D}$ is a* ***regular point*** *of F if the rank of $DF_{\overrightarrow{x}_0}$ is 2.*

2. *It is a* ***critical point*** *of F otherwise—that is, if the rank of $DF_{\overrightarrow{x}_0}$ is 1 or 0.*

If the two gradients $\overrightarrow{\nabla} f_1$ and $\overrightarrow{\nabla} f_2$ are linearly independent, their cross-product is nonzero, so one of its components is nonzero. We will state and prove our theorem assuming the first component of $(\overrightarrow{\nabla} f_1) \times (\overrightarrow{\nabla} f_2)$, sometimes denoted $\left|\frac{\partial(f_1,f_2)}{\partial(y,z)}\right|$, is nonzero:

$$\begin{aligned}\left|\frac{\partial\,(f_1,f_2)}{\partial\,(y,z)}\right| &:= \det\begin{bmatrix}\partial f_1/\partial y & \partial f_1/\partial z\\ \partial f_2/\partial y & \partial f_2/\partial z\end{bmatrix}\\ &= \frac{\partial f_1}{\partial y}\frac{\partial f_2}{\partial z} - \frac{\partial f_1}{\partial z}\frac{\partial f_2}{\partial y}\\ &\neq 0.\end{aligned}$$

Geometrically, this is the condition that the plane spanned by $\overrightarrow{\nabla} f_1(\overrightarrow{x}_0)$ and $\overrightarrow{\nabla} f_2(\overrightarrow{x}_0)$ does not contain the vector $\overrightarrow{\imath}$ or equivalently, that it projects *onto* the yz-plane (Figure 4.10). Note that if F is *continuously* differentiable, then this condition holds at all points sufficiently near $\overrightarrow{x}_0$, as well.

Theorem 4.4.8 (Implicit Mapping Theorem for $\mathbb{R}^3 \to \mathbb{R}^2$)**.** *Suppose*

$$F(\overrightarrow{x}) = \begin{bmatrix} f_1(\overrightarrow{x})\\ f_2(\overrightarrow{x})\end{bmatrix}$$

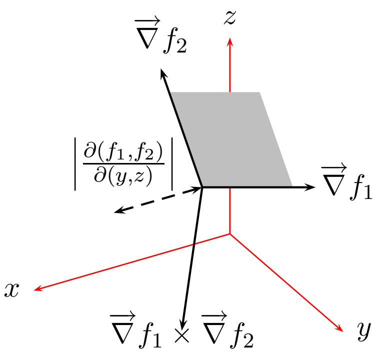

$\overrightarrow{\nabla} f_2$
z
$\left|\frac{\partial(f_1,f_2)}{\partial(y,z)}\right|$
$\overrightarrow{\nabla} f_1$
x
$\overrightarrow{\nabla} f_1 \times \overrightarrow{\nabla} f_2$
y

Figure 4.10: $\left|\frac{\partial(f_1,f_2)}{\partial(y,z)}\right| \neq 0$

is a continuously differentiable mapping from $\mathbb{R}^3$ *to* $\mathbb{R}^2$,

$$\overrightarrow{x}_0 = (x_0, y_0, z_0)$$

is a regular point of F*, and*

$$F(\overrightarrow{x}_0) = (c, d).$$

Then there exists a neighborhood V *of* $\overrightarrow{x}_0 = (x_0, y_0, z_0)$ *and a neighborhood* W *of* $F(\overrightarrow{x}_0) = (c, d)$ *such that* F *maps* V *onto* W*: for every* $\overrightarrow{y}$ *interior to* W *the level set* $\mathcal{L}(F, \overrightarrow{y}) := \{\overrightarrow{x} \,|\, F(\overrightarrow{x}) = \overrightarrow{y}\}$ *intersects* V *in a regular curve.*

Stated more precisely, if the plane spanned by $\overrightarrow{\nabla} f_1(\overrightarrow{x}_0)$ *and* $\overrightarrow{\nabla} f_2(\overrightarrow{x}_0)$ *does not contain* $\overrightarrow{\imath}$*, that is,*

$$\left|\frac{\partial\,(f_1, f_2)}{\partial\,(y, z)}\right| := \frac{\partial f_1}{\partial y}\frac{\partial f_2}{\partial z} - \frac{\partial f_1}{\partial z}\frac{\partial f_2}{\partial y} \neq 0,$$

then there is a rectangular box

$$B = [x_0 - \varepsilon_1, x_0 + \varepsilon_1] \times [y_0 - \varepsilon_2, y_0 + \varepsilon_2] \times [z_0 - \varepsilon_3, z_0 + \varepsilon_3]$$

whose intersection with the level set

$$\mathcal{L}(F, (c, d)) = \{\overrightarrow{x} \,|\, F(\overrightarrow{x}) = (c, d)\}$$

is a regular curve parametrized by $\overrightarrow{p}(t)$, $x_0 - \varepsilon_1 \leq t \leq x_0 + \varepsilon_1$, in the form

$$\begin{aligned} x &= t \\ y &= \gamma_1(t) \\ z &= \gamma_2(t)\,. \end{aligned}$$

The line tangent to this curve at $\overrightarrow{x}_0$ is parallel to $\overrightarrow{\nabla} f_1(\overrightarrow{x}_0) \times \overrightarrow{\nabla} f_2(\overrightarrow{x}_0)$.

(See Figure 4.11.)

When the plane spanned by $\overrightarrow{\nabla} f_1(\overrightarrow{x}_0)$ and $\overrightarrow{\nabla} f_2(\overrightarrow{x}_0)$ does not contain $\overrightarrow{\jmath}$ (*resp.* $\overrightarrow{k}$), then we can locally parametrize $\mathcal{L}(F,(c,d))$ using y (*resp.* z) as a parameter. You are asked to work out the details in Exercise 4.

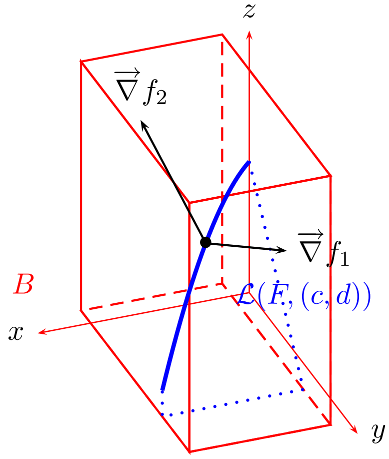


Figure 4.11: Implicit Mapping Theorem

Proof. That F maps some neighborhood V of $\overrightarrow{x}_0$ onto some neighborhood W of $F(\overrightarrow{x}_0)$ follows easily from the Inverse Mapping Theorem for mappings of the plane to itself. The projection of

$$\overrightarrow{\nabla} f_i(\overrightarrow{x}_0) = \left(\frac{\partial f_i}{\partial x}(\overrightarrow{x}_0), \frac{\partial f_i}{\partial y}(\overrightarrow{x}_0), \frac{\partial f_i}{\partial z}(\overrightarrow{x}_0)\right)$$

onto the yz-plane is

$$\operatorname{proj}_{yz} \overrightarrow{\nabla} f_i(\overrightarrow{x}_0) = \left(\frac{\partial f_i}{\partial y}(\overrightarrow{x}_0), \frac{\partial f_i}{\partial z}(\overrightarrow{x}_0)\right)$$

which is the gradient at $\overrightarrow{x}_0$ of the function $f_{x_0,y,z}(o)$f two variables obtained by restricting f_i to the plane through $\overrightarrow{x}_0$ parallel to the yz-plane, and so the projection on the yz-plane of $\overrightarrow{\nabla} F(\overrightarrow{x}_0)$ is just the Jacobian determinant

$$\frac{\partial\left(f_1, f_2\right)}{\partial\left(y, z\right)} = \frac{\partial f_1}{\partial y}\left(\overrightarrow{x}_0\right)\frac{\partial f_2}{\partial z}\left(\overrightarrow{x}_0\right) - \frac{\partial f_1}{\partial z}\left(\overrightarrow{x}_0\right)\frac{\partial f_2}{\partial y}\left(\overrightarrow{x}_0\right)$$

whose nonvanishing, according to the Inverse Mapping Theorem for $\mathbb{R}^2$ (Theorem 4.4.2), guarantees that this restriction maps a neighborhood $\tilde{V}$ of $\overrightarrow{x}_0$ *in the plane* onto a neighborhood W of $F(\overrightarrow{x}_0)$ in $\mathbb{R}^2$. (See Figure 4.12.)

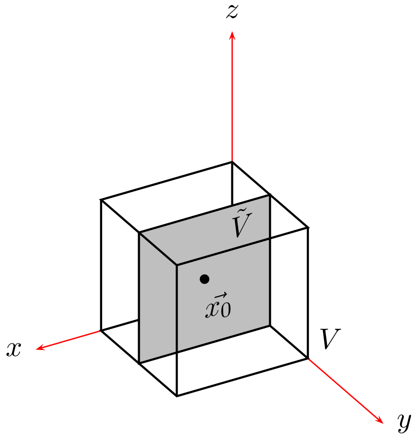


Figure 4.12: The sets $V$and $\tilde{V}$

This of course insures that any neighborhood V of $\overrightarrow{x}_0$ in $\mathbb{R}^3$ containing $\tilde{V}$ also maps onto W.

To see that the level sets are curves, note that if $\left|\frac{\partial(f_1,f_2)}{\partial(y,z)}\right| \neq 0$ then at least one of the two gradients has a nonzero component in the z-direction. Let us assume it is $\overrightarrow{\nabla} f_2(\overrightarrow{x}_0)$. Then the Implicit Function Theorem (Theorem 3.4.3, for $\mathbb{R}^3 \to \mathbb{R}$) tells us that the level set $\mathcal{L}(f_2, d)$ of f_2 through $\overrightarrow{x}_0$ is locally the graph of a $\mathcal{C}^1$ function $z = \varphi(x, y)$ (Figure 4.13).

Now, let

$$g(x, y) = f_1(x, y, \varphi(x, y))$$

be the restriction of f_1 to this graph. A point $\overrightarrow{x}$ on the graph of φ (that is, such that $f_2(\overrightarrow{x}) = d$) belongs to $\mathcal{L}(F, (c, d))$ precisely if $f_1(\overrightarrow{x}) = g(x, y) = c$, or

$$\mathcal{L}(F, (c, d)) = \{(x, y, \varphi(x, y)) \,|\, (x, y) \in \mathcal{L}(g, c)\}.$$

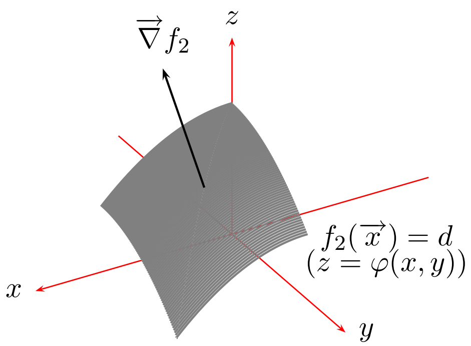


Figure 4.13: $\mathcal{L}(f_2, d)$ =graph of $\varphi(x, y)$

By the Chain Rule

$$\begin{aligned}\frac{\partial g}{\partial y} &= \frac{\partial}{\partial y}\left[\left(f_1(x, y, \varphi(x, y))\right)\right] \\ &= \frac{\partial f_1}{\partial y} + \left(\frac{\partial f_1}{\partial z}\right) \cdot \left(\frac{\partial \varphi}{\partial y}\right),\end{aligned}$$

and by the Implicit Function Theorem,

$$\frac{\partial \varphi}{\partial y} = -\frac{\partial f_1/\partial y}{\partial f_1/\partial z}$$

so

$$\begin{aligned}\frac{\partial g}{\partial y} &= \left(\frac{1}{\partial f_2/\partial z}\right)\left[\left(\frac{\partial f_1}{\partial y}\right) \cdot \left(\frac{\partial f_2}{\partial z}\right) - \left(\frac{\partial f_1}{\partial z}\right) \cdot \left(\frac{\partial f_2}{\partial y}\right)\right] \\ &= \left(\frac{1}{\partial f_2/\partial z}\right)\left|\frac{\partial\left(f_1, f_2\right)}{\partial\left(y, z\right)}\right|\end{aligned}$$

which does not vanish at $\overrightarrow{x}_0$. But then the Implicit Function Theorem applied to g tells us that $\mathcal{L}(g, c)$ is locally a curve through (x_0, y_0) parametrized by

$$\begin{aligned}x &= t \\ y &= \gamma(t)\end{aligned}$$

which means that $\mathcal{L}(F,(c,d))$ is parametrized by

$$\begin{aligned} x &= t \\ y &= \gamma_1(t) := \gamma(t) \\ z &= \gamma_2(t) := \varphi(t,\gamma(t)) \end{aligned}$$

as required (Figure 4.14).

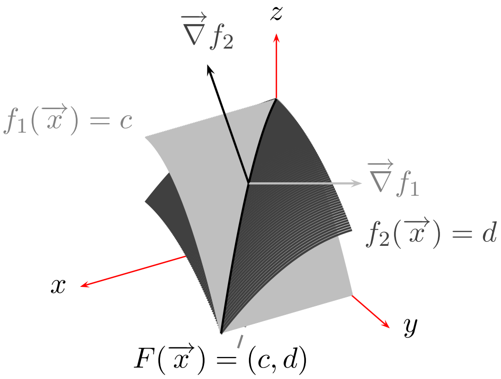

$\overrightarrow{\nabla} f_2$
z
$f_1(\overrightarrow{x}) = c$
$\overrightarrow{\nabla} f_1$
$f_2(\overrightarrow{x}) = d$
x
y
$F(\overrightarrow{x}) = (c,d)$

Figure 4.14: $\mathcal{L}(F,(c,d))$

That the line tangent to $\mathcal{L}(F,(c,d))$ at $\overrightarrow{x}_0$ is parallel to $\overrightarrow{\nabla} f_1 \times \overrightarrow{\nabla} f_2$ is an immediate consequence of the fact that the curve belongs to $\mathcal{L}(f_1,c)$ (and hence its tangent is perpendicular to the normal, $\overrightarrow{\nabla} f_1$) and also to $\mathcal{L}(f_2,c)$ (and hence its tangent is *also* perpendicular to *its* normal, $\overrightarrow{\nabla} f_2$)—but the cross product of the two vectors $\overrightarrow{\nabla} f_1$ and $\overrightarrow{\nabla} f_2$ *also* points in the direction perpendicular to both of them.

□

There are several other ways to understand this result.

Geometrically, the level *curve* $\mathcal{L}(F,(c,d)) = \{\overrightarrow{x} \mid f_1(\overrightarrow{x}) = c \text{ and } f_2(\overrightarrow{x}) = d\}$ is the intersection of the two level *surfaces* $\mathcal{L}(f_1,c)$ and $\mathcal{L}(f_2,d)$. Two regular surfaces are said to **meet transversally**, or are **transverse** at a common point if their normal vectors at that point are linearly independent. This is the same as saying that the tangent planes to the two surfaces at that point are not parallel, and hence intersect in a line. Theorem 4.4.8 is a special case of a general statement (illustrated by Figure 4.14):

> *If two regular surfaces meet transversally at a point, then their intersection near this point is a regular arc, and the line tangent*

to this arc is the intersection of the tangent planes to the two surfaces at the point.

Algebraically, the statement of Theorem 4.4.8 says that if we have a system of two (non-linear) equations in three unknowns and know one solution, and if we can solve the linear system of equations coming from replacing each function with its degree one Taylor polynomial (at this point) for y and z in terms of x, then there is in principle also a solution of the nonlinear system (at least near our point) for y and z in terms of x, and furthermore these solutions have first order contact with the solutions of the linearized system. Row reduction normally gives us the solution of a linear system of this type for x and y in terms of z, and when we have that form we naturally turn to one of the alternative statements of Theorem 4.4.8 from Exercise 4.

For example, one solution of the system

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

is

$$\begin{aligned} x &= 1 \\ y &= -1 \\ z &= 1. \end{aligned}$$

The linearization of the left side of the first equation at $\overrightarrow{x}_0 = (1, -1, 1)$ is

$$\begin{aligned} T_{f_1}\overrightarrow{x}_0(\triangle\overrightarrow{x}) &= f_1(\overrightarrow{x}_0) + \frac{\partial f_1}{\partial x}(\overrightarrow{x}_0)\triangle x + \frac{\partial f_1}{\partial y}(\overrightarrow{x}_0)\triangle y + \frac{\partial f_1}{\partial z}(\overrightarrow{x}_0)\triangle z \\ &= 1 + 3(1)^2\triangle x + 3(-1)^2\triangle y + 3(1)^2\triangle z \\ &= 1 + 3\triangle x - 3\triangle y + 3\triangle z \end{aligned}$$

where

$$\begin{aligned} \triangle x &= x - 1 \\ \triangle y &= y + 1 \\ \triangle z &= z - 1. \end{aligned}$$

Note that the constant term agrees with the right side of the equation (since we are at a solution), so the linearization of the first equation is

$$d_{\overrightarrow{x}_0} f_1(\triangle\overrightarrow{x}) = 0$$

or

$$3\triangle x + 3\triangle y + 3\triangle z = 0.$$

Similarly the linearization of the second equation is

$$\frac{\partial f_2}{\partial x}\left(\overrightarrow{x}_0\right)\triangle x + \frac{\partial f_2}{\partial y}\left(\overrightarrow{x}_0\right)\triangle y + \frac{\partial f_2}{\partial z}\left(\overrightarrow{x}_0\right)\triangle z = 0$$

that is

$$6(1)^2\triangle x + \triangle y + 2(1)^2\triangle z = 0$$

or

$$6\triangle x + \triangle y + 2\triangle z = 0.$$

Thus, the **linearized system** at $(1, -1, 1)$ is

$$\left\{\begin{array}{llllll} 3\triangle x & +3\triangle y & +3\triangle z & = & 0 \\ 6\triangle x & +\triangle y & +2\triangle z & = & 0 \end{array}\right. .$$

Applying row reduction (or Gaussian elimination) to the linearized system, we arrive at

$$\begin{aligned} \triangle x &= -\frac{1}{5}\triangle z \\ \triangle y &= -\frac{4}{5}\triangle z \end{aligned}$$

or equivalently

$$\begin{aligned} \triangle z &= -5\triangle x \\ \triangle y &= 4\triangle x. \end{aligned}$$

So for example, we expect the solution of the *nonlinear* system with $x = 1.1$ (*i.e.*, , $\triangle x = 0.1$) to be approximated by

$$\begin{aligned} x &= 1 + 0.1 = 1.1 \\ y &= -1 + (0.4) = -0.6 \\ z &= 1 + (-0.5) = 0.5. \end{aligned}$$

We can try to see if this is close to a solution of the nonlinear system by evaluating the two functions at (1.1,0.6,0.5): a quick calculation shows that

$$\begin{aligned} f_1(1.1,0.6,0.5) &= (1.1)^3 + (-0.6)^3 + (0.5)^3 \\ &= 1.331 - 0.216 + 0.125 \\ &= 1.24 \\ f_2(1.1,0.6,0.5) &= 2(1.1)^3 - 0.6 + (0.5)^2 \\ &= 2.662 - 0.6 + 0.25 \\ &= 2.312. \end{aligned}$$

Similarly, $\triangle x = 0.01$ leads to $\triangle\overrightarrow{x} = (0.01, 0.04, -0.05)$, so $\overrightarrow{x} = \overrightarrow{x}_0 + \triangle\overrightarrow{x} = (1.01, -0.96, 0.95)$, and

$$\begin{aligned} f_1(1.01,-0.96,0.95) &= 1.003 \\ f_2(1.01,-0.96,0.95) &= 2.003. \end{aligned}$$

Theorem 4.4.8 tells us about level sets *locally* near a regular point. However, if we happen to know that *all* points on a given level set are regular points, then we can combine the local pictures into a global one.

Definition 4.4.9. *Suppose $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ is a differentiable mapping with domain an open set $\mathcal{D} \subset \mathbb{R}^3$.*

1. *A point $\overrightarrow{y} \in \mathbb{R}^2$ is a **critical value** of F if there exists at least one critical point $\overrightarrow{x}$ of F with $F(\overrightarrow{x}) = \overrightarrow{y}$.*
2. *A point $\overrightarrow{y} \in \mathbb{R}^2$ is a **regular value** if it is in the range of F (i.e., there is at least one point $\overrightarrow{x} \in \mathcal{D}$ with $\overrightarrow{y} = F(\overrightarrow{x})$)* [10] *and* every *such point is a regular point of F.*

Recall that a regular curve is one for which every point has a neighborhood with a regular parametrization: in this light, an almost immediate corollary of Theorem 4.4.8 is

Corollary 4.4.10. *If $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ is a $\mathcal{C}^1$ mapping, and (c,d) is a regular value of F, then $\mathcal{L}(F,(c,d))$ is a regular curve in $\mathbb{R}^3$.*

[10] That is, $\overrightarrow{y}$ must be a "value" of F. Some authors use the term "regular value" for *any* point of the target space which is not a critical value, irrespective of whether it is the image of some point under F.

Three Equations in Three Unknowns: The Inverse Mapping Theorem Again

If we think in mapping terms, the situation for *three* equations in *three* unknowns is analogous to that of *two* equations in *two* unknowns: the coefficient matrix of a linear system in which the number of equations equals the number of unknowns is square, and typically is nonsingular. In this case, the linear mapping is invertible, which tells us that it is both one-to-one and onto: every linear system with a nonsingular coefficient matrix (regardless of the right-hand side) has a unique solution. For nonlinear systems, we expect the analogous situation *locally*, provided the derivative mapping is nonsingular.

Definition 4.4.11. *Suppose $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ is differentiable.*

1. *A point $\overrightarrow{x}_0$ in the domain of F is a* ***regular point*** *for F if $DF_{\overrightarrow{x}_0}$ is invertible—that is, the Jacobian determinant is nonzero:*

$$JF(\overrightarrow{x}_0) = \det\left[DF_{\overrightarrow{x}_0}\right] \neq 0.$$

2. *A point in the domain of F which is* not *regular is a* ***critical point*** *for F.*

As before, if F is *continuously* differentiable, then every point sufficiently near a regular point is itself a regular point.

Theorem 4.4.12. *If $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ is $\mathcal{C}^1$ and $\overrightarrow{x}_0$ is a regular point for F, then F is* ***locally invertible*** *at $\overrightarrow{x}_0$: there exists a neighborhood V of $\overrightarrow{x}_0$ and a neighborhood W of $\overrightarrow{y}_0 = F(\overrightarrow{x}_0)$, and a $\mathcal{C}^1$ mapping $G\colon \mathbb{R}^W \to \mathbb{R}^V$ such that*

$$F(G(\overrightarrow{y})) = \overrightarrow{y} \text{ for every } \overrightarrow{y} \in W$$

and

$$G(F(\overrightarrow{x})) = \overrightarrow{x} \text{ for every } \overrightarrow{x} \in V.$$

Furthermore, the linearization at $\overrightarrow{y}_0$ of G is the inverse of the linearization of F at $\overrightarrow{x}_0$:

$$DG_{\overrightarrow{y}_0} = \left(DF_{\overrightarrow{x}_0}\right)^{-1}$$

Proof. In light of Lemma 4.4.3 and Lemma 4.4.4, the only thing we need to prove is the analogue of Claim 1 in the proof of Theorem 4.4.2: the existence of neighborhoods $V \subset \mathbb{R}^3$ and $W \subset \mathbb{R}^3$ of $\overrightarrow{x}_0$ and $\overrightarrow{y}_0$, respectively, such that F maps V onto W in one-to-one fashion.

Let us, as usual, write

$$F(\overrightarrow{x}) = \begin{bmatrix} f_1(\overrightarrow{x}) \\ f_2(\overrightarrow{x}) \\ f_3(\overrightarrow{x}) \end{bmatrix}$$

where $f_j \colon \mathbb{R}^3 \to \mathbb{R}$ for $j = 1, 2, 3$, and set $\overrightarrow{y}_0 = (a, b, c)$. Expanding $JF(\overrightarrow{x}_0)$ by minors, we have (Exercise 5)

$$JF = \frac{\partial f_1}{\partial x}\frac{\partial\,(f_2, f_3)}{\partial\,(y, z)} - \frac{\partial f_1}{\partial y}\frac{\partial\,(f_2, f_3)}{\partial\,(x, z)} + \frac{\partial f_1}{\partial z}\frac{\partial\,(f_2, f_3)}{\partial\,(x, y)} \tag{4.23}$$

so at least one of these terms must be nonzero at $\overrightarrow{x}_0$ assume it is the last: this says that

$$\frac{\partial f_1}{\partial z}\,(\overrightarrow{x}_0) \neq 0$$

and

$$\frac{\partial\,(f_2, f_3)}{\partial\,(y, z)}(\overrightarrow{x}_0) \neq 0.$$

The first inequality implies, by the Implicit Function Theorem (Theorem 3.4.3), that there is a neighborhood V_1 of $\overrightarrow{x}_0$ of the form $[\alpha_1, \alpha_2] \times [\beta_1, \beta_2] \times [\gamma_1, \gamma_2]$ such that $\mathcal{L}(f_1, a) \cap V_1$ is the graph $z = \varphi(x, y)$ of a $\mathcal{C}^1$ function defined on $[\alpha_1, \alpha_2] \times [\beta_1, \beta_2]$ (Figure 4.15). We can also assume (shrinking V_1 if necessary) that the first inequality holds when $\overrightarrow{x}_0$ is replaced by any $\overrightarrow{x} \in V_1$. To use the second inequality, we define the mapping $F_2 \colon \mathbb{R}^3 \to \mathbb{R}^2$ which drops the first component of F:

$$F_2(\overrightarrow{x}) = \begin{bmatrix} f_2(\overrightarrow{x}) \\ f_3(\overrightarrow{x}) \end{bmatrix}.$$

Then our second inequality implies, by the Implicit Mapping Theorem (Theorem 4.4.8), that there is a neighborhood $W_1 \subset \mathbb{R}^2$ of (b, c) such that the intersection of V_1 (reduced further, if necessary) with each level set of the form $\mathcal{L}(F_2, (b', c'))$ for $(b', c') \in W_1$ is a path (or possibly a union of paths) parametrized by z; in particular, the level set $\mathcal{L}(F_2, (b, c))$ through $\overrightarrow{x}_0$ is a curve $\mathcal{C}$ running vertically from the bottom to the top of V_1. Now we mimic the argument for Claim 1 in the proof of the two-dimensional Inverse Mapping Theorem (Theorem 4.4.2):

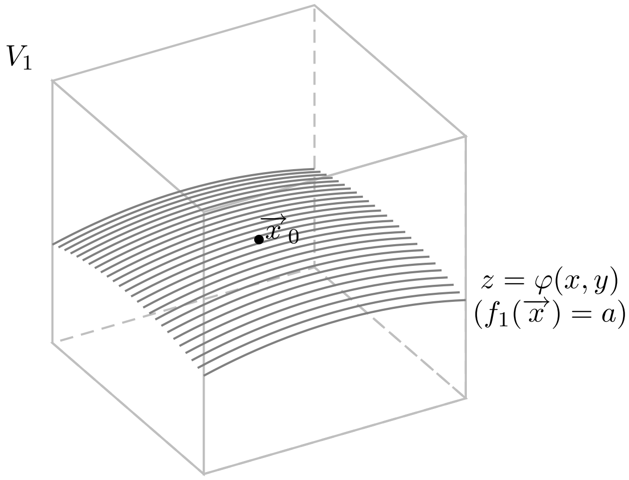


Figure 4.15: $\mathcal{L}(f_1, a)$

1. Pick $a_1 < a < a_2$ between the values of f_1 at the ends of $\mathcal{C}$ and for $i = 1, 2$ let
$$\mathfrak{S}_i = \mathcal{L}(f_1, a_i) \cap V_1.$$
Each of these is the graph of a function $z = \varphi_i(x, y)$ defined on $[\alpha_1, \alpha_2] \times [\beta_1, \beta_2]$.

2. A neighborhood U_i in $\mathfrak{S}_i$ of its intersection with $\mathcal{C}$ maps onto some fixed neighborhood $W_2 \subset W_1$ of (b, c) under F_2; we can assume that $W_2 = [b_1, b_2] \times [c_1, c_2]$, so that U_i is a curvilinear rectangle in $\mathfrak{S}_i$ bounded by its intersection with the four level surfaces $\mathcal{L}(f_2, b_j)$ and $\mathcal{L}(f_3, c_j)$, $j = 1, 2$.

3. Then let V be the curvilinear "parallelepiped" formed by $\mathfrak{S}_1$, $\mathfrak{S}_2$ and these four level surfaces (Figure 4.16).

Then each of the level sets $\mathcal{L}(F_2, (b', c'))$, $(b', c') \in W_2$, of F_2 in V runs from $\mathfrak{S}_1$ to $\mathfrak{S}_2$, which is to say the values of f_1 along each of these curves run from a_1 to a_2, so it follows that each value a' of f_1 between these two occurs in conjunction with the corresponding value (b', c'); in other words, F maps V *onto* the neighborhood of (a, b, c)

$$W = [a_1, a_2] \times W_2 = [a_1, a_2] \times [b_1, b_2] \times [c_1, c_2]\,;$$

furthermore, since z is strictly increasing along each of these curves, and $\frac{\partial f_1}{\partial z} \neq 0$ in V, any such value occurs only once—that is, F is *one-to-one* on V.

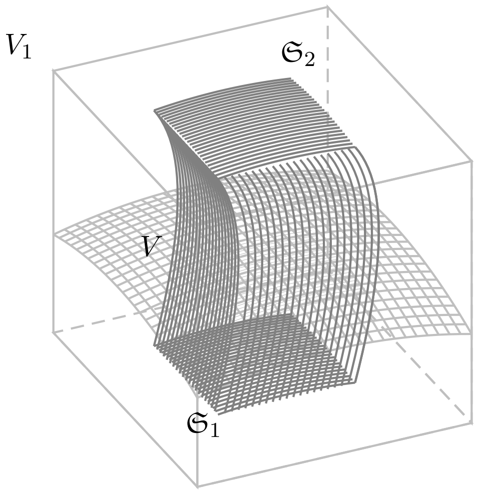


Figure 4.16: V

This, together with Lemmas 4.4.3 and 4.4.4, proves Theorem 4.4.12. □

We should stress that the conclusion of the Inverse Mapping Theorem is strictly local. A mapping $F: \mathbb{R}^3 \to \mathbb{R}^3$ whose derivative DF is invertible at every point of its domain is guaranteed to be *locally* one-to-one and onto, but this says nothing about its global properties. For example, the change-of-coordinates transformation Cyl from cylindrical to rectangular coordinates has an invertible derivative everywhere off the z-axis ($r = 0$); if we compose this with the mapping E which leaves θ and z unchanged but replaces r with its exponential

$$E\begin{pmatrix} r \\ \theta \\ z \end{pmatrix} = \begin{pmatrix} e^r \\ \theta \\ z \end{pmatrix}$$

then the composition

$$Sph \circ E\begin{pmatrix} r \\ \theta \\ z \end{pmatrix} = \begin{pmatrix} e^r \cos\theta \\ e^r \sin\theta \\ z \end{pmatrix}$$

has an invertible derivative everywhere (so is one-to-one and onto near any particular point) but it is neither globally onto (it misses the z-axis) nor globally one-to-one (increasing or decreasing θ by an integer multiple of 2π gives the same image).

In our present context, the main part of Theorem 4.4.12 is the formula for the derivative of the (local) inverse, Equation (4.13), which is the content of Lemma 4.4.4:

$$DF^{-1}_{F\left(\overrightarrow{x}_0\right)} = \left(DF_{\overrightarrow{x}_0}\right)^{-1}.$$

How do we use this? As an example, consider the change-of-coordinates transformation *Sph* from spherical to rectangular coordinates, which we found in § 4.2. While a bit tedious, it was relatively easy to calculate its Jacobian matrix (Equation (4.6)). We can use this to translate differential data back from rectangular to spherical coordinates without having to solve explicitly for the spherical coordinates in terms of the rectangular. For example, consider the curve (Figure 4.17) given in rectangular coordinates by

$$\begin{aligned} x &= t\cos 2\pi t \\ y &= t\sin 2\pi t \\ z &= 1 - t^2 \end{aligned}$$

with derivative

$$\begin{aligned} \dot{x} &= \cos 2\pi t - 2\pi t \sin 2\pi t \\ \dot{y} &= \sin 2\pi t + 2\pi t \cos 2\pi t \\ \dot{z} &= -2t. \end{aligned}$$

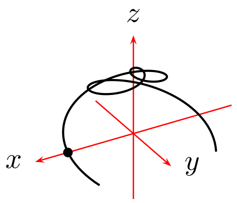


Figure 4.17: $(x, y, z) = (t\cos 2\pi t, t\sin 2\pi t, 1 - t^2)$, $-1.1 < t < 1.1$

Suppose we want to find the rate of change of one or more of the spherical coordinates as the curve crosses the xy-plane at $t = 1$ (so $x = 1$, $y = z = 0$). It is easy to see that this position is given in spherical coordinates by $\rho = 1$, $\theta = 0$, $\phi = \pi/2$ and using this (together with Equation (4.6)) we find that the derivative of Sph at this point has matrix representative

$$JSph((1,0,\pi/2)) = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & -1 & 0 \end{pmatrix}.$$

The inverse of this matrix (found by reducing the super-augmented matrix $[J|I]$) is

$$(JSph((1,0,\pi/2)))^{-1} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & -1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

so we see that

$$\begin{pmatrix} \dot{\rho} \\ \dot{\phi} \\ \dot{\theta} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & -1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} \dot{x} = 1 \\ \dot{y} = 2\pi \\ \dot{z} = -2 \end{pmatrix} = \begin{pmatrix} \frac{2\pi+1}{\sqrt{2}} \\ 2 \\ \frac{2\pi-1}{\sqrt{2}} \end{pmatrix}$$

and we can write

$$\begin{aligned} \frac{d\rho}{dt} &= \pi\sqrt{2} + \frac{1}{\sqrt{2}} \\ \frac{d\phi}{dt} &= 2 \\ \frac{d\theta}{dt} &= \pi\sqrt{2} - \frac{1}{\sqrt{2}}. \end{aligned}$$

Exercises for § 4.4

Practice problems:

1. Find the critical points of each map $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ below.

   (a) $F(x,y) = (x+y, xy)$ (b) $F(x,y) = (x^2+y^2, xy)$

   (c) $F(x,y) = (x+y^{2,x^2+y)}$ (d) $F(x,y) = (e^x y, e^{-y} x)$

   (e) $F(x,y) = (x^2, y^3)$ (f) $F(x,y) = (x^2+y, y^3)$

2. Find the critical points of each map $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ below.

   (a) $F(x,y,z) = (yz, xz, xy)$

   (b) $F(x,y,z) = (y^2+z^2, x^2+z^2, x^2+y^2)$

(c) $F(x,y,z) = (x+y+z, x^2+y^2+z^2, x^3+y^3+z^3)$

(d) $F(x,y,z) = (xe^y, ye^z, ze^x)$

3. Find the critical points of each map $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ below, and identify at which regular points one can locally parametrize the level curve using z as the parameter.

   (a) $F(x,y,z) = (e^z x, e^z y)$ (b) $F(x,y,z) = (xz, yz)$

   (c) $F(x,y,z) = (x^2+yz, xy+z^2)$ (d) $F(x,y,z) = (x+y+z, x^2+y^2+z^2)$

Theory problems:

4. (a) Show that a plane which does not contain $\overrightarrow{\jmath}$ (*resp.* $\overrightarrow{k}$) projects onto the xz-plane (*resp.* yz-plane).

   (b) Use this to discuss how the statement and proof of Theorem 4.4.8, given in case the plane spanned by $\overrightarrow{\nabla} f_1(\overrightarrow{x}_0)$ and $\overrightarrow{\nabla} f_2(\overrightarrow{x}_0)$ does not contain $\overrightarrow{\imath}$, can be adapted to the situations when this plane does not contain $\overrightarrow{\jmath}$ (*resp.* $\overrightarrow{k}$).

5. Prove Equation (4.23).

Challenge problem:

6. Consider the following variant on cylindrical coordinates: the coordinates (R, θ, z) are assigned to the point with rectangular coordinates

$$x = (R+z^2)\cos\theta$$
$$y = (R+z^2)\sin\theta$$
$$z = z.$$

   (a) Find the Jacobian determinant of the map $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ which assigns to the triple (R, θ, z) the triple (x, y, z), according to the equations above.

   (b) Where does the system fail to yield a coordinate system, locally?

   (c) Describe the level sets corresponding to a fixed value of R.

   (d) If a moving object passes the point with rectangular coordinates

$$x = \sqrt{2}$$
$$y = \sqrt{2}$$
$$z = -1$$

so

$$R = 1$$
$$\theta = \frac{\pi}{4}$$

with velocity

$$\frac{dx}{dt} = \sqrt{2}$$
$$\frac{dy}{dt} = \sqrt{2}$$
$$\frac{dz}{dt} = 1,$$

what is the rate of change of R?

7. Consider the mapping $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$F(x, y, z) = (x + y + z, xyz).$$

   (a) Verify that $(1, -1, 1)$ is a regular point of F.

   (b) Show that it is possible to parametrize the level curve through this point using z as a parameter.

   (c) Give an estimate of the x and y coordinates of the point on the level set where $z = 1.1$.

8. Our proof of Theorem 4.4.2 (the Inverse Mapping Theorem for $F\colon \mathbb{R}^2 \to \mathbb{R}^2$) was based on the Implicit Function Theorem for $f\colon \mathbb{R}^2 \to \mathbb{R}$ (Theorem 3.4.2), and similarly, our proof of the Inverse Mapping Theorem for $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ (Theorem 4.4.12) was based on the Implicit Mapping Theorem for $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ (Theorem 4.4.8). In this problem, you will show that in each of these instances we could have used the *Inverse* Mapping Theorem to prove the *Implicit* one (if we had an independent proof of the Inverse one).[11]

[11] The proofs we gave in the text are similar to the original ones given in the 19th century. Nowadays the normal procedure (which you will see in more advanced courses) is to first prove the Inverse Mapping Theorem using the **Banach Contraction Mapping Lemma**— which is beyond the scope of this book—and then to prove the Implicit Mapping Theorem by an argument similar to the one in this problem.

(a) *Inverse Mapping for $\mathbb{R}^2 \to \mathbb{R}^2$ implies Implicit Mapping for $\mathbb{R}^2 \to \mathbb{R}$:* Suppose $f\colon 2 \to \mathbb{R}$ is a $\mathcal{C}^1$ function near the point (a,b), with

$$f(a,b) = c$$

and

$$\frac{\partial f}{\partial y}(a,b) \neq 0.$$

Define $G\colon \mathbb{R}^2 \to \mathbb{R}^2$ by

$$G(x_1,x_2) = (x_1, f(x_1,x_2)).$$

i. Show that the Jacobian determinant of G at (a,b) is nonzero.

ii. Then by Theorem 4.4.2, G has a local inverse. Show that such an inverse must have the form

$$G^{-1}(y_1,y_2) = (y_1, \phi(y_1,y_2)).$$

iii. From this, conclude that if $f(x_1,x_2) = c$ at a point in the domain of G^{-1}, then

$$x_2 = \phi(x_1,c)\,.$$

iv. Explain how this proves Theorem 3.4.2.

(b) *Inverse Mapping for $\mathbb{R}^3 \to \mathbb{R}^3$ implies Implicit Mapping for $\mathbb{R}^3 \to \mathbb{R}^2$:* Suppose $F\colon \mathbb{R}^3 \to \mathbb{R}^2$ is a $\mathcal{C}^1$ function near the point (a_1,a_2,a_3), with

$$F(a_1,a_2,a_3) = (b_1,b_2)$$

i.e.,

$$\begin{aligned} f_1(a_1,a_2,a_3) &= b_1 \\ f_2(a_1,a_2,a_3) &= b_2 \end{aligned}$$

and also assume that

$$\left|\frac{\partial\,(f_1,f_2)}{\partial\,(y,z)}(a_1,a_2,a_3)\right| \neq 0.$$

Define $G\colon \mathbb{R}^3 \to \mathbb{R}^3$ by

$$G(x,y,z) = (x, F(x,y,z)).$$

i. Show that the Jacobian determinant of G at (a_1, a_2, a_3) is nonzero.

ii. Then by Theorem 4.4.12, G has a local inverse. Show that such an inverse must have the form

$$G^{-1}(x', y', z') = (x', \phi_1(y', z'), \phi_2(y', z')).$$

iii. From this, conclude that if $F(x, y, z) = (b_1, b_2)$ at a point in the domain of G^{-1}, then

$$y = \phi_1(x, b_1, b_2)$$

and

$$z = \phi_2(x, b_1, b_2)$$

iv. Explain how this proves Theorem 3.4.3.

5
Integral Calculus for Real-Valued Functions of Several Variables

In this chapter, we consider integrals of functions of several variables.

5.1 Integration over Rectangles

In this section, we will generalize the process of integration from functions of one variable to several variables. As we shall see, the passage from one to several variables presents new difficulties, although the basic underlying ideas from single-variable integration remain.

Integrals in One Variable: A Review

Let us recall the theory behind the Riemann integral for a function of one variable, which is motivated by the idea of finding the area underneath the graph.

Given a function $f(x)$ defined and bounded on the closed interval $[a, b]$, we consider the partition of $[a, b]$ into n subintervals via the partition points

$$\mathcal{P} = \{a = x_0 < x_1 < \cdots < x_n = b\};$$

the j^{th} **component interval** is then

$$I_j = [x_{j-1}, x_j]\,;$$

its length is denoted

$$\triangle x_j = \|I_j\| = x_j - x_{j-1}$$

and the **mesh size** of $\mathcal{P}$ is

$$\operatorname{mesh}(\mathcal{P}) = \max_{j=1,\dots,n} \triangle x_j.$$

From this data we form two sums: the **lower sum**

$$\mathcal{L}(\mathcal{P}, f) = \sum_{j=1}^{n} \left(\inf_{I_j} f\right) \triangle x_j$$

and the **upper sum**

$$\mathcal{U}(\mathcal{P}, f) = \sum_{j=1}^{n} \left(\sup_{I_j} f\right) \triangle x_j.$$

It is clear that the lower sum is less than or equal to the upper sum for $\mathcal{P}$; however, we can also compare the sums obtained for different partitions: we show that *every* lower sum is lower than *every* upper sum: that is, for any pair of partitions $\mathcal{P}$ and $\mathcal{P}'$,

$$\mathcal{L}(\mathcal{P}, f) \le \mathcal{U}(\mathcal{P}', f)$$

by comparing each of these to the lower (*resp.* upper) sum for their mutual refinement $\mathcal{P} \vee \mathcal{P}'$ (the partition points of $\mathcal{P} \vee \mathcal{P}'$ are the union of the partition points of $\mathcal{P}$ with those of $\mathcal{P}'$). In particular this means that we can define the **lower integral**

$$\underline{\int} f(x)\; dx = \sup_{\mathcal{P}} \mathcal{L}(\mathcal{P}, f)$$

and the **upper integral**

$$\overline{\int} f(x)\; dx = \inf_{\mathcal{P}} \mathcal{U}(\mathcal{P}, f);$$

clearly,

$$\underline{\int} f(x)\ dx \leq \overline{\int} f(x)\ dx$$

and if the two are equal, we say that $f(x)$ is **Riemann integrable** (or just *integrable*) over $[a,b]$, and define the **definite integral** (or *Riemann integral*) of $f(x)$ over $[a,b]$ to be the common value of the lower and upper integrals:

$$\int_{[a,b]} f(x)\ dx = \underline{\int} f(x)\ dx = \overline{\int} f(x)\ dx.$$

A few important observations about the Riemann integral are:

- If $f(x)$ is integrable over $[a,b]$ then it is integrable over any subinterval of $[a,b]$.
- If $f(x)$ is continuous on $[a,b]$ (with the possible exception of a finite number of points), then for any sequence of partitions $\mathcal{P}_k$ with $\operatorname{mesh}(\mathcal{P}_k) \to 0$, the corresponding lower (or upper) sums converge to the integral.
- In fact, for any partition we can replace the infimum (*resp.* supremum) of $f(x)$ in the lower (*resp.* upper) sum with an arbitrary sample point $s_j \in I_j$ to form a **Riemann sum**

  $$\mathcal{R}(\mathcal{P}, f) = \mathcal{R}(\mathcal{P}, f, \{s_j\}) := \sum_{j=1}^{n} f(s_j)\ \triangle x_j.$$

  Then, if $f(x)$ is continuous at all but a finite number of points in $[a,b]$, and $\mathcal{P}_k$ is a sequence of partitions with $\operatorname{mesh}(\mathcal{P}_k) \to 0$, the sequence of Riemann sums corresponding to any choice of sample points for each $\mathcal{P}_k$ converges to the integral:

  $$\mathcal{R}(\mathcal{P}_k, f) \to \int_{[a,b]} f(x)\ dx.$$

Integrals over Rectangles

Let us now see how this line of reasoning can be mimicked to define the integral of a function $f(x,y)$ of *two* variables. A major complication arises at the outset: we integrate a function of one variable over an interval: what is the analogue for functions of two variables? In different terms, a "piece" of the real line is, in a natural way, a subinterval[1], but a "piece" of the plane is a *region* whose shape can be quite complicated. We shall start with the simplest regions and then explore the generalization to other regions later. By a **rectangle** in the plane we will mean something more specific: a rectangle whose sides are parallel to the coordinate axes. This is defined by its projections $[a, b]$ onto the x-axis (*resp.* $[c, d]$ onto the y-axis), and is officially referred to as their *product*:[2]

$$[a, b] \times [c, d] := \{(x, y) \mid x \in [a, b] \text{ and } y \in [c, d]\}.$$

A natural way to partition the "product" rectangle $[a, b] \times [c, d]$ is to partition each of the "factors" separately (see Figure 5.1): that is, a **partition** $\mathcal{P}$ of the product rectangle $[a, b] \times [c, d]$ is defined by a partition of $[a, b]$

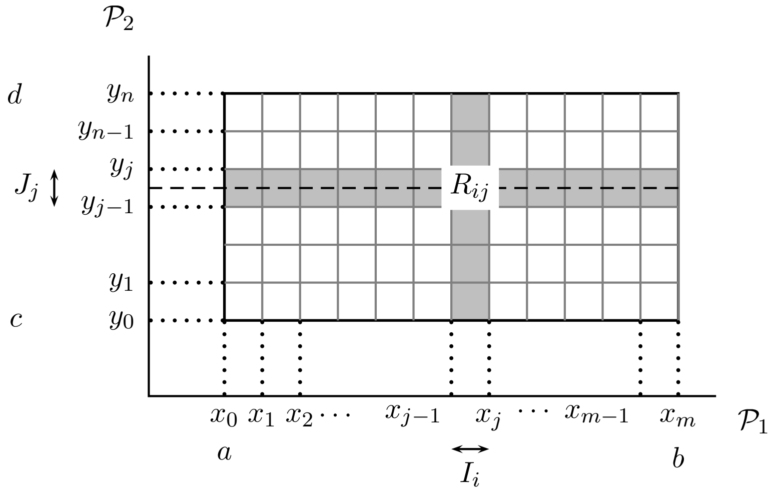


Figure 5.1: Partitioning the Rectangle $[a, b] \times [c, d]$

$$\mathcal{P}_1 = \{a = x_0 < x_1 < \cdots < x_m = b\}$$

and a partition[3] of $[c, d]$

$$\mathcal{P}_2 = \{c = y_0 < y_1 < \cdots < y_n = d\}.$$

[1] or maybe a union of subintervals

[2] In general, the **product** of two sets A and B is the set of pairs (a, b) consisting of an element a of A and an element b of B.

[3] Note that the number of elements in the two partitions is not assumed to be the same.

This defines a subdivision of $[a,b] \times [c,d]$ into mn subrectangles R_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$:

$$\begin{aligned} R_{ij} &= I_i \times J_j \\ &= [x_{i-1}, x_i] \times [y_{j-1}, y_j] \end{aligned}$$

whose respective areas are

$$\begin{aligned} \triangle A_{ij} &= \triangle x_i \triangle y_j \\ &= (x_i - x_{i-1})(y_j - y_{j-1}). \end{aligned}$$

Now we can again form lower and upper sums[4]

$$\begin{aligned} \mathcal{L}(\mathcal{P}, f) &= \sum_{i,j=1}^{i=m,\, j=n} \left(\inf_{R_{ij}} f \right) \triangle A_{ij} \\ \mathcal{U}(\mathcal{P}, f) &= \sum_{i,j=1}^{i=m,\, j=n} \left(\sup_{R_{ij}} f \right) \triangle A_{ij}. \end{aligned}$$

If $f(x,y) > 0$ over $[a,b] \times [c,d]$, we can picture the lower (*resp.* upper) sum as the total volume of the rectilinear solid formed out of rectangles with base R_{ij} and height $h_{ij}^{-} = \inf_{R_{ij}} f(x,y)$ (*resp.* $h_{ij}^{+} = \sup_{R_{ij}} f(x,y)$) (see Figure 5.2 (*resp.* Figure 5.3)).

As before, we can show (Exercise 2) that for any two partitions $\mathcal{P}$ and $\mathcal{P}'$ of $[a,b] \times [c,d]$, $\mathcal{L}(\mathcal{P}, f) \leq \mathcal{U}(\mathcal{P}', f)$ and so can define $f(x,y)$ to be **integrable** if

$$\underline{\iint}_{[a,b]\times[c,d]} f(x,y)\, dA := \sup_{\mathcal{P}} \mathcal{L}(\mathcal{P}, f)$$

and

$$\overline{\iint}_{[a,b]\times[c,d]} f(x,y)\, dA := \inf_{\mathcal{P}} \mathcal{U}(\mathcal{P}, f)$$

are equal, in which case their common value is the **integral**[5] of $f(x,y)$ over the rectangle $[a,b] \times [c,d]$

$$\begin{aligned} \iint_{[a,b]\times[c,d]} f\, dA &= \underline{\iint}_{[a,b]\times[c,d]} f(x,y)\, dA \\ &= \overline{\iint}_{[a,b]\times[c,d]} f(x,y)\, dA. \end{aligned}$$

[4] Of course, to form these sums we must assume that $f(x,y)$ is bounded on $[a,b] \times [c,d]$.

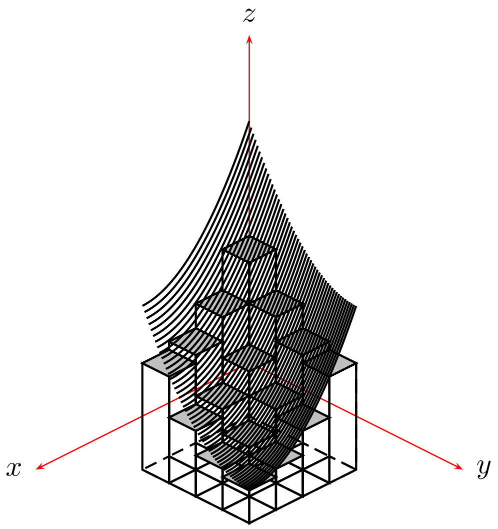


Figure 5.2: Lower sum

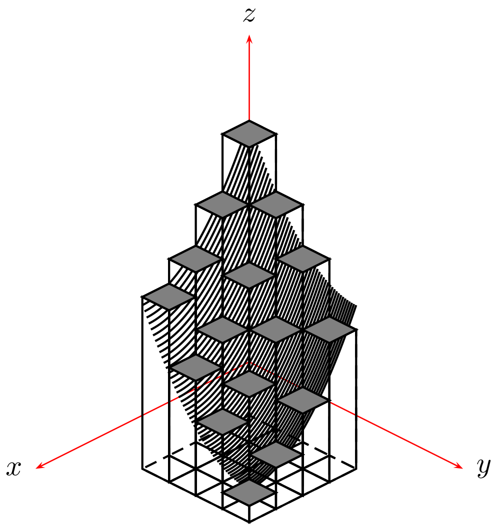


Figure 5.3: Upper sum

Again, given a collection of sample points $\overrightarrow{s}_{ij} = (x^*_{ij}, y^*_{ij}) \in R_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, we can form a Riemann sum

$$\mathcal{R}(\mathcal{P}, f) = \mathcal{R}(\mathcal{P}, f, \{\overrightarrow{s}_{ij}\}) = \sum_{i=1,j=1}^{m,n} f(\overrightarrow{s}_{ij})\ dA_{ij}.$$

Since (Exercise 3) $\mathcal{L}(\mathcal{P}, f) \leq \mathcal{R}(\mathcal{P}, f) \leq \mathcal{U}(\mathcal{P}, f)$ for every partition $\mathcal{P}$ (and every choice of sample points), it follows (Exercise 4) that if $f(x, y)$ is integrable over $[a, b] \times [c, d]$ and $\mathcal{P}_k$ is a sequence of partitions with $\mathcal{L}(\mathcal{P}_k, f) \to \iint_{[a,b]\times[c,d]} f\, dA$ and $\mathcal{U}(\mathcal{P}_k, f) \to \iint_{[a,b]\times[c,d]} f\, dA$, then the Riemann sums converge to the definite integral:

$$\mathcal{R}(\mathcal{P}_k, f) \to \iint_{[a,b]\times[c,d]} f\, dA.$$

Which functions $f(x, y)$ are Riemann integrable? For functions of one variable, there are several characterizations of precisely which functions are Riemann integrable; in particular, we know that every monotone function and every continuous function (in fact, every function with a finite number of points of discontinuity) is Riemann integrable (*Calculus Deconstructed*, §5.2 and §5.9). We shall not attempt such a general characterization in the case of several variables; however, we wish to establish that continuous functions, as well as functions with certain kinds of discontinuity, are integrable. To this end, we need to refine our notion of continuity. There are two ways to define continuity of a function at a point $\overrightarrow{x} \in \mathbb{R}^2$: in terms of sequences converging to $\overrightarrow{x}$, or the "$\varepsilon - \delta$" definition. It is the latter that we wish to refine. Recall this definition:

Definition 5.1.1. *$f(x, y)$ is **continuous** at $\overrightarrow{x}_0 = (x_0, y_0)$ if we can guarantee any required accuracy in the output by requiring some specific, related accuracy in the input: that is, given $\varepsilon > 0$, there exists $\delta > 0$ such that for every point $\overrightarrow{x} = (x, y)$ in the domain of $f(x, y)$ with $\operatorname{dist}(\overrightarrow{x}, \overrightarrow{x}_0) < \delta$, the value of f at $\overrightarrow{x}$ is within ε of the value at $\overrightarrow{x}_0$:*

$$\|\overrightarrow{x} - \overrightarrow{x}_0\| < \delta \Rightarrow |f(\overrightarrow{x}) - f(\overrightarrow{x}_0)| < \varepsilon. \tag{5.1}$$

Suppose we know that f is continuous at every point of some set $S \subset \operatorname{dom}(f)$ in the sense of the definition above. What this says, precisely formulated, is

> *Given a point $\overrightarrow{x}_0$ in S and $\varepsilon > 0$, there exists $\delta > 0$ such that (5.1) holds.*

The thing to note is that the accuracy required of the input—that is, δ—depends on *where* we are trying to apply the definition: that is, continuity at another point $\overrightarrow{x}_1$ may require a different value of δ to guarantee the estimate $|f(\overrightarrow{x}) - f(\overrightarrow{x}_1)| < \varepsilon$, even for the same $\varepsilon > 0$. (An extensive discussion of this issue can be found in (*Calculus Deconstructed*, §3.7).) We say that f is *uniformly* continuous on a set S if δ can be chosen in a way that is independent of the "basepoint" $\overrightarrow{x}_0$; that is,[6]

Definition 5.1.2. *f is* ***uniformly continuous*** *on a set $S \subset \mathrm{dom}(f)$ if, given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(\overrightarrow{x}) - f(\overrightarrow{x}')| < \varepsilon$ whenever $\overrightarrow{x}$ and $\overrightarrow{x}'$ are points of S satisfying $\|\overrightarrow{x} - \overrightarrow{x}'\| < \delta$:*

$$\overrightarrow{x}, \quad \overrightarrow{x}' \in S \text{ and } \left\|\overrightarrow{x} - \overrightarrow{x}'\right\| < \delta \Rightarrow \left|f(\overrightarrow{x}) - f\left(\overrightarrow{x}'\right)\right| < \varepsilon. \tag{5.2}$$

The basic fact that allows us to prove integrability of continuous functions is the following.

Lemma 5.1.3. *If S is a compact set and f is continuous on S, then it is uniformly continuous on S.*

Proof. The proof is by contradiction. Suppose that f is continuous, but not uniformly continuous, on the compact set S. This means that for some required accuracy $\varepsilon > 0$, there is *no* $\delta > 0$ which guarantees (5.2). In other words, no matter how small we pick $\delta > 0$, there is at least one pair of points in S with $\|\overrightarrow{x} - \overrightarrow{x}'\| < \delta$ but $|f(\overrightarrow{x}) - f(\overrightarrow{x}')| \geq \varepsilon$. More specifically, for each positive integer k, we can find a pair of points $\overrightarrow{x}_k$ and $\overrightarrow{x}'_k$ in S with $\|\overrightarrow{x}_k - \overrightarrow{x}'_k\| < \frac{1}{k}$, but $|f(\overrightarrow{x}) - f(\overrightarrow{x}')| \geq \varepsilon$. Now, since S is (sequentially) compact, there exists a subsequence of the $\overrightarrow{x}_k$ (which we can assume is the full sequence) which converges to some point v_0 in S; furthermore, since $\|\overrightarrow{x}_k - \overrightarrow{x}'_k\| \to 0$, the $\overrightarrow{x}'_k$ also converge to the same limit $\overrightarrow{x}_0$. Since f is continuous, this implies that $f(\overrightarrow{x}_k) \to f(\overrightarrow{x}_0)$ and $f(\overrightarrow{x}'_k) \to f(\overrightarrow{x}_0)$. But this is impossible, since $|f(\overrightarrow{x}_k - f(\overrightarrow{x}'_k))| \geq \varepsilon > 0$, and provides us with the contradiction that proves the lemma. □

Using this, we can prove that continuous functions are Riemann integrable. However, we need first to define one more notion generalizing the one-variable situation: the mesh size of a partition. For a partition of an interval, the length of a component interval I_j also controls the distance between points in that interval; however, the *area* of a rectangle R_{ij} can be

[6] Technically, there is a leap here: when $\overrightarrow{x}_0 \in S$, the definition of continuity at $\overrightarrow{x}_0$ given above allows the other point $\overrightarrow{x}$ to be *any* point of the domain, not just a point of S. However, as we use it, this distinction will not matter.

small and still allow some pairs of points in it to be far apart (if, for example, it is tall but extremely thin). Thus, we need to separate out a measure of distances from area. We can define the **diameter** of a rectangle (or of any other set) to be the supremum of the pairwise distances of points in it; for a rectangle, this is the same as the length of the diagonal (Exercise 5). A more convenient definition in the case of a rectangle, though, would be to take the maximum of the lengths of its sides, that is, define

$$\|R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]\| := \max\{\triangle x_i, \triangle y_j\}.$$

This is always less than the diagonal and hence also controls the possible distance between pairs of points in $R_{i,j}$, but has the virtue of being easy to calculate. Then we define the **mesh size** (or just *mesh*) of a partition $\mathcal{P}$ to be the maximum of these "diameters":

$$\text{mesh}(\mathcal{P}) := \max_{i,j} \|R_{ij}\| = \max_{i \leq m, j \leq n} \{\triangle x_i, \triangle y_j\}.$$

With this, we can formulate the following.

Theorem 5.1.4. *Every function f which is continuous on the rectangle $[a,b]\times[c,d]$ is Riemann integrable on it. More precisely, if $\mathcal{P}_k$ is any sequence of partitions of $[a,b] \times [c,d]$ for which mesh($\mathcal{P}_k$) $\to 0$, then the sequence of corresponding lower sums (and upper sums—in fact any Riemann sums) converges to the definite integral:*

$$\lim \mathcal{L}(\mathcal{P}_k, f) = \lim \mathcal{U}(\mathcal{P}_k, f) = \lim \mathcal{R}(\mathcal{P}_k, f) = \iint_{[a,b]\times[c,d]} f\, dA.$$

Proof. Note first that it suffices to show that

$$\mathcal{U}(\mathcal{P}_k, f) - \mathcal{L}(\mathcal{P}_k, f) \to 0 \tag{5.3}$$

since for every k

$$\mathcal{L}(\mathcal{P}_k, f) \leq \sup_{\mathcal{P}} \mathcal{L}(\mathcal{P}, f) \leq \inf_{\mathcal{P}} \mathcal{U}(\mathcal{P}, f) \leq \mathcal{U}(\mathcal{P}_k, f)$$

and for every sample choice

$$\mathcal{L}(\mathcal{P}_k, f) \leq \mathcal{R}(\mathcal{P}_k, f) \leq \mathcal{U}(\mathcal{P}_k, f)$$

(see Exercise 6 for details).

Now let A be the area of $[a,b]\times[c,d]$ (that is, $A=|b-a|\cdot|d-c|$). Given $\varepsilon>0$, use the uniform continuity of f on the compact set $[a,b]\times[c,d]$ to find $\delta>0$ such that

$$\overrightarrow{x}, \overrightarrow{x}' \in S \text{ and } \left\|\overrightarrow{x}-\overrightarrow{x}'\right\| < \delta \Rightarrow \left|f(\overrightarrow{x})-f\left(\overrightarrow{x}'\right)\right| < \frac{\varepsilon}{A}.$$

Then if a partition $\mathcal{P}$ has $\operatorname{mesh}(\mathcal{P})<\delta$, we can guarantee that any two points in the same subrectangle R_{ij} are at distance at most δ apart, which guarantees that the values of f at the two points are at most ε/A apart, so

$$\begin{aligned}\mathcal{U}(\mathcal{P},f)-\mathcal{L}(\mathcal{P},f) &= \sum_{i,j}\left(\sup_{R_{ij}} f - \inf_{R_{ij}} f\right)\triangle A_{ij}\\ &\leq \left(\frac{\varepsilon}{A}\right)\sum_{i,j}\triangle A_{ij}\\ &= \frac{\varepsilon}{A}\cdot A\\ &= \varepsilon.\end{aligned}$$

Thus if $\mathcal{P}_k$ are partitions satisfying $\operatorname{mesh}(\mathcal{P}_k)\to 0$, then for every $\varepsilon>0$ we eventually have $\operatorname{mesh}(\mathcal{P}_k)<\delta$ and hence $\mathcal{U}(\mathcal{P}_k,f)-\mathcal{L}(\mathcal{P}_k,f)<\varepsilon$; that is, Equation (5.3) holds, and f is Riemann integrable. □

Iterated Integrals

After all of this nice theory, we need to come back to Earth. How, in practice, can we compute the integral $\iint_{[a,b]\times[c,d]} f\,dA$ of a given function f over the rectangle $[a,b]\times[c,d]$?

The intuitive idea is to consider how we might calculate a Riemann sum for this integral. If we are given a partition $\mathcal{P}$ consisting of the partition $\mathcal{P}_1=\{a=x_0<x_1<\cdots<x_m=b\}$ of $[a,b]$ and the partition $\mathcal{P}_2=\{c=y_0<y_1<\cdots<y_n=d\}$ of $[c,d]$, the simplest way to pick a sample set $\{\overrightarrow{s}_{ij}\}$ for the Riemann sum $\mathcal{R}(\mathcal{P},f)$ is to pick a sample x-coordinate $x_i'\in I_i$ in each component interval of $\mathcal{P}_1$ and a sample y-coordinate $y_j'\in J_j$ in each component interval of $\mathcal{P}_2$, and then to declare the sample point in the subrectangle $R_{ij}=I_i\times J_j$ to be $\overrightarrow{s}_{ij}=(x_i',y_j')\in R_{ij}$ (Figure 5.4).

Then we can sum $\mathcal{R}(\mathcal{P},f)$ by first adding up along the i^{th} “column” of

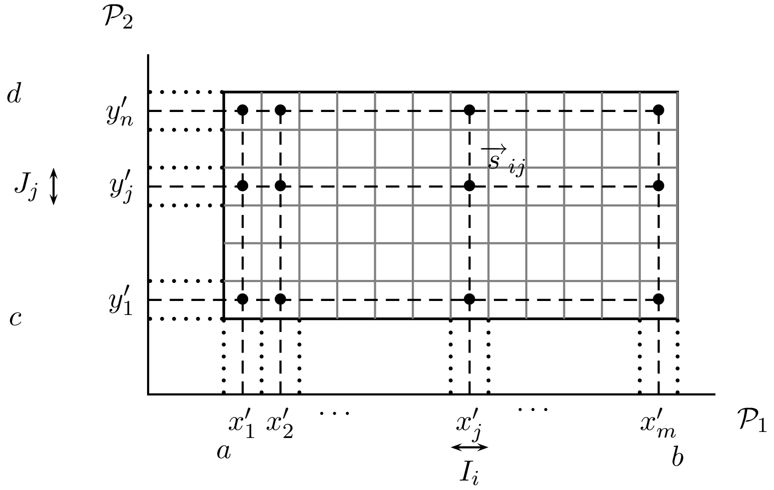


Figure 5.4: Picking a Sample Set

our partition

$$\begin{aligned} S_i &= \sum_{j=1}^{n} f\left(x'_i, y'_j\right)\ \triangle A_{ij} \\ &= \sum_{j=1}^{n} f\left(x'_i, y'_j\right)\ \triangle x_i \triangle y_j \\ &= \triangle x_i \sum_{j=1}^{n} f\left(x'_i, y'_j\right)\ \triangle y_j \end{aligned}$$

and then adding up these column sums:

$$\begin{aligned} \mathcal{R}(\mathcal{P}, f) &= \sum_{i=1}^{m} S_i \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} f\left(x'_i, y'_j\right)\ dA_{ij} \\ &= \sum_{i=1}^{m} \left(\triangle x_i \sum_{j=1}^{n} f\left(x'_i, y'_j\right)\ \triangle y_j \right) \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} f\left(x'_i, y'_j\right)\ \triangle y_j \right) \triangle x_i \end{aligned}$$

Notice that in the sum S_i, the x-value is fixed at $x = x_i'$, and S_i can be viewed as $\triangle x_i$ times a Riemann sum for the integral $\int_c^d g(y)\ dy$, where $g(y)$ is the function of y alone obtained from $f(x,y)$ by fixing the value of x at x_i': we denote this integral by

$$\int_c^d f\big(x_i', y\big)\ dy := \int_c^d g(y)\ dy.$$

This gives us a number that depends on x_i'; call it $G(x_i')$.

For example, if

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

and

$$[a,b] \times [c,d] = [0,1] \times [1,2]$$

then fixing $x = x_i'$ for some $x_i' \in [0,1]$,

$$g(y) = (x_i')^2 + 2x_i'y$$

and

$$\int_c^d g(y)\ dy = \int_1^2 \big((x_i')^2 + 2x_i'y\big)\ dy$$

which, since x_i' is a constant, equals

$$\begin{aligned}\left[(x_i')^2 y + 2x_i'\frac{y^2}{2}\right]_1^2 &= \big[2(x_i')^2 + 4x_i'\big] - \big[(x_i')^2 + x_i'\big]\\ &= (x_i')^2 + 3x_i';\end{aligned}$$

that is,

$$G\big(x_i'\big) = (x_i')^2 + 3x_i'.$$

But now, when we sum over the i^{th} “column”, we add up S_i over all values of i; since S_i is an approximation of $G(x_i')\,\triangle x_i$, we can regard $\mathcal{R}(\mathcal{P}, f)$ as a

Riemann sum for the integral of the function $G(x)$ over the interval $[a, b]$. In our example, this means

$$\begin{aligned}
\mathcal{R}(\mathcal{P}, x^2 + 2xy) &= \sum_{i=1}^{m} \sum_{j=1}^{n} \left[(x_i')^2 + 2x_i' y_j'\right] \triangle x_i \triangle y_j \\
&\approx \sum_{i=1}^{m} \left[\int_1^2 \left((x_i')^2 + 2x_i' y\right)\, dy\right] \triangle x_i \\
&= \sum_{i=1}^{m} \left[(x_i')^2 + 3x_i'\right] \triangle x_i \\
&\approx \int_0^1 \left[x^2 + 3x\right]\, dx \\
&= \left[\frac{x^3}{3} + 3\frac{x^2}{2}\right]_0^1 \\
&= \left(\frac{1}{3} + \frac{3}{2}\right) - (0) \\
&= \frac{11}{6}.
\end{aligned}$$

Ignoring for the moment the fact that we have made two approximations here, the process can be described as: first, we integrate our function treating x as a constant, so that $f(x, y)$ looks like a function of y alone: this is denoted

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

and for each value of x, yields a number—in other words, this **partial integral** is a function of x. Then we integrate *this* function (with respect to x) to get the presumed value

$$\iint_{[a,b]\times[c,d]} f\, dA = \int_a^b \left(\int_c^d f(x, y)\, dy\right)\, dx.$$

We can drop the parentheses, and simply write the result of our computation as the **iterated integral** or **double integral**

$$\int_a^b \int_c^d f(x, y)\, dy\, dx.$$

Of course, our whole process could have started by summing first over the j^{th} "*row*", and then adding up the row sums; the analogous notation would be another iterated integral,

$$\int_c^d \int_a^b f(x,y)\ dx\, dy.$$

In our example, this calculation would go as follows:

$$\begin{aligned}
\int_1^2 \int_0^1 (x^2+2xy)\, dx\, dy &= \int_1^2 \left[\int_0^1 (x^2+2xy)\, dx\right]\ dy \\
&= \int_1^2 \left[\left(\frac{x^3}{3}+x^2y\right)_{x=0}^{1}\right]\ dy \\
&= \int_1^2 \left[\left(\frac{1}{3}+y\right)-(0)\right]\ dy \\
&= \int_1^2 \left[\frac{1}{3}+y\right]\ dy \\
&= \left[\frac{y}{3}+\frac{y^2}{2}\right]_1^2 \\
&= \left[\frac{2}{3}+\frac{4}{2}\right]-\left[\frac{1}{3}+\frac{1}{2}\right] \\
&= \frac{16}{6}-\frac{5}{6} \\
&= \frac{11}{6}.
\end{aligned}$$

Let us justify our procedure.

Theorem 5.1.5 (Fubini's Theorem[7]). *If f is continuous on $[a,b]\times[c,d]$, then its integral can be computed via double integrals:*

$$\iint_{[a,b]\times[c,d]} f\, dA = \int_a^b \int_c^d f(x,y)\ dy\, dx = \int_c^d \int_a^b f(x,y)\ dx\, dy. \tag{5.4}$$

Proof. We will show the first equality above; the proof that the second iterated integral equals the definite integral is analogous.

Define a function $F(x)$ on $[a,b]$ via

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

[7] It is something of an anachronism to call this Fubini's Theorem. The result actually proved by Guido Fubini (1879-1943) [16] is far more general, and far more complicated than this. However, "Fubini's Theorem" is used generically to refer to all such results about expressing integrals over multi-dimensional regions via iterated integrals.

Given a partition $\mathcal{P}_2 = \{c = y_0 < y_1 < \cdots < y_n = d\}$ of $[c, d]$, we can break the integral defining $F(x)$ into the component intervals J_j of $\mathcal{P}_2$

$$F(x) = \sum_{j=1}^{n} \int_{y_{j-1}}^{y_j} f(x, y)\, dy$$

and since $f(x, y)$ is continuous, we can apply the Integral Mean Value Theorem (*Calculus Deconstructed*, Proposition 5.2.10) on J_j to find a point $Y_j(x) \in J_j$ where the value of $f(x, y)$ equals its average (with x fixed) over J_j; it follows that the sum above

$$= \sum_{j=1}^{n} f(x, Y_j(x)) \triangle y_j.$$

Now if $\mathcal{P}_1 = \{a = x_0 < x_1 < \cdots < x_m = b\}$ is a partition of $[a, b]$ then a Riemann sum for the integral $\int_a^b F(x)\, dx$, using the sample coordinates $x_i \in I_i$, is

$$\sum_{i=1}^{m} F(x_i) \triangle x_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} f(x_i, Y_j(x_i)) \triangle y_j \right) \triangle x_i;$$

but this is also a Riemann sum for the integral $\iint_{[a,b]\times[c,d]} f\, dA$ using the "product" partition $\mathcal{P}$ generated by $\mathcal{P}_1$ and $\mathcal{P}_2$, and the sample coordinates

$$s_{ij} = (x_i, Y_i(x_i)).$$

Thus, if we pick a sequence of partitions of $[a, b] \times [c, d]$ with mesh going to zero, the left sum above converges to $\int_a^b F(x)\, dx$ while the right sum converges to $\iint_{[a,b]\times[c,d]} f\, dA$. □

If the function $f(x, y)$ is positive over the rectangle $[a, b] \times [c, d]$, then the definite integral $\iint_{[a,b]\times[c,d]} f\, dA$ is interpreted as the volume between the graph of $f(x, y)$ and the xy-plane. In this case, the calculation via iterated integrals can be interpreted as finding this area by "slicing" parallel to one of the vertical coordinate planes (see Figure 5.5); this is effectively an application of **Cavalieri's Principle**:

> *If two solid bodies intersect each of a family of parallel planes in regions with equal areas, then the volumes of the two bodies are equal.*

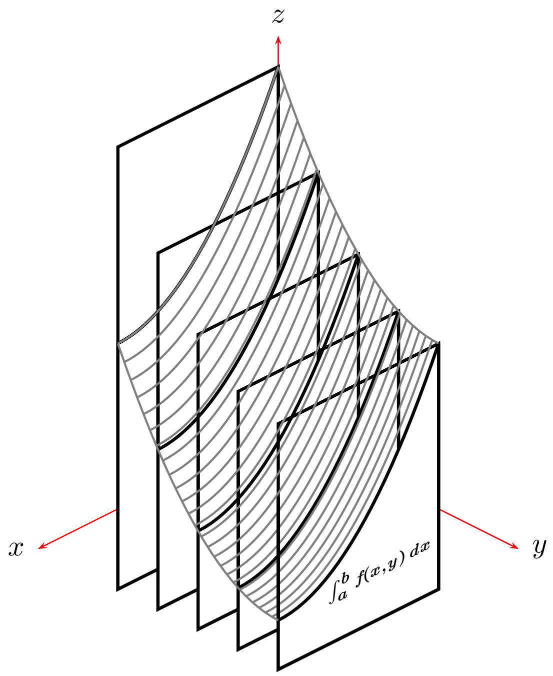


Figure 5.5: Fubini's Theorem: Volume via Slices

Let us consider a few more examples of this process.

The integral

$$\iint_{[-1,1]\times[0,1]} (x^2+y^2)\,dA$$

can be calculated via two different double integrals:

$$\begin{aligned}
\int_0^1 \int_{-1}^1 (x^2+y^2)\,dA &= \int_0^1 \left[\frac{x^3}{3}+xy^2\right]_{x=-1}^1 dy\\
&= \int_0^1 \left[\left(\frac{1}{3}+y^2\right)-\left(-\frac{1}{3}-y^2\right)\right] dy\\
&= \int_0^1 \left[\frac{2}{3}+2y^2\right] dy\\
&= \left[\frac{2}{3}y+\frac{2y^3}{3}\right]_{y=0}^1\\
&= \frac{4}{3}
\end{aligned}$$

or

$$\begin{aligned}\int_{-1} 1 \int_0^1 \left(x^2+y^2\right)\, dy\, dx &= \int_{-1}^1 \left[x^2 y + \frac{y^3}{3}\right]_{y=0}^1 dy \\ &= \int_{-1}^1 \left[\left(x^2+\frac{1}{3}\right) - (0)\right] dy \\ &= \left[\frac{x^3}{3} + \frac{x}{3}\right]_{x=-1}^1 \\ &= \left[\frac{1}{3}+\frac{1}{3}\right] - \left[-\frac{1}{3}-\frac{1}{3}\right] \\ &= \frac{4}{3}.\end{aligned}$$

A somewhat more involved example shows that the order in which we do the double integration can affect the difficulty of the process. The integral

$$\iint_{[1,4]\times[0,1]} y\sqrt{x+y^2}\, dA$$

can be calculated two ways. To calculate the double integral

$$\int_0^1 \int_1^4 y\sqrt{x+y^2}\, dx\, dy$$

we start with the "inner" partial integral, in which y is treated as a constant: using the substitution

$$\begin{aligned} u &= x+y^2 \\ du &= dx \end{aligned}$$

we calculate the indefinite integral as

$$\begin{aligned}\int y\sqrt{x+y^2}\, dx &= \int y\, u^{1/2}\, du \\ &= \frac{2}{3} y\, u^{3/2} + C \\ &= \frac{2}{3} y(x+y^2)^{3/2} + C\end{aligned}$$

so the (inner) definite integral is

$$\begin{aligned}\int_1^4 y\sqrt{x+y^2}\, dx &= \frac{2}{3} y\left(y^2+x\right)_{x=1}^4 \\ &= \frac{2}{3}\left[y(y^2+4)^{3/2} - y(y^2+1)^{3/2}\right].\end{aligned}$$

Thus the "outer" integral becomes

$$\frac{2}{3}\int_0^1 \left[y(y^2+4)^{3/2} - y(y^2+1)^{3/2}\right]\,dy.$$

Using the substitution

$$\begin{aligned} u &= y^2+4 \\ du &= 2y\,dy \end{aligned}$$

we calculate the indefinite integral of the first term as

$$\begin{aligned} \int y(y^2+4)\,dy &= \frac{1}{2}\int u^{3/2}\,du \\ &= \frac{1}{5}u^{5/2}+C \\ &= \frac{1}{5}(y^2+4)^{5/2}+C; \end{aligned}$$

similarly, the indefinite integral of the second term is

$$\int y(y^2+1)\,dy = \frac{1}{5}(y^2+1)^{5/2}+C.$$

It follows that the whole "outer" integral is

$$\begin{aligned} \frac{2}{3}\int_0^1 \left[y(y^2+4)^{3/2} - y(y^2+1)^{3/2}\right]\,dy &= \frac{2}{15}\left[(y^2+4)^{5/2}-(y^2+1)^{5/2}\right]_{y=0}^{1} \\ &= \frac{2}{15}\left[\left(5^{5/2}-2^{5/2}\right)-\left(4^{5/2}-1^{5/2}\right)\right] \\ &= \frac{2}{15}\left[25\sqrt{5}-4\sqrt{2}-31\right]. \end{aligned}$$

If instead we perform the double integration in the opposite order

$$\int_1^4\int_0^1 y\sqrt{x+y^2}\,dy\,dx$$

the "inner" integral treats x as constant; we use the substitution

$$\begin{aligned} u &= x+y^2 \\ du &= 2y\,dy \end{aligned}$$

to find the "inner" indefinite integral

$$\begin{aligned}\int y\sqrt{x+y^2}\,dy &= \int \frac{1}{2}u^{1/2}\,du\\ &= \frac{1}{3}u^{3/2}+C\\ &= \frac{1}{3}(x+y^2)^{3/2}+C\end{aligned}$$

so the definite "inner" integral is

$$\begin{aligned}\int_0^1 y\sqrt{x+y^2}\,dy &= \frac{1}{3}(x+y^2)^{3/2}\Big|_{y=0}^{1}\\ &= \frac{1}{3}\left[(x+1)^{3/2}-(x)^{3/2}\right].\end{aligned}$$

Now the "outer" integral is

$$\begin{aligned}\frac{1}{3}\int_1^4\left[(x+1)^{3/2}-(x)^{3/2}\right]\,dx &= \frac{1}{3}\left[\frac{2}{5}(x+1)^{5/2}-\frac{2}{5}x^{5/2}\right]_{x=1}^{4}\\ &= \frac{2}{15}\left[(5^{5/2}-4^{5/2})-(2^{5/2}-1^{5/2})\right]\\ &= \frac{2}{15}\left[25\sqrt{5}-4\sqrt{2}-31\right].\end{aligned}$$

Finally, let us find the volume of the solid with vertical sides whose base is the rectangle $[0,1]\times[0,1]$ in the xy-plane and whose top is the planar quadrilateral with vertices $(0,0,4)$, $(1,0,2)$, $(0,1,3)$, and $(1,1,1)$ (Figure 5.6).

First, we should find the equation of the top of the figure. Since it is planar, it has the form

$$z = ax+by+c;$$

substituting each of the four vertices into this yields four equations in the three unknowns a, b and c

$$\begin{aligned}4 &= c\\ 2 &= a+c\\ 3 &= b+c\\ 1 &= a+b+c.\end{aligned}$$

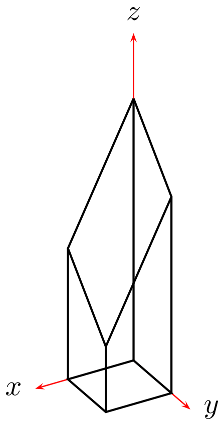


Figure 5.6: A Volume

The first three equations have the solution

$$\begin{aligned} a &= -2 \\ b &= -1 \\ c &= 4 \end{aligned}$$

and you can check that this also satisfies the fourth equation; so the top is the graph of

$$z = 4 - 2x - 3y.$$

Thus, our volume is given by the integral

$$\begin{aligned} \iint_{[0,1]\times[0,1]} (4-2x-3y)\,dA &= \int_0^1 \int_0^1 (4-2x-3y)\,dx\,dy \\ &= \int_0^1 \left[4x - x^2 - 3xy\right]_{x=0}^1 \\ &= \int_0^1 \left[(3-3y)-(0)\right]\,dy \\ &= \left[3y - \frac{3y^2}{2}\right]_{y=0}^1 \\ &= \frac{3}{2}. \end{aligned}$$

Exercises for § 5.1

Practice problems:

1. Calculate each integral below:

   (a) $\iint_{[0,1]\times[0,2]} 4x\,dA$ (b) $\iint_{[0,1]\times[0,2]} 4xy\,dA$

   (c) $\iint_{[0,1]\times[0,\pi]} x\sin y\,dA$ (d) $\iint_{[0,1]\times\left[-\frac{\pi}{2},\frac{\pi}{2}\right]} e^x\cos y\,dA$

   (e) $\iint_{[0,1]\times[1,2]} (2x+4y)\,dA$ (f) $\iint_{[1,2]\times[0,1]} (2x+4y)\,dA$

Theory problems:

2. Let $\mathcal{P}$ and $\mathcal{P}'$ be partitions of the rectangle $[a,b]\times[c,d]$.

   (a) Show that if every partition point of $\mathcal{P}$ is also a partition point of $\mathcal{P}'$ (that is, $\mathcal{P}'$ is a refinement of $\mathcal{P}$) then for any function f

   $$\mathcal{L}(\mathcal{P},f)\le\mathcal{L}(\mathcal{P}',f)\le\mathcal{U}(\mathcal{P}',f)\le\mathcal{U}(\mathcal{P},f).$$

   (b) Use this to show that for *any* two partitions $\mathcal{P}$ and $\mathcal{P}'$,

   $$\mathcal{L}(\mathcal{P},f)\le\mathcal{U}(\mathcal{P}',f).$$

   (*Hint:* Use the above on the mutual refinement $\mathcal{P}\vee\mathcal{P}'$, whose partition points consist of all partition points of $\mathcal{P}$ together with those of $\mathcal{P}'$.)

3. Let $\mathcal{P}$ be a partition of $[a,b]\times[c,d]$ and f a function on $[a,b]\times[c,d]$. Show that the Riemann sum $\mathcal{R}(\mathcal{P},f)$ corresponding to any choice of sample points is between the lower sum $\mathcal{L}(\mathcal{P},f)$ and the upper sum $\mathcal{U}(\mathcal{P},f)$.

4. Show that if $\mathcal{P}_k$ is a sequence of partitions of $[a,b]\times[c,d]$ for which $\mathcal{L}(\mathcal{P}_k,f)$ and $\mathcal{U}(\mathcal{P}_k,f)$ both converge to $\iint_{[a,b]\times[c,d]} f\,dA$ then for any choice of sample points in each partition, the corresponding Riemann sums also converge there.

5. Show that the diameter of a rectangle equals the length of its diagonal, and that this is always greater than the maximum of its sides.

6. Let f be any function on $[a,b] \times [c,d]$.

   (a) Show that for any partition $\mathcal{P}$ of $[a,b]\times[c,d]$, $\mathcal{L}(\mathcal{P},f) \leq \sup_{\mathcal{P}} \mathcal{L}(\mathcal{P},f)$ and $\inf_{\mathcal{P}} \mathcal{U}(\mathcal{P},f) \leq \mathcal{U}(\mathcal{P},f)$.

   (b) Use this, together with Exercise 3, to show that if

   $$\sup_{\mathcal{P}} \mathcal{L}(\mathcal{P},f) = \inf_{\mathcal{P}} \mathcal{U}(\mathcal{P},f)$$

   then there exists a sequence $\mathcal{P}_k$ of partitions such that

   $$\mathcal{U}(\mathcal{P}_k,f) - \mathcal{L}(\mathcal{P}_k,f) \to 0$$

   and conversely that the existence of such a sequence implies that $\sup_{\mathcal{P}} \mathcal{L}(\mathcal{P},f) = \inf_{\mathcal{P}} \mathcal{U}(\mathcal{P},f)$.

   (c) Use this, together with Exercise 4, to show that if f is integrable, then for any such sequence, the Riemann sums corresponding to any choices of sample points converge to the integral of f over $[a,b] \times [c,d]$.

5.2 Integration over General Planar Regions

In this section we extend our theory of integration to more general regions in the plane. By a "region" we mean a bounded set defined by a finite set of inequalities of the form $g_i(x,y) \leq c_i$, $i = 1,\ldots,k$, where the functions g_i are presumed to be reasonably well behaved. When the Implicit Function Theorem (Theorem 3.4.2) applies, the region is bounded by a finite set of (level) curves, each of which can be viewed as a graph of the form $y = \phi(x)$ or $x = \psi(y)$. In fact, the most general kind of "region" over which such an integration can be performed was the subject of considerable study in the 1880's and early 1890's [23, pp. 86-96]. The issue was finally resolved by Camille Marie Ennemond Jordan (1838-1922) in 1892; his solution is well beyond the scope of this book.

Discontinuities and Integration

The basic idea for integrating a function $f(x,y)$ over a general region takes its inspiration from our idea of the area of such a region: we try to "subdivide" the region into rectangles (in the sense of the preceding section) and add up the integrals over them. Of course, this is essentially impossible for most regions, and instead we need to think about two kinds of *approximate* calculations: "inner" ones using rectangles entirely contained inside the region, and "outer" ones over unions of rectangles which contain our region (rather like the inscribed and circumscribed polygons used to find the area of a circle). For the theory to make sense, we need to make sure that these two calculations give rise to the same value for the integral. This is done via the following technical lemma.

Lemma 5.2.1. *Suppose a curve $\mathcal{C}$ is the graph of a continuous function, $y = \phi(x)$, $a \leq x \leq b$. Then given any $\varepsilon > 0$ we can find a finite family of rectangles $B_i = [a_i, b_i] \times [c_i, d_i]$, $i = 1, \ldots, k$, covering the curve (Figure 5.7)*

$$\mathcal{C} \subset \bigcup_{i=1}^{k} B_i$$

such that

1. *Their total area is at most ε*

$$\sum_{i=1}^{k} \mathcal{A}(B_i) \leq \varepsilon.$$

2. *The horizontal edges of each B_i are disjoint from $\mathcal{C}$*

$$c_i < \phi(x) < d_i \text{ for } a_i \leq x \leq b_i.$$

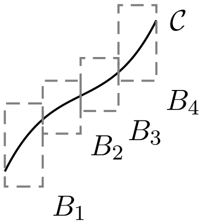


Figure 5.7: Lemma 5.2.1

Proof. Given $\varepsilon > 0$, use the uniform continuity of the function ϕ to pick $\delta > 0$ such that

$$\|x - x'\| < \delta \Rightarrow |\phi(x) - \phi(x')| < \frac{\varepsilon}{3|b-a|},$$

and let $\mathcal{P} = \{a = x_0 < x_1 < \cdots < x_k\}$ be a partition of $[a, b]$ with $\text{mesh}(\mathcal{P}) < \delta$. Then, for each $i = 1, \ldots, k$, let

$$J_i = \left[\min_{I_i} \phi - \frac{\varepsilon}{3\,|b-a|}, \max_{I_i} \phi + \frac{\varepsilon}{3\,|b-a|}\right]$$

and set

$$B_i := I_i \times J_i.$$

We wish to show that these rectangles satisfy the two properties in the statement of the lemma:

1. We know that
$$\mathcal{A}\left(B_i\right) = \|I_i\| \cdot \|J_i\|\,.$$
Since by assumption $\text{mesh}(\mathcal{P}) < \delta$, we also know that
$$\max_{I_i} \phi - \min_{I_i} \phi \leq \frac{\varepsilon}{3\,|b-a|}$$
so
$$\begin{aligned}\|J_i\| &= \left(\max_{I_i} \phi + \frac{\varepsilon}{3\,|b-a|}\right) - \left(\min_{I_i} \phi - \frac{\varepsilon}{3\,|b-a|}\right) \\ &= \left(\max_{I_i} \phi - \min_{I_i} \phi\right) + 2\left(\frac{\varepsilon}{3\,|b-a|}\right) \\ &\leq 3\left(\frac{\varepsilon}{3\,|b-a|}\right) \\ &= \frac{\varepsilon}{|b-a|};\end{aligned}$$
hence
$$\mathcal{A}\left(B_i\right) \leq \frac{\varepsilon}{|b-a|} I_i$$
and summing over $i = 1, \ldots, k$
$$\begin{aligned}\sum_i^k \mathcal{A}\left(B_i\right) &\leq \frac{\varepsilon}{|b-a|} \sum_i^k I_i \\ &= \frac{\varepsilon}{|b-a|}\,|b-a| \\ &= \varepsilon\end{aligned}$$
as required.

2. By construction,

$$\begin{aligned} c_i &= \min_{I_i} \phi - \frac{\varepsilon}{3\,|b-a|} \\ &< \min_{I_i} \phi \end{aligned}$$

and

$$\begin{aligned} \max_{I_i} \phi &< \max_{I_i} \phi + \frac{\varepsilon}{3\,|b-a|} \\ &= d_i \end{aligned}$$

as required.

□

Using this result, we can extend the class of functions which are Riemann integrable beyond those continuous on the whole rectangle (as given in Theorem 5.1.4), allowing certain kinds of discontinuity. This will in turn allow us to define the integral of a function over a more general region in the plane.

Theorem 5.2.2. *If a function f is bounded on $[a,b]\times[c,d]$ and continuous except possibly for some points lying on a finite union of graphs (curves of the form $y=\phi(x)$ or $x=\psi(y)$), then f is Riemann integrable over $[a,b]\times[c,d]$.*

Proof. For ease of notation, we shall assume that f is bounded on $[a,b]\times[c,d]$ and that any points of discontinuity lie on a single graph $\mathcal{C}:\quad y=\phi(x)$, $a\leq x\leq b$.

Given $\varepsilon>0$, we need to find a partition $\mathcal{P}$ of $[a,b]\times[c,d]$ for which

$$\mathcal{U}(\mathcal{P},f)-\mathcal{L}(\mathcal{P},f)<\varepsilon.$$

First, since f is bounded on $[a,b]\times[c,d]$, pick an upper bound for $|f|$ on $[a,b]\times[c,d]$, say

$$M>\max\{1, sup_{[a,b]\times[c,d]}\,|f|\}.$$

Next, use Lemma 5.2.1 to find a finite family B_i, $i=1,\ldots,k$, of rectangles covering the graph $y=\phi(x)$ such that

$$\sum_{i=1}^{k}\mathcal{A}\left(B_i\right)<\frac{\varepsilon}{2M}.$$

Now extend each edge of each B_i to go completely across the rectangle $[a,b] \times [c,d]$ (horizontally or verticaly)—there are finitely many such lines, and they define a partition $\mathcal{P}$ of $[a,b] \times [c,d]$ such that each B_i (and hence the union of all the B_i) is itself a union of subrectangles R_{ij} for $\mathcal{P}$. Note that if we refine this partition further by adding more (horizontal or vertical) lines, it will still be true that $\mathcal{B} = \bigcup_{i=1}^{k} B_i$ is a union of subrectangles, and

$$\begin{aligned}\left(\sum_{R_{ij}\subset\mathcal{B}} \sup_{R_{ij}} f \triangle A_{ij}\right) - \left(\sum_{R_{ij}\subset\mathcal{B}} \inf_{R_{ij}} f \triangle A_{ij}\right) &= \sum_{R_{ij}\subset\mathcal{B}} \left(\sup_{R_{ij}} f - \inf_{R_{ij}} f\right) \triangle A_{ij} \\ &< M \cdot \mathcal{A}\left(\mathcal{B}\right) \\ &< M\left(\frac{\varepsilon}{2M}\right) \\ &= \frac{\varepsilon}{2}.\end{aligned}$$

Finally, consider the union $\mathcal{D}$ of the rectangles of $\mathcal{P}$ which are disjoint from $\mathcal{C}$. This is a compact set on which f is continuous, so f is *uniformly* continuous on $\mathcal{D}$; hence as in the proof of Theorem 5.1.4 we can find $\delta > 0$ such that for any of the subrectangles R_{ij} contained in $\mathcal{D}$ we have

$$\sup_{R_{ij}} f - \inf_{R_{ij}} f < \frac{\varepsilon}{2\mathcal{A}\left([a,b] \times [c,d]\right)}$$

so that

$$\left(\sum_{R_{ij}\subset\mathcal{D}} \sup_{R_{ij}} f \triangle A_{ij}\right) - \left(\sum_{R_{ij}\subset\mathcal{D}} \inf_{R_{ij}} f \triangle A_{ij}\right) < \frac{\varepsilon}{2\mathcal{A}\left([a,b] \times [c,d]\right)}\mathcal{A}\left(\mathcal{D}\right).$$

From this it follows that for our final partition,

$$\begin{aligned}\mathcal{U}(\mathcal{P},f) - \mathcal{L}(\mathcal{P},f) &= \sum_{R_{ij}\subset\mathcal{B}} \left(\sup_{R_{ij}} f - \inf_{R_{ij}} f\right) \triangle A_{ij} + \sum_{R_{ij}\subset\mathcal{D}} \left(\sup_{R_{ij}} f - \inf_{R_{ij}} f\right) \triangle A_{ij} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2\mathcal{A}\left([a,b] \times [c,d]\right)}\mathcal{A}\left(\mathcal{D}\right) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon\end{aligned}$$

as required.

□

Integration over Non-Rectangular Regions

Suppose we have a function $f(x, y)$ defined and positive on a rectangle $[a, b] \times [c, d]$, and we wish to find a volume under its graph—not the volume over the whole rectangle, but only the part above a subregion $D \subset [a, b] \times [c, d]$. One way to do this is to "crush" the part of the graph outside $\mathcal{D}$ down to the xy-plane and integrate the resulting function defined in pieces

$$f \restriction_{\mathcal{D}} (\overrightarrow{x}) := \begin{cases} f(\overrightarrow{x}) & \text{if } \overrightarrow{x} \in \mathcal{D}, \\ 0 & \text{otherwise.} \end{cases}$$

Of course, this definition makes sense even if f is not positive on $[a, b] \times [c, d]$. And this process can be turned around: the definition above extends any function which is defined at least on $\mathcal{D}$ to a function defined on the whole plane.

Definition 5.2.3. *If $f(x, y)$ is defined at every point of the bounded set $\mathcal{D}$, then the integral of f over $\mathcal{D}$ is defined as*

$$\iint_{\mathcal{D}} f\, dA := \iint_{[a,b]\times[c,d]} f \restriction_{\mathcal{D}}\, dA$$

where $[a, b] \times [c, d]$ is any rectangle containing $\mathcal{D}$ (provided this integral exists, i.e., provided $f \restriction_{\mathcal{D}}$ is Riemann integrable on $[a, b] \times [c, d]$).

An immediate consequence of Theorem 5.2.2 is the following.

Remark 5.2.4. *If f is continuous on a region $\mathcal{D}$ bounded by finitely many graphs of continuous functions $y = \phi(x)$ or $x = \psi(y)$, then Theorem 5.2.2 guarantees that $\iint_{\mathcal{D}} f\, dA$ is well-defined.*

We shall only consider these kinds of regions. Any such region can be broken down into regions of a particularly simple type:

Definition 5.2.5. *A region $D \subset \mathbb{R}^2$ is **y-regular** if it can be specified by inequalities on y of the form*

$$\mathcal{D} = \{(x, y) \,|\, c(x) \leq y \leq d(x)\,, \quad a \leq x \leq b\}$$

where $c(x)$ and $d(x)$ are continuous and satisfy $c(x) \leq d(x)$ on $[a, b]$. (See Figure 5.8.)

*It is **x-regular** if it can be specified by inequalities on x of the form*

$$\mathcal{D} = \{(x, y) \,|\, a(y) \leq x \leq b(y)\,, \quad c \leq y \leq d\}$$

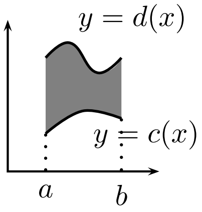


Figure 5.8: y-regular region

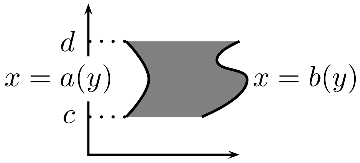


Figure 5.9: x-regular region

where $a(y)$ and $b(y)$ are continuous and satisfy $a(y) \leq b(y)$ on $[c, d]$. (See Figure 5.9.)

A region which is both *x- and y-regular is (simply)* ***regular*** *(see Figure 5.10), and regions of* either *type are called* ***elementary regions****.*

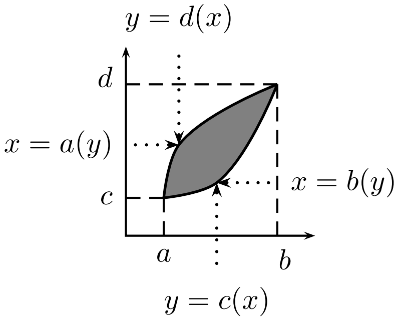


Figure 5.10: Regular region

Basically, a region is y-regular if first, every line parallel to the y-axis intersects the region, if at all, in a closed interval, and second, if each of the endpoints of this interval vary continuously as functions of the x-intercept of the line.

Integration over elementary regions can be done via iterated integrals.

We illustrate with an example.

Let $\mathcal{D}$ be the triangle with vertices $(0,0)$, $(1,0)$ and $(1,1)$. $\mathcal{D}$ is is a regular region, bounded by the line $y = x$, the x-axis ($y = 0$) and the line $x = 1$ (Figure 5.11). Slicing vertically, it can be specified by the pair of inequalities

$$\begin{aligned} 0 \leq & y \leq x \\ 0 \leq & x \leq 1 \end{aligned}$$

or, slicing horizontally, by the inequalities

$$\begin{aligned} y \leq & x \leq 1 \\ 0 \leq & y \leq 1. \end{aligned}$$

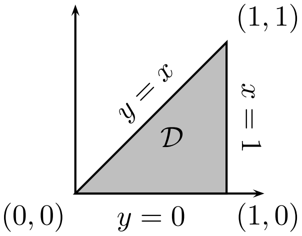


Figure 5.11: The triangle $\mathcal{D}$

To integrate the function $f(x) = 12x^2 + 6y$ over $\mathcal{D}$, we can enclose $\mathcal{D}$ in the rectangle $[0,1] \times [0,1]$ and then integrate $f \restriction_{x,y}$ over this rectangle, slicing vertically. This leads to the double integral

$$\int_0^1 \int_0^1 f \restriction_{\mathcal{D}} \, dy \, dx.$$

Now, since the function $f \restriction_{\mathcal{D}}$ is defined in pieces,

$$f \restriction_{\mathcal{D}} (x,y) = \begin{cases} 12x^2 + 6y & \text{if } 0 \leq y \leq x \text{ and } x \in [0,1], \\ 0 & \text{otherwise,} \end{cases}$$

the "inner" integral $\int_0^1 f \restriction_{\mathcal{D}} \, dy$ (with $x \in [0,1]$ fixed) can be broken into two parts

$$\int_0^1 f \restriction_{\mathcal{D}} \, dy = \int_0^x (12x^2 + 6y) \, dy + \int_x^1 (12x^2 + 6y) \, dy$$

and since the integrand is zero in the second integral, we can write

$$\int_0^1 f\restriction_{\mathcal{D}}\, dy = \int_0^x (12x^2 + 6y)\, dy$$

which is a regular integral when x is treated as a constant:

$$\begin{aligned} &= \left(12x^2y + 3y^2\right)_{y=0}^{y=x} \\ &= (12x^3 + 3x^2) - (0). \end{aligned}$$

Now, we can write the "outer" integral as

$$\begin{aligned} \int_0^1 \left(\int_0^1 f\restriction_{\mathcal{D}}\, dy\right) dx &= \int_0^1 \int_0^x (12x^2 + 6y)\, dy\, dx \\ &= \int_0^1 (12x^3 + 3x^2)\, dx \\ &= (3x^4 + x^3)_{x=0}^{x=1} \\ &= (3 + 1) - (0) \\ &= 4. \end{aligned}$$

Alternatively, if we slice horizontally, we get the "inner" integral (with $y \in [0, 1]$ fixed)

$$\begin{aligned} \int_0^1 f\restriction_{\mathcal{D}}\, dx &= \int_0^y (12x^2 + 6y)\, dy + \int_y^1 (12x^2 + 6y)\, dx \\ &= \int_y^1 (12x^2 + 6y)\, dx \end{aligned}$$

(since $f\restriction_{\mathcal{D}}(x)$ is zero to the left of $x = y$)

$$\begin{aligned} &= (4x^3 + 6xy)_{x=y}^{x=1} \\ &= (4 + 6y) - (4y^3 + 6y^2). \end{aligned}$$

Then the outer integral is

$$\begin{aligned}\int_0^1 \left(\int_0^1 f\restriction_{\mathcal{D}}\, dx\right) dy &= \int_0^1 \int_y^1 (12x^2+6y)\, dx\, dy\\ &= \int_0^1 (4+6y-4y^3-6y^2)\, dy\\ &= (4y+3y^2-y^4-2y^3)_{y=0}^{y=1}\\ &= (4+3-1-2)-(0)\\ &= 4.\end{aligned}$$

The procedure illustrated by this example is only a slight modification of what we do to integrate over a rectangle. In the case of a rectangle, the "inner" integral has fixed limits, and we integrate regarding all but one variable in the integrand as constant. The result is a function of the other variable. When integrating over a y-regular region, the limits of the inner integration may also depend on x, but we regard x as fixed in both the limits and the integrand; this still yields an integral that depends on the value of x—that is, it is a function of x alone—and in the "outer" integral we simply integrate this function with respect to x, with fixed (numerical) limits. The analogous procedure, with the roles of x and y reversed, results from slicing horizontally, when $\mathcal{D}$ is x-regular.

We illustrate with some further examples.

Let $\mathcal{D}$ be the region bounded by the curves

$$y = x+1$$

and

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

to find their intersection, we solve

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

or

$$x^2-x-2=0$$

whose solutions are

$$x=-1,2.$$

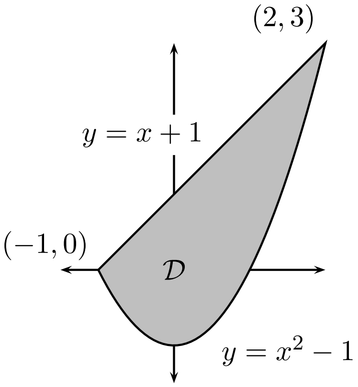


Figure 5.12: The region $\mathcal{D}$

(see Figure 5.12).

To calculate the integral

$$\iint_{\mathcal{D}} (x + 2y)\, dA$$

over this region, which is presented to us in y-regular form, we slice vertically: a vertical slice is determined by an x-value, and runs from $y = x^2 - 1$ up to $y = x + 1$; the possible x-values run from $x = -1$ to $x = 2$ (Figure 5.13).

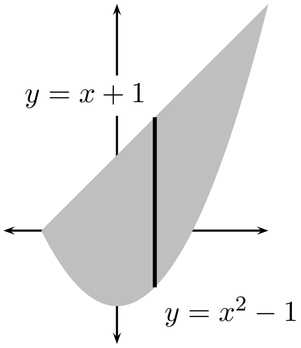


Figure 5.13: Vertical Slice

This leads to the double integral

$$\begin{aligned}
\int_{1}^{2}\int_{x^2-1}^{x+1}(x+2y)\,dy\,dx &= \int_{-1}^{2}(xy+y^2)_{y=x^2-1}^{y=x+1}\,dx\\
&= \int_{-1}^{2}\left[\{x(x+1)+(x+1)^2\}-\{x(x^2-1)+(x^21)^2\}\right]\,dx\\
&= \int_{-1}^{2}\left[\{x^2+x+x^2+2x+1\}-\{x^3-x+x^4-2x^2+1\}\right]\,dx\\
&= \int_{-1}^{2}\left[-x^4-x^3+4x^2+4x\right]\,dx\\
&= \left[-\frac{x^5}{5}-\frac{x^4}{4}+\frac{4x^3}{3}+2x^2\right]_{-1}^{2}\\
&= \left[-\frac{32}{5}-4+\frac{32}{3}+8\right]-\left[\frac{1}{5}-\frac{1}{4}-\frac{4}{3}+2\right]\\
&= \frac{153}{20}\\
&= 7\frac{13}{20}.
\end{aligned}$$

Now technically, the region $\mathcal{D}$ is also x-regular, but horizontal slices are much more cumbersome: horizontal slices *below* the x-axis run between the two solutions of $y = x^2 - 1$ for x in terms of y, which means the horizontal slice at height $-1 \leq y \leq 0$ runs from $x = -\sqrt{y+1}$ to $x = \sqrt{y+1}$, while horizontal slices *above* the x-axis at height $0 \leq y \leq 3$ run from $x = y - 1$ to $x = \sqrt{y+1}$ (Figure 5.14).

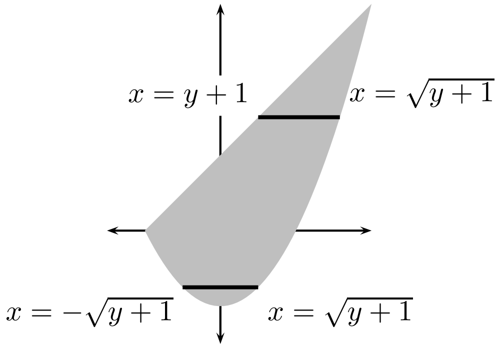


Figure 5.14: Horizontal Slices

This leads to the pair of double integrals

$$\int_{-1}^{0}\int_{-\sqrt{y+1}}^{\sqrt{y+1}}(x+2y)\,dx\,dy+\int_{0}^{3}\int_{y-1}^{\sqrt{y+1}}(x+2y)\,dx\,dy$$

which is a lot messier than the previous calculation.

As another example, let us find the volume of the simplex (or "pyramid") cut off from the first octant by the triangle with vertices one unit out along each coordinate axis (Figure 5.15). The triangle is the graph of a linear function

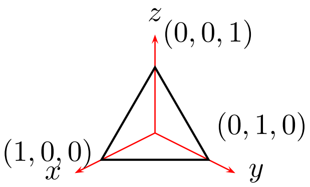


Figure 5.15: Simplex

$$z=ax+by+c$$

and satisfies

$$\begin{aligned}0&=a+c\\0&=b+c\\1&=c\end{aligned}$$

so the graph in question is

$$z=-x-y+1.$$

We are interested in the integral of this function over the triangle T in the xy-plane with vertices at the origin, $(1,0)$ and $(0,1)$ (Figure 5.16). It is fairly easy to see that the upper edge of this triangle has equation $x+y=1$, so T is described by the (y-regular) inequalities

$$\begin{aligned}0\leq&y\leq 1-x\\0\leq&x\leq 1;\end{aligned}$$

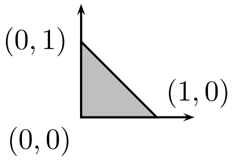


Figure 5.16: The base of the simplex, the triangle T

that is, a vertical slice at $0 \le x \le 1$ runs from $y = 0$ to $y = 1 - x$. Hence the volume in question is given by the integral

$$\begin{aligned}\iint_T (1 - x - y)\, dA &= \int_0^1 \int_0^{1-x} (1 - x - y)\, dy\, dx \\ &= \int_0^1 \left(y - xy - \frac{y^2}{2} \right)_{y=0}^{y=1-x} dx \\ &= \int_0^1 \left((1 - x) - x(1 - x) - \frac{(1 - x)^2}{2} \right) dx \\ &= \int_0^1 \left(\frac{(1 - x)^2}{2} \right) dx \\ &= -\frac{(1 - x)^3}{6}\Big|_0^1 \\ &= \frac{1}{6}.\end{aligned}$$

Sometimes the integral dictates which way we slice. For example, consider the integral

$$\iint_{\mathcal{D}} \sqrt{a^2 - y^2}\, dA$$

where $\mathcal{D}$ is the part of the circle $x^2 + y^2 = a^2$ in the first quadrant (Figure 5.17). The y-regular description of this region is

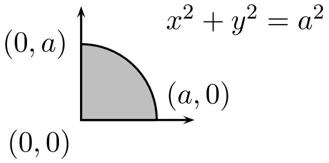


Figure 5.17: The quarter-circle $\mathcal{D}$

$$\begin{aligned} 0 \leq & y \leq \sqrt{a^2-x^2} \\ 0 \leq & x \leq a \end{aligned}$$

which leads to the double integral

$$\iint_{\mathcal{D}} \sqrt{a^2-y^2}\,dA = \int_0^a \int_0^{\sqrt{a^2-x^2}} \sqrt{a^2-y^2}\,dy\,dx;$$

the inner integral can be done, but requires a trigonometric substitution (and the subsequent evaluation at the limits is a real mess). However, if we consider the region as x-regular, with description

$$\begin{aligned} 0 \leq & x \leq \sqrt{a^2-y^2} \\ 0 \leq & y \leq a \end{aligned}$$

we come up with the double integral

$$\iint_{\mathcal{D}} \sqrt{a^2-y^2}\,dA = \int_0^a \int_0^{\sqrt{a^2-y^2}} \sqrt{a^2-y^2}\,dx\,dy;$$

since the integrand is constant as far as the inner integral is concerned, we can easily integrate this:

$$\begin{aligned} \int_0^a \left(\int_0^{\sqrt{a^2-y^2}} \sqrt{a^2-y^2}\,dx \right) dy &= \int_0^a \left(x\sqrt{a^2-y^2} \right)_{x=0}^{x=\sqrt{a^2-y^2}} dy \\ &= \int_0^a \left(\sqrt{a^2-y^2} \right)^2 dy \\ &= \int_0^a \left(a^2-y^2 \right) dy \\ &= \left(a^2 y - \frac{y^3}{3} \right)_{y=0}^{a} \\ &= \left(a^3 - \frac{a^3}{3} \right) - (0) \\ &= \frac{2a^3}{3}. \end{aligned}$$

This illustrates the usefulness of reinterpreting a double integral geometrically and then switching the order of iterated integration. As another example, the double integral

$$\int_0^1 \int_y^1 \frac{\sin x}{x}\, dx\, dy.$$

Here, the inner integral is impossible.[8] However, the double integral is the x-regular version of

$$\iint_{\mathcal{D}} \frac{\sin x}{x}\, dA$$

where $\mathcal{D}$ is the triangle in Figure 5.11, and $\mathcal{D}$ can also be described in y-regular form

$$\begin{aligned} 0 \leq & y \leq x \\ 0 \leq & x \leq 1 \end{aligned}$$

leading to the double integral

$$\iint_{\mathcal{D}} \frac{\sin x}{x}\, dA = \int_0^1 \int_0^x \frac{\sin x}{x}\, dy\, dx.$$

Since the integrand in the *inner* integral is regarded as constant, this can be integrated easily:

$$\begin{aligned} \int_0^1 \left(\int_0^x \frac{\sin x}{x}\, dy \right) dx &= \int_0^1 \left(\frac{\sin x}{x} \cdot y \right)_{y=0}^{y=x} dx \\ &= \int_0^1 \sin x\, dx \\ &= -\cos x \Big|_0^1 \\ &= 1 - \cos 1. \end{aligned}$$

Symmetry Considerations

You may recall from single-variable calculus that some integrals can be simplified with the help of symmetry considerations.

The clearest instance is that of an **odd** function—that is, a function satisfying $f(-x) = -f(x)$ for all x integrated over an interval that is symmetric about the origin, $[-a, a]$: the integral is necessarily zero. To see this,

[8] Of course, it is also an improper integral.

we note that given a partition $\mathcal{P}$ of $[-a, a]$, we can refine it by throwing in the negative of each point of $\mathcal{P}$ together with zero; for this refinement, every component interval $I_j = [p_{j-1}, p_j]$ to the right of zero ($p_j > 0$) is matched by another component interval $I_{j^*} = [-p_j, -p_{j-1}]$ to the left of zero. If we use sample points also chosen symmetrically (the point in I_{j^*} is the negative of the one in I_j), then in the resulting Riemann sum, the contributions of matching intervals will cancel. Thus for example, even though we can't find an antiderivative for $f(x) = \sin x^3$, we know that it is odd, so automatically $\int_{-1}^{1} \sin x^3 \, dx = 0$.

A related argument says that the integral of an **even** function over a symmetric interval $[-a, a]$ equals twice its integral over either of its halves, say $[0, a]$.

These kinds of arguments can be extended to multiple integrals. A planar region $\mathcal{D}$ is ***x*-symmetric** if it is invariant under reflection across the y-axis—that is, if $(-x, y) \in \mathcal{D}$ whenever $(x, y) \in \mathcal{D}$. A function $f(x, y)$ is **odd in *x*** if $f(-x, y) = -f(x, y)$ for all (x, y); it is **even in *x*** if $f(-x, y) = f(x, y)$ for all (x, y). In particular, a polynomial in x and y is odd (*resp.* even) in x if every power of x which appears is odd (*resp.* even).

The one-variable arguments can be applied to an iterated integral (Exercise 7) to give

Remark 5.2.6. *If $\mathcal{D}$ is an x-regular region which is x-symmetric, then*

1. *For any function $f(x, y)$ which is odd in x,*

$$\iint_{\mathcal{D}} f \, dA = 0.$$

2. *If $f(x, y)$ is even in x, then*

$$\iint_{D} f(x, y) \; dA = 2 \iint_{D^+} f(x, y) \; dA$$

where

$$D^+ = \{\overrightarrow{x} = (x, y) \in D \,|\, x \geq 0\}$$

is the part of D to the right of the y-axis.

Of course, x can be replaced by y in the above definitions, and then also in Remark 5.2.6. One can also consider symmetry involving both x and y; see Exercise 10.

Exercises for § 5.2

Practice problems:

1. In the following, specify any region which is elementary by inequalities of the type given in Definition 5.2.5; subdivide any non-elementary region into non-overlapping elementary regions and describe each by such inequalities.

   (a) The region bounded above by $y = 9 - x^2$, below by $y = x^2 - 1$ and on the sides by the y-axis and the line $x = 2$.

   (b) The unit disc $\{(x, y) \,|\, x^2 + y^2 \leq 1\}$.

   (c) The part of the unit disc above the x-axis.

   (d) The part of the unit disc in the first quadrant.

   (e) The part of the unit disc in the second quadrant (to the left of the first).

   (f) The triangle with vertices $(0, 0)$, $(1, 0)$, and $(1, 3)$.

   (g) The triangle with vertices $(-1, 0)$, $(1, 0)$, and $(0, 1)$.

   (h) The region bounded above by $x + y = 5$ and below by $y = x^2 - 1$.

   (i) The region bounded by $y = x^2$ and $x = y^2$.

   (j) The region bounded by the curve $y = x^3 - 4x$ and the x-axis.

2. Each region described below is regular. If it is described as a y-regular (*resp.* x-regular) region, give its description as an x-regular (*resp.* y-regular) region.

   (a)
$$\begin{cases} 0 \leq y \leq 2x \\ 0 \leq x \leq 1 \end{cases}$$

   (b)
$$\begin{cases} 0 \leq y \leq 2 - x \\ 0 \leq x \leq 2 \end{cases}$$

   (c)
$$\begin{cases} x^2 \leq y \leq x \\ 0 \leq x \leq 1 \end{cases}$$

   (d)
$$\begin{cases} -\sqrt{4 - y^2} \leq x \leq \sqrt{4 - y^2} \\ -2 \leq y \leq 2 \end{cases}$$

   (e)
$$\begin{cases} 0 \leq x \leq \sqrt{4 - y^2} \\ -2 \leq y \leq 2 \end{cases}$$

3. Calculate each iterated integral below.

   (a) $\int_0^1 \int_x^1 (x^2y + xy^2)\, dy\, dx$

   (b) $\int_1^e \int_0^{\ln x} x\, dy\, dx$

   (c) $\int_1^2 \int_x^{x^2} (x - 5y)\, dy\, dx$

   (d) $\int_0^2 \int_0^y (2xy - 1)\, dx\, dy$

4. Calculate $\iint_D f\, dA$ as indicated.

   (a) $f(x, y) = 4x^2 - 6y$, D described by

   $$\begin{cases} 0 & \leq & y & \leq & x \\ 0 & \leq & x & \leq & 2 \end{cases}$$

   (b) $f(x, y) = y\sqrt{x^2 + 1}$, D described by

   $$\begin{cases} 0 & \leq & y & \leq & \sqrt{x} \\ 0 & \leq & x & \leq & 1 \end{cases}$$

   (c) $f(x, y) = 4y + 15$, D described by

   $$\begin{cases} y^2 + 2 & \leq & x & \leq & 3y \\ 1 & \leq & y & \leq & 2 \end{cases}$$

   (d) $f(x, y) = x$, D is the region bounded by $y = \sin x$, the x-axis and $x = \pi/2$.

   (e) $f(x, y) = xy$, D is the region bounded by $x+y = 5$ and $y = x^2-1$.

   (f) $f(x, y) = 1$, D is the intersection of the discs given by $x^2+y^2 \leq 1$ and $x^2 + (y - 1)^2 \leq 1$.

5. Rewrite each iterated integral with the order of integration reversed, and calculate it both ways (note that you should get the same answer both ways!)

   (a) $\int_0^2 \int_x^2 xy\, dy\, dx$ (b) $\int_0^1 \int_{x^2}^{\sqrt{x}} x\, dy\, dx$

(c) $\int_0^1 \int_{1-y}^{\sqrt{1-y^2}} y\,dx\,dy$ (d) $\int_{-1}^2 \int_1^{\sqrt{3-y}} x\,dx\,dy$

6. For each region below, decide whether it is x-symmetric, y-symmetric, or neither:

   (a) $\{(x,y)\,|\,x^2+y^2 \le 1\}$

   (b) $\{(x,y)\,|\,\frac{x^2}{a^2}+\frac{y^2}{b^2} \le 1\}$ $(a^2 \neq b^2)$

   (c) $\{(x,y)\,|\,-1 \le xy \le 1\}$

   (d) $\{(x,y)\,|\,0 \le xy \le 1\}$

   (e) $\{(x,y)\,|\,|y| \le |x|\,, \quad |x| \le 1\}$

   (f) The region bounded by the lines $x+y=1$, $x+y=-1$, and the coordinate axes.

   (g) The region bounded by the lines $x+y=1$, $x+y=-1$, $x-y=-1$ and $x-y=1$.

   (h) The region bounded by the lines $y=x+1$, $y=1-x$, and $y=0$.

   (i) The region bounded by the lines $x=1$, $y=2$, $x=-1$, and $y=-2$.

   (j) The triangle with vertices $(-1,0)$, $(0,-1)$, and $(1,1)$.

   (k) The triangle with vertices $(-1,-1)$, $(-1,1)$, and $(1,0)$.

   (l) The inside of the rose $r=\cos 2\theta$

   (m) The inside of the rose $r=\sin 2\theta$

   (n) The inside of the rose $r=\cos 3\theta$

   (o) The inside of the rose $r=\sin 3\theta$

Theory problems:

7. (a) Show that a polynomial in x and y is odd (*resp.* even) in x if and only if each power of x which appears is odd (*resp.* even).

   (b) Prove Remark 5.2.6.

   (c) Formulate the analogous concepts and results for regions symmetric in y, etc.

8. (a) Show that if $f(x,y)$ is even in x, then the region $\{(x,y)\,|\,f(x,y) \le c\}$ for any c is x-symmetric.

(b) Show that if $f(x,y)$ is even, then the region $\{(x,y)\,|\,f(x,y)\leq c\}$ for any c is symmetric with respect to the origin.

9. Use symmetry considerations either to conclude that the given iterated integral is zero, or to rewrite it as twice a different iterated integral.

(a) $\int_{-1}^{1}\int_{x^2-1}^{1-x^2} xy\,dy\,dx$

(b) $\int_{-1}^{1}\int_{-\cos x}^{\cos x}(x+y)\,dy\,dx$

(c) $\int_{-2}^{1}\int_{-x^3-3x^2-1}^{x^3+3x^2+1} x^2y\,dy\,dx$

(d) $\int_{-2}^{2}\int_{x^2-6}^{2-x^2}(xy^2+x^3y)\,dy\,dx$

(e) $\int_{-1}^{1}\int_{1-x^2}^{4-4x^2} x^2y\,dy\,dx$

(f) $\int_{-1}^{1}\int_{x^2-1}^{|1-x|}\sin x^3\,dy\,dx$

Challenge problem:

10. (a) A planar region $\mathcal{D}$ is **symmetric with respect to the origin** if $(-x,-y)\in\mathcal{D}$ whenever $(x,y)\in\mathcal{D}$.

i. Show that a region which is both x-symmetric and y-symmetric is also symmetric with respect to the origin.

ii. Give an example of a region which is symmetric with respect to the origin but neither x-symmetric nor y-symmetric.

(b) For each of the regions in Exercise 6, decide whether or not it is symmetric about the origin.

(c) A function $f(x,y)$ of two variables is **odd** (*resp.* **even**) if

$$f(-x,-y)=-f(x,y)\quad (\textit{resp. } f(-x,-y)=f(x,y))$$

for all (x,y).

i. Show that a function which is both even in x and even in y is even.

ii. What about a function which is both odd in x and odd in y?

iii. Show that a polynomial is odd (*resp.* even) precisely if each term has even (*resp.* odd) degree (the degree of a term is the sum of the powers appearing in it).

(d) Show that the integral of an odd function over an elementary region which is symmetric with respect to the origin equals zero.

5.3 Changing Coordinates

Substitution in a Double Integral.

Recall that when we perform a substitution $x = \varphi(t)$ inside an integral $\int_a^b f(x)\, dx$, it is not enough to just rewrite $f(x)$ in terms of t (as $f(\varphi(t))$); we also need to express the limits of integration as well as the "differential" term dx in terms of t. For double (and triple) integrals, this process is a little more complicated; this section is devoted to understanding what is needed.

A substitution in a double integral $\iint_D f(x,y)\, dA$ consists of a pair of substitutions

$$\begin{aligned} x &= \varphi_1(s,t) \\ y &= \varphi_2(s,t)\,. \end{aligned}$$

This can be viewed as a mapping $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ from the s,t-plane to the x,y-plane. We need, however, to be able to solve these substitution equations for s and t in terms of x and y, which means we must have a mapping φ which is **one-to-one**: different pairs (s,t) of values for the input must lead to different outputs. Furthermore, we will require this mapping to be differentiable and, for technical reasons, to have no critical points. We can think of this as a regular parametrization of the region $D \subset \mathbb{R}^2$ over which we are integrating, a view which will naturally carry over to surface integrals in the next section.

We expect the integrand $f(x,y)$ in our integral to be replaced by a function of s and t; in fact it is pretty clear that the natural choice is

$$f(s,t) = f(\varphi(s,t))\,.$$

It is also pretty clear that we need to take as our new domain of integration the domain of our parametrization. That is, we need to integrate over a region $D_{s,t}$ in the s,t-plane such that every point of D (living in the x,y-plane) is the image $\varphi(s,t)$ of some point of $D_{s,t}$ (which lives in the s,t-plane). When we apply the mapping $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ to a region $\mathcal{D}$, the **image** of $\mathcal{D}$ under φ, denoted $\varphi(\mathcal{D})$, is the set of all points that are hit, under the action of φ, by points in $\mathcal{D}$:

$$\varphi(\mathcal{D}) := \{\varphi(\overrightarrow{x}) \mid \overrightarrow{x} \in \mathcal{D}\}.$$

We say that φ maps $\mathcal{D}$ **onto** a set $E \subset \mathbb{R}^2$ if

$$E = \varphi(\mathcal{D})\,.$$

An example of such a substitution is the switch from rectangular to polar coordinates,

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

provided we stay within a region where θ does not increase by as much as 2π radians. We shall refer to all such mappings as *coordinate transformations.*

Definition 5.3.1. *A* ***coordinate transformation*** *on a region $\mathcal{D} \subset \mathbb{R}^2$ is a $\mathcal{C}^1$ mapping $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ satisfying:*

1. *φ has no critical points in $\mathcal{D}$ (i.e., its Jacobian determinant is nonzero at every point of $\mathcal{D}$).*
2. *φ maps $\mathcal{D}$ onto $\varphi(\mathcal{D})$ in a one-to-one manner.*

Another name for such a mapping is a **diffeomorphism** between $\mathcal{D}$ and $\varphi(\mathcal{D})$. By the Inverse Mapping Theorem (Theorem 4.4.2), a diffeomorphism is invertible: the mapping $\varphi^{-1}\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined on $\varphi(\mathcal{D})$ by

$$\varphi^{-1}(x,y) = (s,t) \Leftrightarrow (x,y) = \varphi(s,t)$$

is also $\mathcal{C}^1$, and its Jacobian matrix is the inverse of $J\varphi$: $(J\varphi)(J\varphi^{-1})$ is the identity matrix.

So far, we have seen how to express the integrand $f(x,y)$ in a double integral, as well as the domain of integration, in terms of s and t. It remains to see what to do with the element of area dA. Recalling that this corresponds to the areas $\triangle A_{ij}$ of the atoms of a partition in the construction of the double integral, we need to see how the change of coordinates affects area. That is, we need to know the relation between the area of a set $\mathcal{D}$ and that of its image $\varphi(\mathcal{D})$.

We begin with linear transformations.

Linear Transformations and Area

Suppose $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ is linear, and $\mathcal{D}$ is a region in the plane. What is the relation between the area of $\mathcal{D}$ and the area of its image $L(\mathcal{D})$?

We already know the answer to our question, at least for a special region $\mathcal{D}$—namely, the unit square (with corners at the origin and $(1,0)$, $(0,1)$, and $(1,1)$). The directed edges of this square emanating from the origin are the basis vectors $\overrightarrow{\imath}$ and $\overrightarrow{\jmath}$, and the image of the square is a parallelogram with directed edges (also emanating from the origin, which is fixed by L) given by the vectors $L(\overrightarrow{\imath})$ and $L(\overrightarrow{\jmath})$. We saw in § 1.6 that the area of this parallelogram is given by the absolute value $|\det [L]|$ of the determinant of $[L]$, the (2×2) matrix representative of L

$$[L] = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

This quantity will play an important role in what follows, so let us denote it by

$$\Delta(L) = |\det [L]|.$$

What about other regions? First we note that if we apply a displacement to our square, moving the origin to (α, β), then the image of the displaced square will, by linearity (check!) be our parallelogram, displaced by $L(\alpha, \beta)$, so the area is still given by $\Delta(L)$. So if $\mathcal{D}$ is any square whose sides are parallel to the axes and of length one, $L(\mathcal{D})$ is a parallelogram with area $\Delta(L)$. What about triangles? Again, displacement does not affect area (of either the region or its image), so we can think about triangles with one vertex at the origin, and directed sides given by vectors $\overrightarrow{v}$ and $\overrightarrow{w}$. Let us write each of these as a combination of $\overrightarrow{\imath}$ and $\overrightarrow{\jmath}$:

$$\begin{aligned} \overrightarrow{v} &= v_1 \overrightarrow{\imath} + v_2 \overrightarrow{\jmath} \\ \overrightarrow{w} &= w_1 \overrightarrow{\imath} + w_2 \overrightarrow{\jmath}. \end{aligned}$$

First of all, we know that the original triangle has, up to sign, area given by

$$\pm \mathcal{A}(\mathcal{D}) = \det \begin{pmatrix} v_1 & v_2 \\ w_1 & w_2 \end{pmatrix}$$

while the image is a triangle with edges given by the images of $\overrightarrow{v}$ and $\overrightarrow{w}$

$$\begin{aligned} L(\overrightarrow{v}) &= (av_1 + bv_2)\overrightarrow{\imath} + (cv_1 + dv_2)\overrightarrow{\jmath} \\ L(\overrightarrow{w}) &= (aw_1 + bw_2)\overrightarrow{\imath} + (cw_1 + dw_2)\overrightarrow{\jmath} \end{aligned}$$

and so has area (up to sign) given by

$$\pm \mathcal{A}(L(\mathcal{D})) = \det \begin{pmatrix} av_1 + bv_2 & cv_1 + dv_2 \\ aw_1 + bw_2 & cw_1 + dw_2 \end{pmatrix}.$$

Using the fact that the determinant is linear in its columns, we can rewrite this as

$$\det \begin{pmatrix} av_1 & aw_1 \\ cv_1 & cw_1 \end{pmatrix} + \det \begin{pmatrix} bv_2 & aw_1 \\ dv_2 & cw_1 \end{pmatrix} + \det \begin{pmatrix} av_1 & bw_2 \\ cv_1 & dw_2 \end{pmatrix} + \det \begin{pmatrix} bv_2 & bw_2 \\ dv_2 & dw_2 \end{pmatrix}$$

and then factoring the v's and w's from individual columns, as

$$v_1w_1 \det\begin{pmatrix} a & a \\ c & c \end{pmatrix} + v_2w_1 \det\begin{pmatrix} b & a \\ d & c \end{pmatrix} + v_1w_2 \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} + v_2w_2 \det\begin{pmatrix} b & b \\ d & d \end{pmatrix}.$$

The first and last determinants are zero (why?), and the two middle ones are negatives of each other (why?), so we can finally write

$$\begin{aligned} \pm\, \mathcal{A}\left(L(\mathcal{D})\right) &= (v_1w_2 - v_2w_1) \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \det\begin{pmatrix} v_1 & v_2 \\ w_1 & w_2 \end{pmatrix} \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \pm\mathcal{A}\left(\mathcal{D}\right) \det\begin{pmatrix} a & b \\ c & d \end{pmatrix}. \end{aligned}$$

From this calculation (and taking absolute values on both sides) we see that *for any triangle in* $\mathbb{R}^2$*, the area of its image is the area of the triangle times the absolute value of the determinant of* $[L]$. In this calculation, we have ignored signs, but you are asked in Exercise 6 to retrace this argument to show that, had we paid attention to signs and talked about *oriented* areas, then we would see that the oriented area of any directed triangle is multiplied, under the action of a linear transformation, by the determinant of its matrix representative.

Actually, as an aside, this calculation also yields another useful result.

Remark 5.3.2. *The* determinant of the product *of two* 2×2 *matrices is the* product of the determinants *of the two matrices: for* A *and* B 2×2*,*

$$\det\left(AB\right) = \left(\det\, A\right)\left(\det\, B\right).$$

To see this, it suffices to note that (1) the original matrix whose determinant gives $\mathcal{A}\left(L(\mathcal{D})\right)$ (up to sign) is the product

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}$$

and (2) that the determinant of the transpose is the same as the determinant of the original matrix. Details in Exercise 7.

Having established that the area of an arbitrary triangle is multiplied by $\Delta\left(L\right)$, we can proceed to general areas. First, since any polygon can be tiled by triangles, the same property holds for their areas. Second, suppose $\mathcal{D}$ is an arbitrary elementary region in $\mathbb{R}^2$: for example, suppose it is x-regular, say

$$\mathcal{D} = \{(x, y) \,|\, a \leq x \leq b, \quad \psi(x) \leq y \leq \varphi(x)\}.$$

We can partition $[a, b]$ so that the inner and outer sums differ by an arbitrarily specified amount; given $\varepsilon > 0$, we make sure that this difference *multiplied by* $\Delta\,(L)$ is less than ε. The (nested) polygons P_- and P_+ formed by the union of the inscribed and circumscribed rectangles, respectively, are taken to (nested) polygons $L(P_-)$ and $L(P_+)$, respectively; we know that

$$\mathcal{A}\,(P_+) \leq \mathcal{A}\,(\mathcal{D}) \leq \mathcal{A}\,(P_-)$$

and also (by nesting)

$$\mathcal{A}\,(L(P_+)) \leq \mathcal{A}\,(L(\mathcal{D})) \leq \mathcal{A}\,(L(P_-))\,;$$

since

$$\begin{aligned}\mathcal{A}\,(L(P_+)) - \mathcal{A}\,(L(P_-)) &= \Delta\,(L)\,(\mathcal{A}\,(P_+) - \mathcal{A}\,(P_-)) \\ &< \varepsilon\end{aligned}$$

and

$$\Delta\,(L)\,\mathcal{A}\,(P_+) \leq \Delta\,(L)\,\mathcal{A}\,(\mathcal{D}) \leq \Delta\,(L)\,\mathcal{A}\,(P_-)$$

we can conclude that

$$|\mathcal{A}\,(L(\mathcal{D})) - \Delta\,(L)\,\mathcal{A}\,(\mathcal{D})| < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the two quantities are equal, and we have proved

Proposition 5.3.3. *For any elementary planar region $\mathcal{D}$, the area of its image under a linear map $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ is the original area multiplied by the absolute value of the determinant of the matrix representative of L:*

$$\mathcal{A}\,(L(\mathcal{D})) = \Delta\,(L)\,\mathcal{A}\,(\mathcal{D})$$

where

$$\Delta\,(L) = |\det\,[L]|\,.$$

While we have concentrated on linear maps, the same results hold for affine maps, since a displacement does not change areas.

Corollary 5.3.4. *If $T\colon\mathbb{R}^2\to\mathbb{R}^2$ is an affine map ($T(\overrightarrow{x}) = \overrightarrow{y}_0 + L(\overrightarrow{x})$, L linear) and $\mathcal{D}$ is an elementary region in $\mathbb{R}^2$,*

$$\mathcal{A}\left(T(\mathcal{D})\right) = \Delta\left(L\right)\cdot\mathcal{A}\left(\mathcal{D}\right).$$

To avoid having to specify the linear part of an affine map, we will often write $\Delta\left(T\right)$ in place of $\Delta\left(L\right)$.

Coming back to substitution in a double integral, we see that an affine mapping $T\colon\mathbb{R}^2\to\mathbb{R}^2$ with nonzero determinant $\Delta\left(T\right)\neq 0$ (which is the same as an affine coordinate transformation) multiplies all areas uniformly by $\Delta\left(T\right)\dot{\text{T}}$hus we expect that, when we make an affine substitution

$$\begin{aligned}x &= a_{11}s + a_{12}t + b_1\\ y &= a_{21}s + a_{22}t + b_2\end{aligned}$$

in the double integral $\iint_{\mathcal{D}} f\,dA$, the element of area $dA_{x,y}$ in the x,y-plane (and hence the "differential" term dA) should be replaced by the element of area $dA_{s,t}$ in the s,t-plane, which should be related to the former by

$$dA_{x,y} = \Delta\left(T\right)dAss, t.$$

To be more precise,

Proposition 5.3.5 (Affine Change of Coordinates)**.** *Suppose $\mathcal{D}$ is an elementary region, $T\colon\mathbb{R}^2\to\mathbb{R}^2$ is an affine coordinate transformation defined on $\mathcal{D}$, and $f\colon\mathbb{R}^2\to\mathbb{R}$ is a real-valued function which is integrable on $T(\mathcal{D})$.*

Then

$$\iint_{T(\mathcal{D})} f(\overrightarrow{x})\ dA = \iint_{\mathcal{D}} f(T(\vec{s}))\,\Delta\left(T\right)\ dA. \tag{5.5}$$

Proof. Note first that the definition of the integral allows us to enclose $\mathcal{D}$ in a rectangle $[a,b]\times[c,d]$ and extend the integrand f to be zero off the set. In effect, this means we can assume $\mathcal{D} = [a,b]\times[c,d]$ ..

We first consider the special case when the matrix representative for the linear part of T is **diagonal**:

$$T(\vec{s}) = L(\vec{s}) + \overrightarrow{C},$$

where

$$[L] = \begin{pmatrix} a_{11} & 0\\ 0 & a_{22}\end{pmatrix};$$

geometrically, this says that horizontal and vertical directions are preserved, with horizontal distances scaled by a_{11} and vertical distances scaled by a_{22}.

Let $\mathcal{P}$ be a partition of $\mathcal{D} = [a,b] \times [c,d]$ defined by a partition of $[a,b]$ and a partition of $[c,d]$, and consider the partition $T(\mathcal{P}) = \mathcal{P}'$ of

$$T([a,b] \times [c,d]) = \left[a',b'\right] \times \left[c',d'\right]$$

where[9]

$$\begin{aligned} a' &= T(a) \\ b' &= T(b) \\ c' &= T(c) \\ d' &= T(d) \end{aligned}$$

defined by the images under L of the points defining $\mathcal{P}$.

The lower (*resp.* upper) sums for these two partitions are

$$\begin{aligned} \mathcal{L}(\mathcal{P}, f \circ T\Delta\,(T)) &= \sum_{i,j} \left(\inf_{S_{ij}}(f \circ T)\Delta\,(T)\right) \triangle A_{ij} \\ \mathcal{U}(\mathcal{P}, f \circ T\Delta\,(T)) &= \sum_{i,j} \left(\sup_{S_{ij}}(f \circ T)\Delta\,(T)\right) \triangle A_{ij} \\ \mathcal{L}(\mathcal{P}', f) &= \sum_{i,j} \left(\inf_{S'_{ij}} f\right) \triangle A'_{ij} \\ \mathcal{U}(\mathcal{P}', f) &= \sum_{i,j} \left(\sup_{S'_{ij}} f\right) \triangle A'_{ij} \end{aligned}$$

where

$$S'_{ij} = T(S_{ij})\,.$$

But the values of f on $S'_{ij} = T(S_{ij})$ (and hence their infimum and supremum) are precisely the same as those of $f \circ T$ on S_{ij}, and the area $\triangle A'_{ij}$ of $T(S_{ij})$ is precisely $\Delta\,(T)$ times the area $\triangle A_{ij}$ of S_{ij}; it follows that the corresponding sums are equal

$$\begin{aligned} \mathcal{L}(\mathcal{P}, (f \circ T)\Delta\,(T)) &= \mathcal{L}(\mathcal{P}', f) \\ \mathcal{U}(\mathcal{P}, (f \circ T)\Delta\,(T)) &= \mathcal{U}(\mathcal{P}', f) \end{aligned}$$

[9] Note that we are using the fact that $[L]$ is diagonalizable here

and hence the two integrals are equal:

$$\iint_{T(\mathcal{D})} f(\overrightarrow{x})\ dA = \iint_{\mathcal{D}} (f \circ T)(\vec{s})\,\Delta\,(T)\ dA$$

which is the same as Equation (5.5), proving it in the diagonal case.

The difficulty with this argument when $[L]$ is non-diagonal case is that the image $T(S_{ij})$ of a rectangle of $\mathcal{P}$ might not be a rectangle with sides parallel to the x- and y-axes: in fact, it is in general a parallelogram, often with no horizontal or vertical sides. In particular, we cannot claim that the images of the subrectangles of a partition $\mathcal{P}$ of $\mathcal{D}$ are themselves the subrectangles of any partition of $T(\mathcal{D})$.

To simplify matters, let us assume that the original partition $\mathcal{P}$ comes from dividing $[a, b]$ (*resp.* $[c, d]$) into m (*resp.* n) equal parts with points s_i (*resp.* t_j), and let us consider the points $T(s_i, t_j)$ in the x, y-plane. We can form a partition of the smallest rectangle containing $T(\mathcal{D})$, $R = [A, B] \times [C, D]$, by drawing horizontal and vertical lines through all of these points; furthermore we can refine this partition by adding more horizontal and vertical lines in such a way that we have a partition $\mathcal{P}'$ of R with arbitrarily small mesh size μ, .

What is the total area of those rectangles of this partition which meet the parallelogram which forms the boundary of $T([a, b] \times [c, d])$? For each non-horizontal (*resp.* non-vertical) edge of the parallelogram, we can slide all the rectangles which meet it across to the y-axis (*resp.* x-axis). Since all the rectangles have width (*resp.* height) at most μ, they will fit inside a vertical (*resp.* horizontal) rectangle whose width (*resp.* height) is μ and whose height (*resp.* width) is the projection of that edge on the y-axis (*resp.* x-axis). This means that the total area of the rectangles meeting the boundary of $T([a, b] \times [c, d])$ will be at most μ times the perimeter of $R = [A, B] \times [C, D]$, which is $2(B - A) + 2(D - C)$. Now we can pick μ sufficiently small to guarantee that the total area of the rectangles of $\mathcal{P}'$ which meet the boundary have total area whose ratio to the area of the parallelogram $T([a, b] \times [c, d])$ is arbitrarily small, say it is less than ε for some specified $\varepsilon > 0$.

Now note that this argument can be scaled to apply to *all* of the parallelograms formed as images of the subrectangles of $\mathcal{P}$, and this can be done simultaneously for all of them, since they are all congruent. This is what we really need: pick μ so small that the total area of the subrectangles of $\mathcal{P}'$ of mesh μ meeting the boundary of any particular parallelogram $T(S_{ij})$ is less than ε times the area of the parallelogram. Now, let us consider the contribution to $\mathcal{L}(\mathcal{P}', f)$ (*resp.* $\mathcal{U}(\mathcal{P}', f)$) of *all* the subrectangles of $\mathcal{P}'$ which meet the boundary of any one of the parallelograms $T(S_{ij})$ (for all possible i,j). For each parallelogram $T(S_{ij})$, the infimum (*resp.* supremum) of f on any subrectangle of $\mathcal{P}'$ *contained* in the region $T(S_{ij})$ is at least (*resp.* at most) equal to the infimum (*resp.* supremum) of f on $T(S_{ij})$, which in turn *equals* the infimum (*resp.* supremum) of $f\circ T$ on S_{ij}. The sum of $\inf f\triangle A_{ij}$ (*resp.* $\sup f\triangle A_{ij}$) over *all* these rectangles (which is to say those that *don't* meet the boundary of any parallelogram) is at least (*resp.* at most) equal to $\inf_{S_{ij}}(f\circ T)$ times the total area of these rectangles, which is at least $1-\varepsilon$ (*resp.* at most $1+\varepsilon$) times the area of $T(S_{ij})$, and this in turn equals $\Delta\,(T)$ times the area of S_{ij}. This shows that the individual terms of $\mathcal{L}(\mathcal{P}',f)$ and $\mathcal{U}(\mathcal{P}',f)$ coming from subrectangles S' not meeting the boundary of any parallelogram $T(S_{ij})$ satisfy

$$\sum_{S'\subset T(S_{ij}),\text{ some }i,j}(\inf_{S'} f)(\Delta\,(T))\,dA' \geq (1-\varepsilon)\sum_{i,j}\inf_{S_{ij}}(f\circ T)(\Delta\,(T))\,dA_{i,j}$$

and

$$\sum_{S'\subset T(S_{ij}),\text{ some }i,j}(\sup_{S'} f)(\Delta\,(T))\,dA' \leq (1+\varepsilon)\sum_{i,j}\sup_{S_{ij}}(f\circ T)(\Delta\,(T))\,dA_{i,j}.$$

Furthermore, the contribution to $\mathcal{L}(\mathcal{P}',f)$ and $\mathcal{U}(\mathcal{P}',f)$ from those subrectangles that *do* intersect the boundary of some $T(S_{ij})$ is between ε times the infimum and ε times the supremum of the function $f\circ T$ over the whole of $\mathcal{D}$. Thus, for each $\varepsilon>0$ we can construct a partition P' for which

$$\begin{aligned}(1-\varepsilon)\mathcal{L}(\mathcal{P}',f)+\varepsilon(\inf_{\mathcal{D}}\circ fT\Delta\,(T))\mathcal{A}\,(T(\mathcal{D}))&\\ &\leq \mathcal{L}(\mathcal{P}',(\circ fT)\Delta\,(T))\\ &\leq \mathcal{U}(\mathcal{P}',(\circ fT)\Delta\,(T))\\ &\leq (1+\varepsilon)\mathcal{U}(\mathcal{P}',f)+\varepsilon(\inf_{\mathcal{D}}\circ fT\Delta\,(T))\mathcal{A}\,(T(\mathcal{D}))\end{aligned}$$

Since this is true for arbitrary $\varepsilon>0$, given $\mathcal{P}$, we see that

$$\sup_{\mathcal{P}'}\mathcal{L}(\mathcal{P}',f)\leq\sup_{\mathcal{P}}\mathcal{L}(\mathcal{P},(\circ fT)\Delta\,(T))\leq\inf_{\mathcal{P}}\mathcal{U}(\mathcal{P},(\circ fT)\Delta\,(T))\leq\sup_{\mathcal{P}'}\mathcal{U}(\mathcal{P}',f)$$

which shows the two integrals are equal, as required.

□

As an example, let us consider the integral $\iint_{\mathcal{D}} x\,dA$, where $\mathcal{D}$ is the parallelogram with vertices $(2,0)$, (3,-1), $(4,0)$, and $(3,1)$.

The region is the image of the unit rectangle $[0,1]\times[0,1]$ under the affine coordinate transformation

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

which has matrix

$$[L] = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

with determinant -2, so

$$\Delta(T) = 2.$$

Thus we replace $dA = dA_{x,y}$ with $dA = 2\,dA_{s,t}$, x with $s+t$, and the domain of integration with $[0,1]\times[0,1]$, leading to the integral

$$\begin{aligned} \iint_{\mathcal{D}} x\,dA &= \int_0^1\int_0^1 (s+t)(2d_{dt}) \\ &= 2\int_0^1 \left(\frac{s^2}{2}+st\right)_{s=0}^1 dt \\ &= 2\int_0^1 \left(\frac{1}{2}+t\right) dt \\ &= 2\left(\frac{t}{2}+\frac{t^2}{2}\right)_0^1 \\ &= 2. \end{aligned}$$

Coordinate Transformations and Area

Our next goal is to decide what happens to areas under differentiable maps. The description can be complicated for differentiable maps which either have critical points or overlap images of different regions. Thus, we will consider only maps covered by the following

In keeping with our general philosophy, we expect the behavior of a coordinate transformation F with respect to area, at least locally, to reflect the behavior of its linearization. To sharpen this expectation, we establish some technical estimates. We know what the linearization map $T_{\vec{x}_0}F$ at a point does to a square: it maps it to a parallelogram whose area is the original area times $\Delta\left(T_{\overrightarrow{x}_0}F\right)$. We would like to see how far the image of the same square under the nonlinear transformation F deviates from this parallelogram. Of course, we only expect to say something when the square is small.

Suppose P is a parallelogram whose sides are generated by the vectors $\overrightarrow{v}$ and $\overrightarrow{w}$. We will say the **center** is the intersection of the line joining the midpoints of the two edges parallel to $\overrightarrow{v}$ (this line is parallel to $\overrightarrow{w}$) with the line (parallel to $\overrightarrow{v}$) joining the midpoints of the other two sides (Figure 5.18. If the center of P is $\overrightarrow{x}_0$, then it is easy to see that

$$P = \{\overrightarrow{x}_0 + \alpha\overrightarrow{v} + \beta\overrightarrow{w} \mid |\alpha|, |\beta| \leq 0.5\}.$$

Now we can **scale** P by a factor $\lambda > 0$ simply by multiplying all distances by λ. The scaled version will be denoted

$$\lambda P := \{\overrightarrow{x}_0 + \alpha\overrightarrow{v} + \beta\overrightarrow{w} \mid |\alpha|, |\beta| \leq 0.5\lambda\}.$$

When we scale a parallelogram by a factor λ, its area scales by λ^2; in particular, if λ is close to 1, then the area of the scaled parallelogram is close to that of the original. Our immediate goal is to establish that if a square is small enough, then its image under F is contained between two scalings of its image under the linearization $TF = T_{\overrightarrow{x}_0}F$ of F at some point in the square—that is, for some $\varepsilon > 0$, $F(\mathcal{D})$ *contains* $(1-\varepsilon)TF(\mathcal{D})$ and is *contained in* $(1+\varepsilon)TF(\mathcal{D})$.[10] (See Figure 5.19.) Note that scaling commutes

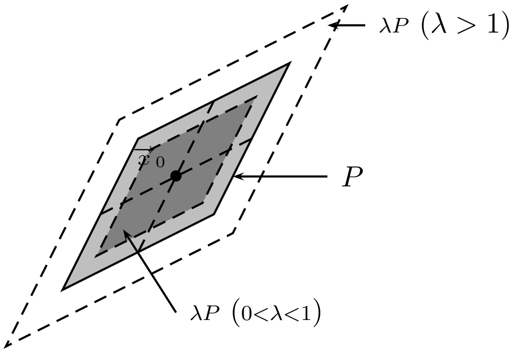
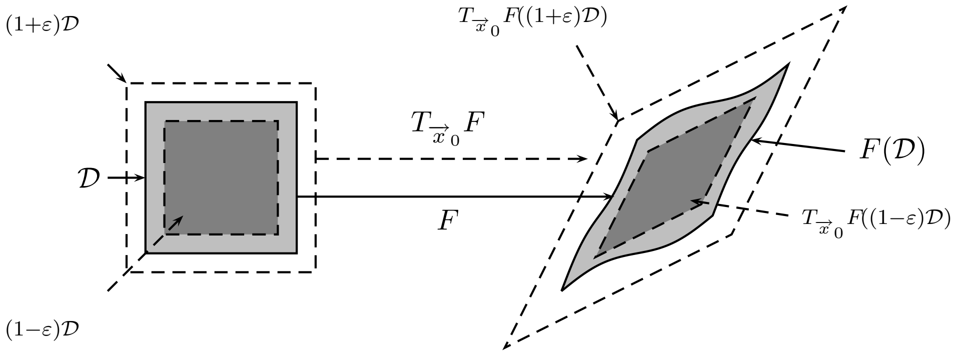


Figure 5.18: Center of a Parallelogram, Scaling.

[10] Our argument here is motivated by [11, pp. 178-9, 248-51].

with an affine map: the image of a scaled square is the same scaling of the image parallelogram.



$(1+\varepsilon)\mathcal{D}$
$\mathcal{D}$
$(1-\varepsilon)\mathcal{D}$
$T_{\overrightarrow{x}_0}F((1+\varepsilon)\mathcal{D})$
$T_{\overrightarrow{x}_0}F$
F
$F(\mathcal{D})$
$T_{\overrightarrow{x}_0}F((1-\varepsilon)\mathcal{D})$

Figure 5.19: Nonlinear Image Between two scaled affine images

The argument is easiest to see when the linear part of $T_{\overrightarrow{x}_0}F$ is the identity map and the region is a square; after working this through, we will return to the general case. Given a point $\overrightarrow{x}_0 = (x_0, y_0)$, we will refer to the square $[x_0 - r, x_0 + r] \times [y_0 - r, y_0 + r]$ as the **square of radius r centered at $\overrightarrow{x}_0$.**

Remark 5.3.6. *If $\mathcal{D}$ is a square of radius r centered at $\overrightarrow{x}_0$, then any point $\overrightarrow{x}$ whose distance from the boundary of $\mathcal{D}$ is less than $r\varepsilon$ is* inside $(1+\varepsilon)\mathcal{D}$ *and* outside $(1-\varepsilon)\mathcal{D}$.

(See Figure 5.20.)

Lemma 5.3.7. *Suppose $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is differentiable at $\overrightarrow{x}_0$ and its derivative at $\overrightarrow{x}_0$ is the identity map. If $\mathcal{D}$ is a square of radius r, centered at $\overrightarrow{x}_0$, such that for all $\overrightarrow{x} \in \mathcal{D}$ the first-order contact condition*

$$\left|F(\overrightarrow{x}) - T_{\overrightarrow{x}_0}F(\overrightarrow{x})\right| < \delta\,|\overrightarrow{x} - \overrightarrow{x}_0| \tag{5.6}$$

holds, where

$$0 < \delta < \frac{\varepsilon}{\sqrt{2}}. \tag{5.7}$$

Then (provided $0 < \varepsilon < 1$) $F(\mathcal{D})$ is between $T_{\overrightarrow{x}_0}F((1-\varepsilon)\mathcal{D})$ and $T_{\overrightarrow{x}_0}F((1+\varepsilon)\mathcal{D})$:

$$T_{\overrightarrow{x}_0}F((1-\varepsilon)\mathcal{D}) \subset F(\mathcal{D}) \subset T_{\overrightarrow{x}_0}F((1+\varepsilon)\mathcal{D})\,.$$

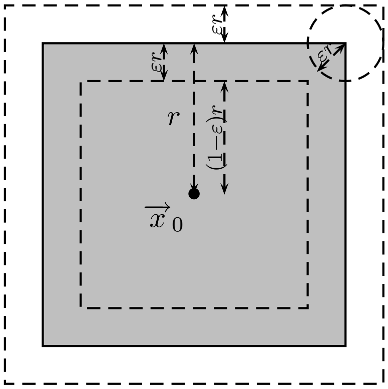


Figure 5.20: Remark 5.3.6

Proof. The main observation here is that the distance from the center to any point on the boundary of a square of radius r is between r and $r\sqrt{2}$; the latter occurs at the corners. Thus, for any point $\vec{x}$ on the boundary of $\mathcal{D}$, Equation (5.6) tells us that

$$\begin{aligned}\left|F(\vec{x}) - T_{\vec{x}_0}F(\vec{x})\right| \delta(r\sqrt{2}) \\ &< \left(\frac{\varepsilon}{\sqrt{2}}\right)(r\sqrt{2}) \\ &= r\varepsilon\end{aligned}$$

and since $T_{\vec{x}_0}F(\vec{x}) = \vec{x}$ by assumption, it follows from Remark 5.3.6 that the boundary of $F(\mathcal{D})$ (which is the image of the boundary of $\mathcal{D}$) lies entirely inside $T_{\vec{x}_0}F((1+\varepsilon)\mathcal{D})$ and entirely outside $T_{\vec{x}_0}F((1-\varepsilon)\mathcal{D})$, from which the desired conclusion follows. □

To remove the assumption that $DF_{\vec{x}_0}$ is the identity map in Lemma 5.3.7, suppose we are given F with arbitrary invertible derivative mapping $L = DF_{\vec{x}_0}$ consider the mapping

$$G = L^{-1} \circ F.$$

By the Chain Rule, $DG_{\vec{x}_0}$ is the identity map, so Lemma 5.3.7 says that if the first-order contact condition $\left|G(\vec{x}) - T_{\vec{x}_0}G(\vec{x})\right| < \delta\left|\vec{x} - \vec{x}_0\right|$ applies on $\mathcal{D}$ with $0 < \delta < \frac{\varepsilon}{\sqrt{2}}$, then

$$T_{\overrightarrow{x}_0}G((1-\varepsilon)\mathcal{D}) \subset G(\mathcal{D}) \subset T_{\overrightarrow{x}_0}G((1+\varepsilon)\mathcal{D})\,.$$

Since $F = L \circ G$, we can simply apply L to all three sets above to see that this conclusion implies the corresponding conclusion for F:

$$\begin{aligned}T_{\overrightarrow{x}_0}F((1-\varepsilon)\mathcal{D}) = L\Big(T_{\overrightarrow{x}_0}G((1-\varepsilon)\mathcal{D})\Big)& \\ \subset F(\mathcal{D}) = L(G(\mathcal{D}))& \\ \subset T_{\overrightarrow{x}_0}F((1+\varepsilon)\mathcal{D}) = L\Big(T_{\overrightarrow{x}_0}G((1+\varepsilon)\mathcal{D})\Big)&\,.\end{aligned}$$

To formulate the hypotheses in terms of F, we note that what is required is

$$\begin{aligned}\Big|G(\overrightarrow{x}) - T_{\overrightarrow{x}_0}G(\overrightarrow{x})\Big| &= \Big|L^{-1}\Big(F(\overrightarrow{x}) - T_{\overrightarrow{x}_0}F(\overrightarrow{x})\Big)\Big| \\ &\leq \| L^{-1} \| \Big|F(\overrightarrow{x}) - T_{\overrightarrow{x}_0}F(\overrightarrow{x})\Big| \\ &< \delta\,|\overrightarrow{x} - \overrightarrow{x}_0|\end{aligned}$$

(*caution*: $\| L^{-1} \|$ is *not* the same as $\| L \|^{-1}$). So dividing both sides of the last inequality by $\| L^{-1} \|$, we see that our hypothesis should be

$$\Big|F(\overrightarrow{x}) - T_{\overrightarrow{x}_0}F(\overrightarrow{x})\Big| < \frac{\delta}{\| L^{-1} \|}\,|\overrightarrow{x} - \overrightarrow{x}_0| \tag{5.8}$$

where δ satisfies (5.7). So we can say, without any assumptions on $DF_{\overrightarrow{x}_0}$, the following:

Lemma 5.3.8. *Suppose $F: \mathbb{R}^2 \to \mathbb{R}^2$ has invertible derivative at $\overrightarrow{x}_0$ and R is a square of radius r, centered at $\overrightarrow{x}_0$, such that for all $\overrightarrow{x} \in \mathcal{D}$ the first-order contact condition*

$$\Big|F(\overrightarrow{x}) - T_{\overrightarrow{x}_0}F(\overrightarrow{x})\Big| < \delta\,|\overrightarrow{x} - \overrightarrow{x}_0| \tag{5.9}$$

holds, where

$$0 < \delta < \frac{\varepsilon}{\sqrt{2}\,\| \left(DF_{\overrightarrow{x}_0}\right)^{-1} \|}. \tag{5.10}$$

Then (provided $0 < \varepsilon < 1$) $F(\mathcal{D})$ is between $T_{\overrightarrow{x}_0}F((1-\varepsilon)\mathcal{D})$ and $T_{\overrightarrow{x}_0}F((1+\varepsilon)\mathcal{D})$:

$$T_{\overrightarrow{x}_0}F((1-\varepsilon)\mathcal{D}) \subset F(\mathcal{D}) \subset T_{\overrightarrow{x}_0}F((1+\varepsilon)\mathcal{D})\,.$$

In particular, under these conditions, we have an estimate of area

$$(1-\varepsilon)\Delta\left(DF_{\overrightarrow{x}_0}\right)\mathcal{A}(R) \leq \mathcal{A}(F(R)) \leq (1+\varepsilon)\Delta\left(DF_{\overrightarrow{x}_0}\right)\mathcal{A}(R)\,. \tag{5.11}$$

So far, what we have is a *local* result: it only applies to a square that is small enough to guarantee the first-order contact condition (5.12). To globalize this, we need to approximate $\mathcal{D}$ with a non-overlapping union of squares small enough to guarantee condition (5.12), relative to its center, on each of them individually. So far, though, we only know that if F is differentiable at a given point $\overrightarrow{x}_0$, the first-order contact condition holds on *some* sufficiently small square about $\overrightarrow{x}_0$; *a priori* the required size may vary with the point. We would like to get a *uniform* condition: to guarantee (5.12) for *any* square whose radius is less than some fixed number that depends only on the desired δ. When F is $\mathcal{C}^1$, this can be established by an argument similar to that used to show that continuity on a compact region guarantees *uniform* continuity there (*Calculus Deconstructed*, Theorem 3.7.6).

Lemma 5.3.9. *Suppose $F\colon\mathbb{R}^2\to\mathbb{R}^2$ is $\mathcal{C}^1$ on a compact region $\mathcal{D}$. Given $\delta>0$, there exists $\delta'>0$ such that the first-order contact condition*

$$\left|F(\overrightarrow{x})-T_{\overrightarrow{x}'}F(\overrightarrow{x})\right|<\delta\left|\overrightarrow{x}-\overrightarrow{x}'\right| \tag{5.12}$$

holds for any pair of points $\overrightarrow{x},\overrightarrow{x}'\in\mathcal{D}$ whose distance apart is less than δ'.

Proof. The proof is by contradiction. Suppose no such δ' exists; then for each choice of δ', there exists a pair of points $\overrightarrow{x},\overrightarrow{x}'\in\mathcal{D}$ with

$$\left|\overrightarrow{x}-\overrightarrow{x}'\right|<\delta'$$

but

$$\left|F(\overrightarrow{x})-T_{\overrightarrow{x}'}F(\overrightarrow{x})\right|\geq\delta\left|\overrightarrow{x}-\overrightarrow{x}'\right|.$$

We pick a sequence of such pairs, $\overrightarrow{x}_k, \overrightarrow{x}'_k\in\mathcal{D}$ corresponding to $\delta'=\frac{1}{k}$, $k=1,\ldots$.

By the Bolzano-Weierstrass Theorem, the sequence $\overrightarrow{x}'_k$ has a convergent subsequence, which we can assume to be the full sequence: say $\overrightarrow{x}'_k\to\overrightarrow{x}_0$. Since F is $\mathcal{C}^1$, we can also say that the Jacobian matrices of F at $\overrightarrow{x}'_k$ converge to the matrix at $\overrightarrow{x}_0$, which means that for k sufficiently high,

$$\left|DF_{\overrightarrow{x}'_k}\left(\overrightarrow{x}-\overrightarrow{x}'_k\right)-DF_{\overrightarrow{x}_0}\left(\overrightarrow{x}-\overrightarrow{x}_0\right)\right|\leq\frac{\delta}{2}\left|\overrightarrow{x}-\overrightarrow{x}_0\right|$$

for all $\overrightarrow{x}$. In particular, the points $\overrightarrow{x}_k$ converge to $\overrightarrow{x}_0$, but

$$\left|F(\overrightarrow{x}_k)-T_{\overrightarrow{x}_0}F(\overrightarrow{x}_k)\right|\geq\delta\left|\overrightarrow{x}_k-\overrightarrow{x}_0\right|$$

contradicting the definition of differentiability at $\overrightarrow{x}_0$. □

Combining this with Lemma 5.3.8 (or more accurately its rectangular variant), we can prove:

Proposition 5.3.10. *Suppose $F: \mathbb{R}^2 \to \mathbb{R}^2$ is a coordinate transformation on the (compact) elementary region $\mathcal{D}$. Then given $\varepsilon > 0$ there exists $\delta > 0$ such that if $R \subset \mathcal{D}$ is any square of radius $r < \delta$,*

$$(1-\varepsilon)\Delta\left(T_{\overrightarrow{x}_0}F\right)\mathcal{A}\left(R\right) < \mathcal{A}\left(F(R)\right) < (1+\varepsilon)\Delta\left(T_{\overrightarrow{x}_0}F\right)\mathcal{A}\left(R\right)$$

where $\overrightarrow{x}_0$ is the center of R.

Proof. Note that since F is $\mathcal{C}^1$ on $\mathcal{D}$, there is a uniform upper bound on the norm $\|\left(DF_{\overrightarrow{x}_0}\right)\|$ for all $\overrightarrow{x}_0 \in \mathcal{D}$. Then we can use Lemma 5.3.9 to find a bound on the radius which insures that the first-order contact condition (5.12) needed to guarantee (5.10) holds on any square whose radius satisfies the bound. But then Lemma 5.3.8 gives us Equation (5.11), which is precisely what we need. □

Finally, we can use this to find the effect of coordinate transformations on the area of elementary regions. Note that $\Delta\left(T_{\overrightarrow{x}}F\right) = \Delta\left(DF_{\overrightarrow{x}}\right)$ is just the absolute value of the Jacobian determinant:

$$\Delta\left(T_{\overrightarrow{x}}F\right) = \Delta\left(DF_{\overrightarrow{x}}\right) = \left|\det\, JF(\overrightarrow{x})\right|;$$

we will, for simplicity of notation, abuse notation and denote this by $|JF(\overrightarrow{x})|$.

Theorem 5.3.11. *Suppose $\mathcal{D} \subset \mathbb{R}^2$ is an elementary region and $F: \mathbb{R}^2 \to \mathbb{R}^2$ is a coordinate transformation defined on a neighborhood of $\mathcal{D}$. Then*

$$\mathcal{A}\left(F(\mathcal{D})\right) = \iint_{\mathcal{D}} |JF(\overrightarrow{x})|\, dA \tag{5.13}$$

Proof. Let us first prove this when $\mathcal{D}$ is a square S (of any size) with sides parallel to the coordinate axes. Note that by subdividing the sides into intervals of equal length, we can get a partition of the square into subsquares of arbitrarily small radius. In particular, we can, given $\varepsilon > 0$, subdivide it into subsquares R_{ij} such that Proposition 5.3.10 guarantees for each R_{ij} that

$$(1-\varepsilon)\Delta\left(T_{\overrightarrow{x}_{ij}}F\right)\mathcal{A}\left(R_{ij}\right) < \mathcal{A}\left(F(R_{ij})\right) < (1+\varepsilon)\Delta\left(T_{\overrightarrow{x}_{ij}}F\right)\mathcal{A}\left(R_{ij}\right)$$

where $\overrightarrow{x}_{ij}$ is the center of R_{ij}. Summing over all i and j, we have

$$(1-\varepsilon)\sum_{i,j}\Delta\left(T_{\overrightarrow{x}_{ij}}F\right)\mathcal{A}\left(R_{ij}\right) < \mathcal{A}\left(F\left(\bigcup_{i,j}R_{ij}\right)\right) < (1+\varepsilon)\sum_{i,j}\Delta\left(T_{\overrightarrow{x}_{ij}}F\right)\mathcal{A}\left(R_{ij}\right).$$

But the sum appearing at either end

$$\sum_{i,j} \Delta \left(T_{\overrightarrow{x}_{ij}} F \right) \mathcal{A}\left(R_{ij}\right)$$

is a Riemann sum for the integral $\iint_S \Delta \left(T_{\overrightarrow{x}} F \right) \, dA$, while

$$\bigcup_{i,j} R_{ij} = S$$

so we have, for arbitrary $\varepsilon > 0$,

$$(1-\varepsilon) \iint_S \Delta \left(T_{\overrightarrow{x}} F \right) \, dA \leq \mathcal{A}\left(F(S)\right) \leq (1+\varepsilon) \iint_S \Delta \left(T_{\overrightarrow{x}} F \right) \, dA;$$

thus the area equals the integral in this case.

Now in general, when $\mathcal{D}$ is an elementary region, we can find two polygons P_i, $i = 1, 2$, such that

1. $P_1 \subset \mathcal{D} \subset P_2$;
2. P_i is a non-overlapping union of squares (with sides parallel to the axes);
3. $\mathcal{A}\left(P_2 \setminus P_1\right)$ is arbitrarily small.

(See Figure 5.21.) Since F is $\mathcal{C}^1$, the quantity $|JF(\overrightarrow{x})|$ is bounded above for $\overrightarrow{x} \in P_2$, say by M. Hence, given $\varepsilon > 0$, we can pick P_1 and P_2 so that the area between them is less than $\varepsilon/$; from this it follows easily that, say

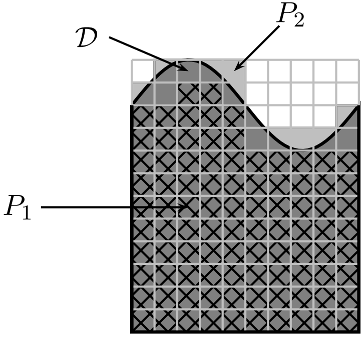


Figure 5.21: Approximating $\mathcal{D}$ with unions of squares

$$\left| \iint_{\mathcal{D}} |JF(\overrightarrow{x})| \, dA - \iint_{P_i} |JF(\overrightarrow{x})| \, dA \right| < \varepsilon;$$

but by the first property above,

$$\iint_{P_1} \left|JF(\overrightarrow{x})\right| \, dA = \mathcal{A}\left(F(P_1)\right) \leq \mathcal{A}\left(F(\mathcal{D})\right) \leq \mathcal{A}\left(F(P_2)\right) = \iint_{P_2} \left|JF(\overrightarrow{x})\right| \, dA$$

and since $\varepsilon > 0$ is arbitrary, this gives the desired equation.

□

Change of Coordinates in Double Integrals

A consequence of Theorem 5.3.11 is the following important result.

Theorem 5.3.12 (Change of Coordinates)**.** *Suppose $\mathcal{D}$ is an elementary region, $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a coordinate transformation defined on $\mathcal{D}$, and $f\colon \mathbb{R}^2 \to \mathbb{R}$ is a real-valued function which is integrable on $F(\mathcal{D})$.*

Then

$$\iint_{F(\mathcal{D})} f(\overrightarrow{x}) \, dA = \iint_{\mathcal{D}} f(F(\overrightarrow{x})) \left|JF(\overrightarrow{x})\right| \, dA. \tag{5.14}$$

Proof. [11] For notational convenience, let us write

$$g = f \circ F.$$

Since both g and $\left|JF(\overrightarrow{x})\right|$ are continuous, they are bounded and *uniformly* continuous on $\mathcal{D}$. Let M be an upper bound for both $|g|$ and $\left|JF(\overrightarrow{x})\right|$ on $\mathcal{D}$. By uniform continuity, given $\varepsilon > 0$, pick $\delta > 0$ so that $\left|\overrightarrow{x} - \overrightarrow{x}'\right| < \delta$, $\overrightarrow{x}, \overrightarrow{x}' \in \mathcal{D}$ guarantees that

$$\left|g(\overrightarrow{x}) - g(\overrightarrow{x}')\right| < \frac{\varepsilon}{2}$$

and

$$\left|\left|JF(\overrightarrow{x})\right| - \left|JF(\overrightarrow{x}')\right|\right| < \frac{\varepsilon}{2}.$$

Let R be a square contained in $\mathcal{D}$; take a partition $\mathcal{P}$ into subsquares of R with mesh size less than δ, and consider a single subsquare R_{ij}. Then

$$\begin{aligned} 0 &\leq \sup_{\overrightarrow{y} \in F(R_{ij})} f(\overrightarrow{y}) - \inf_{\overrightarrow{y} \in F(R_{ij})} f(\overrightarrow{y}) \\ &= \sup_{\overrightarrow{x} \in R_{ij}} g(\overrightarrow{x}) - \inf_{\overrightarrow{x} \in R_{ij}} g(\overrightarrow{x}) \\ &< \frac{\varepsilon}{2} \end{aligned}$$

[11] We note in passing a slight technical difficulty here: the image of an elementary region under a coordinate transformation may no longer be an elementary region. However, under very mild assumptions (see Exercise 9) it can be tiled by elementary regions, and so we can perform integration over it. We will ignore this problem in the following proof.

from which it follows that

$$\left|\iint_{F(R_{ij})} f(\overrightarrow{y})\;dA - f(\overrightarrow{y}_{ij})\,\mathcal{A}\left(F(R_{ij})\right)\right| < \frac{\varepsilon}{2}\mathcal{A}\left(F(R_{ij})\right)$$

where $\overrightarrow{x}_0$ is the center of R_{ij}, and $\overrightarrow{y}_{ij} = F(\overrightarrow{x}_0)$, and

$$\mathcal{A}\left(F(R_{ij})\right) = \iint_{R_{ij}} |JF(\overrightarrow{x}_{ij})|\;dA \leq M\mathcal{A}\left(R_{ij}\right).$$

Similarly,

$$\left|\iint_{R_{ij}} |JF(\overrightarrow{x})|\;dA - |JF(\overrightarrow{x}_{ij})\,\mathcal{A}\left(R_{ij}\right)|\right| < \left(\frac{\varepsilon}{2}\right)\mathcal{A}\left(R_{ij}\right).$$

It follows that

$$\begin{aligned}\left|\iint_{F(R_{ij})} f(\overrightarrow{y})\;dA - g(\overrightarrow{x}_{ij})\,JF(\overrightarrow{x}_{ij})\,\mathcal{A}\left(R_{ij}\right)\right|&\\ \leq \left|\iint_{F(R_{ij})} f(\overrightarrow{y})\;dA - f(\overrightarrow{y}_{ij})\,\mathcal{A}\left(F(R_{ij})\right)\right|&\\ + |g(\overrightarrow{x}_{ij})|\left|\iint_{R_{ij}} |JF(\overrightarrow{x})|\;dA - |JF(\overrightarrow{x}_{ij})|\right|&\\ < \left(\frac{\varepsilon}{2}M + \frac{\varepsilon}{2}M\right)\mathcal{A}\left(R_{ij}\right) = \varepsilon M\mathcal{A}\left(R_{ij}\right).&\end{aligned}$$

Adding up over all the component squares of R_{ij}, we get

$$\left|\iint_{F(R)} f(y)\;dA - \sum_{i,j} g(\overrightarrow{x}_{ij})\,|JF(\overrightarrow{x}_{ij})|\,\mathcal{A}\left(R_{ij}\right)\right| < \varepsilon M\mathcal{A}\left(R\right).$$

Now the sum on the right is a Riemann sum for $\iint_R g(\overrightarrow{x})\,|JF(\overrightarrow{x})|\;dA$ corresponding to the partition $\mathcal{P}$; as the mesh size of $\mathcal{P}$ goes to zero, this converges to $\iint_R g(\overrightarrow{x})\,|JF(\overrightarrow{x})|\;dA$ while $\varepsilon \to 0$. It follows that for any square R in $\mathcal{D}$,

$$\iint_{F(R)} f(y)\;dA = \iint_R g(\overrightarrow{x})\,|JF(\overrightarrow{x})|\;dA,$$

proving our result for a square.

In general, as in the proof of Theorem 5.3.11, given $\varepsilon > 0$, we can find polygons P_1 and P_2, each a non-overlapping union of squares, bracketing $\mathcal{D}$ and with areas differing by less than ε. Then

$$\begin{aligned}\mathcal{A}\left(\mathcal{D}\right) - \mathcal{A}\left(P_1\right) &< \varepsilon \\ \mathcal{A}\left(F(\mathcal{D})\right) - \mathcal{A}\left(F(P_1)\right) &= \iint_{\mathcal{D}\setminus P_1} \left|JF(\overrightarrow{x})\right| \, dA \\ &< \varepsilon M\end{aligned}$$

so

$$\begin{aligned}&\left|\iint_{F(\mathcal{D})} f(y)\, dA - \iint_{\mathcal{D}} g(\overrightarrow{x}_{ij}) \left|JF(\overrightarrow{x}_{ij})\right| \, dA\right| \\ &\qquad \leq \left|\iint_{F(\mathcal{D})} f(y)\, dA - \iint_{F(P_1)} f(y)\, dA\right| \\ &\qquad + \left|\iint_{F(P_1)} f(y)\, dA - \iint_{\mathcal{D}} g(\overrightarrow{x}_{ij}) \left|JF(\overrightarrow{x}_{ij})\right| \, dA\right| \\ &\qquad = \left|\iint_{F(\mathcal{D})} f(y)\, dA - \iint_{F(P_1)} f(y)\, dA\right| \\ &\qquad + \left|\iint_{P_1} g(\overrightarrow{x}) \left|JF(\overrightarrow{x})\right| \, dA - \iint_{\mathcal{D}} g(\overrightarrow{x}) \left|JF(\overrightarrow{x})\right| \, dA\right| \\ &\qquad < M\varepsilon + \varepsilon M = 2M\varepsilon\end{aligned}$$

and as $\varepsilon > 0$ is arbitrarily small, the two integrals in the first line are equal, as required.

□

The most frequent example of the situation in the plane handled by Theorem 5.3.12 is calculating an integral in polar instead of rectangular coordinates. You may already know how to integrate in polar coordinates, but here we will see this as part of a larger picture.

Consider the mapping F taking points in the (r, θ)-plane to points in the (x, y)-plane (Figure 5.22)

$$F\left(\begin{bmatrix} r \\ \theta \end{bmatrix}\right) = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix};$$

this takes horizontal lines (θ constant) in the (r, θ)-plane to rays in the

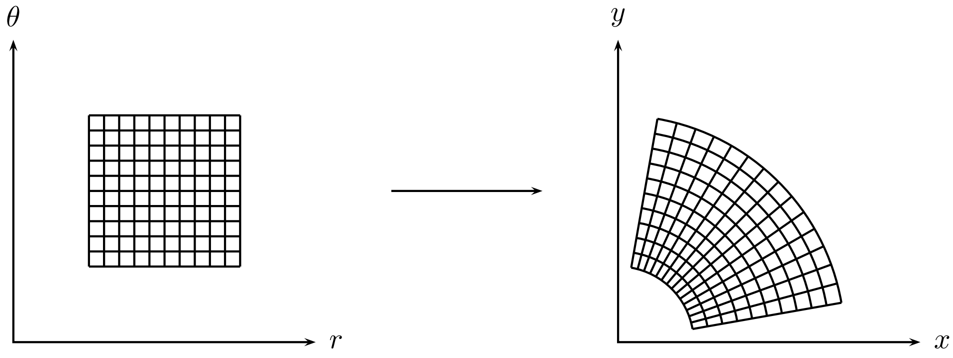


Figure 5.22: The Coordinate Transformation from Polar to Rectangular Coordinates

(x, y)-plane and vertical lines (r constant) to circles centered at the origin. Its Jacobian determinant is

$$JF(r,\theta) = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r,$$

so every point except the origin is a regular point. It is one-to-one on any region $\mathcal{D}$ in the (r, θ) plane for which r is always positive and θ does not vary by 2π or more, so on such a region it is a coordinate transformation.

Thus, to switch from a double integral expressing $\iint_{\mathcal{D}} f\,dA$ in rectangular coordinates to one in polar coordinates, we need to find a region $\mathcal{D}'$ in the r, θ-plane on which F is one-to-one, and then calculate the alternate integral $\iint_{\mathcal{D}'} (f \circ F)\, r\, dA$. this amounts to expressing the quantity $f(x, y)$ in polar coordinates and then using $r\, dr\, d\theta$ in place of $dx\, dy$.

For example, suppose we want to integrate the function

$$f(x, y) = 3x + 16y^2$$

over the region in the first quadrant between the circles of radius 1 and 2, respectively (Figure 5.23). In rectangular coordinates, this is fairly difficult to describe. Technically, it is x-simple (every vertical line crosses it in an interval), and the top is easily viewed as the graph of $y = \sqrt{4 - x^2}$; however, the bottom is a function defined in pieces:

$$y = \begin{cases} \sqrt{1 - x^2} & \text{for } 0 \le x \le 1, \\ 0 & \text{for } 1 \le x \le 2. \end{cases}$$

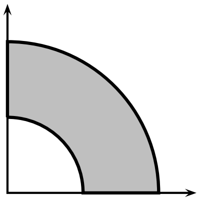


Figure 5.23: Region Between Concentric Circles in the First Quadrant

The resulting specification of $\mathcal{D}$ in effect views this as a union of two regions:

$$\begin{cases} \sqrt{1-x^2} & \leq y \leq & \sqrt{4-x^2} \\ 0 & \leq x \leq & 1 \end{cases}$$

and

$$\begin{cases} 0 & \leq y \leq & \sqrt{4-x^2} \\ 1 & \leq x \leq & 2 \end{cases} ;$$

this leads to the pair of double integrals

$$\iint_{\mathcal{D}} 3x + 16x^2\, dA = \int_0^1 \int_{\sqrt{1-x^2}}^{\sqrt{4-x^2}} + \int_1^2 \int_0^{\sqrt{4-x^2}} .$$

However, the description of our region in polar coordinates is easy:

$$\begin{aligned} 1 &\leq r \leq 2 \\ 0 &\leq \theta \leq \frac{p}{2} \end{aligned}$$

and (using the formal equivalence $dx\,dy = r\,dr\,d\theta$) the integral is

$$\begin{aligned}\iint_{\mathcal{D}} 3x + 16y^2\,dA &= \int_0^{\pi/2}\int_1^2 (3r\cos\theta + 16r^2\sin^2\theta)\,r\,dr\,d\theta\\ &= \int_0^{\pi/2}\int_1^2 (3r^2\cos\theta + 16r^3\sin^2\theta)\,dr\,d\theta\\ &= \int_0^{\pi/2} (r^3\cos\theta + 4r^4\sin^2\theta)_1^2\,d\theta\\ &= \int_0^{\pi/2} (7\cos\theta + 60\sin^2\theta)\,d\theta\\ &= 7\sin\theta\Big|_0^{\pi/2} + 30\int_0^{\pi/2}(1-\cos 2\theta)\,d\theta\\ &= 7 + 30\left(\theta - \frac{1}{2}\sin 2\theta\right)_0^{\pi/2}\\ &= 7 + 15\pi.\end{aligned}$$

We note in passing that the requirement that the coordinate transformation be regular and one-to-one on the whole domain can be relaxed slightly: we can allow critical points on the boundary, and also we can allow the boundary to have multiple points for the map. In other words, we need only require that every *interior* point of $\mathcal{D}$ is a regular point of F, and that the *interior* of $\mathcal{D}$ maps in a one-to-one way to its image.

Remark 5.3.13. *Suppose $\mathcal{D}$ is an elementary region (or is tiled by a finite union of elementary regions) and $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a $\mathcal{C}^1$ mapping defined on $\mathcal{D}$ such that*

1. *Every point interior to $\mathcal{D}$ is a regular point of F*

2. *F is one-to-one on the interior of $\mathcal{D}$.*

Then for any function f which is integrable on $F(\mathcal{D})$, Equation (5.14) still holds.

To see this, Let $P_k \subset \mathcal{D}$ be polygonal regions formed as nonoverlapping unions of squares inside $\mathcal{D}$ whose areas converge to that of $\mathcal{D}$. Then Theorem 5.3.12 applies to each, and the integral on either side of Equation (5.14) over P_k converges to the same integral over $\mathcal{D}$ (because the function is bounded, and the difference in areas goes to zero).

For example, suppose we want to calculate the volume of the upper hemisphere of radius R. One natural way to do this is to integrate the function $f(x,y)=\sqrt{x^2+y^2}$ over the disc $\mathcal{D}$ of radius R, which in rectangular coordinates is described by

$$\begin{aligned} -\sqrt{R^2-x^2} \leq & y \leq \sqrt{R^2-x^2} \\ -R \leq & x \leq R \end{aligned}$$

leading to the double integral

$$\iint_{\mathcal{D}} \sqrt{x^2+y^2}\,dA = \int_{-R}^{R}\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \sqrt{x^2+y^2}\,dy\,dx.$$

This is a fairly messy integral. However, if we describe $\mathcal{D}$ in polar coordinates (r,θ), we have the much simpler description

$$\begin{aligned} 0 \leq & r \leq R \\ 0 \leq & \theta \leq 2\pi. \end{aligned}$$

Now, the coordinate transformation F has a critical point at the origin, and identifies the two rays $\theta=0$ and $\theta=2\pi$; however, this affects only the boundary of the region, so we can apply our remark and rewrite the integral in polar coordinates. The quantity $\sqrt{x^2+y^2}$ expressed in polar coordinates is

$$\begin{aligned} f(r,\theta) &= \sqrt{(r\cos\theta)^2+(r\sin\theta)^2} \\ &= r. \end{aligned}$$

Then, replacing $dx\,dy$ with $r\,dr\,d\theta$, we have the integral

$$\begin{aligned} \iint_{\mathcal{D}} (r)(r\,dr\,d\theta)\,dA &= \int_0^{2\pi}\int_0^1 r^2\,dr\,d\theta \\ &= \int_0^{2\pi}\left(\frac{r^3}{3}\right)_0^R d\theta \\ &= \int_0^{2\pi}\left(\frac{R^3}{3}\right) d\theta \\ &= \frac{2\pi R^3}{3}. \end{aligned}$$

Exercises for § 5.3

Practice problems:

1. Use polar coordinates to calculate each integral below:

   (a) $\iint_D (x^2+y^2)\,dA$ where D is the annulus specified by
   $$1 \leq x^2+y^2 \leq 4.$$

   (b) The area of one "petal" of the "rose" given in polar coordinates as
   $$r = \sin n\theta,$$
   where n is a positive integer.

   (c) The area of the lemniscate given in rectangular coordinates by
   $$(x^2-y^2)^2 = 2a^2(x^2-y^2)$$
   where a is a constant. (*Hint:* Change to polar coordinates, and note that there are two equal "lobes"; find the area of one and double.)

2. Calculate the area of an ellipse in terms of its semiaxes. (*Hint:* There is a simple linear mapping taking a circle centered at the origin to an ellipse with center at the origin and horizontal and vertical axes.)

3. Calculate the integral
   $$\iint_{[0,1]\times[0,1]} \frac{1}{\sqrt{1+2x+3y}}\,dA$$
   using the mapping
   $$\varphi(x,y) = (2x,3y),$$
   that is, using the substitution
   $$\left\{\begin{array}{rcl} u &=& 2x, \\ v &=& 3y. \end{array}\right.$$

4. Calculate
   $$\iint_D (x^2+y^2)\,dA,$$
   where D is the parallelogram with vertices $(0,0)$, $(2,1)$, $(3,3)$, and $(1,2)$, by noting that D is the image of the unit square by the linear map
   $$\varphi(s,t) = (2s+t, s+2t).$$

5. Calculate
$$\iint_D \frac{1}{(x+y)^2}\,dA,$$
where D is the region in the first quadrant cut off by the lines
$$\begin{aligned} x+y&=1\\ x+y&=2, \end{aligned}$$
using the substitution
$$\begin{cases} x &= s-st, \\ y &= st \end{cases}.$$

Theory problems:

6. Suppose $L:\mathbb{R}^2\to\mathbb{R}^2$ is a linear mapping and the determinant of its matrix representative $[L]$ is positive. Suppose $\triangle ABC$ is a positively oriented triangle in the plane.
   (a) Show that the image $L(\triangle ABC)$ is a triangle with vertices $L(A)$, $L(B)$, and $L(C)$.
   (b) Show that $\sigma(L(A)\,L(B)\,L(C))$ is positive. (*Hint:* Consider the effect of L on the two vectors $\overrightarrow{v}=\overrightarrow{AB}$ and $\overrightarrow{w}=\overrightarrow{AC}$.)
   (c) Show that if the determinant of $[L]$ were *negative*, then $\sigma(L(A)\,L(B)\,L(C))$ would be *negative*.
   (d) Use this to show that the signed area of $[L(A)\,,L(B)\,,L(C)]$ equals det $[L]$ times the signed area of $[A,B,C]$.

7. Prove Remark 5.3.2 as follows: suppose $A=[L]$ and $B=[L']$. Then $AB=[L\circ L']$. Consider the unit square S with vertices $(0,0)$, $(1,0)$, $(1,0)$, and $(1,0)$. (In that order, it is positively oriented.) Its signed area is
$$\det\begin{pmatrix} 1 & 0 \\ 0 & v_2 \end{pmatrix}=1.$$
Now consider the parallelogram $S'=L'(S)$. The two directed edges $\overrightarrow{\imath}$ and $\overrightarrow{\jmath}$ of S map to the directed edges of S', which are $\overrightarrow{v}=L'(\overrightarrow{\imath})$ and $\overrightarrow{w}=L'(\overrightarrow{\jmath})$. **Show** that the first column of B is $[\overrightarrow{v}]$ and its second column is $[\overrightarrow{w}]$, so the signed area of $L'(S)$ is det B. Now, consider $L(S')$: its directed edges are $L(\overrightarrow{v})$ and $L(\overrightarrow{w})$. **Show** that

the coordinate columns of these two vectors are the columns of AB, so the signed area of $L(S')$ is $\det AB$. But it is also (by Exercise 6) $\det A$ times the area of S', which in turn is $\det B$. Combine these operations to **show** that $\det AB = \det A \det B$.

Challenge problem:

8. Calculate
$$\iint_D xy^3(1-x^2)\,dA,$$
where D is the region in the first quadrant between the circle
$$x^2+y^2=1$$
and the curve
$$x^4+y^4=1.$$
(*Hint:* Start with the substitution $u=x^2$, $v=y^2$. Note that this is possible only because we are restricted to the first quadrant, so the map is one-to-one.)

9. (a) Show that any trianglular region with two sides parallel to the coordinate axes is regular.

   (b) Show that a region with two sides parallel to the coordinate axes, and whose third side is the graph of a strictly monotone continuous function, is regular.

   (c) Show that a triangular region with at least one horizontal (*resp.* vertical) side can be subdivided into regular regions.

   (d) Show that a trapezoidal region with two horizontal (*resp.* two vertical) sides can be subdivided into regular regions.

   (e) Show that any polygonal region can be subdivided into non-overlapping regular regions.

   (f) Suppose the curve $\mathcal{C}$ bounding D is a simple closed curve consisting of a finite number of pieces, each of which is either a horizontal or vertical line segment or the graph of a $\mathcal{C}^2$ function with nowhere vanishing derivative[12] Show that the region D can be subdivided into non-overlapping regular regions.

[12] Note that the graph $y=\varphi(x)$ of a $\mathcal{C}^2$ function φ with nowhere-vanishing derivative can also be expressed as $x=\psi(y)$, where ψ is $\mathcal{C}^2$ with nowhere-vanishing derivative.

(g) Show that in the previous item it is enough to assume that each piece of $\mathcal{C}$ is either a horizontal or vertical line segment or the graph of a $\mathcal{C}^2$ function with finitely many critical points.

5.4 Integration Over Surfaces

Surface Area

In trying to define the area of a surface in $\mathbb{R}^3$, it is natural to try to mimic the procedure we used in § 2.5 to define the length of a curve: recall that we define the length of a curve $\mathcal{C}$ by partitioning it, then joining successive partition points with straight line segments, and considering the total length of the resulting polygonal approximation to $\mathcal{C}$ as an underestimate of its length (since a straight line gives the shortest distance between two points). The length of $\mathcal{C}$ is defined as the supremum of these underestimates, and $\mathcal{C}$ is *rectifiable* if the length is finite. Unfortunately, an example found (simultaneously) in 1892 by H. A. Schwartz and G. Peano says that if we try to define the area of a surface analogously, by taking the supremum of areas of polygonal approximations to the surface, we get the nonsense result that an ordinary cylinder has infinite area. The details are given in Appendix G.

As a result, we need to take a somewhat different approach to defining surface area. A number of different theories of area were developed in the period 1890-1956 by, among others, Peano, Lebesgue, Gœczes, Radó, and Cesari. We shall not pursue these general theories of area, but will instead mimic the arclength formula for *regular* curves. All of the theories of area agree on the formula we obtain this way in the case of *regular* surfaces.

Recall that in finding the circumference of a circle, Archimedes used two kinds of approximation: *inscribed* polygons and *circumscribed* polygons. The naive approach above is the analogue of the inscribed approximation: in approximating a (differentiable) planar curve, the Mean Value Theorem ensures that a line segment joining two points on the curve is parallel to the tangent at some point in between, and this insures that the projection of the arc onto this line segment joining them does not distort distances too badly (provided the arc is not too long). However, as the Schwarz-Peano example shows, this is no longer true for polygons inscribed in surfaces: inscribed triangles, even small ones, can make a large angle (near perpendicularity) with the surface, so projection distorts areas badly, and our intuition that the "area" of a piece of the surface projects nicely onto an inscribed polygon fails. But by definition, *circumscribed* polygons will be tangent to the surface at some point; this means that the projection of every curve in the surface that stays close to the point of tangency onto the tangent plane will make a relatively small angle with the surface, so that projection will not distort lengths or angles on the surface too badly. This is of course just an intuitive justification, but it suggests that we regard the projection of a (small) piece of surface onto the tangent plane at one of its points as a good approximation to the "actual" area.

To be more specific, let us suppose for the moment that our surface $\mathfrak{S}$ is the graph of a differentiable function $z = f(x, y)$ over the planar region $\mathcal{D}$, which for simplicity we take to be a rectangle $[a, b] \times [c, d]$. A partition of $[a, b] \times [c, d]$ divides $\mathcal{D}$ into subrectangles R_{ij}, and we denote the part of the graph above each such subrectangle as a subsurface $\mathfrak{S}_{ij}$ (Figure 5.24). Now we pick a sample point $(x_i, y_j) \in R_{ij}$ in each subrectangle of $\mathcal{D}$, and consider the plane tangent to $\mathfrak{S}$ at the corresponding point (x_i, y_j, z_{ij}) ($z_{ij} = f(x_i, x_j)$) of $\mathfrak{S}_{ij}$. The part of this plane lying above R_{ij} is a parallelogram whose area we take as an approximation to the area of $\mathfrak{S}_{ij}$, and we take these polygons as an approximation to the area of $\mathfrak{S}$ (Figure 5.25).

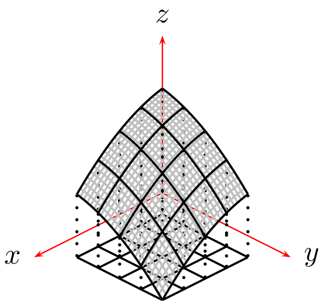


Figure 5.24: Subdividing a Graph

To find the area $\triangle S_{ij}$ of the parallelogram over R_{ij}, we can take as our sample point in R_{ij} its lower left corner; the sides of R_{ij} are parallel to the coordinate axes, so can be denoted by the vectors $\triangle x_i \overrightarrow{\imath}$ and $\triangle y_j \overrightarrow{\jmath}$. The edges of the parallelogram over R_{ij} are then given by vectors $\overrightarrow{v}_x$ and $\overrightarrow{v}_y$ which project down to these two, but lie in the tangent plane, which means their slopes are the two partial derivatives of f at the sample point (Figure 5.26). Thus,

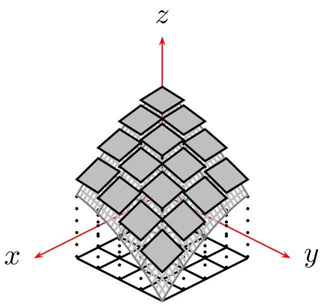


Figure 5.25: Approximating the Area of a Graph

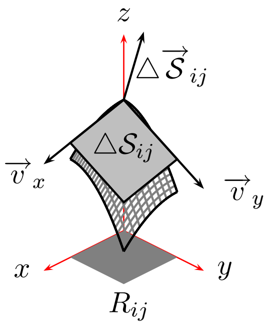

z
$\triangle\overrightarrow{\mathcal{S}}_{ij}$
$\triangle\mathcal{S}_{ij}$
$\overrightarrow{v}_x$
$\overrightarrow{v}_y$
x
y
R_{ij}

Figure 5.26: Element of Surface Area for a Graph

$$\begin{aligned}\overrightarrow{v}_x &= \left(\overrightarrow{\imath} + \frac{\partial f}{\partial x}\overrightarrow{k}\right)\triangle x_i \\ &= (1, 0, f_x)\triangle x_i \\ \overrightarrow{v}_y &= \left(\overrightarrow{\jmath} + \frac{\partial f}{\partial y}\overrightarrow{k}\right)\triangle y_j \\ &= (0, 1, f_x)\triangle y_j\end{aligned}$$

and the signed area of the parallelogram is

$$\begin{aligned}\triangle\overrightarrow{\mathcal{S}}_{ij} &= \overrightarrow{v}_x \times \overrightarrow{v}_y \\ &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix} \triangle x_i \triangle y_j \\ &= \left(-f_x\overrightarrow{\imath} - f_y\overrightarrow{\jmath} + \overrightarrow{k}\right)\triangle x_i\,\triangle y_j\end{aligned}$$

while the unsigned area is the length of this vector

$$\triangle\mathcal{S}_{ij} = \sqrt{f_x^2 + f_y^2 + 1}\,\triangle x_i\,\triangle y_j.$$

An alternate interpretation of this is to note that when we push a piece of $\mathcal{D}$ "straight up" onto the tangent plane at (x_i, y_j), its area gets multiplied by the factor $\sqrt{f_x^2 + f_y^2 + 1}$.

Adding up the areas of our parallelograms, we get as an approximation to the area of $\mathfrak{S}$

$$\sum_{i,j}\triangle\mathcal{S}_{ij}\,\triangle x_i\,\triangle y_j = \sum_{i,j}\sqrt{f_x^2 + f_y^2 + 1}\,\triangle x_i\,\triangle y_j.$$

But this is clearly a Riemann sum for an integral, which we take to be the definition of the area

$$\mathcal{A}\left(\mathfrak{S}\right) := \iint_{\mathcal{D}} d\mathcal{S} \tag{5.15}$$

where

$$d\mathcal{S} := \sqrt{f_x^2 + f_y^2 + 1}\,\triangle x_i\,\triangle y_j \tag{5.16}$$

is called the **element of surface area** for the graph.

For example, to find the area of the surface (Figure 5.27)

$$z = \frac{2}{3}\left(x^{3/2} + y^{3/2}\right)$$

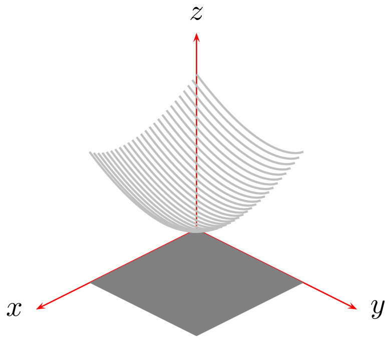


Figure 5.27: $z = \frac{2}{3}(x^{3/2} + y^{3/2})$

over the rectangle

$$\mathcal{D} = [0,1] \times [0,1]$$

we calculate the partials of $f(x,y) = \frac{2}{3}\left(x^{3/2} + y^{3/2}\right)$ as

$$\begin{aligned} f_x &= x^{1/2} \\ f_y &= y^{1/2} \end{aligned}$$

so

$$d\mathcal{S} = \sqrt{x+y+1}\, dx\, dy$$

and

$$\begin{aligned} \mathcal{A}(\mathfrak{S}) &= \iint_{\mathcal{D}} dS \\ &= \int_0^1 \int_0^1 \sqrt{x+y+1}\, dx\, dy \\ &= \int_0^1 \frac{2}{3}\left((x+y+1)^{3/2}\right)_{x=0}^{x=1} dy \\ &= \int_0^1 \frac{2}{3}\left((y+2)^{3/2} - (y+1)^{3/2}\right) dy \\ &= \frac{2}{3} \cdot \frac{2}{5}\left((y+2)^{5/2} - (y+1)^{5/2}\right)_0^1 \\ &= \frac{2}{15}\left((3^{5/2} - 2^{5/2}) - (2^{5/2} - 1^{5/2})\right) \qquad = \frac{2}{15}(9\sqrt{3} - 8\sqrt{2} + 1). \end{aligned}$$

We wish to extend our analysis to a general parametrized surface. The starting point of this analysis is the fact that if $\overrightarrow{p}(s,t)$ is a regular parametrization of the surface $\mathfrak{S}$,

$$\begin{aligned} x &= x(s,t) \\ y &= y(s,t) \\ z &= z(s,t) \end{aligned}$$

then a parametrization of the tangent plane to $\mathfrak{S}$ at $P = \overrightarrow{p}(s_0, t_0)$ is $T_P\overrightarrow{p}(s,t) = P + \frac{\partial \overrightarrow{p}}{\partial s}\triangle s + \frac{\partial \overrightarrow{p}}{\partial t}\triangle t$, that is,

$$\begin{aligned} x &= x(s_0,t_0) + \frac{\partial x}{\partial s}(P)(s - s_0) + \frac{\partial x}{\partial t}(P)(t - t_0) \\ y &= y(s_0,t_0) + \frac{\partial y}{\partial s}(P)(s - s_0) + \frac{\partial y}{\partial t}(P)(t - t_0) \\ z &= z(s_0,t_0) + \frac{\partial z}{\partial s}(P)(s - s_0) + \frac{\partial z}{\partial t}(P)(t - t_0). \end{aligned}$$

This defines the **tangent map** $T_P\overrightarrow{p}$ of the parametrization, which by analogy with the case of the graph analyzed above corresponds to "pushing" pieces of $\mathcal{D}$, the domain of the parametrization, to the tangent plane. To understand its effect on areas, we note that the edges of a rectangle in the domain of $\overrightarrow{p}$ with sides parallel to the s-axis and t-axis, and lengths $\triangle s$ and $\triangle t$, respectively, are taken by the tangent map to the vectors $\frac{\partial \overrightarrow{p}}{\partial s}\triangle s$ and $\frac{\partial \overrightarrow{p}}{\partial t}\triangle t$, which play the roles of $\overrightarrow{v}_x$ and $\overrightarrow{v}_y$ from the graph case. Thus, the signed area of the corresponding parallelogram in the tangent plane is given by the cross product (Figure 5.28)

$$\triangle\overrightarrow{\mathcal{S}} = \left(\frac{\partial \overrightarrow{p}}{\partial s}\triangle s\right) \times \left(\frac{\partial \overrightarrow{p}}{\partial t}\triangle t\right) = \left(\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right)\triangle s\triangle t.$$

The (unsigned) area is the length of this vector

$$\triangle\mathcal{S} = \left\|\triangle\overrightarrow{\mathcal{S}}\right\| = \left\|\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right\|\triangle s\triangle t.$$

Again, if we partition the domain of $\overrightarrow{p}$ into such rectangles and add up their areas, we are forming a Riemann sum, and as the mesh size of the partition goes to zero, these Riemann sums converge to the integral, over the domain $\mathcal{D}$ of $\overrightarrow{p}$, of the function $\left\|\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right\|$:

$$\mathcal{A}(\mathfrak{S}) = \iint_{\mathcal{D}} \left\|\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right\| dA. \tag{5.17}$$

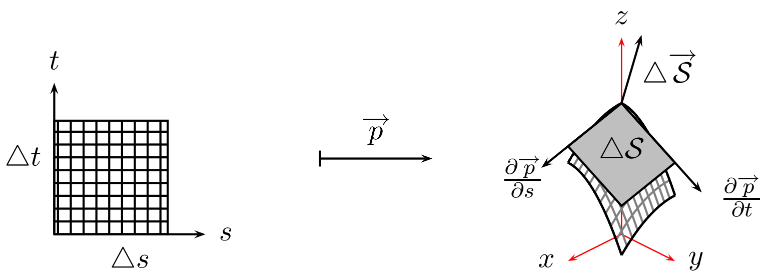

t
$\triangle t$
s
$\triangle s$
$\overrightarrow{p}$
z
$\triangle\overrightarrow{\mathcal{S}}$
$\triangle\mathcal{S}$
$\frac{\partial\overrightarrow{p}}{\partial s}$
$\frac{\partial\overrightarrow{p}}{\partial t}$
x
y

Figure 5.28: Element of Surface Area for a Parametrization

By analogy with the element of arclength $d\mathfrak{s}$, we denote the integrand above $d\mathcal{S}$; this is the **element of surface area**:

$$d\mathcal{S} = \left\| \frac{\partial\overrightarrow{p}}{\partial s} \times \frac{\partial\overrightarrow{p}}{\partial t} \right\| dA = \left\| \frac{\partial\overrightarrow{p}}{\partial s} \times \frac{\partial\overrightarrow{p}}{\partial t} \right\| ds\, dt.$$

We shall see below that the integral in Equation (5.17) is independent of the (regular) parametrization $\overrightarrow{p}$ of the surface $\mathfrak{S}$, and we write

$$\mathcal{A}(\mathfrak{S}) = \iint_{\mathfrak{S}} d\mathcal{S}..$$

For future reference, we also set up a vector-valued version of $d\mathcal{S}$, which could be called the **element of oriented surface area**

$$d\overrightarrow{\mathcal{S}} = \left(\frac{\partial\overrightarrow{p}}{\partial s} \times \frac{\partial\overrightarrow{p}}{\partial t} \right) ds\, dt.$$

To see that the definition of surface area given by Equation (5.17) is independent of the parametrization, it suffices to consider two parametrizations of the same coordinate patch, say $\overrightarrow{p}(u,v)$ and $\overrightarrow{q}(2,t)$. By Corollary 4.4.6, we can write

$$\overrightarrow{q} = \overrightarrow{p} \circ T$$

where

$$T(s,t) = (u(s,t)\,,v(s,t)).$$

By the Chain Rule,

$$\begin{aligned} \frac{\partial\overrightarrow{q}}{\partial s} &= \frac{\partial\overrightarrow{p}}{\partial u}\frac{\partial u}{\partial s} + \frac{\partial\overrightarrow{p}}{\partial v}\frac{\partial v}{\partial s} \\ \frac{\partial\overrightarrow{q}}{\partial t} &= \frac{\partial\overrightarrow{p}}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial\overrightarrow{p}}{\partial v}\frac{\partial v}{\partial t} \end{aligned}$$

so the cross product is

$$\begin{aligned}
\frac{\partial \overrightarrow{q}}{\partial s} \times \frac{\partial \overrightarrow{q}}{\partial t} &= \left(\frac{\partial \overrightarrow{p}}{\partial u}\frac{\partial u}{\partial s} + \frac{\partial \overrightarrow{p}}{\partial v}\frac{\partial v}{\partial s}\right) \times \left(\frac{\partial \overrightarrow{p}}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial \overrightarrow{p}}{\partial v}\frac{\partial v}{\partial t}\right) \\
&= \left(\frac{\partial u}{\partial s}\frac{\partial u}{\partial t}\right)\left(\frac{\partial \overrightarrow{p}}{\partial u} \times \frac{\partial \overrightarrow{p}}{\partial u}\right) + \left(\frac{\partial u}{\partial s}\frac{\partial v}{\partial t}\right)\left(\frac{\partial \overrightarrow{p}}{\partial u} \times \frac{\partial \overrightarrow{p}}{\partial v}\right) \\
&\quad + \left(\frac{\partial v}{\partial s}\frac{\partial u}{\partial t}\right)\left(\frac{\partial \overrightarrow{p}}{\partial v} \times \frac{\partial \overrightarrow{p}}{\partial u}\right) + \left(\frac{\partial v}{\partial s}\frac{\partial v}{\partial t}\right)\left(\frac{\partial \overrightarrow{p}}{\partial v} \times \frac{\partial \overrightarrow{p}}{\partial v}\right) \\
&= \left(\frac{\partial u}{\partial s}\frac{\partial v}{\partial t} - \frac{\partial v}{\partial s}\frac{\partial u}{\partial t}\right)\left(\frac{\partial \overrightarrow{p}}{\partial u} \times \frac{\partial \overrightarrow{p}}{\partial v}\right) \\
&= (\det JT)\left(\frac{\partial \overrightarrow{p}}{\partial u} \times \frac{\partial \overrightarrow{p}}{\partial v}\right).
\end{aligned}$$

Now, by Theorem 5.3.12 (or, if necessary, Remark 5.3.13) we see that the integral over the domain of $\overrightarrow{p}$ of the first cross product equals the integral over the domain of $\overrightarrow{q}$ of the last cross product, which is to say the two surface area integrals are equal.

As an example, let us find the surface area of the cylinder

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

We use the natural parametrization (writing θ instead of s)

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

with domain

$$\mathcal{D} = [0, 2\pi] \times [0, 1]\,.$$

The partial derivatives of the parametrization

$$\overrightarrow{p}(\theta, t) = (\cos\theta, \sin\theta, t)$$

are

$$\begin{aligned}
\frac{\partial \overrightarrow{p}}{\partial \theta} &= (-\sin\theta, \cos\theta, 0) \\
\frac{\partial \overrightarrow{p}}{\partial t} &= (0, 0, 1);
\end{aligned}$$

their cross-product is

$$\begin{aligned}\frac{\partial \overrightarrow{p}}{\partial \theta} \times \frac{\partial \overrightarrow{p}}{\partial t} &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} \\ &= (\cos\theta)\overrightarrow{\imath} + (\sin\theta)\overrightarrow{\jmath}\end{aligned}$$

so the element of area is

$$\begin{aligned}d\mathcal{S} &= \|(\cos\theta)\overrightarrow{\imath} + (\sin\theta)\overrightarrow{\jmath}\|\, d\theta\, dt \\ &= d\theta\, dt\end{aligned}$$

and its integral, giving the surface area, is

$$\begin{aligned}\mathcal{A}\left(\mathfrak{S}\right) &= \iint_{\mathfrak{S}} d\mathcal{S} \\ &= \iint_{[0,2\pi]\times[0,1]} d\theta\, dt \\ &= \int_0^1 \int_0^{2\pi} d\theta\, dt \\ &= \int_0^1 2\pi\, dt \\ &= 2\pi\end{aligned}$$

which is what we would expect (you can form the cylinder by rolling the rectangle $[0, 2\pi] \times [0, 1]$ into a "tube").

As a second example, we calculate the surface area of a sphere $\mathcal{S}$ of radius R; we parametrize via spherical coordinates:

$$\begin{aligned}\overrightarrow{p}(\phi, \theta) &= (R\sin\phi\cos\theta, R\cos\phi\sin\theta, R\cos\phi); \\ \frac{\partial \overrightarrow{p}}{\partial \phi} &= (R\cos\phi\cos\theta, R\cos\phi\sin\theta, -R\sin\phi) \\ \frac{\partial \overrightarrow{p}}{\partial \theta} &= (-R\sin\phi\sin\theta, R\sin\phi\cos\theta, 0) \\ \frac{\partial \overrightarrow{p}}{\partial \phi} \times \frac{\partial \overrightarrow{p}}{\partial \theta} &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ R\cos\phi\cos\theta & R\cos\phi\sin\theta & -R\sin\phi \\ -R\sin\phi\sin\theta & R\sin\phi\cos\theta & 0 \end{vmatrix} \\ &= R^2(\sin^2\phi\cos\theta)\overrightarrow{\imath} + R^2(\sin^2\phi\sin\theta)\overrightarrow{\jmath} \\ &\quad + R^2(\sin\phi\cos\phi\cos^2\theta + \sin\phi\cos\phi\sin^2\theta)\overrightarrow{k}\end{aligned}$$

so the element of oriented area is

$$d\overrightarrow{\mathcal{S}} = R^2(\sin^2\phi\cos\theta, \sin^2\phi\sin\theta, \sin\phi\cos\phi)\,d\phi\,d\theta$$

and the element of area is

$$\begin{aligned} d\mathcal{S} &= R^2\sqrt{\sin^4\phi\cos^2\theta + \sin^4\phi\sin^2\theta + \sin^2\phi\cos^2\phi}\,d\phi\,d\theta \\ &= R^2\sqrt{\sin^4\phi + \sin^2\phi\cos^2\phi}\,d\phi\,d\theta \\ &= R^2\sqrt{\sin^2\phi}\,d\phi\,d\theta \\ &= R^2\sin\phi\,d\phi\,d\theta \end{aligned}$$

(where the last equality is justified by the fact that $0 \leq \phi \leq \pi$, so $\sin\phi$ is always non-negative). From this, we have the area integral

$$\begin{aligned} \mathcal{A}(\mathcal{S}) &= \iint_{\mathcal{S}} d\mathcal{S} \\ &= \int_0^{2\pi}\int_0^{\pi} R^2\sin\phi\,d\phi\,d\theta \\ &= \int_0^{2\pi} (-R^2\cos\theta)_0^{\pi}\,d\theta \\ &= \int_0^{2\pi} 2R^2\,d\theta \\ &= 4\pi R^2. \end{aligned}$$

Finally, let us calculate the area of the helicoid (Figure 5.29)

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

with domain

$$\begin{aligned} 0 &\leq r \leq 1 \\ 0 &\leq \theta \leq 2\pi. \end{aligned}$$

The partials of the parametrization

$$\overrightarrow{p}(r,\theta) = (r\cos\theta, r\sin\theta, \theta)$$

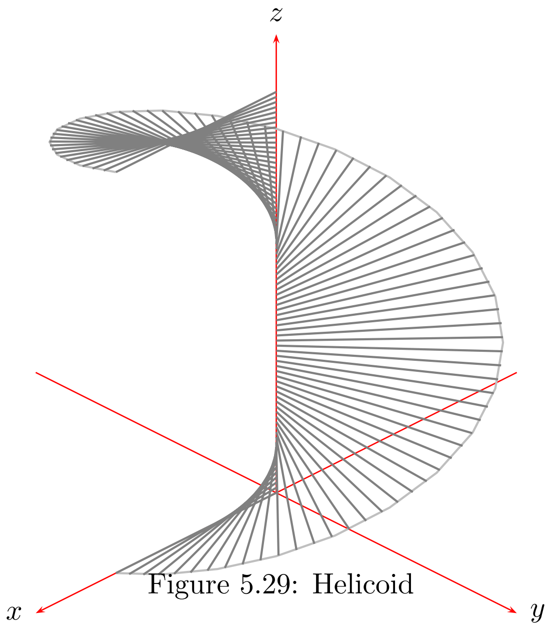


Figure 5.29: Helicoid

are

$$\begin{aligned}
\frac{\partial \overrightarrow{p}}{\partial r} &= (\cos\theta, \sin\theta, 0)\\
\frac{\partial \overrightarrow{p}}{\partial \theta} &= (-r\sin\theta, r\cos\theta, 1)
\end{aligned}$$

so

$$\begin{aligned}
d\overrightarrow{\mathcal{S}} &= (r\cos\theta, r\sin\theta, \theta) \times (-r\sin\theta, r\cos\theta, 1)\, dr\, d\theta\\
&= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k}\\ \cos\theta & \sin\theta & 0\\ -r\sin\theta & r\cos\theta & 1 \end{vmatrix} dr\, d\theta\\
&= (\sin\theta)\overrightarrow{\imath} - (\cos\theta)\overrightarrow{\jmath} + r\,\overrightarrow{k}
\end{aligned}$$

and

$$\begin{aligned}
d\mathcal{S} &= \left\|(\sin\theta)\overrightarrow{\imath} - (\cos\theta)\overrightarrow{\jmath} + r\,\overrightarrow{k}\right\| dr\, d\theta\\
&= \sqrt{1+r^2}\, dr\, d\theta.
\end{aligned}$$

The surface area is given by the integral

$$\iint_{\mathfrak{S}} d\mathcal{S} = \int_0^{2\pi}\int_0^1 \sqrt{1+r^2}\,dr\,d\theta;$$

using the substitutuion

$$\begin{aligned} r &= \tan\alpha \\ dr &= \sec^2\alpha\,d\alpha \\ \sqrt{1+r^2} &= \sec\alpha \\ r = 0 &\leftrightarrow \alpha = 0 \\ r = 1 &\leftrightarrow \alpha = \frac{\pi}{4} \end{aligned}$$

the inner integral becomes

$$\begin{aligned} \int_0^1 \sqrt{1+r^2}\,dr &= \int_0^{\pi/4} \sec^3\alpha\,d\alpha \\ &= \frac{1}{2}\left(\sec\alpha\tan\alpha + \ln|\sec\alpha + \tan\alpha|\right) \\ &= \frac{1}{2}\left(\sqrt{2} + \ln(\sqrt{2}+1)\right) \end{aligned}$$

turning the outer integral into

$$\begin{aligned} \int_0^{2\pi}\int_0^1 \sqrt{1+r^2}\,dr\,d\theta &= \int_0^{2\pi}\int_0^{\pi/4} \sec^3\alpha\,d\alpha\,d\theta \\ &= \int_0^{2\pi} \frac{1}{2}\left(\sqrt{2} + \ln(\sqrt{2}+1)\right)\,d\theta \\ &= \pi\left(\sqrt{2} + \ln(\sqrt{2}+1)\right). \end{aligned}$$

We note that for a surface given as the graph $z = f(x,y)$ of a function over a domain $\mathcal{D}$, the natural parametrization is

$$\overrightarrow{p}(s,t) = (s,t,f(s,t))$$

with partials

$$\begin{aligned} \frac{\partial\overrightarrow{p}}{\partial s} &= (1,0,f_x) \\ \frac{\partial\overrightarrow{p}}{\partial t} &= (0,1,f_y) \end{aligned}$$

so the element of oriented surface area is

$$\begin{aligned}
d\overrightarrow{\mathcal{S}} &= (1,0,f_x)\times(0,1,f_y)\,ds\,dt\\
&= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix}\,ds\,dt\\
&= -f_x\overrightarrow{\imath} - f_y\overrightarrow{\jmath} + \overrightarrow{k}
\end{aligned}$$

and in particular the element of (unoriented) surface area is

$$\begin{aligned}
d\mathcal{S} &= \left\|-f_x\overrightarrow{\imath} - f_y\overrightarrow{\jmath} + \overrightarrow{k}\right\|\,ds\,dt\\
&= \sqrt{(f_x)^2 + (f_y)^2 + 1}\,ds\,dt.
\end{aligned}$$

That is, we recover the formula (5.16) we obtained earlier for this special case.

Another special situation in which the element of surface area takes a simpler form is that of a **revolute** or **surface of revolution**—that is, the surface formed from a plane curve $\mathcal{C}$ when the plane is rotated about an axis that does not cross $\mathcal{C}$ (Figure 5.30). Let us assume that the axis of rotation

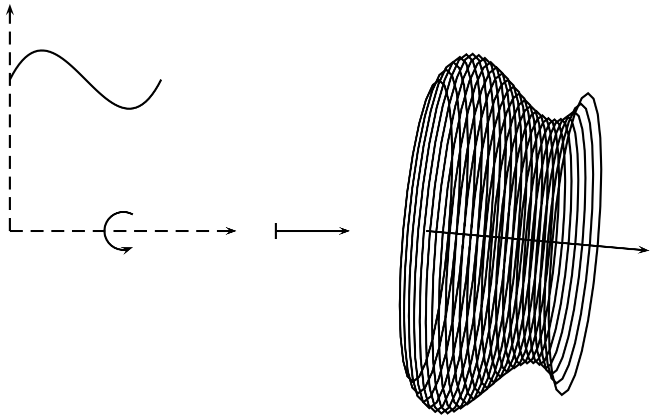


Figure 5.30: Revolute

is the x-axis, and that the curve $\mathcal{C}$ is parametrized by

$$\begin{aligned} x &= x(t) \\ y &= y(t)\,, \\ a \leq & t \leq b. \end{aligned}$$

Then our assumption that the axis does not cross $\mathcal{C}$ is $y(t) \geq 0$. A natural parametrization of the surface of revolution is obtained by replacing the point $(x(t)\,, y(t))$ with a circle, centered at $(x(t)\,, 0, 0)$ and parallel to the yz-plane, of radius $y(t)$; this yields the parametrization $\overrightarrow{p}(t, \theta)$ of the revolute

$$\begin{aligned} x &= x(t) \\ y &= y(t)\cos\theta \\ z &= y(t)\sin\theta, \\ a \leq & t \leq b, \\ 0 \leq & \theta \leq 2\pi \end{aligned}$$

The partials are

$$\begin{aligned} \frac{\partial \overrightarrow{p}}{\partial t} &= (x'(t)\,, y'(t)\cos\theta, y'(t)\sin\theta) \\ \frac{\partial \overrightarrow{p}}{\partial \theta} &= (0, -y(t)\sin\theta, y(t)\cos\theta) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \overrightarrow{p}}{\partial t} \times \frac{\partial \overrightarrow{p}}{\partial \theta} &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ x'(t) & y'(t)\cos\theta & y'(t)\sin\theta) \\ 0 & -y(t)\sin\theta & y(t)\cos\theta) \end{vmatrix} \\ &= (yy')\overrightarrow{\imath} + (yx')\left[-(\cos\theta)\overrightarrow{\jmath} + (\sin\theta)\overrightarrow{k}\right] \end{aligned}$$

with length

$$\left\| \frac{\partial \overrightarrow{p}}{\partial t} \times \frac{\partial \overrightarrow{p}}{\partial \theta} \right\| = y\sqrt{(y')^2 + (x')^2}.$$

Thus the element of surface area for a surface of revolution is

$$d\mathcal{S} = \left[y\sqrt{(y')^2 + (x')^2} \right] dt\, d\theta \tag{5.18}$$

and the surface area is

$$\begin{aligned}\mathcal{A}(\mathfrak{S}) &= \int_0^{2\pi}\int_a^b \left[y\sqrt{(y')^2+(x')^2}\right]\,dt\,d\theta \\ &= 2\pi\int_a^b \left[y\sqrt{(y')^2+(x')^2}\right]\,dt.\end{aligned} \tag{5.19}$$

More generally, we should replace $x(t)$ with the projection of a point $\overrightarrow{p}(t)$ on the curve $\mathcal{C}$ onto the axis of rotation, and $y(t)$ with its distance from that axis.

For example, the area of the surface obtained by rotating the curve $y = x^2$, $0 \leq x \leq 1$ about the x-axis, using the natural parametrization

$$\begin{aligned} x &= t \\ y &= t^2, \\ 0 \leq& t \leq 1\end{aligned}$$

is

$$\begin{aligned}2\pi\int_0^1 t^2\sqrt{t^2+1}\,dt &= 2\pi\left[\frac{t}{8}(1+2t^2\sqrt{t^2+1}) - \frac{t^4}{8}\ln(t+\sqrt{t^2+1})\right]_0^1 \\ &= \frac{\pi}{4}(3\sqrt{2}-\ln(1+\sqrt{2}),\end{aligned}$$

while for the surface obtained by rotating the same surface about the y-axis (using the same parametrization) is

$$\begin{aligned}2\pi\int_0^1 t\sqrt{1+t^2}\,dt &= 2\pi\left[\frac{1}{3}(1+t^2)^{3/2}\right]_0^1 \\ &= \frac{2\pi}{3}\left(2^{3/2}-1^{3/2}\right) \\ &= \frac{2\pi}{3}\left(2\sqrt{2}-1\right).\end{aligned}$$

Surface Integrals

Just as we could use the element of arclength to integrate a function f on $\mathbb{R}^3$ over a curve, so can we integrate this function over a (regular) surface. This can be thought of in terms of starting from a (possibly negative as well as positive) density function to calculate the total mass. Going through the same approximation process as we used to define the surface area itself, this time we sum up the area of small rectangles in the tangent plane at partition points *multiplied by the values of the function* there; this gives a Riemann sum for the **surface integral** of f over the surface

$$\iint_{\mathfrak{S}} f\ d\mathcal{S}.$$

Given a parametrization $\overrightarrow{p}(s,t)$ $((s,t) \in \mathcal{D})$ of the surface $\mathfrak{S}$, the process of calculating the surface integral above is exactly the same as before, except that we also throw in the value $f(\overrightarrow{p}(s,t))$ of the function.

For example, to calculate the integral of $f(x,y,z) = \sqrt{x^2+y^2+1}$ over the helicoid

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= \theta, \\ 0 \leq& r \leq 1 \\ 0 \leq& \theta \leq 2\pi \end{aligned}$$

which we studied earlier, we recall that

$$d\mathcal{S} = \sqrt{1+r^2}\,dr\,d\theta$$

and clearly

$$\begin{aligned} f(r\cos\theta, r\sin\theta, \theta) &= \sqrt{r^2\cos^2\theta + r^2\sin^2\theta + 1} \\ &= \sqrt{r^2+1} \end{aligned}$$

so our integral becomes

$$\begin{aligned} \iint_{\mathfrak{S}} \sqrt{x^2+y^2+1}\ d\mathcal{S} &= \int_0^{2\pi}\int_0^1 \left(\sqrt{r^2+1}\right)\left(\sqrt{r^2+1}\,dr\,d\theta\right) \\ &= \int_0^{2\pi}\int_0^1 \left(r^2+1\right)\ dr\,d\theta \\ &= \int_0^{2\pi} \left(\frac{r^3}{3}+r\right)_0^1\ d\theta \\ &= \int_0^{2\pi} \left(\frac{4}{3}\right)\ d\theta \\ &= \frac{8\pi}{3}. \end{aligned}$$

As another example, let us calculate the surface integral

$$\iint_{\mathcal{S}^2} z^2 \, d\mathcal{S}$$

where $\mathcal{S}^2$ is the unit sphere in $\mathbb{R}^3$. We can parametrize the sphere via spherical coordinates

$$\begin{aligned} x &= \sin\phi\cos\theta \\ y &= \sin\phi\sin\theta \\ z &= \cos\phi, \\ 0 \leq &\phi \leq \pi \\ 0 \leq &\theta \leq 2\pi; \end{aligned}$$

the partials are

$$\begin{aligned} \frac{\partial\overrightarrow{p}}{\partial\phi} &= (\cos\phi\cos\theta, \cos\phi\sin\theta, -\sin\phi) \\ \frac{\partial\overrightarrow{p}}{\partial\theta} &= (-\sin\phi\sin\theta, \sin\phi\cos\theta, 0) \end{aligned}$$

so

$$\begin{aligned} d\overrightarrow{\mathcal{S}} &= \frac{\partial\overrightarrow{p}}{\partial\phi} \times \frac{\partial\overrightarrow{p}}{\partial\theta} \, d\phi \, d\theta \\ &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ \cos\phi\cos\theta & \cos\phi\sin\theta & , -\sin\phi \\ -\sin\phi\sin\theta & \sin\phi\cos\theta & 0 \end{vmatrix} d\phi \, d\theta \\ &= \left(\sin^2\phi\cos\theta\right)\overrightarrow{\imath} + \left(\sin^2\phi\sin\theta\right)\overrightarrow{\jmath} + (\sin\phi\cos\phi)\overrightarrow{k} \; d\phi \, d\theta \end{aligned}$$

and

$$\begin{aligned} d\mathcal{S} &= \left\| d\overrightarrow{\mathcal{S}} \right\| \\ &= \sqrt{\sin^4\phi\cos^2\theta + \sin^4\phi\sin^2\theta + \sin^2\phi\cos^2\phi} \, d\phi \, d\theta \\ &= \sqrt{\sin^4\phi + \sin^2\phi\cos^2\phi} \, d\phi \, d\theta \\ &= |\sin\phi| \; d\phi \, d\theta \end{aligned}$$

which, in the range $0 \leq \phi \leq \pi$, is

$$= \sin\phi \, d\phi \, d\theta.$$

Now, our function $f(x, y, z) = z^2$ is

$$z^2 = \cos^2\phi$$

so the integral becomes

$$\begin{aligned} \iint_{\mathcal{S}^2} z^2 \; d\mathcal{S} &= \int_0^{2\pi} \int_0^{\pi} \left(\cos^2\phi\right) \left(\sin\phi \, d\phi \, d\theta\right) \\ &= \int_0^{2\pi} -\frac{\cos^3\phi}{3}\bigg|_0^{\pi} d\theta \\ &= \frac{2}{3}(2\pi) \\ &= \frac{4\pi}{3}. \end{aligned}$$

Exercises for § 5.4

Practice problems:

1. Find the area of each surface below.

   (a) The graph of $f(x, y) = 1 - \frac{x^2}{2}$ over the rectangle $[-1, 1] \times [-1, 1]$.

   (b) The graph of $f(x, y) = xy$ over the unit disc $x^2 + y^2 \leq 1$.

   (c) The part of the paraboloid $z = a^2 - x^2 - y^2$ above the xy-plane.

   (d) The part of the saddle surface $z = x^2 - y^2$ inside the cylinder $x^2 + y^2 = 1$.

   (e) The cone given in cylindrical coordinates by $z = mr$, $r \leq R$.

   (f) The part of the sphere $x^2 + y^2 + z^2 = 8$ cut out by the cone $z = \sqrt{x^2 + y^2}$.

   (g) The part of the sphere $x^2 + y^2 + z^2 = 9$ outside the cylinder $4x^2 + 4y^2 = 9$.

   (h) The surface parametrized by

$$\begin{cases} x &= & s^2 & +t^2 \\ y &= & s & -t \\ z &= & s & +t \end{cases}, \quad \begin{cases} -s & \leq t & \leq s \\ s^2 & +t^2 & \leq 1. \end{cases}$$

2. Evaluate each surface integral $\iint_{\mathfrak{S}} f \, d\mathcal{S}$ below.

   (a) $f(x, y, z) = x^2 + y^2$, $\mathfrak{S}$ is the part of the plane $z = x + 2y$ lying over the square $[0, 1] \times [0, 1]$.

   (b) $f(x, y, z) = xy + z$, $\mathfrak{S}$ is the part of the hyperboloid $z = xy$ over the square $[0, 1] \times [0, 1]$.

   (c) $f(x, y, z) = xyz$, $\mathfrak{S}$ is the triangle with vertices $(1, 0, 0)$, $(0, 2, 0)$, and $(0, 0, 1)$.

   (d) $f(x, y, z) = z$, $\mathfrak{S}$ is the upper hemisphere of radius R centered at the origin.

   (e) $f(x, y, z) = x^2+y^2$, $\mathfrak{S}$ is the surface of the cube $[0, 1]\times[0, 1]\times[0, 1]$. (*Hint:* Calculate the integral over each face separately, and add.)

   (f) $f(x, y, z) = \sqrt{x^2 + y^2 + 1}$, $\mathfrak{S}$ is the part of the surface $z = xy$ inside the cylinder $x^2 + y^2 = 1$.

   (g) $f(x, y, z) = z$, $\mathfrak{S}$ is the cone given in cylindrical coordinates by $z = 2r$, $0 \le z \le 2$.

Theory problems:

3. (a) Give a parametrization of the ellipsoid

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

   (b) Set up, but do not attempt to evaluate, an integral giving the surface area of the ellipsoid.

4. (a) Let $\mathfrak{S}$ be the surface obtained by rotating a curve $\mathcal{C}$ which lies in the upper half-plane about the x-axis. Show that the surface area as given by Equation (5.19) is just the path integral

$$\mathcal{A}(\mathfrak{S}) = 2\pi \int_{\mathcal{C}} y \, d\mathfrak{s}.$$

   (b) The **centroid** of a curve $\mathcal{C}$ can be defined as the "average" position of points on the curve with respect to arc length; that is, the x-coordinate of the centroid is given by

$$\bar{x} := \frac{\int_{\mathcal{C}} x \, d\mathfrak{s}}{\mathfrak{s}(\mathcal{C})}$$

with analogous definitions for the other two coordinates. This is the "center of gravity" of the curve if it has constant density.

Pappus' First Theorem, given in the *Mathematical Collection* of Pappus of Alexandria (*ca.* 300 AD) and rediscovered in the sixteenth century by Paul Guldin (1577-1643), says that the area of a surface of revolution equals the length of the curve being rotated times the distance travelled by its centroid. Prove this result from the preceding.

5. Suppose $f(x, y, z)$ is a $\mathcal{C}^1$ function for which the partial derivative $\partial f/\partial z$ is nonzero in the region $\mathfrak{D} \subset \mathbb{R}^3$, so that the part of any level surface in $\mathfrak{D}$ can be expressed as the graph of a function $z = \phi(x, y)$ over a region $\mathcal{D}$ in the x, y-plane. Show that the area of such a level surface is given by

$$\mathcal{A}\left(\mathcal{L}(f, c) \cap \mathfrak{D}\right) = \iint_{\mathcal{D}} \frac{\left\|\overrightarrow{\nabla} f\right\|}{|\partial f/\partial z|}\, dx\, dy.$$

Challenge problems:

6. (a) Use the parametrization of the torus given in Equation (3.26) to find its surface area.

   (b) Do the same calculation using Pappus' First Theorem.

7. Given a surface $\mathfrak{S}$ parametrized by $\overrightarrow{p}(u, v)$, $(u, v) \in \mathcal{D}$, define the functions

$$E = \left\|\frac{\partial \overrightarrow{p}}{\partial u}\right\|^2$$
$$F = \frac{\partial \overrightarrow{p}}{\partial u} \cdot \frac{\partial \overrightarrow{p}}{\partial v}$$
$$G = \left\|\frac{\partial \overrightarrow{p}}{\partial v}\right\|^2.$$

   (a) Suppose $\mathcal{C}$ is a curve in $\mathfrak{S}$ given in (u, v) coordinates as $\overrightarrow{g}(t) = (u(t), v(t))$, $t_0 \leq t \leq t_1$—that is, it is parametrized by

$$\gamma(t) = \overrightarrow{p}(\overrightarrow{g}(t)) = \overrightarrow{p}(u(t), v(t)).$$

Show that the speed of $\gamma(t)$ is given by

$$\begin{aligned}\frac{d\mathfrak{s}}{dt} &= \|\gamma'(t)\| \\ &= Q\big(g'(t)\big)\end{aligned}$$

where Q is the quadratic form with matrix representative

$$[Q] = \begin{bmatrix} E & F \\ F & G \end{bmatrix}.$$

This means the length of $\mathcal{C}$ is

$$\begin{aligned}\mathfrak{s}\,(\mathcal{C}) &= \int_{\mathcal{C}} d\mathfrak{s} \\ &= \int_{t_0}^{t_1} \sqrt{E(u')^2 + 2F(u')(v') + G(v')^2}\, dt.\end{aligned}$$

The quadratic form Q s called the **first fundamental form** of $\mathfrak{S}$.

(b) Show that the surface area of $\mathfrak{S}$ is given by

$$\mathcal{A}\,(\mathfrak{S}) = \iint_{\mathcal{D}} \sqrt{EG - F^2}\, du\, dv$$

or

$$d\mathcal{S} = \det\,[Q]\;du\,dv.$$

5.5 Integration in Three Variables

In theory, the extension of integration from two to three variables is a simple matter: the role of rectangles $[a,b] \times [c,d]$ is now played by rectangular solids with faces parallel to the coordinate planes

$$[a_1,b_1] \times [a_2,b_2] \times [a_3,b_3] := \{(x_1,x_2,x_3) \,|\, x_i \in [a_i,b_i] \;\text{ for } i = 1,2,3\};$$

a partition $\mathcal{P}$ of $[a_1,b_1] \times [a_2,b_2] \times [a_3,b_3]$ is determined by three coordinate partitions

$$\begin{aligned}\mathcal{P}_1 &:= \{a_1 = x_0 < x_1 < \cdots < x_m = b_1\} \\ \mathcal{P}_2 &:= \{a_2 = y_0 < y_1 < \cdots < y_n = b_2\} \\ \mathcal{P}_3 &:= \{a_3 = z_0 < z_1 < \cdots < z_p = b_3\}\end{aligned}$$

which subdivide $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ into $m \cdot n \cdot p$ subsolids R_{ijk}, $i = 1, \dots, m$, $j = 1, \dots, n$, $k = 1, \dots, p$

$$R_{ijk} = [x_{i-1}, x_i] \times [y_{j-1}, y_j] \times [z_{k-1}, z_k]$$

with respective volumes

$$\begin{aligned}\triangle V_{ijk} &= \triangle x_i \triangle y_j \triangle z_k \\ &= (x_i - x_{i-1})(y_j - y_{j-1})(z_k - z_{k-1}).\end{aligned}$$

Now given a function f bounded on $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ we can form the lower and upper sums

$$\begin{aligned}\mathcal{L}(\mathcal{P}, f) &= \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \left(\inf_{R_{ijk}} f \right) \triangle V_{ijk} \\ \mathcal{U}(\mathcal{P}, f) &= \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \left(\sup_{R_{ijk}} f \right) \triangle V_{ijk}.\end{aligned}$$

If the lower integral

$$\underline{\iiint}_{[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]} f(x, y, z)\ dV := \sup_{\mathcal{P}} \mathcal{L}(\mathcal{P}, f)$$

equals the upper integral

$$\overline{\iiint}_{[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]} f(x, y, z)\ dV := \inf_{\mathcal{P}} \mathcal{U}(\mathcal{P}, f)$$

then the function is **integrable** over $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$, with **integral**

$$\begin{aligned}\iiint_{[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]} f\, dV &= \underline{\iiint}_{[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]} f(x, y, z)\ dV \\ &= \overline{\iiint}_{[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]} f(x, y, z)\ dV.\end{aligned}$$

We shall not retrace all the details of the theory beyond this. As before, in practice such integrals are calculated via iterated integrals, but (not surprizingly) they are *triple* integrals. We shall call a region $\mathfrak{D} \subset \mathbb{R}^3$ in space ***z*-regular** if (see Figure 5.31)

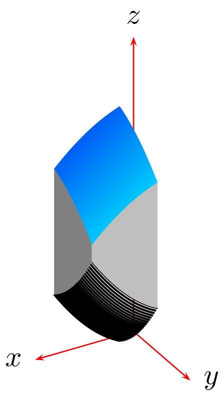


Figure 5.31: A z-regular region $\mathfrak{D}$

- a line parallel to the z-axis (*i.e.*, the set $\ell(x,y)$ defined by fixing the x- and y-coordinates and allowing the z-coordinate to vary) intersects $\mathfrak{D}$ in an interval $[\alpha(x,y)\,,\beta(x,y)]$:
$$\mathfrak{D}\cap\ell(x,y)=\{(x,y,z)\,|\,z\in[\alpha(x,y)\,,\beta(x,y)]\,,\ x,y\text{ fixed}\}$$
- The set of pairs (x,y) for which $\ell(x,y)$ intersects $\mathfrak{D}$ forms an elementary region $\mathcal{D}$ in the x,y-plane.
- the endpoints $\alpha(x,y)$ and $\beta(x,y)$ are continuous functions of $(x,y)\in\mathcal{D}$.

If in turn the region $\mathcal{D}$ is y-regular, then we can specify $\mathfrak{D}$ via three inequalities of the form

$$\begin{cases}\alpha(x,y) & \leq z\leq & \beta(x,y)\\ c(x) & \leq y\leq & d(x)\\ a & \leq x\leq & b,\end{cases}\tag{5.20}$$

while if $\mathcal{D}$ is x-regular, we can specify it via

$$\begin{cases}\alpha(x,y) & \leq z\leq & \beta(x,y)\\ a(y) & \leq x\leq & b(y)\\ c & \leq y\leq & d.\end{cases}\tag{5.21}$$

Note the pattern here: the limits in the first inequality (for z) are functions of x and y, the limits in the second inequality (for y, respectively x) are functions of x (*resp.* y), and the limits in the third inequality (for x, respectively y) are just (constant) numbers. Analogous definitions can be formulated for x-regular or y-regular regions in $\mathbb{R}^3$ (Exercise 7).

When the region $\mathfrak{D}$ is z-regular in the sense of the definition above, and f is integrable over $\mathfrak{D}$, then the integral can be calculated in terms of the partial integral

$$\int_{\alpha(x,y)}^{\beta(x,y)} f(x,y,z)\;dz$$

in which x and y (so also the limits of integration $\alpha(x,y)$ and $\beta(x,y)$) are treated as constant, as far as the integration is concerned; this results in a function of x and y (defined over $\mathcal{D}$) and the full triple integral is the (double) integral of this function over $\mathcal{D}$. Thus, from the specification (5.20) we obtain the triple integral

$$\begin{aligned}\iiint_{[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]} f\,dV &= \iint_{\mathcal{D}} \left(\int_{\alpha(x,y)}^{\beta(x,y)} f(x,y,z)\;dz\right)\,dA \\ &= \int_a^b \int_{c(x)}^{d(x)} \int_{\alpha(x,y)}^{\beta(x,y)} f(x,y,z)\;dz\,dy\,dx\end{aligned}$$

while from (5.21) we obtain

$$\begin{aligned}\iiint_{[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]} f\,dV &= \iint_{\mathcal{D}} \left(\int_{\alpha(x,y)}^{\beta(x,y)} f(x,y,z)\;dz\right)\,dA \\ &= \int_c^d \int_{a(y)}^{b(y)} \int_{\alpha(x,y)}^{\beta(x,y)} f(x,y,z)\;dz\,dx\,dy.\end{aligned}$$

As a first example, let us find the integral of $f(x,y,z) = 3x^2 - 3y^2 + 2z$ over the rectangular solid $\mathfrak{D} = [1,3] \times [1,2] \times [0,1]$ shown in Figure 5.32.

The region is specified by the inequalities

$$\begin{aligned}0 \leq& z \leq 1\\ 1 \leq& y \leq 2\\ 1 \leq& x \leq 3\end{aligned}$$

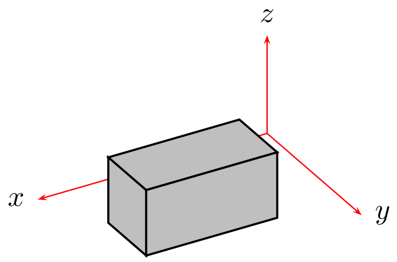


Figure 5.32: The rectangular region $\mathfrak{D} = [1,3] \times [1,2] \times [0,1]$

yielding the triple integral

$$\begin{aligned}\iiint_{[1,3]\times[1,2]\times[0,1]} 3(x^2 - 3y^2 + 2z)\,dV &= \int_1^3 \int_1^2 \int_0^1 (3x^2 - 2y^2 + 2z)\,dz\,dy\,dx \\ &= \int_1^3 \int_1^2 (3x^2 z + 3y^2 z + z^2)_{z=0}^{z=1}\,dy\,dx \\ &= \int_1^3 \int_1^2 (3x^2 + 3y^2 + 1)\,dy\,dx \\ &= \int_1^3 (3x^2 y + y^3 + y)_{y=1}^2\,dx \\ &= \int_1^3 (\{6x^2 + 8 + 2\} - \{3x^2 + 1 + 1\}]\,dx \\ &= \int_1^3 (3x^2 + 8)\,dx \\ &= (x^3 + 8x)_1^3 \\ &= (27 + 24) - (1 + 8) \\ &= 42.\end{aligned}$$

As a second example, let us integrate the function $f(x,y,z) = x + y + 2z$ over the region $\mathfrak{D}$ bounded by the xz-plane, the yz-plane, the plane $z = x + y$, and the plane $z = 2$ (Figure 5.33).

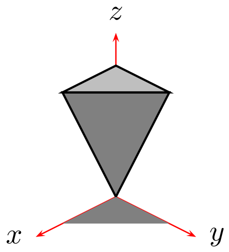


Figure 5.33: The region $\mathfrak{D}$ (and its shadow, $\mathcal{D}$)

The "shadow" of $\mathfrak{D}$, that is, its projection onto the xy-plane, is the triangular region $\mathcal{D}$ determined by the inequalities

$$\begin{aligned} 0 \leq & y \leq 1 \\ 0 \leq & x \leq 1 \\ 0 \leq x & + y \leq 1 \end{aligned}$$

which is a y-regular region; the corresponding specification is

$$\begin{aligned} 0 \leq & y \leq 1 - x \\ 0 \leq & x \leq 1. \end{aligned}$$

A vertical line intersects the three-dimensional region $\mathfrak{D}$ if and only if it goes through this shadow, and then it runs from $z = x + y$ to $z = 2$. Thus, $\mathfrak{D}$ is z-regular, with corresponding inequalities

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

leading to the triple integral

$$\begin{aligned}\iiint_{\mathfrak{D}} (x+y+z)\,dV &= \int_0^1 \int_0^{1-x} \int_{x+y} 2(x+y+2z)\,dz\,dy\,dx \\ &= \int_0^1 \int_0^{1-x} \left\{(x+y)z + z^2\right\}_{z=x+y}^{z=2}\,dy\,dx \\ &= \int_0^1 \int_0^{1-x} \left\{[2(x+y)+4] - \left[(x+y)^2 + (x+y)^2\right]\right\}\,dy\,dx \\ &= \int_0^1 \int_0^{1-x} \left\{2(x+y)^2 - 2(x+y) + 4\right\}\,dy\,dx;\end{aligned}$$

the inner integral is best done using the substitution

$$\begin{aligned}u &= x+y \\ du &= dy \\ y = 0 &\leftrightarrow u = x \\ y = 1-x &\leftrightarrow u = 1\end{aligned}$$

leading to the inner integral

$$\begin{aligned}\int_0^{1-x} \left\{2(x+y)^2 - 2(x+y) + 4\right\}\,dy &= \int_x^1 \left\{2u - 2u^2 + 4\right\}\,du \\ &= \left\{u^2 - \frac{2u^3}{3} + 4u\right\}_{u=x}^1 \\ &= \left\{1 - \frac{2}{3} + 4\right\} - \left\{x^2 - \frac{2x^3}{3} + 4x\right\} \\ &= \frac{13}{3} - x^2 + \frac{2x^3}{3} - 4x;\end{aligned}$$

substituting this into the outer integral yields

$$\begin{aligned}\int_0^1 \int_0^{1-x} \left\{2(x+y)^2 - 2(x+y) + 4\right\}\,dy\,dx &= \int_0^3 \left(\frac{13}{3} - x^2 + \frac{2x^3}{3} - 4x\right)\,dx \\ &= \left(\frac{13}{3}x - \frac{x^3}{3} + \frac{x^4}{6} - 2x^2\right)_0^1 \\ &= \frac{23}{6}.\end{aligned}$$

As a final example, let us integrate $f(x,y,z) = x+y+1$ over the region $\mathfrak{D}$ bounded below by the surface $z = x^2 + 3y^2$ and above by the surface $z = 8 - x^2 - y^2$ (Figure 5.34).

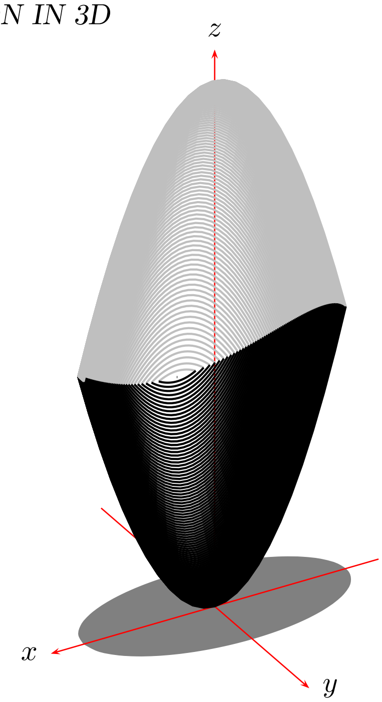


Figure 5.34: The region $\mathfrak{D}$

The two surfaces intersect where

$$8 - x^2 - 5y^2 = x^2 + 3y^2$$

or

$$x^2 + 4y^2 = 4.$$

This defines the shadow $\mathcal{D}$. This can be specified in the y-regular form

$$\begin{aligned} -\frac{1}{4}\sqrt{4 - x^2} \leq & y \leq \frac{1}{4}\sqrt{4 - x^2} \\ -2 \leq & x \leq 2 \end{aligned}$$

or in the x-regular form

$$\begin{aligned} -\sqrt{4 - 4y^2} \leq & x \leq \sqrt{4 - 4y^2} \\ -1 \leq & y \leq 1. \end{aligned}$$

We choose the latter, so our integral becomes

$$\begin{aligned}\iiint_{\mathfrak{D}} f\,dV &= \int_{-1}^{1}\int_{-\sqrt{4-4y^2}}^{\sqrt{4-4y^2}}\int_{x^2+3y^2}^{8-x^2-5y^2(x+y+1)} dz\,dx\,dy\\
&= \int_{-1}^{1}\int_{-\sqrt{4-4y^2}}^{\sqrt{4-4y^2}}(x+y++1)z\Big|_{z=x^2+3y^2}^{z=8-x^2-5y^2} dx\,dy\\
&= \int_{-1}^{1}\int_{-\sqrt{4-4y^2}}^{\sqrt{4-4y^2}}(x+y+1)(8-2x^2-8y^2)\,dx\,dy\\
&= \int_{-1}^{1}\int_{-\sqrt{4-4y^2}}^{\sqrt{4-4y^2}}[-2x^3-(2y+1)x^2+8(1-y^2)x+8(1+y-y^2-y^3)]\,dx\,dy\\
&= \int_{-1}^{1}\int_{-\sqrt{4-4y^2}}^{\sqrt{4-4y^2}}\left[-\frac{x^4}{2}-\frac{(2y+1)x^3}{3}+4(1-y^2)x^2\right.\\
&\qquad\left.+8(1+y-y^2-y^3)\right]_{x=-\sqrt{4-4y^2}}^{x=\sqrt{4-4y^2}} dy\\
&= \int_{-1}^{1}\left[-\frac{2(2y+1)}{3}(4-4y^2)^{3/2}+16(1+y-y^2-y^3)\sqrt{4-4y^2}\right] dy\\
&= \int_{-1}^{1}\left[-\frac{4}{3}(2y+1)(4-4y^2)+16(1+y-y^2-y^3)\right]\sqrt{4-4y^2}\,dy\\
&= \int_{-1}^{1}\left[\frac{32}{3}(1-y^2)+\frac{16}{3}y(1-y^2)\right]\sqrt{4-4y^2}\,dy\\
&= \frac{16\sqrt{2}}{3}\int_{-1}^{1}(2+y)(1-y^2)^{3/2}\,dy.\end{aligned}$$

Using the substitution

$$\begin{aligned}x &= \sin\theta \quad (\theta = \arcsin x)\\
dx &= \cos\theta\,d\theta\\
x = -1 &\leftrightarrow \theta = -\frac{\pi}{2}\\
x = 1 &\leftrightarrow \theta = \frac{\pi}{2}\end{aligned}$$

leads to the integral

$$\frac{16\sqrt{2}}{3}\int_{-\pi/2}^{\pi/2}(2+\sin\theta)(\cos^4\theta)\,d\theta = \frac{32\sqrt{2}}{3}\int_{-\pi/2}^{\pi/2}2\cos^4\theta\,d\theta+\frac{16\sqrt{2}}{3}\int_{-\pi/2}^{\pi/2}\cos^4\theta\sin\theta\,d\theta.$$

The first of these two integrals is done via the half-angle identities

$$\begin{aligned}\frac{32\sqrt{2}}{3}\int_{-\pi/2}^{\pi/2} 2\cos^4\theta\,d\theta &= \frac{8}{3}\sqrt{2}int_{-\pi/2}^{\pi/2}(1+\cos 2\theta)^2\,d\theta\\ &= \frac{8}{3}\sqrt{2}int_{-\pi/2}^{\pi/2}(1+2\cos 2\theta+\frac{1}{2}(1+\cos 4\theta))\,d\theta\\ &= 4\sqrt{2}\theta\Big|_{-\pi/2}^{\pi/2}\\ &= 4\pi\sqrt{2}.\end{aligned}$$

The second integral is an easy substitution of the form $u = \cos\theta$, yielding

$$\begin{aligned}\frac{16\sqrt{2}}{3}\int_{-\pi/2}^{\pi/2}\cos^4\theta\sin\theta\,d\theta &= \frac{16}{15}\sqrt{2}\cos^5\theta\Big|_{-\pi/2}^{\pi/2}\\ &= 0.\end{aligned}$$

Combining these, we have the full integral

$$\begin{aligned}\int_{-1}^{1}\int_{-\sqrt{4-4y^2}}^{\sqrt{4-4y^2}}\int_{x^2+3y^2}^{8-x^2-5y^2(x+y+1)} &dz\,dx\,dy\\ &= \frac{32\sqrt{2}}{3}\int_{-\pi/2}^{\pi/2}2\cos^4\theta\,d\theta+\frac{16\sqrt{2}}{3}\int_{-\pi/2}^{\pi/2}\cos^4\theta\sin\theta\,d\theta\\ &= 4\pi\sqrt{2}+0\\ &= 4\pi\sqrt{2}.\end{aligned}$$

Exercises for § 5.5

Practice problems:

1. Calculate each triple integral $\iiint_{\mathfrak{D}} f\,dV$ below:

   (a) $f(x,y,z) = x^3$, $\mathcal{D}$ is $[0,1]\times[0,1]\times[0,1]$.

   (b) $f(x,y,z) = 3x^3y^2z$, $\mathcal{D}$ is $[0,2]\times[2,3]\times[1,2]$.

   (c) $f(x,y,z) = e^{x-2y+3z}$, $\mathcal{D}$ is $[0,1]\times[0,1]\times[0,1]$.

   (d) $f(x,y,z) = 1$, $\mathcal{D}$ is the region bounded by the coordinate planes and the plane $x+y+2z=2$.

   (e) $f(x,y,z) = x+y+z$, $\mathcal{D}$ is the region bounded by the planes $x=0$, $y=0$, $z=0$, $x+y=1$, and $x+z=2-y$.

(f) $f(x,y,z)=1$, $\mathcal{D}$ is the region bounded by the two surfaces $z=24-5x^2-2y^2$ and $z=x^2+y^2$.

(g) $f(x,y,z)=1$, $\mathcal{D}$ is the region bounded by $2x^2+y^2=4$ and $x+y+2z=6$.

(h) $f(x,y,z)=x+yz$, $\mathcal{D}$ is specified by

$$\begin{aligned}0 \leq & z \leq y\\ 0 \leq & y \leq x\\ 0 \leq & x \leq 1.\end{aligned}$$

(i) $f(x,y,z)=z+2y$, $\mathcal{D}$ is the pyramid with top vertex $(0,0,1)$ and base vertices $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, and $(1,1,0)$.

2. Express each region below by inequalities of the form

$$\begin{aligned}a_1(x,y) \leq z \leq a_2(x,y)\\ b_1(x) \leq y \leq b_2(x)\\ c_1 \leq x \leq c_2.\end{aligned}$$

(a) $\mathcal{D}=\{(x,y,z)\,|\,x^2+y^2+z^2\leq 4, \quad z\geq 2\}$

(b) $\mathcal{D}=\{(x,y,z)\,|\,x^2+y^2+z^2\leq 1, \quad |x|\leq y\}$

(c) $\mathcal{D}=\{(x,y,z)\,|\,x^2+y^2+z^2\leq z\leq \sqrt{x^2+y^2+z^2}\}$

(d) $\mathcal{D}$ is the region bounded by the surfaces $z=6x^2-6y^2$ and $10x^2+10y^2+z=4$.

3. Show that the region in the first octant in which $x+y\leq 1$ and $x\leq z\leq y$ is the simplex with vertices $(0,0,0)$, $(0,1,0)$, $(0,1,1)$, and $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$. Find its volume.

4. Consider the region specified by

$$\begin{aligned}0 \leq z \leq y\\ 0 \leq y \leq x\\ 0 \leq x \leq 1.\end{aligned}$$

Give inequalities expressing the same region in the form

$$\begin{aligned}a_1(y,z) \leq x \leq a_2(y,z)\\ b_1(z) \leq y \leq b_2(z)\\ c_1 \leq z \leq c_2.\end{aligned}$$

5. Express the volume of the pyramid with base $[-1,1] \times [-1,1]$ and vertex $(0,0,1)$ in two ways:

   (a) As an iterated integral of the form $\iiint dy\,dx\,dz$

   (b) As a sum of four iterated integrals of the form $\iiint dz\,dy\,dx$.

   Then evaluate one of these expressions.

6. (a) Let $\mathcal{D}$ be the intersection of the two regions $x^2 + y^2 \leq 1$ and $x^2 + z^2 \leq 1$. Sketch the part of $\mathcal{D}$ lying in the first octant, and set up a triple integral expressing the volume of $\mathcal{D}$.

   (b) Do the same for the intersection of the *three* regions $x^2 + y^2 \leq 1$, $x^2 + z^2 \leq 1$, and $y^2 + z^2 \leq 1$. (*Hint:* First consider the part of $\mathcal{D}$ in the first octant, and in particular the two parts into which it is divided by the vertical plane $x = y$.)

Theory problems:

7. (a) Formulate a definition of x-regular and y-regular regions in $\mathbb{R}^3$ parallel to that given on p. 543 for z-regular regions.

   (b) For each of these give the possible ways such a region can be specified by inequalities.

8. **Symmetry in Three Dimensions:**

   (Refer to Exercise 7 in § 5.2.)

   (a) Formulate a definition of x-symmetric (*resp.* y-symmetric or z-symmetric) regions in $\mathbb{R}^3$.

   (b) Define what it means for a function $f(x,y,z)$ of three variables to be odd (*resp.* even) in x (*resp.* y or z). function!odd .

   (c) Show that if $f(x,y)$ is odd in x on an x-symmetric, x-regular region in $\mathbb{R}^3$, its integral is zero.

   (d) Show that if $f(x,y)$ is even in x on an x-symmetric, x-regular region in $\mathbb{R}^3$, its integral is twice its integral in the part of the region on the positive side of the yz-plane.

   (e) Suppose $f(x,y,z)$ is even in all three variables, and D is regular and symmetric in all three variables. Then the integral of f over D is a multiple of its integral over the intersection of D with the first octant: what multiple?

Challenge problem:

9. Suppose f is continuous on $\mathbb{R}^3$, and let B_δ be the ball of radius $\varepsilon > 0$ centered at (x_0, y_0, z_0), and let $\mathcal{V}(B_\delta)$ denote the volume of the ball. Show that

$$\lim_{\varepsilon \to 0} \frac{1}{\mathcal{V}(B_\delta)} \iiint_{B_\delta} f(x, y, z)\ dV = f(x_0, y_0, z_0)\,.$$

6
Integral Calculus for Vector Fields and Differential Forms

In this chapter, we will consider a family of results known collectively as **Generalized Stokes' Theorem**, which can be regarded as a far-reaching generalization of the Fundamental Theorem of Calculus. These results can be formulated in several languages; we shall follow two of these: the language of *vector fields* and the language of *differential forms*; along the way, we shall develop a dictionary for passing from one of these languages to the other.

6.1 Line Integrals of Vector Fields and 1-Forms

In Chapter 4 we considered mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$, or vector-valued functions of a vector variable. We begin here by looking at a special case of this from a different point of view.

A **vector field** on $D \subset \mathbb{R}^n$ is simply a mapping $\overrightarrow{F}\colon D \to \mathbb{R}^n$ assigning to each point $p \in D$ a vector $\overrightarrow{F}(p)$. However, our point of view is somewhat different from that in Chapter 4. We think of the domain and range of a *mapping* as essentially separate collections of vectors or points (even when they are the same space), whereas in the *vector field* setting we think of the *input* as a *point*, and the *output* as a *vector*; we picture this vector as an arrow "attached" to the point. The distinction is emphasized by our use of an arrow over the name of the vector field, and dropping the arrow over the input point.

One way to formalize this point of view is to take a leaf from our study of surfaces in space (particularly Lagrange multipliers in § 3.6). If a curve $\overrightarrow{p}(t)$ lies on the surface $\mathfrak{S}$, then its velocity is everywhere tangent to the surface; turning this around, we can think of the tangent plane to $\mathfrak{S}$ at $p \in \mathfrak{S}$ as consisting of all the possible velocity vectors for points moving in $\mathfrak{S}$ through p. Analogously, we can formulate the **tangent space to $\mathbb{R}^{n}$** at $p \in \mathbb{R}^n$ as the set $T_p\mathbb{R}^n$ of all velocity vectors for points moving in $\mathbb{R}^n$ through p. This is of course a copy of $\mathbb{R}^n$, but we think of these vectors as all "attached" to p. Examples of physical quantities for which this interpretation is appropriate include forces which vary from point to point (such as interplanetary gravitation), velocity of fluids (such as wind velocity on weather maps), and forces acting on rigid bodies.

We can visualize vector fields in the plane and in 3-space as, literally, "fields of arrows". For example, the vector field in the plane given by

$$\overrightarrow{F}(x,y) = y\overrightarrow{\imath} + x\overrightarrow{\jmath}$$

assigns to every point $(x,0)$ on the x-axis a vertical arrow of length $|x|$ (pointing up for $x > 0$ and down for $x < 0$) and similarly a horizontal arrow of length $|y|$ to every point $(0,y)$ on the y-axis; at a generic point (x,y), the arrow is the sum of these. The resulting field is pictured in Figure 6.1. Note that when $y = \pm x$, the vector points along the diagonal (or antidiagonal).

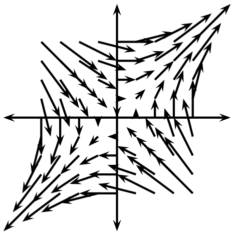


Figure 6.1: The vector field $\vec{F}(x,y) = y\vec{\imath} + x\vec{\jmath}$

By contrast, the vector field

$$\overrightarrow{F}(x,y) = y\overrightarrow{\imath} - x\overrightarrow{\jmath}$$

is everywhere perpendicular to the position vector (x,y), so $\overrightarrow{F}(x,y)$ is tangent to the circle through (x,y) centered at the origin (Figure 6.2).

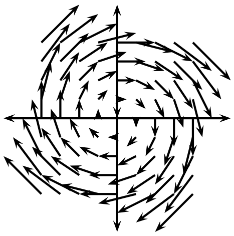


Figure 6.2: The vector field $\vec{F}(x,y) = y\vec{\imath} + x\vec{\jmath}$

Work and Line Integrals

If you have to push your stalled car a certain distance, the work you do is intuitively proportional to how hard you need to push, and also to how far you have to push it. This intuition is formalized in the physics concept of **work**: if a (constant) force of magnitude F is applied to move an object over a straight line distance $\triangle\mathfrak{s}$, then the work W is given by

$$W = F\triangle\mathfrak{s};$$

more generally, if the force is not directed parallel to the direction of motion, we write the force and the displacement as vectors $\overrightarrow{F}$ and $\triangle\overrightarrow{\mathfrak{s}}$, respectively, and consider only the component of the force in the direction of the displacement:

$$\begin{aligned} W &= \left(\operatorname{comp}_{\triangle\overrightarrow{\mathfrak{s}}}\overrightarrow{F}\right)\triangle\mathfrak{s} \\ &= \overrightarrow{F}\cdot\triangle\overrightarrow{\mathfrak{s}}. \end{aligned}$$

When the displacement occurs over a curved path $\mathcal{C}$, and the force varies along the path, then we need to go through a process of integration. We pick partition points p_j, $j = 0, \ldots, n$, along the curve and make two approximations. First, since the force should vary very little along a short piece of curve, we replace the varying force by its value $\overrightarrow{F}(x_j)$ at some representative point x_j between p_{j-1} and p_j along $\mathcal{C}$. Second, we use the vector $\triangle\overrightarrow{\mathfrak{s}_j} = \overrightarrow{p}_j - \overrightarrow{p}_{j-1}$ as the displacement. Thus, the work done along one piece is approximated by the quantity

$$\triangle_j W = \overrightarrow{F}(x_j)\cdot\triangle\overrightarrow{\mathfrak{s}_j}$$

and the total work over $\mathcal{C}$ is approximated by the sum

$$\begin{aligned} W &\approx \sum_{j=1}^{n} \triangle_j W \\ &= \sum_{j=1}^{n} \vec{F}(x_j) \cdot \triangle \vec{\mathfrak{s}_j}. \end{aligned}$$

As usual, we consider progressively finer partitions of $\mathcal{C}$, and expect the approximations to converge to an integral

$$W = \int_{\mathcal{C}} \vec{F} \cdot d\vec{\mathfrak{s}}.$$

This might look like a new kind of integral, but we can see it as a path integral of a function over $\mathcal{C}$, as in § 2.5. For this, it is best to think in terms of a parametrization of $\mathcal{C}$, say $\vec{p}(t)$, $a \leq t \leq b$. We can write

$$p_j = \vec{p}(t_j).$$

Then the vector $\triangle\vec{\mathfrak{s}_j}$ is approximated by the vector $\vec{v}(t_j)\,\triangle t_j$ where

$$\triangle t_j = t_j - t_{j-1}$$

and

$$\vec{v}(t) = \frac{d\vec{p}}{dt}$$

is the velocity of the parametrization. As in § 2.5, we can write

$$\vec{v}(t) = \|\vec{v}(t)\|\,\vec{T}(t)$$

where $\vec{T}$ is a unit vector tangent to $\mathcal{C}$ at $\vec{p}(t)$. Thus, we can write

$$\triangle\vec{\mathfrak{s}_j} \approx \|\vec{v}(t)\|\,\vec{T}(t_j)\,\triangle t_j$$

and the integral for work can be rewritten

$$W = \int_{\mathcal{C}} \vec{F} \cdot d\vec{\mathfrak{s}} = \int_a^b \vec{F} \cdot \vec{T}\,\|\vec{v}(t)\|\,dt$$

which we can recognize as a line integral

$$W = \int_{\mathcal{C}} \overrightarrow{F} \cdot \overrightarrow{T}\, d\mathfrak{s}$$

of the function given by the tangential component of $\overrightarrow{F}$, that is

$$W = \int_{\mathcal{C}} f\, d\mathfrak{s}$$

where

$$\begin{aligned} f(\overrightarrow{p}(t)) &= \overrightarrow{F}(\overrightarrow{p}(t)) \cdot \overrightarrow{T}(\overrightarrow{p}(t)) \\ &= \operatorname{comp}_{\overrightarrow{v}(t)} \overrightarrow{F}(\overrightarrow{p}(t)) \,. \end{aligned}$$

Let us work this out for an example. Suppose our force is given by the planar vector field

$$\overrightarrow{F}(x, y) = \overrightarrow{\imath} + y\,\overrightarrow{\jmath}$$

and $\mathcal{C}$ is the semicircle $y = \sqrt{1 - x^2}$, $-1 \leq x \leq 1$. We can write

$$\overrightarrow{p}(t) = t\,\overrightarrow{\imath} + \sqrt{1 - t^2}\,\overrightarrow{\jmath},$$

or equivalently,

$$\begin{aligned} x &= t \\ y &= \sqrt{1 - t^2}, \\ -1 &\leq t \leq 1. \end{aligned}$$

Then

$$\begin{aligned} \frac{dx}{dt} &= 1 \\ \frac{dy}{dt} &= -\frac{t}{\sqrt{1 - t^2}} \end{aligned}$$

or equivalently

$$\overrightarrow{v}(t) = \overrightarrow{\imath} - \frac{t}{\sqrt{1 - t^2}}\,\overrightarrow{\jmath}$$

and

$$\begin{aligned}\|\vec{v}(t)\| &= \sqrt{1+\frac{t^2}{1-t^2}}\\ &= \frac{1}{\sqrt{1-t^2}}\end{aligned}$$

so

$$\begin{aligned}\vec{T} = \frac{\vec{v}}{\|\vec{v}\|} &= (\sqrt{1-t^2})\left(\vec{\imath} - \frac{t}{\sqrt{1-t^2}}\vec{\jmath}\right)\\ &= \sqrt{1-t^2}\,\vec{\imath} - t\,\vec{\jmath}.\end{aligned}$$

The value of the vector field along the curve is

$$\begin{aligned}\vec{F}(t) &= \vec{F}\left(t, \sqrt{1-t^2}\right)\\ &= \vec{\imath} + \sqrt{1-t^2}\,\vec{\jmath}\end{aligned}$$

so the function we are integrating is

$$\begin{aligned}f(t) &= \vec{F}\cdot\vec{T}\\ &= \sqrt{1-t^2} - t\sqrt{1-t^2}\\ &= (1-t)\sqrt{1-t^2};\end{aligned}$$

meanwhile,

$$\begin{aligned}d\mathfrak{s} &= \|\vec{v}\|\,dt\\ &= \frac{1}{\sqrt{1-t^2}}\,dt\end{aligned}$$

and our integral becomes

$$\begin{aligned}\int_{\mathcal{C}} \vec{F}\cdot\vec{T}\,d\mathfrak{s} &= \int_{-1}^{1} [(1-t)\sqrt{1-t^2}][\frac{1}{\sqrt{1-t^2}}\,dt]\\ &= \int_{-1}^{1} (1-t)\,dt\\ &= -\frac{(1-t)^2}{2}\Big|_{-1}^{1}\\ &= -\frac{(0)^2}{2} + \frac{(2)^2}{2}\\ &= 2.\end{aligned}$$

In the calculation above, you undoubtedly noticed that the factor $\|\overrightarrow{v}\| = \sqrt{1-t^2}$, which appeared in the numerator when calculating the unit tangent, also appeared in the denominator when calculating the differential of arclength, so they cancelled. A moment's thought should convince you that this is always the case: formally,

$$\overrightarrow{T} = \frac{\overrightarrow{v}}{\|\overrightarrow{v}\|}$$

and

$$d\mathfrak{s} = \|\overrightarrow{v}\| \; dt$$

means that

$$\begin{aligned} \overrightarrow{T} \, d\mathfrak{s} &= \left(\frac{\overrightarrow{v}}{\|\overrightarrow{v}\|}\right)(\|\overrightarrow{v}\| \; dt) \\ &= \overrightarrow{v} \, d\mathfrak{s} \end{aligned}$$

so

$$\begin{aligned} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} &= \overrightarrow{F} \cdot \overrightarrow{T} \, d\mathfrak{s} \\ &= \overrightarrow{F} \cdot (\overrightarrow{v} \, d\mathfrak{s}) \, ; \end{aligned}$$

in other words, we can write, formally,

$$\begin{aligned} d\overrightarrow{\mathfrak{s}} &= \overrightarrow{v} \, dt \\ &= \left(\frac{dx}{dt}\overrightarrow{\imath} + \frac{dy}{dt}\overrightarrow{\jmath}\right) \, dt. \end{aligned}$$

If we allow ourselves the indulgence of formal differentials, we can use the relations

$$\begin{aligned} dx &= \frac{dx}{dt} \, dt \\ dy &= \frac{dy}{dt} \, dt \end{aligned}$$

to write

$$d\overrightarrow{\mathfrak{s}} = dx \overrightarrow{\imath} + dy \overrightarrow{\jmath}.$$

Now, if the vector field $\overrightarrow{F}$ is given by

$$\overrightarrow{F}(x,y) = P(x,y)\,\overrightarrow{\imath} + Q(x,y)\,\overrightarrow{\jmath}$$

then (again formally)

$$\overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} = P(x,y)\ dx + Q(x,y)\ dy$$

leading us to the formal integral

$$\int_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} = \int_{\mathcal{C}} P(x,y)\ dx + Q(x,y)\ dy.$$

While the geometric interpretation of this is quite murky at the moment, this way of writing things leads, via the rules of formal integrals, to a streamlined way of calculating our integral. Let us apply it to the example considered earlier.

The vector field

$$\overrightarrow{F}(x,y) = \overrightarrow{\imath} + y\overrightarrow{\jmath}$$

has components

$$\begin{aligned} P(x,y) &= 1 \\ Q(x,y) &= y, \end{aligned}$$

so our integral can be written formally as

$$\int_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} = \int_{\mathcal{C}} (\,dx + y\,dy).$$

Using the parametrization from before

$$\begin{aligned} x &= t \\ y &= \sqrt{1-t^2} \\ -1 \leq t &\leq 1 \end{aligned}$$

we use the rules of formal differentials to write

$$\begin{aligned} dx &= \frac{dx}{dt}\,dt \\ &= dt \\ dy &= \frac{dy}{dt}\,dt \\ &= -\frac{t}{\sqrt{1-t^2}}\,dt \end{aligned}$$

so

$$\begin{aligned} P\,dx + Q\,dy &= \,dx + y\,dy \\ &= (1)(\,dt) + (\sqrt{1-t^2})\left(-\frac{t}{\sqrt{1-t^2}}\,dt\right) \\ &= (1-t)\,dt \end{aligned}$$

and the integral becomes

$$\begin{aligned} \int_{\mathcal{C}} P\,dx + Q\,dy &= \int_{\mathcal{C}} \,dx + y\,dy \\ &= \int_{-1}^{1} (1-t)\,dt \\ &= 2 \end{aligned}$$

as before.

But there is another natural parametrization of the upper half-circle:

$$\begin{aligned} x &= \cos\theta \\ y &= \sin\theta \\ 0 &\leq \theta \leq \pi. \end{aligned}$$

This leads to the differentials

$$\begin{aligned} dx &= -\sin\theta\,d\theta \\ dy &= \cos\theta\,d\theta. \end{aligned}$$

The components of the vector field, expressed in terms of our parametrization, are

$$\begin{aligned} P &= 1 \\ Q &= \sin\theta \end{aligned}$$

so

$$\begin{aligned} P\,dx + Q\,dy &= (-\sin\theta)(\,d\theta) + (\sin\theta)(\cos\theta\,d\theta) \\ &= (-\sin\theta + \sin\theta\cos\theta)\,d\theta \end{aligned}$$

and our integral becomes

$$\begin{aligned}\int_{\mathcal{C}} P\,dx + Q\,dy &= \int_0^{\pi} (-\sin\theta + \sin\theta\cos\theta)\,d\theta \\ &= \left(\cos\theta + \frac{\sin^2\theta}{2}\right)_0^{\pi} \\ &= (-1+0) - (1+0) \\ &= -2.\end{aligned}$$

Note that this has the opposite sign from our previous calculation. Why?

The answer becomes clear if we think in terms of the expression for the work integral

$$W = \int_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} = \int_{\mathcal{C}} \overrightarrow{F} \cdot \overrightarrow{T}\, d\mathfrak{s}.$$

Clearly, the vector field $\overrightarrow{F}$ does not change when we switch parametrizations for $\mathcal{C}$. However, our first parametrization (treating $\mathcal{C}$ as the graph of the function $y = \sqrt{1-t^2}$) traverses the semicircle *clockwise*, while the second one traverses it *counterclockwise*. This means that the unit tangent vector $\overrightarrow{T}$ determined by the first parametrization is the negative of the one coming from the second, so the two parametrizations yield path integrals of functions that differ in sign. Thus, even though the path integral of a *scalar-valued function* $\int_{\mathcal{C}} f\,d\mathfrak{s}$ depends only on the geometric curve $\mathcal{C}$ and not on how we parametrize it, the work integral $\int_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}}$ depends also on the direction in which we move along the curve: in other words, it depends on the *oriented* curve given by $\mathcal{C}$ together with the direction along it—which determines a choice between the two unit tangents at each point of $\mathcal{C}$. To underline this distinction, we shall refer to *path* integrals of (scalar-valued) *functions*, but *line* integrals of *vector fields*.

Definition 6.1.1.

1. *An **orientation** of a curve $\mathcal{C}$ is a continuous unit vector field $\overrightarrow{T}$ defined on $\mathcal{C}$ and tangent to $\mathcal{C}$ at every point. Each regular curve has two distinct orientations.*

2. *An **oriented curve**[1] is a curve $\mathcal{C}$ together with a choice of orientation $\overrightarrow{T}$ of $\mathcal{C}$.*

3. *The **line integral** of a vector field $\overrightarrow{F}$ defined on $\mathcal{C}$ over the oriented*

[1] This is also sometimes called a **directed curve**.

curve determined by the unit tangent field $\overrightarrow{T}$ is the work integral

$$\int_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} = \int_{\mathcal{C}} \overrightarrow{F} \cdot \overrightarrow{T}\, d\mathfrak{s}.$$

Since the function $\overrightarrow{F} \cdot \overrightarrow{T}$ determined by a vector field along an oriented curve is the same for all parametrizations yielding the orientation $\overrightarrow{T}$, we have the following invariance principle.

Remark 6.1.2. *The line integral of a vector field $\overrightarrow{F}$ over an oriented curve is the same for any parametrization whose velocity points in the same direction as the unit tangent field $\overrightarrow{T}$ determined by the orientation. Switching orientation switches the sign of the line integral.*

Differential Forms

So far, we have treated expressions like dx as purely formal expressions, sometimes mysteriously related to each other by relations like $dy = y'\,dx$. An exception has been the notation df for the derivative of a real-valued function $f\colon \mathbb{R}^n \to \mathbb{R}$ on $\mathbb{R}^n$. This exception will be the starting point of a set of ideas which makes sense of other expressions of this sort.

Recall that the derivative d_pf of $f\colon \mathbb{R}^n \to \mathbb{R}$ at a point p in its domain is itself a linear function—that is, it respects linear combinations:

$$d_pf(a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2) = a_1 d_pf(\overrightarrow{v}_1) + a_2 d_pf(\overrightarrow{v}_2)\,.$$

Furthermore, if we consider the way it is used, this linear function is applied only to velocity vectors of curves as they pass through the point p. In other words, we should think of the derivative as a linear function $d_pf\colon T_p\mathbb{R}^n \to \mathbb{R}$ acting on the tangent space to $\mathbb{R}^n$ at p. To keep straight the distinction between the underlying function f, which acts on $\mathbb{R}^n$, and its derivative at p, which acts on the tangent space $T_p\mathbb{R}^n$, we refer to the latter as a **linear functional** on $T_p\mathbb{R}^n$. Now, as we vary the basepoint p, the derivative gives us different linear functionals, acting on different tangent spaces. We abstract this notion in

Definition 6.1.3. *A* ***differential form*** *on $\mathbb{R}^n$ is a rule ω assigning to each point $p \in \mathbb{R}^n$ a linear functional $\omega_p\colon T_p\mathbb{R}^n \to \mathbb{R}$ on the tangent space to $\mathbb{R}^n$ at p.*

We will in the future often deal with differential forms defined only at points in a subregion $D \subset \mathbb{R}^n$, in which case we will refer to a differential form on D.

Derivatives of functions aside, what do other differential forms look like?

Let us consider the case $n = 2$. We know that a linear functional on $\mathbb{R}^2$ is just a homogeneous polynomial of degree 1; since the functional can vary from basepoint to basepoint, the coefficients of this polynomial are actually functions of the basepoint. To keep the distinction between $\mathbb{R}^2$ and $T_p\mathbb{R}^2$, we will denote points in $\mathbb{R}^2$ by $p = (x, y)$ and vectors in $T_p\mathbb{R}^2$ by $\overrightarrow{v} = (v_1, v_2)$; then a typical form acts on a tangent vector $\overrightarrow{v}$ at p via

$$\omega_p(\overrightarrow{v}) = P(x, y)\, v_1 + Q(x, y)\, v_2.$$

To complete the connection between formal differentials and differential forms, we notice that the first term on the right above is a multiple (by the scalar P, which depends on the basepoint) of the component of $\overrightarrow{v}$ parallel to the x-axis. This component is a linear functional on $T_p\mathbb{R}^n$, which we can think of as the derivative of the function on $\mathbb{R}^2$ that assigns to a point p its x-coordinate; we denote it[2] by $\boldsymbol{dx}$. Similarly, the linear functional on $T_p\mathbb{R}^2$ assigning to each tangent vector its y-component is denoted $\boldsymbol{dy}$. We call these the **coordinate forms**:

$$\begin{aligned} dx(\overrightarrow{v}) &= v_1 \\ dy(\overrightarrow{v}) &= v_2. \end{aligned}$$

Then, using this notation, we can write any form on $\mathbb{R}^2$ as

$$\omega = P\, dx + Q\, dy.$$

(Of course, it is understood that ω, P and Q all depend on the basepoint p at which they are applied.)

Using this language, we can systematize our procedure for finding work integrals using forms. Given a curve $\mathcal{C}$ parametrized by

$$\overrightarrow{p}(t) = x(t)\, \overrightarrow{\imath} + y(t)\, \overrightarrow{\jmath}, \quad t_0 \leq t \leq t_1$$

and a form defined along $\mathcal{C}$

$$\omega = P\, dx + Q\, dy$$

[2] Strictly speaking, we should include a subscript indicating the basepoint p, but since the action on any tangent space is effectively the same, we suppress it.

we apply the form to the velocity vector $\vec{p}'(t) = (x'(t), y'(t))$ of the parametrization. The result can be expressed as a function of the parameter alone

$$\begin{aligned} w(t) &= \omega_{\overrightarrow{p}(t)}(\vec{p}'(t)) \\ &= P(x(t), y(t))\, x'(t) + Q(x(t), y(t))\, y'(t); \end{aligned}$$

we then integrate this over the domain of the parametrization:

$$\begin{aligned} \int_{\mathcal{C}} \omega &= \int_{t_0}^{t_1} \left(\omega_{\overrightarrow{p}(t)}(\vec{p}'(t))\right) dt \\ &= \int_{t_0}^{t_1} \left[P(x(t), y(t))\, x'(t) + Q(x(t), y(t))\, y'(t)\right] dt. \end{aligned} \tag{6.1}$$

The expression appearing inside either of the two integrals itself looks like a form, but now it "lives" on the real line. In fact, we can also regard it as a coordinate form on $\mathbb{R}^1$ in the sense of Definition 6.1.3, using the convention that dt acts on a velocity along the line (which is now simply a real number) by returning the number itself. At this stage—when we have a form on $\mathbb{R}$ rather than on a curve in $\mathbb{R}^2$—we simply interpret our integral in the normal way, as the integral of a function over an interval.

However, the interpretation of this expression as a form can still play a role, when we compare different parametrizations of the same curve. We will refer to the form on parameter space obtained from a parametrization of a curve by the process above as the **pullback** of ω by $\overrightarrow{p}$:

$$[\overrightarrow{p}^*(\omega)]_t = \omega_{\overrightarrow{p}(t)}(\vec{p}'(t))\ dt. \tag{6.2}$$

Then we can summarize our process of integrating a form along a curve by saying *the integral of a form ω along a parametrized curve is the integral, over the domain in parameter space, of the pullback $\overrightarrow{p}^*(\omega)$ of the form by the parametrization.*

Suppose now that $\overrightarrow{q}(s)$, $s_0 \leq s \leq s_1$ is a reparametrization of the same curve. By definition, this means that there is a continuous, strictly monotone function $\mathfrak{t}(s)$ such that

$$\overrightarrow{q}(s) = \overrightarrow{p}(\mathfrak{t}(s)).$$

In dealing with regular curves, we assume that $\mathfrak{t}(s)$ is differentiable, with non-vanishing derivative. We shall call this an **orientation-preserving reparametrization** if $\frac{d\mathfrak{t}}{ds}$ is positive at every point, and **orientation-reversing** if $\frac{d\mathfrak{t}}{ds}$ is negative.

Suppose first that our reparametrization is order-preserving. To integrate ω over our curve using $\overrightarrow{q}(s)$ instead of $\overrightarrow{p}(t)$, we take the pullback of ω by $\overrightarrow{q}$,

$$[\overrightarrow{q}^*(\omega)]_s = \omega_{\overrightarrow{q}(s)}\big(\vec{q}'(s)\big)\ ds.$$

By the Chain Rule, setting $t = \mathsf{t}(s)$,

$$\begin{aligned}
\vec{q}'(s) &= \frac{d}{ds}\,[\overrightarrow{q}(s)] \\
&= \frac{d}{ds}\,[\overrightarrow{p}(\mathsf{t}(s))] \\
&= \frac{d}{dt}\,[\overrightarrow{p}(\mathsf{t}(s))]\,\frac{dt}{ds} \\
&= \vec{p}'(\mathsf{t}(s))\,\mathsf{t}'(s)\ ds.
\end{aligned}$$

Now if we think of the change-of-variables map $\mathsf{t}\colon\mathbb{R}\to\mathbb{R}$ as describing a point moving along the t-line, parametrized by $t = \mathsf{t}(s)$, we see that the pullback of any form $\alpha_t = P(t)\ dt$ by t is given by

$$\begin{aligned}
[\mathsf{t}^*(\alpha_t)]_s &= \alpha_{\mathsf{t}(s)}\big(\mathsf{t}'(s)\big)\ ds \\
&= P(\mathsf{t}(s))\,\mathsf{t}'(s)\ ds.
\end{aligned}$$

Applying this to

$$\alpha_t = [\overrightarrow{p}^*(\omega)]_t$$

we see that

$$\begin{aligned}
[\mathsf{t}^*(\overrightarrow{p}^*(\omega))]_s &= [\overrightarrow{p}^*(\omega)]_{\mathsf{t}(s)}\mathsf{t}'(s)\ ds \\
&= \omega_{\overrightarrow{p}(\mathsf{t}(s))}\big(\vec{p}'(\mathsf{t}(s))\big)\,\mathsf{t}'(s)\ ds \\
&= \omega_{\overrightarrow{q}(s)}\big(\vec{p}'(\mathsf{t}(s))\big)\,\mathsf{t}'(s)\ ds \\
&= \omega_{\overrightarrow{q}(s)}\big(\vec{p}'(\mathsf{t}(s))\,\mathsf{t}'(s)\big)\ ds \\
&= \omega_{\overrightarrow{q}(s)}\big(\vec{q}'(s)\big)\ ds \\
&= [\overrightarrow{q}^*(\omega)]_s;
\end{aligned}$$

in other words,

$$\overrightarrow{q}^*(\omega) = \mathsf{t}^*(\overrightarrow{p}^*(\omega))\,. \tag{6.3}$$

Clearly, the two integrals coming from pulling ω back by $\overrightarrow{p}$ and $\overrightarrow{q}$, respectively, are the same:

$$\int_{s_0}^{s_1}[\overrightarrow{q}^*(\omega)]_s = \int_{t_0}^{t_1}[\overrightarrow{p}^*(\omega)]_t.$$

In other words, the definition of $\int_{\mathcal{C}} \omega$ via Equation (6.1) yields the same quantity for a given parametrization as for any orientation-preserving reparametrization.

What changes in the above argument when $\mathfrak{t}$ has *negative* derivative? The integrand in the calculation using $\overrightarrow{q}$ is the same: we still have Equation (6.3). However, since the reparametrization is order-*reversing*, $\mathfrak{t}$ is strictly *decreasing*, which means that it interchanges the endpoints of the domain: $\mathfrak{t}(s_0) = t_1$ and $\mathfrak{t}(s_1) = t_0$. Thus,

$$\int_{t_0}^{t_1} [\overrightarrow{p}^*(\omega)]_t = \int_{s_1}^{s_0} [\overrightarrow{q}^*(\omega)]_s = -\int_{s_0}^{s_1} [\overrightarrow{q}^*(\omega)]_s :$$

the integral given by applying Equation (6.1) to $\overrightarrow{q}$ has the same *integrand*, but the limits of integration are reversed: the resulting integral is the negative of what we would have gotten had we used $\overrightarrow{p}$.

Now let us relate this back to our original formulation of work integrals in terms of vector fields. Recall from § 3.2 that a linear functional on $\mathbb{R}^n$ can be represented as taking the dot product with a fixed vector. In particular, the form

$$\omega = P\,dx + Q\,dy$$

corresponds to the vector field

$$\overrightarrow{F} = P\overrightarrow{\imath} + Q\overrightarrow{\jmath}$$

in the sense that

$$\begin{aligned} \omega_p(\overrightarrow{v}) &= P(p)\,v_1 + Q(p)\,v_2 \\ &= \overrightarrow{F}(p) \cdot \overrightarrow{v}. \end{aligned}$$

In fact, using the formal vector

$$d\overrightarrow{\mathfrak{s}} = dx\,\overrightarrow{\imath} + dy\,\overrightarrow{\jmath}$$

which can itself be thought of as a “vector-valued” form, we can write

$$\omega = \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}}.$$

Our whole discussion carries over practically *verbatim* to $\mathbb{R}^3$. A vector field $\overrightarrow{F}$ on $\mathbb{R}^3$ can be written

$$\overrightarrow{F}(x,y,z) = P(x,y,z)\,\overrightarrow{\imath} + Q(x,y,z)\,\overrightarrow{\jmath} + R(x,y,z)\,\overrightarrow{k}$$

and the corresponding form on $\mathbb{R}^3$ is

$$\begin{aligned}\omega &= \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} \\ &= P\,dx + Q\,dy + R\,dz.\end{aligned}$$

Let us see an example of how the line integral works out in this case.

The vector field

$$\overrightarrow{F}(x,y,z) = z\,\overrightarrow{\imath} - y\,\overrightarrow{\jmath} + x\,\overrightarrow{k}$$

corresponds to the form

$$\omega = \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} = z\,dx - y\,dy + x\,dz.$$

Let us integrate this over the curve given parametrically by

$$\begin{aligned}\overrightarrow{p}(t) &= \cos t\,\overrightarrow{\imath} + \sin t\,\overrightarrow{\jmath} + \sin 2t\,\overrightarrow{k}, \\ &0 \le t \le 2\pi.\end{aligned}$$

The velocity of this parametrization is given by

$$\vec{p}\,'(t) = -\sin t\,\overrightarrow{\imath} + \cos t\,\overrightarrow{\jmath} + 2\cos 2t\,\overrightarrow{k}$$

and its pullback by the form ω is

$$\begin{aligned}[\omega^*(\vec{p}\,')]_t &= \omega_{\overrightarrow{p}(t)}\big(\vec{p}\,'(t)\big)\;dt \\ &= [(\sin 2t)(-\sin t) - (\sin t)(\cos t) + (\cos t)(2\cos 2t)]\,dt \\ &= [-2\sin^2 t\cos t - \sin t\cos t + 2(1 - 2\sin^2 t)\cos t]\,dt \\ &= [-6\sin^2 t - \sin t + 2]\cos t\,dt.\end{aligned}$$

Thus,

$$\begin{aligned}
\int_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} &= \int_{\mathcal{C}} z\,dx - y\,dy + x\,dz \\
&= \int_{\mathcal{C}} \omega \\
&= \int_0^{2\pi} \omega^*(\vec{p}\,') \\
&= \int_0^{2\pi} [-6\sin^2 t - \sin t + 2]\cos t\,dt \\
&= [-3\sin^3 t - \frac{1}{2}\sin^2 t + 2t]_0^{2\pi} \\
&= 4\pi.
\end{aligned}$$

Exercises for § 6.1

Practice problems:

1. Sketch each vector field below, in the style of Figures 6.1 and 6.2.

   (a) $x\overrightarrow{\imath}$

   (b) $x\overrightarrow{\jmath}$

   (c) $y\overrightarrow{\imath} - y\overrightarrow{\jmath}$

   (d) $x\overrightarrow{\imath} + y\overrightarrow{\jmath}$

   (e) $x\overrightarrow{\imath} - y\overrightarrow{\jmath}$

   (f) $-y\overrightarrow{\imath} + x\overrightarrow{\jmath}$

2. Evaluate $\int_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}}$:

   (a) $\overrightarrow{F}(x,y) = x\overrightarrow{\imath} + y\overrightarrow{\jmath}$, $\mathcal{C}$ is the graph $y = x^2$ from $(-2,4)$ to $(1,1)$.

   (b) $\overrightarrow{F}(x,y) = y\overrightarrow{\imath} + x\overrightarrow{\jmath}$, $\mathcal{C}$ is the graph $y = x^2$ from $(1,1)$ to $(-2,4)$.

   (c) $\overrightarrow{F}(x,y) = (x+y)\overrightarrow{\imath} + (x-y)\overrightarrow{\jmath}$, $\mathcal{C}$ is given by $x = t^2$, $y = t^3$, $-1 \leq t \leq 1$.

   (d) $\overrightarrow{F}(x,y) = x^2\overrightarrow{\imath} + y^2\overrightarrow{\jmath}$, $\mathcal{C}$ is the circle $x^2 + y^2 = 1$ traversed counterclockwise.

   (e) $\overrightarrow{F}(x,y,z) = x^2\overrightarrow{\imath} + xz\overrightarrow{\jmath} - y^2\overrightarrow{k}$, $\mathcal{C}$ is given by $x = t$, $y = t^2$, $z = t^3$, $-1 \leq t \leq 1$.

   (f) $\overrightarrow{F}(x,y,z) = y\overrightarrow{\imath} - x\overrightarrow{\jmath} + ze^x\overrightarrow{k}$, $\mathcal{C}$ is the line segment from $(0,0,0)$ to $(1,1,1)$.

   (g) $\overrightarrow{F}(x,y,z) = yz\overrightarrow{\imath} + xz\overrightarrow{\jmath} + xy\overrightarrow{k}$, $\mathcal{C}$ is given by $\overrightarrow{p}(t) = (t^2, t, -t^2)$, $-1 \leq t \leq 1$.

(h) $\overrightarrow{F}(x,y,z) = yz\overrightarrow{\imath} + xz\overrightarrow{\jmath} + xy\overrightarrow{k}$, $\mathcal{C}$ is the polygonal path from $(1,-1,1)$ to $(2,1,3)$ to $(-1,0,0)$.

3. Evaluate $\int_{\mathcal{C}} P\,dx + Q\,dy$:

   (a) $P(x,y) = y$, $Q(x,y) = -x$, $\mathcal{C}$ is given by $x = \cos t$, $y = \sin t$, $0 \leq t \leq 2\pi$.

   (b) $P(x,y) = xy$, $Q(x,y) = y^2$, $\mathcal{C}$ is $y = \sqrt{1-x^2}$ from $(-1,0)$ to $(1,0)$.

   (c) $P(x,y) = xy$, $Q(x,y) = y^2$, $\mathcal{C}$ is given by $x = t^2$, $y = t$, $-1 \leq t \leq 1$

   (d) $P(x,y) = -x$, $Q(x,y) = y$, $\mathcal{C}$ is given by $\overrightarrow{p}(t) = (\cos^3 t, \sin^3 t))$, $0 \leq t \leq 2\pi$.

   (e) $P(x,y,z) = xy$, $Q(x,y,z) = xz$, $R(x,y,z) = yz$, $\mathcal{C}$ is given by $x = \cos t$, $y = \sin t$, $z = -\cos t$, $0 \leq t \leq \frac{\pi}{2}$.

   (f) $P(x,y,z) = z$, $Q(x,y,z) = x^2 + y^2$, $R(x,y,z) = x + z$, $\mathcal{C}$ is given by $x = t^{1/2}$, $y = t$, $z = t^{3/2}$, $1 \leq t \leq 2$.

   (g) $P(x,y,z) = y + z$, $Q(x,y,z) = -x$, $R(x,y,z) = -x$, $\mathcal{C}$ is given by $x = \cos t$, $y = \sin t$, $z = \sin t + \cos t$.

4. Let $\mathcal{C}$ be the closed curve consisting of the upper semicircle $x^2 + y^2 = 1$ from $(1,0)$ to $(-1,0)$, followed by the x-axis back to $(1,0)$. For each 1-form below, set up the integral $\int_{\mathcal{C}} \omega$ two ways:

   - using the parametrization $(x,y) = (\cos\theta, \sin\theta)$, $0 \leq \theta \leq \pi$;
   - using the fact that the upper semicircle is the graph $y = \sqrt{1-x^2}$, $-1 \leq x \leq 1$ (*Caution:* make sure your curve goes in the right direction!);

   Then evaluate one of these versions.

   (a) $\omega = (x^2 + y)\,dx + (x + y^2)\,dy$

   (b) $\omega = x\,dy + y\,dx$

6.2 The Fundamental Theorem for Line Integrals

The Fundamental Theorem for Line Integrals in the Plane

Recall the *Fundamental Theorem of Calculus*, which says in part that if a function f is continuously differentiable on the interior of an interval (a,b)

(and continuous at the endpoints), then the integral over $[a,b]$ of its derivative is the difference between the values of the function at the endpoints:

$$\int_a^b \frac{df}{dt}\,dt = f\Big|_a^b := f(b) - f(a)\,.$$

The analogue of this for functions of several variables is called the *Fundamental Theorem for Line Integrals.* The derivative of a real-valued function on $\mathbb{R}^2$ is our first example of a form;

$$d_{(x,y)}f(v_1, v_2) = \left(\frac{\partial f}{\partial x}(x,y)\right)v_1 + \left(\frac{\partial f}{\partial y}(x,y)\right)v_2.$$

We shall call a form ω **exact** if it equals the differential of some function f: $\omega = df$. Let us integrate such a form over a curve $\mathcal{C}$, parametrized by $\overrightarrow{p}(t) = x(t)\,\overrightarrow{\imath} + y(t)\,\overrightarrow{\jmath}$, $a \le t \le b$. We have

$$\begin{aligned}
[\omega^*(\vec{p}\,')]_t &= \omega_{\overrightarrow{p}(t)}\left(\frac{dx}{dt}\overrightarrow{\imath} + \frac{dy}{dt}\overrightarrow{\jmath}\right)\,dt \\
&= \left[\left(\frac{\partial f}{\partial x}(x(t)\,, y(t))\right)\frac{dx}{dt} + \left(\frac{\partial f}{\partial y}(x(t)\,, y(t))\right)\frac{dy}{dt}\right]dt
\end{aligned}$$

which, by the Chain Rule, is

$$\begin{aligned}
&= \frac{d}{dt}\left[f(x(t)\,, y(t))\right]\,dt \\
&= g'(t)\,\,dt,
\end{aligned}$$

where

$$\begin{aligned}
g(t) &= f(\overrightarrow{p}(t)) \\
&= f(x(t)\,, y(t))\,.
\end{aligned}$$

Thus,

$$\begin{aligned}
\int_{\mathcal{C}} df &= \int_a^b \frac{d}{dt}\left[f(x(t)\,, y(t))\right]\,dt \\
&= \int_a^b g'(t)\,\,dt.
\end{aligned}$$

Provided this integrand is continuous (that is, the partials of f are continuous), the Fundamental Theorem of Calculus tells us that this equals

$$g(t)\Big|_a^b = g(b) - g(a)$$

or, writing this in terms of our original function,

$$f(\overrightarrow{p}(t))\Big|_a^b = f(\overrightarrow{p}(b)) - f(\overrightarrow{p}(a))\,.$$

Let us see how this translates to the language of vector fields. The vector field corresponding to the differential of a function is its gradient

$$df = \frac{\partial f}{\partial x}\,dx + \frac{\partial f}{\partial y}\,dy = \overrightarrow{\nabla} f \cdot d\overrightarrow{\mathfrak{s}}.$$

A vector field $\overrightarrow{F}$ is called **conservative** if it equals the gradient of some function f; the function f is then a **potential** for $\overrightarrow{F}$.

The bilingual statement (that is, in terms of both vector fields and forms) of this fundamental result is

Theorem 6.2.1 (Fundamental Theorem for Line Integrals)**.** *Suppose $\mathcal{C}$ is an oriented curve starting at p_{start} and ending at p_{end}, and f is a continuously differential function defined along $\mathcal{C}$. Then the integral of its differential df (resp. the line integral of its gradient vector field $\overrightarrow{\nabla} f$) over $\mathcal{C}$ equals the difference between the values of f at the endpoints of $\mathcal{C}$:*

$$\int_{\mathcal{C}} \overrightarrow{\nabla} f \cdot d\overrightarrow{\mathfrak{s}} = \int_{\mathcal{C}} df = f(x)\Big|_{p_{start}}^{p_{end}} = f(p_{end}) - f(p_{start})\,. \tag{6.4}$$

This result leads to a rather remarkable observation. We saw that the line integral of a vector field over an oriented curve $\mathcal{C}$ depends only on the curve (as a set of points) and the direction of motion along $\mathcal{C}$—it does not change if we reparametrize the curve before calculating it. But the Fundamental Theorem for Line Integrals tells us that if the vector field is conservative, then the line integral depends only on where the curve starts and where it ends, *not on how we get from one to the other.* Saying this a little more carefully,

Corollary 6.2.2. *Suppose f is a $\mathcal{C}^1$ function defined on the region D.*

Then the line integral $\int_{\mathcal{C}} \overrightarrow{\nabla} f \cdot d\overrightarrow{\mathfrak{s}} = \int_{\mathcal{C}} df$ is independent of the curve $\mathcal{C}$—that is, if $\mathcal{C}_1$ and $\mathcal{C}_2$ are two curves in D with a common starting point and a common ending point, then

$$\int_{\mathcal{C}_1} \overrightarrow{\nabla} f \cdot d\overrightarrow{\mathfrak{s}} = \int_{\mathcal{C}_1} \overrightarrow{\nabla} f \cdot d\overrightarrow{\mathfrak{s}}.$$

A second consequence of Equation (6.4) concerns a **closed** curve—that is, one that starts and ends at the same point ($p_{start} = p_{end}$). In this case,

$$\int_{\mathcal{C}} \vec{\nabla} f \cdot d\vec{\mathfrak{s}} = \int_{\mathcal{C}} df = f(x)\Big|_{p_{start}}^{p_{end}} = f(p_{end}) - f(p_{start}) = 0$$

Corollary 6.2.3. *Suppose f is a $\mathcal{C}^1$ function defined on the region D. Then the line integral of df around any closed curve $\mathcal{C}$ is zero:*

$$\int_{\mathcal{C}} \vec{\nabla} f \cdot d\vec{\mathfrak{s}} = \int_{\mathcal{C}} df = 0.$$

Sometimes, the integral of a vector field $\vec{F}$ over a *closed* curve is denoted $\oint_{\mathcal{C}} \vec{F} \cdot d\vec{\mathfrak{s}}$, to emphasize the fact that the curve is closed.

Actually, Corollary 6.2.2 and Corollary 6.2.3 are easily shown to be equivalent, using the fact that reversing orientation switches the sign of the integral (Exercise 4).

How do we decide whether or not a given vectorfield $\vec{F}$ is conservative?

The most direct way is to try to find a potential function f for $\vec{F}$. Let us investigate a few examples.

An easy one is

$$\vec{F}(x, y) = y\,\vec{\imath} + x\,\vec{\jmath}.$$

The condition that

$$\begin{aligned} \vec{F} &= \vec{\nabla} f \\ &= \frac{\partial f}{\partial x}\vec{\imath} + \frac{\partial f}{\partial y}\vec{\jmath} \end{aligned}$$

consists of the two equations

$$\frac{\partial f}{\partial x}(x, y) = y$$

and

$$\frac{\partial f}{\partial y}(x, y) = x.$$

The first is satisfied by

$$f(x, y) = xy$$

and we see that it also satisfies the second. Thus, we know that one potential for $\overrightarrow{F}$ is

$$f(x,y) = xy.$$

However, things are a bit more complicated if we consider

$$\overrightarrow{F}(x,y) = (x+y)\overrightarrow{\imath} + (x+y)\overrightarrow{\jmath}.$$

It is easy enough to guess that a function satisfying the first condition

$$\frac{\partial f}{\partial x}(x,y) = x+y$$

is

$$f(x,y) = \frac{x^2}{2} + xy,$$

but when we try to fit the second condition, which requires

$$\frac{\partial}{\partial y}\left[\frac{x^2}{2} + xy\right] = x+y$$

we come up with the impossible condition

$$x = x+y.$$

Does this mean our vector field is not conservative? Well, no. We need to think more systematically.

Note that our guess for $f(x,y)$ is not the *only* function satisfying the condition

$$\frac{\partial f}{\partial x} = x+y;$$

we need a function which is an antiderivative of $x+y$ *when y is treated as a constant.* This means that a complete list of antiderivatives consists of our specific antiderivative *plus an arbitray "constant"*–which in our context means any expression that does not depend on x. So we should write the "constant" as a function of y:

$$f(x,y) = \frac{x^2}{2} + xy + C(y)\,.$$

Now, when we try to match the second condition, we come up with

$$x+y=\frac{\partial f}{\partial y}=x+C'(y)$$

or

$$C'(y)=y$$

which leads to

$$C(y)=\frac{y^2}{2}+C$$

(where this time, C is an honest constant—it depends on neither x nor y). Thus the list of all functions satisfying *both* conditions is

$$f(x,y)=\frac{x^2}{2}+xy+\frac{y^2}{2}+C,$$

showing that indeed $\overrightarrow{F}$ is conservative.

This example illustrates the general procedure. If we seek a potential $f(x,y)$ for the vector field

$$\overrightarrow{F}(x,y)=P(x,y)\,\overrightarrow{\imath}+Q(x,y)\,\overrightarrow{\jmath},$$

we first look for a complete list of functions satisfying the first condition

$$\frac{\partial f}{\partial x}=P(x,y)\,;$$

this is a process much like taking the "inner" integral in an iterated integral, but without specified "inner" limits of integration: we treat y as a constant, and (provided we can do the integration) end up with an expression that looks like

$$f(x,y)=f_1(x,y)+C(y)$$

as a list of all functions satisfying the first condition. To decide which of these *also* satisfy the second condition, we take the partial with respect to y of our expression above, and match it to the second component of $\overrightarrow{F}$:

$$\frac{\partial}{\partial y}\left[f_1(x,y)\right]+C'(y)=Q(x,y)\,.$$

If this match is possible (we shall see below how this might fail), then we end up with a list of all potentials for $\overrightarrow{F}$ that looks like

$$f(x,y) = f_1(x,y) + f_2(y) + C$$

where $f_2(y)$ does not involve y, and C is an arbitrary constant.

Let's try this on a slightly more involved vector field,

$$\overrightarrow{F}(x,y) = (2xy + y^3 + 2)\overrightarrow{\imath} + (x^2 + 3xy^2 - 3)\overrightarrow{\jmath}.$$

The list of functions satisfying

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

is obtained by integrating, treating y as a constant:

$$f(x,y) = x^2y + xy^3 + 2x + C(y);$$

differentiating with respect to y (and of course now treating x as constant) we obtain

$$\frac{\partial f}{\partial y} = x^2 + 3xy^2 + C'(y).$$

Matching this with the second component of $\overrightarrow{F}$ gives

$$x^2 + 3xy^2 - 3 = x^2 + 3xy^2 + C'(y)$$

or

$$-3 = C'(y)$$

so

$$C(y) = -3y + C$$

and our list of potentials for $\overrightarrow{F}$ is

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

Now let us see how such a procedure can fail. If we look for potentials of

$$\overrightarrow{F}(x,y) = (x+2xy)\overrightarrow{\imath} + (x^2+xy)\overrightarrow{\jmath}$$

the first condition

$$\frac{\partial f}{\partial x} = x + 2xy$$

means

$$f(x,y) = \frac{x^2}{2} + x^2y + C(y)\,;$$

the partial with respect of y of such a function is

$$\frac{\partial f}{\partial y} = x^2 + C'(y)\,.$$

But when we try to match this to the second component of $\overrightarrow{F}$, we require

$$x^2 + xy = x^2 + C'(y)$$

or, cancelling the first term on both sides,

$$xy = C'(y)\,,$$

requiring $C(y)$, which is explicitly a function not involving x, to equal something that *does* involve x. This is impossible, and so no function can satisfy *both* of the conditions required to be a potential for $\overrightarrow{F}$; thus $\overrightarrow{F}$ is *not* conservative.

It is hardly obvious at first glance why our last example failed when the others succeeded. So we might ask if there is another way to decide whether a given vector field $\overrightarrow{F}(x,y) = P(x,y)\,\overrightarrow{\imath} + Q(x,y)\,\overrightarrow{\jmath}$ is conservative.

A necessary condition follows from the equality of cross-partials (Theorem 3.7.1). If $\overrightarrow{F}(x,y) = P(x,y)\,\overrightarrow{\imath} + Q(x,y)\,\overrightarrow{\jmath}$ is the gradient of the function $f(x,y)$, that is,

$$\begin{aligned} P(x,y) &= \frac{\partial f}{\partial x}(x,y) \\ Q(x,y) &= \frac{\partial f}{\partial y}(x,y) \end{aligned}$$

then

$$\frac{\partial P}{\partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

and

$$\frac{\partial Q}{\partial x} = \frac{\partial^2 f}{\partial x \partial y}$$

and equality of cross-partials then says that these are equal:

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}.$$

Technically, Theorem 3.7.1 requires that the two second-order partials be continuous, which means that the components of $\overrightarrow{F}$ (or of the form $\omega = P\,dx + Q\,dy$) have have $\frac{\partial P}{\partial y}$ and $\frac{\partial Q}{\partial x}$ continuous. In particular, it applies to any continuously differentiable, or $\mathcal{C}^1$, vector field.

Remark 6.2.4. *For any conservative $\mathcal{C}^1$ vector field $\overrightarrow{F}(x,y) = P(x,y)\,\overrightarrow{\imath} + Q(x,y)\,\overrightarrow{\jmath}$ (resp. $\mathcal{C}^1$ exact form $\omega_{(x,y)} = P(x,y)\ dx + Q(x,y)\ dy$),*

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}. \tag{6.5}$$

A vector field $\overrightarrow{F} = P\overrightarrow{\imath} + Q\overrightarrow{\jmath}$ (*resp.* differential form $\omega = P\,dx + Q\,dy$) is called **irrotational**[3] (*resp.* **closed**) if it satisfies Equation (6.5); thus Remark 6.2.4 says that every conservative vector field (*resp.* exact form) is irrotational (*resp.* closed).

How about the converse—if this condition holds, is the vector field (*resp.* form) necessarily conservative (*resp.* exact)? Well...almost.

We shall explore this in a sequence of technical lemmas.

Lemma 6.2.5. *Suppose D is a right triangle whose legs are parallel to the coordinate axes, and $P(x,y)$ and $Q(x,y)$ are $\mathcal{C}^1$ functions which satisfy Equation (6.5) on D:*

$$\frac{\partial Q}{\partial x}(x,y) = \frac{\partial P}{\partial y}(x,y) \ \textit{for all}\ (x,y) \in D.$$

[3] The reason for this terminology will become clear later.

Let $\mathcal{C}_1$ be the curve formed by the legs of the triangle, and $\mathcal{C}_2$ its hypotenuse, both oriented so that they start at at a common vertex of the triangle (and end at a common vertex: Figure 6.3). Then

$$\int_{\mathcal{C}_1} P\,dx + Q\,dy = \int_{\mathcal{C}_2} P\,dx + Q\,dy.$$

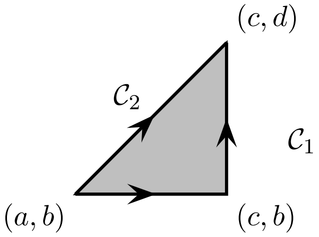


Figure 6.3: Integrating along the sides of a triangle

Note that the statement of the theorem allows either the situation in Figure 6.3 or the complementary one in which $\mathcal{C}_1$ goes up to (a,d) and then across to (c,d). We give the proof in the situation of Figure 6.3 below, and leave to you the modifications necessary to prove the complementary case. (Exercise 5a).

Proof. The integral along $\mathcal{C}_1$ is relatively straightforward: on the horizontal part, y is constant (so, formally, $dy = 0$), while on the vertical part, x is constant ($dx = 0$); it follows that

$$\int_{\mathcal{C}_1} P\,dx + Q\,dy = \int_a^c P(x,b)\;dx + \int_b^d Q(c,y)\;dy.$$

To integrate $P\,dx$ over $\mathcal{C}_2$, we write the curve as the graph of an affine function $y = \varphi(x)$, then use this to write

$$\int_{\mathcal{C}_2} P\,dx = \int_a^c P(x,\varphi(x))\;dx.$$

Similarly, to integrate $Q\,dy$ over $\mathcal{C}_2$ we write it as $x = \psi(y)$, to obtain

$$\int_{\mathcal{C}_2} Q\,dy = \int_b^d Q(\psi(y)\,,y)\;dy.$$

Combining these three expressions, we can express the difference between the two integrals as

$$\int_{\mathcal{C}_1} P\,dx + Q\,dy - \int_{\mathcal{C}_2} P\,dx + Q\,dy$$
$$= \int_a^c [P(x,b) - P(x,\varphi(x))]\,dx + \int_b^d [Q(c,y) - Q(\psi(y)\,,y)]\,dy.$$

We can apply Fundamental Theorem of Calculus to the integrand in the second integral to write the difference of integrals as an iterated integral and then interpret it as a double integral:

$$\begin{aligned} \int_{\mathcal{C}_1} Q\,dy - \int_{\mathcal{C}_2} Q\,dy &= \int_b^d \int_{\psi(y)}^c \frac{\partial Q}{\partial x}(x,y)\;dy \\ &= \iint_D \frac{\partial Q}{\partial x}\,dA \end{aligned} \tag{6.6}$$

Similarly, we can apply the Fundamental Theorem of Calculus to the integrand in the first integral to write the difference as an iterated integral; note however that the orientation of the inner limits of integration is backward, so this gives the negative of the appropriate double integral:

$$\begin{aligned} \int_{\mathcal{C}_1} P\,dx - \int_{\mathcal{C}_2} P\,dx &= \int_a^c \int_{\varphi(x)}^b \frac{\partial P}{\partial y}(x,y)\;dy \\ &= -\iint_D \frac{\partial P}{\partial y}\,dA. \end{aligned} \tag{6.7}$$

But our hypothesis says that these two integrands are equal, so we have

$$\int_{\mathcal{C}_1} P\,dx + Q\,dy - \int_{\mathcal{C}_2} P\,dx + Q\,dy = \iint_D \frac{\partial Q}{\partial x}\,dA - \iint_D \frac{\partial P}{\partial} dAy = 0.$$

□

An immediate corollary of Lemma 6.2.5 is the following:

Corollary 6.2.6. *Suppose Equation (6.5) holds on the rectangle $D = [a,b] \times [c,d]$; then*

$$\int_{\mathcal{C}_1} P\,dx + Q\,dy = \int_{\mathcal{C}_2} P\,dx + Q\,dy$$

for any two polygonal curves in D going from (a,c) to (b,d) (Figure 6.4).

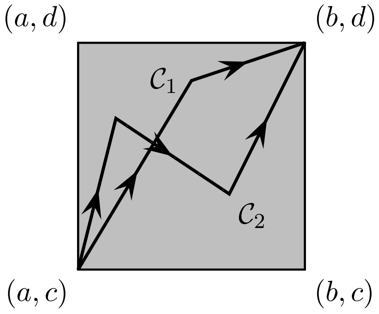


Figure 6.4: Polygonal curves with common endpoints in D.

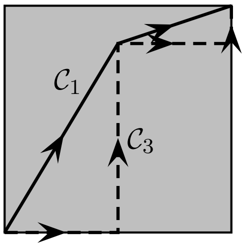


Figure 6.5: $\int_{\mathcal{C}_1} P\,dx + Q\,dy = \int_{\mathcal{C}_3} P\,dx + Q\,dy$

Proof. First, by Lemma 6.2.5, we can replace each straight segment of $\mathcal{C}_1$ with a broken line curve consisting of a horizontal and a vertical line segment (Figure 6.5) yielding $\mathcal{C}_3$.

Then, we can replace $\mathcal{C}_3$ with $\mathcal{C}_4$, the diagonal of the rectangle (Figure 6.6).

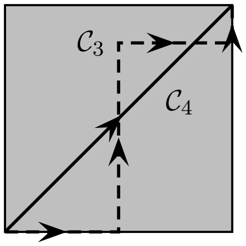


Figure 6.6: $\int_{\mathcal{C}_3} P\,dx + Q\,dy = \int_{\mathcal{C}_4} P\,dx + Q\,dy$

Applying the same argument to $\mathcal{C}_2$, we end up with

$$\int_{\mathcal{C}_1} P\,dx + Q\,dy = \int_{\mathcal{C}_4} P\,dx + Q\,dy = \int_{\mathcal{C}_2} P\,dx + Q\,dy.$$

□

We note that the statement of Corollary 6.2.6 can be loosened to allow the rectangle $[a,b] \times [c,d]$ to be replaced by any polygonal region containing both points, and then allow any polygonal curves $\mathcal{C}_i$ in this polygonal region which join these points (Exercise 5b).

Using this, we can prove our main result.

Proposition 6.2.7 (Poincaré Lemma)**.** *Suppose $P(x,y)$ and $Q(x,y)$ are $\mathcal{C}^1$ functions on the disk centered at (a,b)*

$$D := \{(x,y) \,|\, \operatorname{dist}((x,y),(a,b)) < r\}$$

satisfying Equation (6.5):

$$\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}.$$

Then there exists a function f defined on D such that

$$\frac{\partial f}{\partial x}(x,y) = P(x,y) \tag{6.8}$$

and

$$\frac{\partial f}{\partial y}(x,y) = Q(x,y) \tag{6.9}$$

at every point $(x,y) \in D$.

Proof. Define a function on the disc by

$$f(x,y) = \int_{\mathcal{C}} P\,dx + Q\,dy \tag{6.10}$$

where $\mathcal{C}$ is any polygonal curve in D from (a,b) to (x,y); by Corollary 6.2.6, this is well-defined.

We need to show that equations (6.8) and (6.9) both hold.

To this end, fix a point $(x_0,y_0) \in D$; we shall interpret the definition of f at points on a short horizontal line segment centered at (x_0,y_0) $\{(x_0 + t, y_0) \mid -\varepsilon \leq t \leq \varepsilon\}$ as given by the curve $\mathcal{C}$ consisting of a fixed curve from (a,b) to (x_0,y_0), followed by the horizontal segment $H(t)$ to (x_0+t,y_0). Then we can write

$$\begin{aligned} f(x_0+t,y_0) - f(x_0,y_0) &= \int_{H(t)} P\,dx + Q\,dy \\ &= \int_0^t P(x_0+x,y_0)\;dx; \end{aligned}$$

then we can apply the Fundamental Theorem of Calculus to this last integral to see that

$$\begin{aligned} \frac{\partial f}{\partial x}(x_0,y_0) &= \frac{\partial}{\partial t}\bigg|_{t=0}\left[\int_0^t P(x_0+x,y_0)\;dx\right] \\ &= P(x_0,y_0)\,, \end{aligned}$$

proving Equation (6.8). The proof of Equation (6.9) is analogous (Exercise 6). □

This shows that if Equation (6.5) holds everywhere inside some disc, then there is a function f defined on this disc satisfying

$$df = P\,dx + Q\,dy$$

or equivalently,

$$\vec{\nabla} f = P\vec{\imath} + Q\vec{\jmath}$$

at every point of this disc. So if ω (*resp.* $\vec{F}$) is an exact form (*resp.* irrotational vector field) in some planar region D, then given any point in D, there is a function defined *locally* (that is, on some disc around that point) which acts as a potential.

There is, however, a subtle problem with extending this conclusion *globally*—that is, to the whole region—illustrated by the following example.

Recall that the polar coordinates of a point in the plane are not unique—distinct values of (r, θ) can determine the same geometric point. In particular, the angular variable θ is determined only up to adding an integer multiple of π (an *odd* multiple corresponds to changing the sign of the other polar coordinate, r) . Thus, θ is not really a function on the complement of the origin, since its value at any point is ambiguous. However, once we pick out one value $\theta(x, y)$ at a particular point $(x, y) \neq (0, 0)$, then there is only one way to define a *continuous* function that gives a legitimate value for θ at nearby points. Any such function will have the form

$$\theta(x, y) = \arctan \frac{y}{x} + n\pi$$

for some (constant) integer n (why?). When we take the differential of this, the constant term disappears, and we get

$$\begin{aligned} d\theta &= \frac{\frac{dy}{x} - \frac{y\,dx}{x^2}}{1 + \left(\frac{y}{x}\right)^2} \\ &= \frac{x\,dy - y\,dx}{x^2 + y^2}. \end{aligned}$$

So even though the "function" $\theta(x, y)$ is not uniquely defined, its "differential" *is*. Furthermore, from the preceding discussion, Equation (6.5) holds (you should check this directly, at least once in your life).

Now let us try integrating $d\theta$ around the unit circle $\mathcal{C}$, oriented counterclockwise. The parametrization

$$\begin{aligned} x &= \cos t \\ y &= \sin t \\ 0 &\leq t \leq 2\pi \end{aligned}$$

leads to

$$\begin{aligned} dx &= -\sin t\,dt \\ dy &= \cos t\,dt \end{aligned}$$

so

$$\begin{aligned} d\theta &= \frac{(\cos t)(\cos t\,dt) - (\sin t)(-\sin t\,dt)}{\cos^2 t + \sin^2 t} \\ &= \frac{\cos^2 t + \sin^2 t}{\cos^2 t + \sin^2 t}\,dt \\ &= dt \end{aligned}$$

and thus

$$\begin{aligned} \int_{\mathcal{C}} d\theta &= \int_0^{2\pi} dt \\ &= 2\pi \end{aligned}$$

which of course would contradict Corollary 6.2.3, if $d\theta$ were exact. In fact, we can see that integrating $d\theta$ along the curve $\mathcal{C}$ amounts to continuing θ along the circle: that is, starting from the value we assign to θ at the starting point $(1,0)$, we use the fact that there is only one way to continue θ along a short arc through this point; when we get to the end of that arc, we *still* have only one way of continuing θ along a further arc through *that* point, and so on. But when we have come all the way around the circle, the angle has steadily increased, and is now at 2π more than it was when we started!

Another way to look at this phenomenon is to cut the circle into its upper and lower semicircles, and consider the continuation of θ along each from $(1,0)$ to $(-1,0)$. Supposing we start with $\theta = 0$ at $(1,0)$, the continuation along the upper semicircle lands at $\theta = \pi$ at $(-1,0)$. However, when we continue it along the lower semicircle, our angle goes negative, and we end up with $\theta = -\pi$ at $(-1,0)$. Thus, the two continuations do not agree.

Now, the continuation of θ is determined not just along an arc through a point, but on a whole *neighborhood* of that point. In particular, we can deform our original semicircle continuously—so long as we keep the two endpoints $(1,0)$ and $(-1,0)$, and as long as our deformation never goes through the origin—without changing the effect of the continuation along the curve: continuing θ along any of these deformed curves will still lead to the value π for θ at the end (Figure 6.7; see Exercise 7).

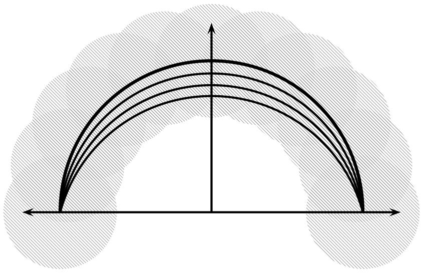


Figure 6.7: Continuation along deformed curves

We see, then, that our problem with continuing θ (or equivalently, integrating $d\theta$) around the upper and lower semicircles is related to the fact that we cannot deform the upper semicircle into the lower semicircle without going through the origin—where our form is undefined. A region in which this problem does not occur is called *simply connected*:

Definition 6.2.8. *A region $D \subset \mathbb{R}^n$ is* ***simply connected*** *if any pair of curves in D with a common start point and a common end point can be deformed into each other through a family of curves in D (without moving the start point and end point).*

An equivalent definition is: D is simply connected if any closed *curve in D can be deformed (through a family of closed curves in D) to a single point.*[4]

From the discussion above, we can construct a proof of the following:

Proposition 6.2.9. *If $D \subset \mathbb{R}^2$ is a simply connected region, then any differential form $\omega = P\,dx + Q\,dy$ (resp. vector field $\overrightarrow{F}$) on D is exact precisely if it is closed (resp. irrotational).*

Line Integrals in Space

The situation for forms and vector fields in $\mathbb{R}^3$ is completely analogous to that in the plane.

A vector field on $\mathbb{R}^3$ assigns to a point $(x, y, z) \in \mathbb{R}^3$ a vector

$$\overrightarrow{F}(x,y,z) = P(x,y,z)\,\overrightarrow{\imath} + Q(x,y,z)\,\overrightarrow{\jmath} + R(x,y,z)\,\overrightarrow{k}$$

[4] That is, to a curve defined by a constant vector-valued function.

while a form on $\mathbb{R}^3$ assigns to $(x, y, z) \in \mathbb{R}^3$ the functional

$$\omega_{(x,y,z)} = P(x, y, z)\ dx + Q(x, y, z)\ dy + R(x, y, z)\ dz.$$

The statement of Theorem 6.2.1 that we gave holds in $\mathbb{R}^3$: *the line integral of the gradient of a function (resp. of the differential of a function) over any curve equals the difference between the values of the function at the endpoints of the curve.*

It is instructive to see how the process of finding a potential function for a vector field or form works in $\mathbb{R}^3$. Let us consider the vector field

$$\overrightarrow{F}(x, y, z) = (y^2 + 2xz + 2)\overrightarrow{\imath} + (2xy + z^3)\overrightarrow{\jmath} + (x^2 + 3yz^2 + 6z)\overrightarrow{k}$$

or equivalently, the form

$$\omega_{(x,y,z)} = (y^2 + 2xz + 2)\, dx + (2xy + z^3)\, dy + (x^2 + 3yz^2 + 6z)\, dz.$$

A potential for either one is a function $f(x, y, z)$ satisfying the three conditions

$$\begin{aligned}
\frac{\partial f}{\partial x}(x, y, z) &= P(x, y, z) = y^2 + 2xz + 2\\
\frac{\partial f}{\partial y}(x, y, z) &= Q(x, y, z) = 2xy + z^3\\
\frac{\partial f}{\partial x}(x, y, z) &= R(x, y, z) = x^2 + 3yz^2 + 6z.
\end{aligned}$$

The first condition leads to

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

or, more accurately, the list of *all* functions satisfying the first condition consists of this function plus any function depending only on y and z:

$$f(x, y, z) = xy^2 + x^2z + 2x + C(y, z)\,.$$

Differentiating this with respect to y

$$\frac{\partial f}{\partial y}(x, y, z) = 2xy + \frac{\partial C}{\partial y}$$

turns the second condition into

$$2xy + z^3 = 2xy + \frac{\partial C}{\partial y}$$

so the function $C(y, z)$ must satisfy

$$z^3 = \frac{\partial C}{\partial y}.$$

this tells us that

$$C(y, z) = yz^3 + C(z)$$

(since a term depending only on z will not show up in the partial with respect to y). Substituting back, we see that the list of all functions satisfying the *first two* conditions is

$$f(x, y, z) = xy^2 + x^2 z + 2x + yz^3 + C(z)\,.$$

Now, taking the partial with respect to z and substituting into the third condition yields

$$x^2 + 3yz^2 + \frac{dC}{dz} = \frac{\partial f}{\partial z}\,(x, y, z) = x^2 + 3yz^2 + 6z$$

or

$$\frac{dC}{dz} = 6z;$$

hence

$$C(z) = 3z^2 + C$$

where this time C is an honest constant. Thus, the list of all functions satisfying all three conditions—that is, all the potential functions for $\overrightarrow{F}$ or ω—is

$$f(x, y, z) = xy^2 + x^2 z + 2x + yz^3 + 3z^2 + C$$

where C is an arbitrary constant.

If we recall that Equation (6.5)—that every conservative vectorfield (*resp.* exact form) must be irrotational (*resp.* closed)—came from the equality of cross-partials (Theorem 3.7.1), it is natural that the corresponding condition in $\mathbb{R}^3$ consists of three equations (Exercise 8):

$$\begin{aligned}\frac{\partial P}{\partial y} &= \frac{\partial Q}{\partial x}\\ \frac{\partial P}{\partial z} &= \frac{\partial R}{\partial x}\\ \frac{\partial Q}{\partial z} &= \frac{\partial R}{\partial y}.\end{aligned} \tag{6.11}$$

The version of Proposition 6.2.7 remains true—condition (6.11) implies the existence of a potential function, provided the region in question is simply-connected. However, simple-connectedness in $\mathbb{R}^3$ is a bit more subtle than in the plane. In the plane, a closed simple curve encloses a simply-connected region, and a region fails to be simply connected precisely if it has a "hole". In $\mathbb{R}^3$, a hole need not destroy simple connectedness: for example, any curve in a ball with the center excised can be shrunk to the point without going through the origin (Figure 6.8); the kind of hole that *does* destroy this property is more like a tunnel through the ball (Figure 6.9).

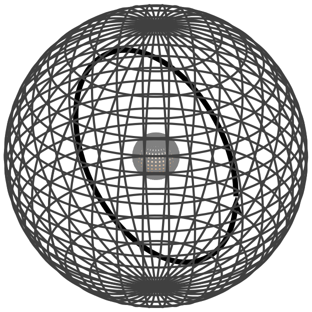


Figure 6.8: Simply Connected

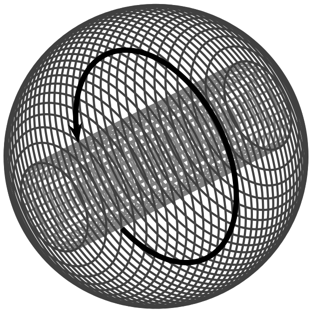


Figure 6.9: Not Simply Connected

We shall not prove the version of Proposition 6.2.7 for $\mathbb{R}^3$ here, but it will follow from Stokes' Theorem in the next section.

Exercises for § 6.2

Practice problems:

1. For each vectorfield below, determine whether it is conservative, and if it is, find a potential function; in either case, evaluate $\int_{\mathcal{C}} \overrightarrow{F} \cdot \overrightarrow{T}\, d\mathfrak{s}$ over the given curve:

   (a) $\overrightarrow{F}(x,y) = (x^2 + x + y)\overrightarrow{\imath} + (x + \sin y)\overrightarrow{\jmath}$, $\mathcal{C}$ is the straight line segment from $(0,0)$ to $(1,1)$.

   (b) $\overrightarrow{F}(x,y) = (2xy + y^2)\overrightarrow{\imath} + (2xy + x^2)\overrightarrow{\jmath}$, $\mathcal{C}$ is the straight line segment from $(0,0)$ to $(1,1)$.

   (c) $\overrightarrow{F}(x,y) = (x^2 - y^2)\overrightarrow{\imath} + (x^2 - y^2)\overrightarrow{\jmath}$, $\mathcal{C}$ is the circle $x^2 + y^2 = 1$, traversed counterclockwise.

2. Each vector field below is conservative. Find a potential function, and evaluate $\int_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}}$.

(a) $\overrightarrow{F}(x,y,z) = (2xy+z)\overrightarrow{\imath} + (x^2+z)\overrightarrow{\jmath} + (x+y)\overrightarrow{k}$, $\mathcal{C}$ is the straight line segment from $(1,0,1)$ to $(1,2,2)$.

(b) $\overrightarrow{F}(x,y,z) = y\cos xy\,\overrightarrow{\imath} + (x\cos xy - z\sin yz)\overrightarrow{\jmath} - y\sin yz\,\overrightarrow{k}$, $\mathcal{C}$ is the straight line segment from $(0,\pi,-1)$ to $(1,\frac{\pi}{2},4)$.

(c) $\overrightarrow{F}(x,y,z) = y^2z^3\overrightarrow{\imath} + (2xy + z^3 + 2z)\overrightarrow{\jmath} + 2(y+z)\overrightarrow{k}$, $\mathcal{C}$ is given by $\overrightarrow{p}(t) = (\sin\frac{\pi t}{2}, te^t, te^t\sin\frac{\pi t}{2})$, $0 \leq t \leq 1$.

(d) $\overrightarrow{F}(x,y,z) = (2xy - y^2z))\overrightarrow{\imath} + (x^2 - 2xyz)\overrightarrow{\jmath} + xy^2vk$, $\mathcal{C}$ is given by $x = \cos\pi t$, $y = t$, $z = t^2$, $0 \leq t \leq 2$.

(e) $\overrightarrow{F}(x,y,z) = ze^x cosy\,\overrightarrow{\imath} - ze^x\sin y\,\overrightarrow{\jmath} + e^x\cos y\,\overrightarrow{k}$, $\mathcal{C}$ is the broken line curve from $(0,0,0)$ to $(2,\pi,1)$ to $(1,\pi,1)$.

3. For each 1-form ω, determine whether it is exact, and if so, find a potential function. In either case, evaluate $\int_{\mathcal{C}}\omega$, where $\mathcal{C}$ is the straight-line segment from $(-1,1,-1)$ to $(1,2,2)$.

(a) $\omega = 2xyz^3\,dx + x^2z^3\,dy + 3x^2yz^2\,dz$

(b) $\omega = (2xy + yz)\,dx + (x^2 + xz + 2y)\,dy + (xy + 2z)\,dz$

(c) $\omega = (y - z)\,dx + (x - z)\,dy + (x - y)\,dz$

Theory problems:

4. **Show** that for a continuous vector field $\overrightarrow{F}$ defined in the region $D \subset \mathbb{R}^2$, the following are equivalent:

- The line integral of $\overrightarrow{F}$ over any closed curve in D is zero;
- For any two paths in D with a common starting point and a starting endpoint, the line integrals of $\overrightarrow{F}$ over the two paths are equal.

5. (a) Mimic the proof given for Lemma 6.2.5 to prove the complementary case when the curve goes up to (a,d) and then across to (c,d).

(b) Extend the proof given for Corollary 6.2.6 when the rectangle is replaced by an arbitrary polygonal region.

6. Mimic the proof of Equation (6.8) in the Poincaré Lemma (Proposition 6.2.7) to prove Equation (6.9).

7. **Show** that the line integral of the form $d\theta$ over the upper semicircle is unchanged if we replace the semicircle with a curve obtained by deforming the semicircle, keeping the endpoints fixed, as long as the curve doesn't go through the origin during the deformation.

8. Prove that the equations

$$\begin{aligned}\frac{\partial P}{\partial y} &= \frac{\partial Q}{\partial x}\\ \frac{\partial P}{\partial z} &= \frac{\partial R}{\partial x}\\ \frac{\partial Q}{\partial z} &= \frac{\partial R}{\partial y}.\end{aligned} \tag{6.12}$$

are satisfied by any conservative vector field in $\mathbb{R}^3$.

6.3 Green's Theorem

We saw in § 6.1 that the line integral of a conservative vector field (or of an exact form) around a closed curve is zero. Green's Theorem tells us what happens when a planar vector field is not conservative. This is related to Equations (6.6) and (6.7) which occurred in the course of proving Lemma 6.2.5. In these two equations, the difference between integrating the form $Q\,dx$ (*resp.* $P\,dy$) along the sides of a right triangle and integrating it along the hypotenuse was related to the integral of the partial $\frac{\partial Q}{\partial x}$ (*resp.* $\frac{\partial P}{\partial y}$) over the inside of the triangle. Here, we need to reformulate this more carefully, and do so in terms of a closed curve.

Recall that in § 1.6 we defined the orientation of a triangle in the plane, and its associated signed area. A triangle or other polygon has **positive orientation** if its vertices are traversed in counterclockwise order. We now extend this notion to a closed, simple curve[5] An intuitively plausible observation, but one which is very difficult to prove rigorously, is known as the **Jordan Curve Theorem**: it says that a simple, closed curve $\mathcal{C}$ in the plane divides the plane into two regions (the "inside" and the "outside"): any two points in the same region can be joined by a curve disjoint from $\mathcal{C}$, but it is impossible to join a point *inside* the curve to one *outside* the

[5] Recall that a curve is **closed** if it starts and ends at the same point. A curve is **simple** if it does not intersect itself: that is, if it can be parametrized over a closed interval, say by $\overrightarrow{p}(t)$, $t_0 \leq t \leq t_1$ so that the only instance of $\overrightarrow{p}(s) = \overrightarrow{p}(t)$ with $s \neq t$ is $s = a$, $t = b$. A simple, closed curve can also be thought of as parametrized over a circle, in such a way that distinct points correspond to distinct parameter values on the circle.

curve without crossing $\mathcal{C}$. The "inside" is a bounded set, referred to as the region **bounded** by $\mathcal{C}$; the "outside" is unbounded. This result was formulated by Camille Jordan (1838-1922) in 1887 [29, 1st ed., Vol. 3, p. 593], but first proved rigorously by the American mathematician Oswald Veblen (1880-1960) in 1905 [52].

We shall formulate the notion of positive orientation first for a *regular* simple closed curve. Recall from Definition 6.1.1 that an *orientation* of a regular curve $\mathcal{C}$ is a continuous choice of unit tangent vector $\overrightarrow{T}$ at each point of $\mathcal{C}$; there are exactly two such choices.

Definition 6.3.1. *1. If $\overrightarrow{T} = (\cos\theta, \sin\theta)$ is a unit vector in $\mathbb{R}^2$, then the **leftward normal** to $\overrightarrow{T}$ is the vector*

$$\overrightarrow{N}_+ = \left(\cos(\theta + \frac{\pi}{2}), \sin(\theta + \frac{\pi}{2})\right) = (-\sin\theta, \cos\theta).$$

*2. Suppose $\mathcal{C}$ is a regular, simple, closed curve in the plane. The **positive orientation** of $\mathcal{C}$ is the choice $\overrightarrow{T}$ for which the leftward normal points into the region bounded by $\mathcal{C}$—in other words, if $\overrightarrow{p}$ is the position vector for the basepoint of $\overrightarrow{T}$, then for small $\varepsilon > 0$, the point $\overrightarrow{p} + \varepsilon\overrightarrow{N}_+$ belongs to the* inside *region. (Figure 6.10). The other orientation (for which $\overrightarrow{N}_+$ points into the* unbounded *region) is the **negative orientation**.*

Recall also, from § 5.2, that a region D in the plane is *regular* if it is both x-regular and y-regular—meaning that any horizontal or vertical line intersects D in either a single point or a single interval. The theory of multiple integrals we developed in Chapter 5 was limited to regions which are either regular or can be subdivided into regular subregions.

Unfortunately, a regular region need not be bounded by a regular curve. For example, a polygon such as a triangle or rectangle is not a regular curve, since it has "corners" where there is no well-defined tangent line. As another example, if D is defined by the inequalities

$$\begin{aligned} x^2 \leq & y \leq \sqrt{x} \\ 0 \leq & x \leq 1 \end{aligned}$$

then its boundary consists of two pieces: the lower edge is part of the graph of $y = x^2$, while the upper edge is part of the graph of $y = \sqrt{x}$. Each piece is naturally parametrized as a regular curve. The natural parametrization of the lower edge, $x = t$, $y = t^2$, $0 \leq t \leq 1$, is clearly regular. If we try to

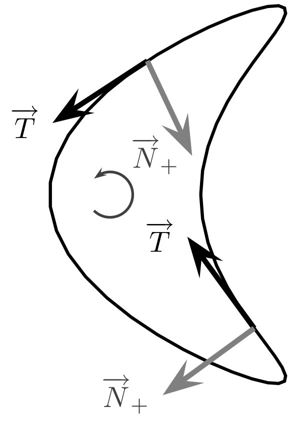


Figure 6.10: Positive Orientation of a Simple Closed Curve

parametrize the upper edge analogously as $x = t$, $y = \sqrt{t}$, we have a problem at $t = 0$, since $\sqrt{t}$ is not differentiable there. We can, however, treat it as the graph of $x = y^2$, leading to the regular parametrization $x = t^2$, $y = t$, $0 \leq t \leq 1$. Unfortunately, these two regular parametrizations do not fit together in a "smooth" way: their velocity vectors at the two points where they meet—$(0, 0)$ and $(1, 1)$—point in different directions, and there is no way of "patching up" this difference to get a regular parametrization of the full boundary curve.

But for our purposes, this is not a serious problem: we can allow this kind of discrepancy at finitely many points, and extend our definition to this situation:

Definition 6.3.2. *A locally one-to-one curve $\mathcal{C}$ in $\mathbb{R}^2$ is **piecewise regular** if it can be partitioned into finitely many arcs $\mathcal{C}_i$, $i = 1, \ldots, k$ such that*

1. *Each $\mathcal{C}_i$ is the image of a regular parametrization $\overrightarrow{p}_i$ defined on a closed interval $[a_i, b_i]$ (in particular, the tangent vectors $\vec{p_i}\,'(a_i)$ and $\vec{p_i}\,'(b_i)$ at the endpoints are nonzero, and each is the limit of the tangent vectors at nearby points of $\mathcal{C}_i$, and*

2. *the arcs abut at endpoints: for $i = 1, \ldots, k-1$, $\overrightarrow{p}_i(1) = \overrightarrow{p}_{i+1}(0)$.*

Thus, we allow, at each of the finitely many common endpoints of these arcs, that there are two "tangent" directions, each defined in terms of one of the *two arcs that abut there. We will refer to points where such a discrepancy occurs as* ***corners*** *of the curve.*

The discrepancy between tangent vectors at a corner can amount to as much as π radians; see Figure 6.11. This means that Definition 6.3.1 cannot be applied at such points; however, we can still apply it at all other points, and have a coherent definition.

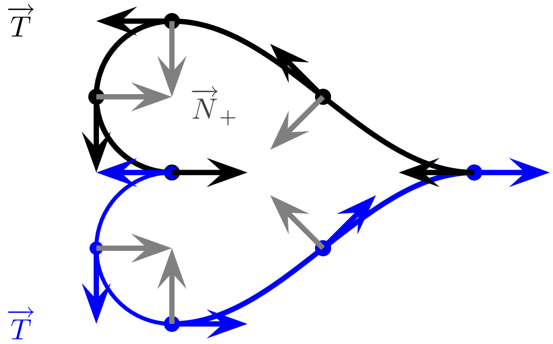


Figure 6.11: Positive Orientation for a Piecewise-Regular Curve with Corners

Definition 6.3.3. *Suppose $\mathcal{C}$ is a piecewise regular, simple, closed curve in $\mathbb{R}^2$. Then the* ***positive orientation*** *is the choice of unit tangent vector $\overrightarrow{T}$ at all non-corners such that the leftward normal $\overrightarrow{N}_+$ points into the region bounded by $\mathcal{C}$.*

With these definitions, we can formulate Green's Theorem. This was originally formulated and proved by George Green (1793-1841), a self-taught mathematician whose exposition was contained in a self-published pamphlet on the use of mathematics in the study of electricity and magnetism [19] in 1828.[6]

[6] The son of a successful miller in Nottingham, he entered his father's business instead of going to university, but studied privately. He finally went to Cambridge at the age of 40, obtaining his degree in 1837, and subsequently published six papers. Interest in his 1828 Essay on the part of William Thomson (later Lord Kelvin) got him a Fellowship at Caius College in 1839. He remained for only two terms, then returned home, dying the following year. [1, p. 202]

Theorem 6.3.4 (Green's Theorem)**.** *Suppose $\mathcal{C}$ is a piecewise regular, simple, closed curve with positive orientation in the plane, bounding the regular region D.*

Then for any pair of $\mathcal{C}^1$ functions P and Q defined on D,

$$\oint_{\mathcal{C}} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA. \tag{6.13}$$

Proof. First, let us describe D as a y-regular region (Figure 6.12)

$$\begin{aligned} \varphi(x) \leq & y \leq \psi(x) \\ a \leq & x \leq b. \end{aligned}$$

and use it to calculate $\oint_{\mathcal{C}} P\,dx$.

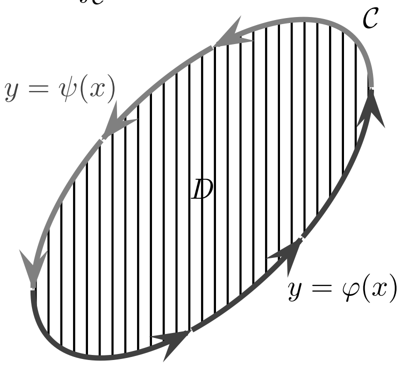


Figure 6.12: y-regular version of D

Note that while the *bottom* edge ($y = \varphi(x)$) is traversed with x *increasing*, the *top* edge ($y = \psi(x)$) has x *decreasing*, so the line integral of $P\,dx$ along the bottom edge has the form

$$\int_{y=\varphi(x)} P(x, y)\ dx = \int_a^b P(x, \varphi(x))\ dx,$$

the integral along the *top* edge is reversed, so it has the form

$$\int_{y=\psi(x)} P(x, y)\ dx = \int_a^b -P(x, \psi(x))\ dx.$$

Also, if $\varphi(a) < \psi(a)$ (*resp.* $\varphi(b) < \psi(b)$)—so that $\mathcal{C}$ has a vertical segment corresponding to $x = a$ (*resp.* $x = b$)—then since x is constant, $dx = 0$ along these pieces, and they contribute nothing to $\oint_{\mathcal{C}} P\,dx$. Thus we can write

$$\begin{aligned}\oint_{\mathcal{C}} P\,dx &= \int_a^b P(x,\varphi(x))\;dx + \int_a^b -P(x,\psi(x))\;dx\\ &= \int_a^b \big(-P(x,\psi(x)) + P(x,\varphi(x))\big)\,dx.\end{aligned}$$

But for each fixed value of x, the quantity in parentheses above is the difference between the values of P at the ends of the vertical slice of D corresponding to that x-value. Thus we can write

$$-P(x,\psi(x)) + P(x,\varphi(x)) = \int_{\varphi(x)}^{\psi(x)} -\frac{\partial P}{\partial y}\,dy$$

and hence we have the analogue of Equation (6.7) in § 6.1:

$$\oint_{\mathcal{C}} P\,dx = \int_a^b \int_{\varphi(x)}^{\psi(x)} \left(-\frac{\partial P}{\partial y}\right)dy\,dx = \iint_D \left(-\frac{\partial P}{\partial y}\right)dA. \tag{6.14}$$

Now, to handle $\oint_{\mathcal{C}} Q\,dy$, we revert to the description of D as an x-regular region (Figure 6.13):

$$\begin{aligned}\alpha(y) \leq & x \leq \beta(y)\\ c \leq & y \leq d.\end{aligned}$$

The argument is analogous to that involving $P\,dx$: this time, y is *increasing* on the *right* edge ($x = \beta(y)$) of D and *decreasing* on the *left* ($x = \alpha(y)$). There is no contribution to $\oint_{\mathcal{C}} Q\,dy$ from horizontal segments in $\mathcal{C}$. This leads to the calculation

$$\begin{aligned}\oint_{\mathcal{C}} Q\,dy &= \int_{x=\beta(y)} Q\,dy + \int_{x=\alpha(y)} -Q\,dy\\ &= \int_c^d \big(Q(\beta(y)\,,y) - Q(\alpha(y)\,,y)\big)\,dy\\ &= \int_c^d \left(\int_{\alpha(y)}^{\beta(y)} \frac{\partial Q}{\partial x}\,dx\right)dy\end{aligned}$$

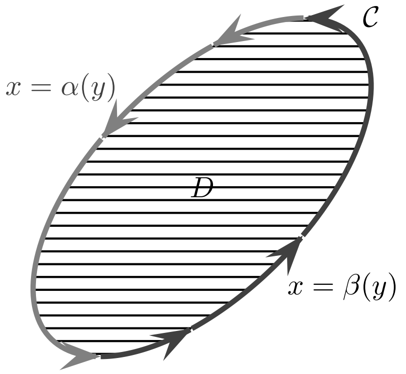


Figure 6.13: x-regular version of D

from which we have the analogue of Equation (6.6) in § 6.1:

$$\oint_{\mathcal{C}} Q\,dy = \int_a^b \int_{\alpha(y)}^{\beta(y)} \frac{\partial Q}{\partial x}\,dx\,dy = \iint_D \frac{\partial Q}{\partial x}\,dA. \tag{6.15}$$

Combining these, we get Green's Theorem

$$\oint_{\mathcal{C}} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA$$

when D is a regular region. □

If a region is not regular, it can often be subdivided into regular regions. One approach is to draw a grid (Figure 6.14): most of the interior is cut into rectangles (which are certainly regular) and what is left are regions with some straight sides and others given by pieces of the bounding curve. With a careful choice of grid lines, these regions will also be regular[7]

Clearly, the double integral of $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ over all of D equals the sum of its integrals over each of the regular subregions, and Equation (6.16) applies to each of these individually, so that we can replace each such double integral in this sum with the corresponding line integral of $P\,dx + Q\,dy$ over the edge of that piece, oriented positively. Note that positive orientation of two adjacent

[7] If the curve has vertical and horizontal tangents at only finitely many points, and only finitely many "corners", then it suffices to make sure the grid lines go through all of these points. The only difficulty is when there are infinitely many horizontal or vertical tangents; in this case we can try to use a slightly rotated grid system. This is always possible if the curve is $\mathcal{C}^2$; the proof of this involves a sophisticated result in differential topology, the Morse-Sard Theorem.

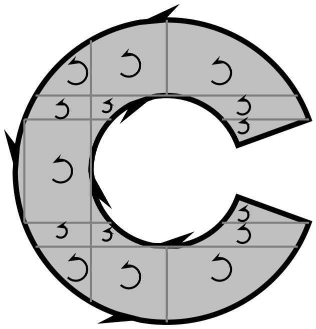


Figure 6.14: Subdivision of a Region into Regular Ones

pieces induces *opposite* directions along their common boundary segment, so when we sum up all these line integrals, the ones corresponding to pieces of the grid cancel, and we are left with only the sum of line integrals along pieces of our original bounding curve, $\mathcal{C}$. This shows that Equation (6.16) holds for the region bounded by a single closed curve—even if it is not regular—as long as it can be subdivided into regular regions.

We can take this one step further. Consider for example the region between two concentric circles[8] (Figure 6.15). This is bounded by not by one, but *two* closed curves.

If we subdivide this region into regular subregions via a grid, and orient the edge of each subregion positively, we can apply the same reasoning as above to conclude that the sum of the line integrals of $P\,dx + Q\,dy$ over the edges of the pieces (each oriented positively) equals the integral of $\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)$ over the whole region, and that furthermore each piece of edge coming from the grid appears twice in this sum, but with opposite directions, and hence is cancelled. Thus, the only line integrals contributing to the sum are those coming from the two boundary curves. We know that the positive orientation of the outer circle is *counterclockwise*—but we see from Figure 6.15 that the *inner* circle is directed *clockwise.* However, this is exactly the orientation we get if we adopt the phrasing in Definition 6.3.1:

[8] This is called an **annulus**.

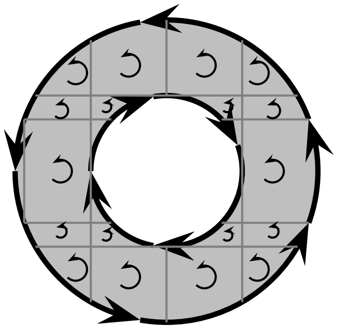


Figure 6.15: Subdivision of an Annulus into Regular Ones

that the leftward normal must point into the region. Thus we see that the appropriate orientation for a boundary curve is determined by where the region lies relative to that curve. To avoid confusion with our earlier definition, we formulate the following:

Definition 6.3.5. *Suppose $D \subset \mathbb{R}^2$ is a region whose boundary ∂D consists of finitely many piecewise regular closed curves. Then for each such curve, the* ***boundary orientation*** *is the one for which the leftward normal at each non-corner points into the region D.*

With this definition, we see that Green's Theorem can be extended as follows:

Theorem 6.3.6 (Green's Theorem, Extended Version)**.** *Suppose $D \subset \mathbb{R}^2$ is a region whose boundary ∂D consists of a finite number of piecewise regular closed curves, and which can be decomposed into a finite number of regular regions.*

Then for any pair of $\mathcal{C}^1$ functions P and Q defined on D,

$$\oint_{\partial D} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA \tag{6.16}$$

where the line integral over the boundary ∂D is interpreted as the sum of line integrals over its constituent curves, each with boundary orientation.

As an example, consider the case

$$\begin{aligned} P(x,y) &= x+y \\ Q(x,y) &= -x \end{aligned}$$

and take as our region the annulus $D = \{(x,y) \mid 1 \leq x^2 + y^2 \leq 4\}$. This has two boundary components, the *outer* circle $\mathcal{C}_1 = \{(x,y) \mid x^2 + y^2 = 4\}$, for which the boundary orientation is *counterclockwise*, and the *inner* circle, $\mathcal{C}_2 = \{(x,y) \mid x^2 + y^2 = 4\}$, for which the boundary orientation is *clockwise.*

We parametrize the *outer* $\mathcal{C}_1$ circle via

$$\left\{ \begin{array}{rcl} x &=& 2\cos\theta \\ y &=& 2\sin\theta \end{array} \right. \qquad \left\{ \begin{array}{rcr} dx &=& -2\sin\theta\, d\theta \\ dy &=& 2\cos\theta\, d\theta \end{array} \right., \qquad 0 \leq \theta \leq 2\pi.$$

Also, along $\mathcal{C}_1$,

$$\begin{aligned} P(2\cos\theta, 2\sin\theta) &= 2(\cos\theta + \sin\theta) \\ Q(2\cos\theta, 2\sin\theta) &= -2\cos\theta \end{aligned}$$

so along $\mathcal{C}_1$, the form is

$$\begin{aligned} P\,dx + Q\,dy &= 2(\cos\theta + \sin\theta)(-2\sin\theta\, d\theta) + (-2\cos\theta)(2\cos\theta\, d\theta) \\ &= (-4\cos\theta\sin\theta - 4)\, d\theta \end{aligned}$$

leading to the integral

$$\begin{aligned} \int_{\mathcal{C}_1} P\,dx + Q\,dy &= \int_0^{2\pi} -4(\sin\theta\cos\theta + 1)\, d\theta \\ &= -8\pi. \end{aligned}$$

Now, the *inner* circle $\mathcal{C}_2$ needs to be parametrized *clockwise*; one way to do this is to reverse the two functions:

$$\left\{ \begin{array}{rcl} x &=& \sin\theta \\ y &=& \cos\theta \end{array} \right. \qquad \left\{ \begin{array}{rcr} dx &=& \cos\theta\, d\theta \\ dy &=& -\sin\theta\, d\theta \end{array} \right., \qquad 0 \leq \theta \leq 2\pi.$$

Then

$$\begin{aligned} P(\sin\theta, \cos\theta) &= (\sin\theta + \cos\theta) \\ Q(\sin\theta, \cos\theta) &= -\sin\theta \end{aligned}$$

so along $\mathcal{C}_2$, the form is

$$\begin{aligned} P\,dx + Q\,dy &= (\sin\theta + \cos\theta)(\cos\theta\,d\theta) + (-\sin\theta)(-\sin\theta\,d\theta) \\ &= (\sin\theta\cos\theta + 1)\,d\theta \end{aligned}$$

with integral

$$\begin{aligned} \int_{\mathcal{C}_2} P\,dx + Q\,dy &= \int_0^{2\pi} (\sin\theta\cos\theta + 1)\,d\theta \\ &= 2\pi. \end{aligned}$$

Combining these, we have

$$\oint_{\partial D} P\,dx + Q\,dy = \int_{\mathcal{C}_1} P\,dx + Q\,dy + \int_{\mathcal{C}_2} P\,dx + Q\,dy = -8\pi + 2\pi = -6\pi.$$

Let us compare this to the double integral:

$$\begin{aligned} \frac{\partial Q}{\partial x} &= \frac{\partial}{\partial x}[-x] \\ &= -1; \\ -\frac{\partial P}{\partial y} &= -\frac{\partial}{\partial y}[x+y] \\ &= -1 \end{aligned}$$

so

$$\begin{aligned} \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA &= \iint_D (-2) dA \\ &= -2\mathcal{A}(D) \\ &= -2(4\pi - \pi) \\ &= -6\pi. \end{aligned}$$

Green's Theorem in the Language of Vector Fields

If we think of the planar vector field

$$\vec{F}(x,y) = P(x,y)\,\vec{\imath} + Q(x,y)\,\vec{\jmath}$$

as the velocity of a fluid, then the line integral

$$\oint_{\mathcal{C}} P\,dx + Q\,dy = \oint_{\mathcal{C}} \vec{F} \cdot d\vec{\mathfrak{s}}$$

around a closed curve $\mathcal{C}$ is the integral of the tangent component of $\overrightarrow{F}$: thus it can be thought of as measuring the tendency of the fluid to flow around the curve; it is sometimes referred to as the **circulation** of $\overrightarrow{F}$ around $\mathcal{C}$.

The double integral in Green's Theorem is a bit more subtle. One way is to consider a "paddle wheel" immersed in the fluid, in the form of two line segments through a given point (a, b)—one horizontal, the other vertical (Figure 6.16.

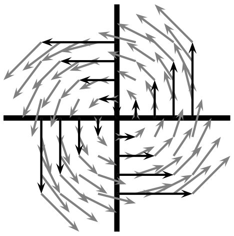


Figure 6.16: Rotation of a Vector Field: the "Paddle Wheel"

When will the wheel tend to turn? Let us first concentrate on the horizontal segment. Intuitively, the horizontal component of velocity will have no turning effect (rather it will tend to simply displace the paddle horizontally). Similarly, a vertical velocity field which is *constant* along the length of the paddle will result in a vertical (parallel) displacement. A *turning* of the paddle will result from a monotone *change* in the vertical component of the velocity as one moves left-to-right along the paddle. In particular, counterclockwise turning requires that the vertical component Q of the velocity *increases* as we move left-to-right: that is, the horizontal paddle will tend to turn counterclockwise around (a, b) if $\frac{\partial Q}{\partial x}(a, b) > 0$. This is sometimes referred to as a **shear** effect of the vector field.

Now consider the vertical "paddle". Again, the velocity component tangent to the paddle, as well as a *constant* horizontal velocity will effect a parallel displacement: to obtain a shear effect, we need the *horizontal* component of velocity to be changing monotonically as we move vertically. Note that in this case, a *counterclockwise* rotation results from a vertical velocity component that is *decreasing* as we move up along the paddle: $\frac{\partial P}{\partial y}(a, b) < 0$.

Since the paddle wheel is rigid, the effect of these two shears will be cumulative,and the net counterclockwise rotation effect of the two shears will be given by $\frac{\partial Q}{\partial x}(a, b) - \frac{\partial P}{\partial y}(a, b)$.

This discussion comes with an immediate disclaimer: it is purely intuitive; a more rigorous derivation of this expression as representing the tendency to turn is given in Exercise 6 using Green's Theorem. However, it helps motivate our designation of this as the (planar) **curl**[9] of the vector field $\overrightarrow{F}$:

$$\operatorname{curl} \overrightarrow{F} = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}. \tag{6.17}$$

With this terminology, we can formulate Green's Theorem in the language of vector fields as follows:

Theorem 6.3.7 (Green's Theorem: Vector Version)**.** *If $\overrightarrow{F} = P\overrightarrow{\imath} + Q\overrightarrow{\jmath}$ is a $\mathcal{C}^1$ vector field on the planar region $D \subset \mathbb{R}^2$, and D has a piecewise regular boundary and can be subdivided into regular regions, then the circulation of $\overrightarrow{F}$ around the boundary of D (each constituent simple closed curve of ∂D given the boundary orientation) equals the integral over the region D of the (planar) curl of $\overrightarrow{F}$:*

$$\oint_{\partial D} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} = \oint_{\partial D} \overrightarrow{F} \cdot \overrightarrow{T}\, d\mathfrak{s} = \iint_D \operatorname{curl} \overrightarrow{F}\, dA.$$

Exercises for § 6.3

Practice problems:

1. Evaluate $\oint_{\mathcal{C}} \omega$ for each form below, where $\mathcal{C}$ is the circle $x^2 + y^2 = R^2$ traversed counterclockwise, two ways: directly, and using Green's Theorem:

   (a) $\omega = y\,dx + x\,dy$

   (b) $\omega = x\,dx + y\,dy$

   (c) $\omega = xy^2\,dx + x^2y\,dy$

   (d) $\omega = (x - y)\,dx + (x + y)\,dy$

   (e) $\omega = xy\,dx + xy\,dy$

2. Evaluate $\oint_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}}$ for each vector field below, where $\mathcal{C}$ is the circle $x^2 + y^2 = 1$ traversed counterclockwise, two ways: directly, and using Green's Theorem:

   (a) $\overrightarrow{F} = x\overrightarrow{\imath} - (x + y)\overrightarrow{\jmath}$

[9] We shall see in § 6.5 that the "true" curl of a vector field is a vector in $\mathbb{R}^3$; the present quantity is just one of its components.

(b) $\overrightarrow{F} = 3y\overrightarrow{\imath} - x\overrightarrow{\jmath}$

(c) $\overrightarrow{F} = 3x\overrightarrow{\imath} - y\overrightarrow{\jmath}$

(d) $\overrightarrow{F} = -x^2y\overrightarrow{\imath} + xy^2\overrightarrow{\jmath}$

(e) $\overrightarrow{F} = y^3\overrightarrow{\imath} - x^3\overrightarrow{\jmath}$

(f) $\overrightarrow{F} = A\overrightarrow{x}$, where

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

3. Calculate the line integral $\oint_{\mathcal{C}}(e^x - y)\,dx + (e^y + x)\,dy$, where

   (a) $\mathcal{C}$ is the polygonal path from $(0,0)$ to $(1,0)$ to $(1,1)$ to $(0,1)$ to $(0,0)$.

   (b) $\mathcal{C}$ is the circle $x^2 + y^2 = R^2$, traversed counterclockwise.

Theory problems:

4. (a) Show that the area of the region bounded by a simple closed curve $\mathcal{C}$ is given by any one of the following three integrals:

$$\begin{aligned} A &= \int_{\mathcal{C}} x\,dy \\ &= \int_{\mathcal{C}} y\,dx \\ &= \frac{1}{2}\int_{\mathcal{C}} x\,dy - y\,dx. \end{aligned}$$

   (b) Use this to find the area bounded by one arc of the cycloid

$$\begin{aligned} x &= a(\theta - \sin\theta) \\ y &= a(1 - \cos\theta). \end{aligned}$$

5. (a) Show that the area of the region bounded by a curve $\mathcal{C}$ expressed in polar coordinates as

$$r = f(\theta)$$

   is given by

$$A = \frac{1}{2}\int_{\mathcal{C}} (f(\theta))^2\,d\theta.$$

(b) Use this to find the area of the rose

$$r = \sin n\theta.$$

(Caution: the answer is different for n even and n odd; in particular, when n is even, the curve traverses the $2n$ leaves once as θ goes from 0 to 2π, while for n odd, it traverses the n leaves twice in that time interval.)

Challenge problems:

6. (a) Show that a rotation of the plane (about the origin) with angular velocity ω gives a (spatial) velocity vector field which at each point away from the origin is given by

$$r\overrightarrow{\omega}(x,y) = r\omega\overrightarrow{T}$$

where

$$\overrightarrow{T}(x,y) = -\frac{y}{r}\overrightarrow{\imath} + \frac{x}{r}\overrightarrow{\jmath}$$

is the unit vector, perpendicular to the ray through (x, y), pointing counterclockwise, and r is the distance from the origin.

(b) Show that the circulation of $r\overrightarrow{\omega}(x,y)$ around the circle of radius r centered at the origin is $2\pi r^2\omega$.

(c) Now suppose the vector field

$$\overrightarrow{F} = P\overrightarrow{\imath} + Q\overrightarrow{\jmath}$$

is the velocity vector field of a planar fluid. Given a point $\overrightarrow{p}_0$ and $\overrightarrow{p} \neq \overrightarrow{p}_0$, let $\overrightarrow{T}_0(\overrightarrow{p})$ be the unit vector perpendicular to the ray from $\overrightarrow{p}_0$ to $\overrightarrow{p}$, pointing counterclockwise, and define $\omega_0(\overrightarrow{p})$ by

$$r\omega_0(\overrightarrow{p}) = \overrightarrow{F}(\overrightarrow{p}) \cdot \overrightarrow{T}_0(\overrightarrow{p})$$

where $r = \|\overrightarrow{p} - \overrightarrow{p}_0\|$ is the distance of $\overrightarrow{p}$ from $\overrightarrow{p}_0$; in other words,

$$\begin{aligned} r\overrightarrow{\omega}_0(\overrightarrow{p}) &:= r\omega_0(\overrightarrow{p})\,\overrightarrow{T}_0(v') \\ &= \overrightarrow{T}_0(\overrightarrow{p}) \circ \overrightarrow{F} \end{aligned}$$

is the component of $\overrightarrow{F}$ perpendicular to the ray from $\overrightarrow{p}_0$ to $\overrightarrow{p}$. Show that

$$r\omega_0(\overrightarrow{p}) = (Q(x,y)\,\overrightarrow{\imath} - P(x,y)\,\overrightarrow{\jmath}) \cdot \overrightarrow{T}_0(\overrightarrow{p})\,.$$

(d) Let $\mathcal{C}_r$ be the circle of radius r about $\overrightarrow{p}_0$. Show that the circulation of $\overrightarrow{F}$ around $\mathcal{C}_r$ equals the circulation of $r\overrightarrow{\omega}_0(\overrightarrow{p})$ around $\mathcal{C}_r$, and hence the average value of $\omega_0(\overrightarrow{p})$ around $\mathcal{C}_r$ is

$$\omega(r) = \frac{1}{2\pi r^2}\oint_{\mathcal{C}_r} Q\,dx - P\,dy.$$

(e) Use Green's Theorem to show that this equals half the average value of the scalar curl of $\overrightarrow{F}$ over the disc of radius r centered at $\overrightarrow{p}_0$.

(f) Use the Integral Mean Value Theorem to show that

$$\lim_{r\to 0}\omega(r) = \frac{1}{2}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right).$$

7. Given the region $D \subset \mathbb{R}^2$ bounded by a simple closed curve $\mathcal{C}$ (with positive orientation) and a vector field $\overrightarrow{F} = P\overrightarrow{\imath} + Q\overrightarrow{\jmath}$ on $\mathcal{C}$, show that

$$\oint_{\mathcal{C}} \overrightarrow{F} \cdot \overrightarrow{N}\,d\mathfrak{s} = \iint_D \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dA$$

where $\overrightarrow{N}$ is the outward pointing unit normal to $\mathcal{C}$.

(*Hint:* Rotate $\overrightarrow{F}$ and $\overrightarrow{N}$ in such a way that $\overrightarrow{N}$ is rotated into the tangent vector $\overrightarrow{T}$. What happens to $\overrightarrow{F}$? Now apply Green's Theorem.)

8. **Green's identities**: Given a $\mathcal{C}^2$ function, define the **Laplacian** of f as

$$\nabla^2 f := \operatorname{div}\overrightarrow{\nabla} f = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y}$$

.

Furthermore, if $D \subset \mathbb{R}^2$ is a regular region, define $\frac{\partial f}{\partial n} = \overrightarrow{\nabla} f \cdot \overrightarrow{N}$ on ∂D, where $\overrightarrow{N}$ is the outward unit normal to ∂D.

Suppose f and g are $\mathcal{C}^2$ functions on D.

(a) Prove

$$\iint_D (f\nabla^2 g + \overrightarrow{\nabla} f \cdot \overrightarrow{\nabla} g)\, dA = \oint_{\partial D} f \frac{dg}{dn}\, d\mathfrak{s}$$

(*Hint:* Use Exercise 7 with $P = -f\frac{\partial g}{\partial y}$, $Q = f\frac{\partial g}{\partial x}$.)

(b) Use this to prove

$$\iint_D (f\nabla^2 g - g\nabla^2 f)\, dA = \oint_{\partial D} \left(f\frac{\partial g}{\partial n} - g\frac{\partial f}{\partial n} \right) d\mathfrak{s}.$$

6.4 Green's Theorem and 2-forms in $\mathbb{R}^2$

Bilinear Functions and 2-Forms on $\mathbb{R}^2$

In § 6.1 we defined a differential form on $\mathbb{R}^2$ as assigning to each point $p \in \mathbb{R}^2$ a linear functional on the tangent space $T_p\mathbb{R}^2$ at p; we integrate these objects over curves. Green's Theorem (Theorem 6.3.4) relates the line integral of such a form over the *boundary* of a region to an integral over the region itself. In the language of forms, the objects we integrate over two-dimensional regions are called 2-*forms*. These are related to *bilinear functions.*

Definition 6.4.1. *A* ***bilinear function*** *on* $\mathbb{R}^2$ *is a function of two vector variables* $B(\overrightarrow{v}, \overrightarrow{w})$ *such that fixing one of the inputs results in a linear function of the other input:*

$$\begin{aligned} B(a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2, \overrightarrow{w}) &= a_1B(\overrightarrow{v}_1, \overrightarrow{w}) + a_2B(\overrightarrow{v}_2, \overrightarrow{w}) \\ B(\overrightarrow{v}, b_1\overrightarrow{w}_1 + b_2\overrightarrow{w}_2) &= b_1B(\overrightarrow{v}, \overrightarrow{w}_1) + b_2B(\overrightarrow{v}, \overrightarrow{w}_2) \end{aligned} \tag{6.18}$$

for arbitrary vectors in $\mathbb{R}^2$ *and real scalars.*

One example of a bilinear function, by Proposition 1.4.2, is the dot product: $B(\overrightarrow{v}, \overrightarrow{w}) = \overrightarrow{v} \cdot \overrightarrow{w}$. More generally, a bilinear function on $\mathbb{R}^2$ is a special kind of homogeneous degree two polynomial in the coordinates of its entries: using Equation (6.18), we see that if

$$\begin{aligned} \overrightarrow{v} &= (x_1, y_1) \\ &= x_1\overrightarrow{\imath} + y_1\overrightarrow{\jmath} \end{aligned}$$

and

$$\overrightarrow{w} = (x_2, y_2) = x_2\overrightarrow{\imath} + y_2\overrightarrow{\jmath},$$

then

$$\begin{aligned} B(\overrightarrow{v},\overrightarrow{w}) &= B(x_1\overrightarrow{\imath}+y_1\overrightarrow{\jmath},\overrightarrow{w}) \\ &= x_1B(\overrightarrow{\imath},\overrightarrow{w})+y_1B(\overrightarrow{\jmath},\overrightarrow{w}) \\ &= x_1B(\overrightarrow{\imath},x_2\overrightarrow{\imath}+y_2\overrightarrow{\jmath})+y_1B(\overrightarrow{\jmath},x_2\overrightarrow{\imath}+y_2\overrightarrow{\jmath}) \\ &= x_1x_2B(\overrightarrow{\imath},\overrightarrow{\imath})+x_1y_2B(\overrightarrow{\imath},\overrightarrow{\jmath})+y_1x_2B(\overrightarrow{\jmath},\overrightarrow{\imath})+y_1y_2B(\overrightarrow{\jmath},\overrightarrow{\jmath})\,. \end{aligned}$$

So if we write the values of B on the four pairs of basis vectors as

$$\begin{aligned} b_{11} &= B(\overrightarrow{\imath},\overrightarrow{\imath}) \\ b_{12} &= B(\overrightarrow{\imath},\overrightarrow{\jmath}) \\ b_{21} &= B(\overrightarrow{\jmath},\overrightarrow{\imath}) \\ b_{22} &= B(\overrightarrow{\jmath},\overrightarrow{\jmath}) \end{aligned}$$

then we can write B as the homogeneous degree two polynomial

$$B(\overrightarrow{v},\overrightarrow{w}) = b_{11}x_1x_2+b_{12}x_1y_2+b_{21}y_1x_2+b_{22}y_1y_2. \tag{6.19}$$

As an example, the dot product satisfies Equation (6.19) with $b_{ij}=1$ when $i=j$ and $b_{ij}=0$ when $i\neq j$. The fact that in this case the coefficient for v_iw_j is the same as that for v_jw_i ($b_{ij}=b_{ji}$) reflects the additional property of the dot product, that it is **commutative** ($\overrightarrow{v}\cdot\overrightarrow{w}=\overrightarrow{w}\cdot\overrightarrow{v}$).

By contrast, the bilinear functions which come up in the context of 2-forms are **anti-commutative**: for every pair of vectors $\overrightarrow{v}$ and $\overrightarrow{w}$, we require

$$B(\overrightarrow{w},\overrightarrow{v}) = -B(\overrightarrow{v},\overrightarrow{w})\,.$$

An anti-commutative, bilinear function on $\mathbb{R}^2$ will be referred to as a **2-form** on $\mathbb{R}^2$.

Note that an immediate consequence of anti-commutativity is that $B(\overrightarrow{v},\overrightarrow{w})=0$ if $\overrightarrow{v}=\overrightarrow{w}$ (Exercise 4). Applied to the basis vectors, these conditions tell us that

$$\begin{aligned} b_{11} &= 0 \\ b_{21} &= -b_{12} \\ b_{22} &= 0. \end{aligned}$$

Thus, a 2-form on $\mathbb{R}^2$ is determined by the value $b=B(\overrightarrow{\imath},\overrightarrow{\jmath})$:

$$B(\overrightarrow{v},\overrightarrow{w}) = b(x_1y_2-x_2y_1).$$

You might recognize the quantity in parentheses as the determinant

$$\Delta(\vec{v},\vec{w}) = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$$

from § 1.6, which gives the signed area of the parallelogram with sides $\vec{v}$ and $\vec{w}$: this is in fact a 2-form on $\mathbb{R}^2$, and every other such function is a constant multiple of it:

$$B(\vec{v},\vec{w}) = B(\vec{\imath},\vec{\jmath})\,\Delta(\vec{v},\vec{w})\,. \tag{6.20}$$

To bring this in line with our notation for 1-forms as $P\,dx + Q\,dy$, we reinterpret the entries in the determinant above as the values of the 1-forms dx and dy on $\vec{v}$ and $\vec{w}$; in general, we define the **wedge product** of two 1-forms α and β to be the determinant formed from applying them to a pair of vectors:

$$(\alpha\wedge\beta)(\vec{v},\vec{w}) := \begin{vmatrix} \alpha(\vec{v}) & \beta(\vec{v}) \\ \alpha(\vec{w}) & \beta(\vec{w}) \end{vmatrix}. \tag{6.21}$$

You should check that this is bilinear and anti-commutative in $\vec{v}$ and $\vec{w}$—that is, it is a 2-form (Exercise 5)—and that as a product, $\wedge$ is anti-commutative: for any two 1-forms α and β,

$$\beta\wedge\alpha = -\alpha\wedge\beta. \tag{6.22}$$

Using the wedge product notation, we see

Remark 6.4.2. *Every 2-form on $\mathbb{R}^2$ is a scalar multiple of $dx\wedge dy$.*

Now, we define a **differential 2-form** on a region $D\subset\mathbb{R}^2$ to be a mapping ω which assigns to each point $p\in D$ a 2-form ω_p on the tangent space $T_p\mathbb{R}^2$. From Remark 6.4.2, a differential 2-form on $D\subset\mathbb{R}^2$ can be written

$$\omega_p = b(p)\ dx\wedge dy$$

for some function b on D.

Finally, we define the integral of a differential 2-form ω over a region $D\subset\mathbb{R}^2$ to be

$$\iint_D \omega := \iint_D \omega_p(\vec{\imath},\vec{\jmath})\,dA;$$

that is,

$$\iint_D b(p)\ dx\wedge dy = \iint_D b\,dA.$$

Green's Theorem in the Language of Forms

To formulate Theorem 6.3.4 in terms of forms, we need two more definitions.

First, we define the **exterior product** of two differential 1-forms α and β on $D \subset \mathbb{R}^2$ to be the mapping $\alpha \wedge \beta$ assigning to $p \in D$ the wedge product of α_p and β_p:

$$(\alpha \wedge \beta)_p = \alpha_p \wedge \beta_p.$$

Second, we define the **exterior derivative** $d\omega$ of a 1-form ω. There are two basic kinds of 1-form on $\mathbb{R}^2$: $P\,dx$ and $Q\,dy$, where P (*resp.* Q) is a function of x and y. The differential of a function is a 1-form, and we take the exterior derivative of one of our basic 1-forms by finding the differential of the function and taking its exterior product with the coordinate 1-form it multiplied. This yields a 2-form:

$$\begin{aligned}
d(P\,dx) &= (dP) \wedge dx \\
&= \left(\frac{\partial P}{\partial x}\,dx + \frac{\partial P}{\partial y}\,dy\right) \wedge dx \\
&= \frac{\partial P}{\partial x}\,dx \wedge dx + \frac{\partial P}{\partial y}\,dy \wedge dx \\
&= 0 - \frac{\partial P}{\partial y}\,dx \wedge dy \\
d(Q\,dy) &= (dQ) \wedge dy \\
&= \left(\frac{\partial Q}{\partial x}\,dx + \frac{\partial Q}{\partial y}\,dy\right) \wedge dy \\
&= \frac{\partial Q}{\partial x}\,dx \wedge dy + \frac{\partial Q}{\partial y}\,dy \wedge dy \\
&= \frac{\partial Q}{\partial x}\,dx \wedge dy + 0.
\end{aligned}$$

We extend this definition to arbitrary 1-forms by making the derivative respect sums: if

$$\omega = P(x,y)\ dx + Q(x,y)\ dy,$$

then

$$d\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy.$$

To complete the statement of Theorem 6.3.4 in terms of forms, we recall the notation ∂D for the boundary of a region $D \subset \mathbb{R}^2$. Then we can restate Green's Theorem as

Theorem 6.4.3 (Green's Theorem, Differential Form)**.** *Suppose $D \subset \mathbb{R}^2$ is a region bounded by the curve*

$$\mathcal{C} = \partial D$$

and D and ∂D are both positively oriented. Then for any differential 1-form ω on D,

$$\oint_{\partial D} \omega = \iint_D d\omega. \tag{6.23}$$

Exercises for § 6.4

Practice problems:

1. Evaluate $\omega_p(\overrightarrow{\imath}, \overrightarrow{\jmath})$ and $\omega(p)\overrightarrow{v},\overrightarrow{w}$, where ω, $\overrightarrow{p}$, $\overrightarrow{v}$, and $\overrightarrow{w}$ are as given.

   (a) $\omega = dx \wedge dy$, $p = (2,1)$, $\overrightarrow{v} = 2\overrightarrow{\imath} - 3\overrightarrow{\jmath}$, $\overrightarrow{w} = 3\overrightarrow{\imath} - 2\overrightarrow{\jmath}$.

   (b) $\omega = x^2\, dx \wedge dy$, $p = (2,1)$, $\overrightarrow{v} = \overrightarrow{\imath} + \overrightarrow{\jmath}$, $\overrightarrow{w} = 2\overrightarrow{\imath} - \overrightarrow{\jmath}$.

   (c) $\omega = x^2\, dx \wedge dy$, $p = (-2,1)$, $\overrightarrow{v} = 2\overrightarrow{\imath} + \overrightarrow{\jmath}$, $\overrightarrow{w} = 4\overrightarrow{\imath} + 2\overrightarrow{\jmath}$.

   (d) $\omega = (x^2 + y^2)\, dx \wedge dy$, $p = (1,-1)$, $\overrightarrow{v} = 3\overrightarrow{\imath} - \overrightarrow{\jmath}$, $\overrightarrow{w} = \overrightarrow{\jmath} - \overrightarrow{\imath}$.

   (e) $\omega = (x\, dx) \wedge (y\, dy)$, $p = (1,1)$, $\overrightarrow{v} = 2\overrightarrow{\imath} - \overrightarrow{\jmath}$, $\overrightarrow{w} = 2\overrightarrow{\imath} + \overrightarrow{\jmath}$.

   (f) $\omega = (y\, dy) \wedge (y\, dx)$, $p = (1,1)$, $\overrightarrow{v} = 2\overrightarrow{\imath} - \overrightarrow{\jmath}$, $\overrightarrow{w} = 2\overrightarrow{\imath} + \overrightarrow{\jmath}$.

   (g) $\omega = (y\, dy) \wedge (x\, dx)$, $p = (1,1)$, $\overrightarrow{v} = 2\overrightarrow{\imath} - \overrightarrow{\jmath}$, $\overrightarrow{w} = 2\overrightarrow{\imath} + \overrightarrow{\jmath}$.

   (h) $\omega = (x\, dx + y\, dy) \wedge (x\, dx - y\, dy)$, $p = (1,1)$, $\overrightarrow{v} = 2\overrightarrow{\imath} - \overrightarrow{\jmath}$, $\overrightarrow{w} = 2\overrightarrow{\imath} + \overrightarrow{\jmath}$.

2. Evaluate $\iint_{[0,1]\times[0,1]} \omega$:

   (a) $\omega = x\, dx \wedge y\, dy$

   (b) $\omega = x\, dy \wedge y\, dx$

   (c) $\omega = y\, dx \wedge x\, dy$

   (d) $\omega = y\, dy \wedge y\, dx$

   (e) $\omega = (x\, dx + y\, dy) \wedge (x\, dy - y\, dx)$

3. Find the exterior derivative of each differential 1-form below.

   (a) $\omega = xy\,dx + xy\,dy$

   (b) $\omega = x\,dx + y\,dy$

   (c) $\omega = y\,dx + x\,dy$

   (d) $\omega = (x^2 + y^2)\,dx + 2xy\,dy$

   (e) $\omega = \cos xy\,dx + \sin x\,dy$

   (f) $\omega = y\sin x\,dx + \cos x\,dy$

   (g) $\omega = y\,dx - x\,dy$

   (h) $\omega = \frac{y\,dx - x\,dy}{\sqrt{x^2+y^2}}$

   (i) $\omega = \frac{y\,dx + x\,dy}{\sqrt{x^2+y^2}}$

Theory problems:

4. Show that if B is an anti-commutative 2-form, then for any vector $\vec{v}$, $B(\vec{v}, \vec{v}) = 0$.

5. (a) Show that Equation (6.21) defines a 2-form: that is, the wedge product of two 1-forms is a bilinear and anti-commutative functional.

   (b) Show that, as a product, the wedge product is anti-commutative (*i.e.*, Equation (6.22)).

6. Show that if $f(x, y)$ is a $\mathcal{C}^2$ function, and $\omega = df$ is its differential, then $d\omega = 0$.

6.5 Stokes' Theorem

Oriented Surfaces and Flux Integrals

We saw in § 6.1 that a vector field can be usefully integrated over a curve in $\mathbb{R}^2$ or $\mathbb{R}^3$ by taking the path integral of its component tangent to the curve; the resulting line integral (Definition 6.1.1) depends on the orientation of the curve, but otherwise depends only on the curve as a point-set.

There is an analogous way to integrate a vector field in $\mathbb{R}^3$ over a *surface*, by taking the surface integral of the component *normal* to the surface. There are two choices of normal vector at any point of a surface; if one makes a choice continuously at all points of a surface, one has an orientation of the surface.

Definition 6.5.1. *Suppose $\mathfrak{S}$ is a regular surface in $\mathbb{R}^3$.*

*An **orientation** of $\mathfrak{S}$ in $\mathbb{R}^3$ is a vector field $\overrightarrow{n}$ defined at all points of $\mathfrak{S}$ such that*

1. *$\overrightarrow{n}(p) \in T_p\mathbb{R}^3$ is normal to $\mathfrak{S}$ (that is, it is perpendicular to the plane tangent to $\mathfrak{S}$ at p);*
2. *$\overrightarrow{n}(p)$ is a unit vector ($\|\overrightarrow{n}(p)\| = 1$ for all $p \in \mathfrak{S}$);*
3. *$\overrightarrow{n}(p)$ varies continuously with $p \in \mathfrak{S}$.*

*An **oriented surface** is a regular surface $\mathfrak{S} \subset \mathbb{R}^3$, together with an orientation $\overrightarrow{n}$ of $\mathfrak{S}$.*

Recall (from § 3.5) that a **coordinate patch** is a regular, one-to-one mapping $\overrightarrow{p}\colon \mathbb{R}^2 \to \mathbb{R}^3$ of a plane region D into $\mathbb{R}^3$; by abuse of terminology, we also refer to the image $\mathfrak{S} \subset \mathbb{R}^3$ of such a mapping as a coordinate patch. If we denote the parameters in the domain of $\overrightarrow{p}$ by $(s,t) \in D$, then since by regularity $\frac{\partial \overrightarrow{p}}{\partial s}$ and $\frac{\partial \overrightarrow{p}}{\partial t}$ are linearly independent at each point of D, their cross product gives a vector normal to $\mathfrak{S}$ at $\overrightarrow{p}(s,t)$. Dividing this vector by its length gives an orientation of $\mathfrak{S}$, determined by the order of the parameters: the cross product in reverse order gives the "opposite" orientation of $\mathfrak{S}$.

At any point of $\mathfrak{S}$, there are only two directions normal to $\mathfrak{S}$, and once we have picked this direction at *one* point, there is only one way to extend this to a continuous vector field normal to $\mathfrak{S}$ at *nearby* points of $\mathfrak{S}$. Thus

Remark 6.5.2. *A coordinate patch $\overrightarrow{p}\colon \mathbb{R}^2 \to \mathbb{R}^3$ with domain in (s,t)-space and image $\mathfrak{S} \in \mathbb{R}^3$ has two orientations. The orientation*

$$\overrightarrow{n} = \frac{\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}}{\left\|\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right\|} \tag{6.24}$$

*is the **local orientation** of $\mathfrak{S}$ **induced** by the mapping $\overrightarrow{p}$, while the opposite orientation is*

$$-\overrightarrow{n} = \frac{\frac{\partial \overrightarrow{p}}{\partial t} \times \frac{\partial \overrightarrow{p}}{\partial s}}{\left\|\frac{\partial \overrightarrow{p}}{\partial t} \times \frac{\partial \overrightarrow{p}}{\partial s}\right\|}$$

In general, a regular surface in $\mathbb{R}^3$ is a union of (overlapping) coordinate patches, and each can be given a local orientation; if two patches overlap, we say the two corresponding local orientations are **coherent** if at each overlap point the normal vectors given by the two local orientations are the same.

In that case we have an orientation on the *union* of these patches. If we have a family of coordinate patches such that on any overlap the orientations are coherent, then we can fit these together to give a **global orientation** of the surface. Conversely, if we have an orientation of a regular surface, then we can cover it with overlapping coordinate patches for which the induced local orientations are coherent (Exercise 5).

However, not every regular surface in $\mathbb{R}^3$ can be given a global orientation. Consider the **Möbius band**, obtained by taking a rectangle and joining a pair of parallel sides but with a twist (Figure 6.17). This was named after A. F. Möbius (1790-1860). [10]

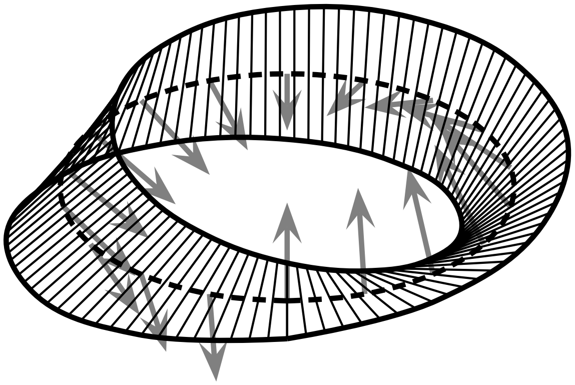


Figure 6.17: A Möbius Band

One version of the Möbius band is the image of the mapping defined by

$$\begin{aligned} x(s,t) &= \left(3 + t\cos\frac{s}{2}\right)\cos s \\ y(s,t) &= \left(3 + t\cos\frac{s}{2}\right)\sin s \\ z(s,t) &= t\sin\frac{s}{2} \end{aligned}$$

where t is restricted to $|t| \leq 1$. Geometrically, the central circle corresponding to $t = 0$ is a horizontal circle of radius 3 centered at the origin. For a fixed value of s, the interval $-1 \leq t \leq 1$ is mapped to a line segment,

[10] See footnote on p. 46.

centered on this circle: as s increases over an interval of length 2π, this segment rotates in the plane perpendicular to the circle by an angle of π. This means that the two intervals corresponding to s and to $s+2\pi$ are mapped to the same line segment, but in opposite directions. In different terms, the vector $\frac{\partial \overrightarrow{p}}{\partial s}(s,0)$ always points along the central circle, in a counterclockwise direction (viewed from above); the vector $\frac{\partial \overrightarrow{p}}{\partial t}(s,0)$ is always perpendicular to it: the two points $\overrightarrow{p}(s,0)$ and $\overrightarrow{p}(s+2\pi,0)$ are the same, but the two vectors $\frac{\partial \overrightarrow{p}}{\partial t}(s,0)$ and $\frac{\partial \overrightarrow{p}}{\partial t}(s+2\pi,0)$ point in *opposite* directions. Now, if we start at $\overrightarrow{p}(s,0)$ with a normal parallel to the cross product $\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}$, then a continuation of this normal field along the central circle continues to point in the direction of $\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}$; however, when we return to the same position (but corresponding to an s-value 2π higher), this direction is opposite to the one we have already chosen there. This surface is **non-orientable**: it is impossible to give it a global orientation.

We shall henceforth consider only **orientable** surfaces in our theory.

With this definition, we can proceed to define the flux integral of a vector field over an oriented surface. Recall that in § 5.4, to define the surface integral $\iint_{\mathfrak{S}} f \, d\mathcal{S}$ of a function f over a regular surface $\mathfrak{S}$, we subdivided the domain of a parametrization into rectangles, approximating the area of each rectangle by the area $\triangle\mathcal{S}$ of a corresponding parallelogram in the tangent space, then multiplied each such area by the value of the function at a representative point of the rectangle, and finally added these to form a Riemann sum; as the mesh size of our subdivision went to zero, these Riemann sums converged to an integral independent of the parametrization from which we started.

To define the flux integral of a vector field $\overrightarrow{F}$ on an oriented regular surface $\mathfrak{S}$, we replace the element of surface area

$$d\mathcal{S} = \left\| \frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t} \right\| ds\, dt$$

with the **element of oriented surface area**

$$d\overrightarrow{\mathcal{S}} = \left(\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t} \right) ds\, dt.$$

We know that the vector $\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}$ is perpendicular to $\mathfrak{S}$, so either it points in the same direction as the unit normal $\overrightarrow{n}$ defining the orientation of $\mathfrak{S}$ or it points in the opposite direction. In the latter case, we modify our definition of $d\overrightarrow{\mathcal{S}}$ by taking the cross product in the opposite order. With this modification (if necessary) we can write

$$d\overrightarrow{\mathcal{S}} = \overrightarrow{n}\, d\mathcal{S} \tag{6.25}$$

and instead of multiplying the (scalar) element of surface area by the (scalar) function f, we take the *dot* product of the vector field $\overrightarrow{F}$ with the (vector) element of oriented surface area $d\overrightarrow{\mathcal{S}}$; the corresponding limit process amounts to taking the surface integral of the function obtained by dotting the vector field with the unit normal giving the orientation:

Definition 6.5.3. *Suppose $\overrightarrow{F}$ is a $\mathcal{C}^1$ vector field on $\mathbb{R}^3$, defined on a region $D \subset \mathbb{R}^3$, and $\mathfrak{S}$ is an oriented surface contained in D.*

*The **flux integral** of $\overrightarrow{F}$ over $\mathfrak{S}$ is defined as*

$$\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} = \iint_{\mathfrak{S}} \overrightarrow{F} \cdot \overrightarrow{n}\; d\mathcal{S}$$

where $\overrightarrow{n}$ is the unit normal defining the orientation of $\mathfrak{S}$, and $d\overrightarrow{\mathcal{S}}$ is the element of oriented surface area defined by Equation (6.25).

If we think of the vector field $\overrightarrow{F}$ as the velocity field of a fluid (say with constant density), then the flux integral is easily seen to express the amount of fluid crossing the surface $\mathfrak{S}$ per unit time. This also makes clear the fact that reversing the orientation of $\mathfrak{S}$ reverses the sign of the flux integral. On a more formal level, replacing $\overrightarrow{n}$ with its negative in the flux integral means we are taking the surface integral of the negative of our original function, so the integral also switches sign.

We saw in Corollary 4.4.6 that two different regular parametrizations $\overrightarrow{p}$ and $\overrightarrow{q}$ of the same surface $\mathfrak{S}$ differ by a change-of-coordinates transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ whose Jacobian determinant is nowhere zero, and from this we argued that the surface integral of a function does not depend on the parametrization. Thus, provided we pick the correct unit normal $\overrightarrow{n}$, the flux integral is independent of parametrization.

Note that calculating the unit normal vector $\overrightarrow{n}$ in the surface-integral version of the flux integral involves finding the cross product $\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}$ and then dividing by its length; but then the element of surface area $d\mathcal{S}$ equals that same length times $ds\, dt$, so these lengths cancel and at least the calculation of the length is redundant. If we just use the formal definition of the element of oriented area

$$d\overrightarrow{\mathcal{S}} = \left(\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right) ds\, dt$$

and take its formal dot product with the vector field $\overrightarrow{F}$ (expressed in terms of the parametrization), we get the correct integrand without performing the redundant step.

However, we do need to pay attention to the direction of the unit normal $\overrightarrow{n}$, which is the same as the direction of the vector $\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}$. It is usually a fairly simple matter to decide whether this cross product points in the correct direction; if it does not, we simply use its negative, which is the same as the cross product in the opposite order.

To see how this works, consider the vector field

$$\overrightarrow{F}(x,y,z) = x^2y\overrightarrow{\imath} + yz^2\overrightarrow{\jmath} + xyz\,\overrightarrow{k}$$

over the surface

$$z = xy, \quad \begin{cases} 0 \leq & x \leq & 1 \\ 0 \leq & y \leq & 1 \end{cases},$$

with *upward* orientation—that is, we want $\overrightarrow{n}$ to have a positive z-component. (See Figure 6.18.)

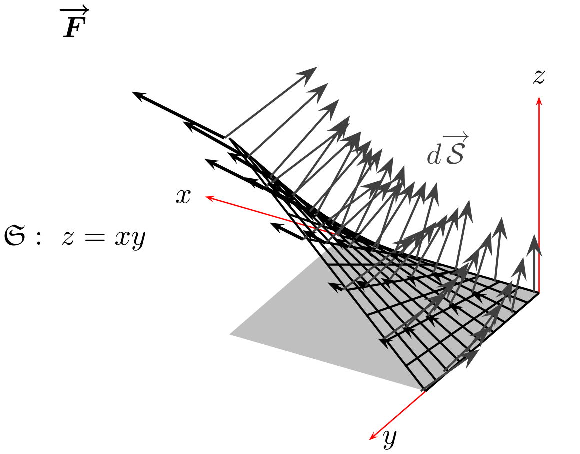


Figure 6.18: $\vec{F}(x,y,z) = x^2y\vec{\imath} + yz^2\vec{\jmath} + xyz\vec{k}$ on $z = xy$

Since this is the graph of a function, it is a coordinate patch, with the natural parametrization

$$\begin{cases} x &= & s \\ y &= & t \\ z &= & st \end{cases}, \quad \begin{cases} 0 \leq & s & \leq 1 \\ 0 \leq & t & \leq 1 \end{cases}.$$

In vector terms, this is

$$\overrightarrow{p}(s,t) = (s,t,st)$$

so

$$\frac{\partial \overrightarrow{p}}{\partial s} = (1,0,t)$$

and

$$\frac{\partial \overrightarrow{p}}{\partial t} = (0,1,s).$$

Then

$$\begin{aligned}\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t} &= (1,0,t) \times (0,1,s) \\ &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ 1 & 0 & t \\ 0 & 1 & s \end{vmatrix} \\ &= -t\overrightarrow{\imath} - s\overrightarrow{\jmath} + \overrightarrow{k}\end{aligned}$$

Note that this has an upward vertical component, so corresponds to the correct (upward) orientation. Thus we can write

$$d\overrightarrow{\mathcal{S}} = (-t\overrightarrow{\imath} - s\overrightarrow{\jmath} + \overrightarrow{k})\, ds\, dt.$$

In terms of the parametrization, the vector field along $\mathfrak{S}$ becomes

$$\begin{aligned}\overrightarrow{F}(\overrightarrow{p}(s,t)) &= (s)^2(t)\overrightarrow{\imath} + (t)(st)^2\overrightarrow{\jmath} + (s)(t)(st)\overrightarrow{k} \\ &= s^2t\overrightarrow{\imath} + s^2t^3\overrightarrow{\jmath} + s^2t^2\overrightarrow{k}\end{aligned}$$

giving

$$\begin{aligned}\overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} &= \left[\left(s^2t\right)(-t) + \left(s^2t^3\right)(-s) + \left(s^2t^2\right)(1)\right] ds\, dt \\ &= \left[-s^2t^2 - s^3t^3 + s^2t^2\right] ds\, dt \\ &= -s^3t^3\, ds\, dt\end{aligned}$$

and the flux integral becomes

$$\begin{aligned}\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} &= \int_0^1 \int_0^1 \left(-s^3t^3\right)\,ds\,dt\\ &= \int_0^1 -\frac{s^4}{4}\bigg|_0^1 t^3\,dt\\ &= -\frac{1}{4}\int_0^1 t^3\,dt\\ &= -\frac{t^4}{16}\bigg|_0^1\\ &= -\frac{1}{16}.\end{aligned}$$

We note in passing that for a surface given as the graph of a function, $z = f(x, y)$, the natural parametrization using the input to the function as parameters

$$\begin{cases} x &= & s\\ y &= & t\\ z &= & f(s,t)\end{cases}$$

leads to a particularly simple form for the element of oriented surface area $d\overrightarrow{\mathcal{S}}$. The proof is a straightforward calculation, which we leave to you (Exercise 6):

Remark 6.5.4. *If $\mathfrak{S}$ is the graph of a function $z = f(x, y)$, then the natural parametrization*

$$\overrightarrow{p}(s,t) = s\,\overrightarrow{\imath} + t\,\overrightarrow{\jmath} + f(s,t)\,\overrightarrow{k}$$

with orientation upward *has element of surface area*

$$\begin{aligned} d\overrightarrow{\mathcal{S}} &= \left(\frac{\partial \overrightarrow{p}}{\partial x} \times \frac{\partial \overrightarrow{p}}{\partial y}\right) dx\,dy\\ &= \left(-f_x\,\overrightarrow{\imath} - f_y\,\overrightarrow{\jmath} + \overrightarrow{k}\right) dx\,dy.\end{aligned}$$

As a second example, let

$$\overrightarrow{F}(x,y,z) = 2x\,\overrightarrow{\imath} + 2y\,\overrightarrow{\jmath} + 8z\,\overrightarrow{k}$$

and take as $\mathfrak{S}$ the portion of the sphere of radius 1 about the origin lying between the xy-plane and the plane $z = 0.5$, with orientation *into* the sphere (Figure **??**).

The surface is most naturally parametrized using spherical coordinates:

$$\left\{\begin{array}{rcr} x & = & \sin\phi\cos\theta \\ y & = & \sin\phi\sin\theta \\ z & = & \cos\phi \end{array}\right. , \quad \left\{\begin{array}{lll} \frac{\pi}{3} & \leq \phi & \leq \frac{\pi}{2} \\ 0 & \leq \theta & \leq 2\pi \end{array}\right. ;$$

the partial derivatives of this parametrization are

$$\begin{aligned} \frac{\partial\overrightarrow{p}}{\partial\phi} &= \cos\phi\cos\theta\overrightarrow{\imath} + \cos\phi\sin\theta\overrightarrow{\jmath} - \sin\phi\overrightarrow{k} \\ \frac{\partial\overrightarrow{p}}{\partial\theta} &= -\sin\phi\sin\theta\overrightarrow{\imath} + \sin\phi\cos\theta\overrightarrow{\jmath} \end{aligned}$$

leading to

$$\begin{aligned} \frac{\partial\overrightarrow{p}}{\partial\phi} \times \frac{\partial\overrightarrow{p}}{\partial\theta} &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ \cos\phi\cos\theta & \cos\phi\sin\theta & -\sin\phi \\ -\sin\phi\sin\theta & \sin\phi\cos\theta & 0 \end{vmatrix} \\ &= \sin^2\phi\cos\theta\overrightarrow{\imath} + \sin^2\phi\sin\theta\overrightarrow{\jmath} + \sin\phi\cos\phi(\cos^2\theta + \sin^2\theta)\overrightarrow{k} \\ &= \sin^2\phi(\cos\theta\overrightarrow{\imath} + \sin\theta\overrightarrow{\jmath}) + \sin\phi\cos\phi\overrightarrow{k}. \end{aligned}$$

Does this give the appropriate orientation? Since the sphere is orientable, it suffices to check this at one point: say at $\overrightarrow{p}\left(\frac{\pi}{2},0\right) = (1,0,0)$: here $\frac{\partial\overrightarrow{p}}{\partial\phi}\times\frac{\partial\overrightarrow{p}}{\partial\theta} = \overrightarrow{\imath}$, which points *outward* instead of inward. Thus we need to use the cross product in the other order (which means the negative of the vector above) to set

$$d\overrightarrow{\mathcal{S}} = -\{\sin^2\phi(\cos\theta\overrightarrow{\imath} + \sin\theta\overrightarrow{\jmath}) - \sin\phi\cos\phi\overrightarrow{k}\}\,d\phi\,d\theta.$$

In terms of this parametrization,

$$\overrightarrow{F} = 2\sin\phi\cos\theta\overrightarrow{\imath} + 2\sin\phi\sin\theta\overrightarrow{\jmath} + 8\cos\phi\overrightarrow{k}$$

so

$$\begin{aligned} \overrightarrow{F}\cdot d\overrightarrow{\mathcal{S}} &= (-2\sin^3\phi\cos^2\theta - 2\sin^3\phi\sin^2\theta - 8\sin\phi\cos^2\phi)\,d\phi\,d\theta \\ &= (-2\sin^3\phi - 8\sin\phi\cos^2\phi)\,d\phi\,d\theta \end{aligned}$$

and the integral becomes

$$\begin{aligned}
\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} &= \int_0^{2\pi} \int_{\pi/3}^{\pi/2} (-2\sin^3\phi - 8\sin\phi\cos^2\phi)\, d\phi\, d\theta \\
&= \int_0^{2\pi} \int_{\pi/3}^{\pi/2} -2\sin\phi(1-\cos^2\phi + 4\cos^2\phi)\, d\phi\, d\theta \\
&= 2\int_0^{2\pi} \int_{\pi/3}^{\pi/2} (1+3\cos^2\phi)(d(\cos\phi))\, d\theta \\
&= 2\int_0^{2\pi} \left(\cos\phi + \cos^3\phi\right)_{\pi/3}^{\pi/2} d\theta \\
&= 2\int_0^{2\pi} -\left(\frac{\sqrt{3}}{2} + \frac{3\sqrt{3}}{8}\right) d\theta \\
&= -4\pi\left(\frac{7\sqrt{3}}{8}\right) \\
&= -\frac{7\pi\sqrt{3}}{2}.
\end{aligned}$$

The Curl of a Vector Field

Let us revisit the discussion of (planar) curl for a vector field in the plane, from the end of § 6.3. There, we looked at the effect of a local shear in a vector field, which tends to rotate a line segment around a given point. The main observation was that for a segment parallel to the x-axis, the component of the vector field in the direction of the x-axis, as well as the actual value of the component in the direction of the y-axis, are irrelevant: the important quantity is the *rate of change* of the y-component in the x-direction. A similar analysis applies to a segment parallel to the y-axis: the important quantity is then the rate of change in the y-direction of the x-component of the vector field—more precisely, of the component of the vector field in the direction of the unit vector which is a right angle *counterclockwise* from the unit vector $\overrightarrow{\jmath}$. That is, we are looking at the directional derivative, in the direction of $\overrightarrow{\jmath}$, of the component of our vector field in the direction of $-\overrightarrow{\imath}$.

How do we extend this analysis to a segment and vector field in 3-space? Fix a point $\overrightarrow{p} \in \mathbb{R}^3$, and consider a vector field

$$\overrightarrow{F}(x,y,z) = P(x,y,z)\,\overrightarrow{\imath} + Q(x,y,z)\,\overrightarrow{\jmath} + R(x,y,z)\,\overrightarrow{k}$$

acting on points near $\overrightarrow{p}$. If a given segment through $\overrightarrow{p}$ rotates under the influence of $\overrightarrow{F}$, its angular velocity, following the ideas at the end of

§ 1.7, will be represented by the vector $\overrightarrow{\omega}$ whose direction gives the axis of rotation, and whose magnitude is the angular velocity. Now, we can try to decompose this vector into components. The vertical component of $\overrightarrow{\omega}$ represents precisely the rotation about a vertical axis through $\overrightarrow{p}$, with an *upward* direction corresponding to *counterclockwise* rotation. We can also think of this in terms of the projection of the line segment onto the horizontal plane through $\overrightarrow{p}$ and its rotation about $\overrightarrow{p}$. We expect the vertical component, $R(x, y, z)$, of $\overrightarrow{F}$ to have no effect on this rotation, as it is "pushing" along the length of the vertical axis. So we expect the planar curl $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ to be the correct measure of the tendency of the vector field to produce rotation about a vertical axis. As with directed area in § 1.7, we make this into a *vector* pointing along the axis of rotation (more precisely, pointing along that axis toward the side of the horizontal plane from which the rotation being induced appears counterclockwise). This leads us to multiply the (scalar) planar curl by the vertical unit vector $\overrightarrow{k}$:

$$\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\overrightarrow{k}.$$

Now we extend this analysis to the other two components of rotation. To analyze the tendency for rotation about the x-axis, we stare at the yz-plane from the positive x-axis: the former role of the x-axis (*resp.* y-axis) is now played by the y-axis (*resp.* z-axis), and in a manner completely analogous to the argument in § 6.3 and its reinterpretation in the preceding paragraph, we represent the tendency toward rotation about the x-axis by the vector

$$\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\overrightarrow{\imath}.$$

Finally, the tendency for rotation about the y-axis requires us to look from the direction of the *negative* y-axis, and we represent this tendency by the vector

$$\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)(-\overrightarrow{\jmath}).$$

This way of thinking may remind you of our construction of the cross product from oriented areas in § 1.6; however, in this case, instead of multiplying certain components of two vectors, we seem to be taking different partial derivatives. We can formally recover the analogy by creating an abstract "vector" whose components are differentiations

$$\overrightarrow{\nabla} := \overrightarrow{\imath}\frac{\partial}{\partial x} + \overrightarrow{\jmath}\frac{\partial}{\partial y} + \overrightarrow{k}\frac{\partial}{\partial z} \tag{6.26}$$

and interpreting "multiplication" by one of these components as performing the differentiation it represents: it is a **differential operator**—a "function of functions", whose input is a function, and whose output depends on derivatives of the input. This formal idea was presented by William Rowan Hamilton (1805-1865) in his *Lectures on Quaternions* (1853) [21, Lecture VII, pp. 610-11]. We pronounce the symbol $\overrightarrow{\nabla}$ as "del".[11]

At the most elementary formal level, when we "multiply" a function of three variables by this, we get the gradient vector:

$$\overrightarrow{\nabla} f = \left(\overrightarrow{\imath} \frac{\partial}{\partial x} + \overrightarrow{\jmath} \frac{\partial}{\partial y} + \overrightarrow{k} \frac{\partial}{\partial z} \right) f = \frac{\partial f}{\partial x} \overrightarrow{\imath} + \frac{\partial f}{\partial y} \overrightarrow{\jmath} + \frac{\partial f}{\partial z} \overrightarrow{k}.$$

However, we can also apply this operator to a vector field in several ways. For present purposes, we can take the formal cross product of this vector with a vector field, to get a different operator: if

$$\overrightarrow{F} = P \overrightarrow{\imath} + Q \overrightarrow{\jmath} + R \overrightarrow{k}$$

then the **curl** of $\overrightarrow{F}$ is

$$\begin{aligned} \overrightarrow{curl}\, \overrightarrow{F} &= \overrightarrow{\nabla} \times \overrightarrow{F} \\ &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ P & Q & R \end{vmatrix} \\ &= \overrightarrow{\imath} \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z} \right) - \overrightarrow{\jmath} \left(\frac{\partial R}{\partial x} - \frac{\partial P}{\partial z} \right) + \overrightarrow{k} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right). \end{aligned}$$

The expression on the right in the first line above is pronounced "del cross F". Note that if $R = 0$ and P and Q depend only on x and y—that is, $\overrightarrow{F} = P(x,y)\,\overrightarrow{\imath} + Q(x,y)\,\overrightarrow{\jmath}$ is essentially a planar vector field—then the only nonzero component of $\overrightarrow{\nabla} \times \overrightarrow{F}$ is the vertical one, and it equals what we called the planar curl of the associated planar vector field in § 6.3. When necessary, we distinguish between the vector $\overrightarrow{\nabla} \times \overrightarrow{F}$ and the planar curl (a scalar) by calling this the *vector* curl.

[11] This symbol appears in Maxwell's *Treatise on Electricity and Magnetism* [38, vol. 1, p. 16]—as well as an earlier paper [37]—but it is not given a name until Wilson's version of Gibbs' Lectures in 1901 [54, p. 138]: here he gives the "del" pronunciation, and mentions that "Some use the term *Nabla* owing to its fancied resemblance to an Assyrian harp..." (*nabla* is the Hebrew word for harp). In fact, Hamilton already has the operations we call *curl* and *divergence* in a combined (quaternion-valued) operator which he denotes $\lhd$ [21, p. 610].

Boundary Orientation

Suppose $\mathfrak{S}$ is an oriented surface in space, with orientation defined by the unit normal vector $\overrightarrow{n}$, and bounded by one or more curves. We would like to formulate an orientation for these curves which corresponds to the boundary orientation for ∂D when D is a region in the plane. Recall that in that context, we took the unit vector $\overrightarrow{T}$ tangent to a boundary curve and rotated it by $\frac{\pi}{2}$ radians counterclockwise to get the "leftward normal" $\overrightarrow{N}_+$; we then insisted that $\overrightarrow{N}_+$ point into the region D. It is fairly easy to see that such a rotation of a vector in the plane is accomplished by setting $\overrightarrow{N}_+ = \overrightarrow{k} \times \overrightarrow{T}$, and we can easily mimic this by replacing $\overrightarrow{k}$ with the unit normal $\overrightarrow{n}$ defining our orientation (that is, we rotate $\overrightarrow{T}$ counterclockwise when viewed from the direction of $\overrightarrow{n}$). However, when we are dealing with a surface in space, the surface might "curl away" from the plane in which this vector sits, so that it is harder to define what it means for it to "point into" $\mathfrak{S}$.

One way to do this is to invoke Proposition 4.4.5, which tells us that we can always parametrize a surface as the graph of a function, locally. If a surface is the graph of a function, then its boundary is the graph of the restriction of this function to the boundary of its domain. Thus we can look at the projection of $\overrightarrow{N}_+$ onto the plane containing the domain of the function, and ask that *it* point into the domain. This is a particularly satisfying formulation when we use the second statement in Proposition 4.4.5, in which we regard the surface as the graph of a function whose domain is in the tangent plane of the surface—which is to say the plane perpendicular to the normal vector $\overrightarrow{n}$— since it automatically contains $\overrightarrow{N}_+$.

We will adopt this as a definition.

Definition 6.5.5. *Given an oriented surface $\mathfrak{S}$ with orientation given by the unit normal vector field $\overrightarrow{n}$, and $\gamma(t)$ a boundary curve of $\mathfrak{S}$, with unit tangent vector $\overrightarrow{T}$ (parallel to the velocity), we say that $\gamma(t)$ has the **boundary orientation** if for every boundary point $\gamma(t)$ the leftward normal $\overrightarrow{N}_+ = \overrightarrow{n} \times \overrightarrow{T}$ points into the projection of $\mathfrak{S}$ on its tangent plane at $\gamma(t)$.*

Stokes' Theorem in the Language of Vector Fields

Using the terminology worked out above, we can state Stokes' Theorem as an almost verbatim restatement, in the context of 3-space, of Theorem 6.3.7:

Theorem 6.5.6 (Stokes' Theorem)**.** *If $\overrightarrow{F} = P\overrightarrow{\imath} + Q\overrightarrow{\jmath} + R\overrightarrow{k}$ is a $\mathcal{C}^1$ vector field defined in a region of* 3*-space containing the oriented surface with boundary $\mathfrak{S}$, then the circulation of $\overrightarrow{F}$ around the boundary of $\mathfrak{S}$ (each constituent piecewise regular, simple, closed curve of $\partial\mathfrak{S}$ given the boundary orientation) equals the flux integral over $\mathfrak{S}$ of the (vector) curl of $\overrightarrow{F}$:*

$$\oint_{\partial\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}} = \iint_{\mathfrak{S}} (\overrightarrow{\nabla} \times \overrightarrow{F}) \cdot d\overrightarrow{\mathcal{S}}.$$

Proof. This proof involves a calculation which reduces to Green's Theorem (Theorem 6.3.4). By an argument similar to the proof used there, it suffices to prove the result for a coordinate patch with one boundary component: that is, we will assume that $\mathfrak{S}$ is parametrized by a regular, $\mathcal{C}^2$ function $\overrightarrow{p}\colon \mathbb{R}^2 \to \mathbb{R}^3$ which is one-to-one on its boundary. Instead of using s and t for the names of the parameters, we will use u and v (so as not to conflict with the parameter t in the parametrization of $\partial\mathfrak{S}$):

$$\overrightarrow{p}(u, v) = (x(u, v)\,, y(u, v)\,, z(u, v)), \quad (u, v) \in D \subset \mathbb{R}^2$$

and assume that the boundary ∂D of the domain D is given by a curve

$$\gamma(t) = \overrightarrow{p}(u(t)\,, v(t))\,, \quad t \in [t_0, t_1]\,.$$

Let us begin by computing the integrand on the right of the statement of the theorem: this is a dot product of two formal determinants:

$$d\overrightarrow{\mathcal{S}} = \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ \partial x/\partial u & \partial y/\partial u & \partial z/\partial u \\ \partial x/\partial v & \partial y/\partial v & \partial z/\partial v \end{vmatrix} du\,dv$$

$$\overrightarrow{\nabla} \times \overrightarrow{F} = \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ P & Q & R \end{vmatrix}$$

where

$$\overrightarrow{F}(x, y, z) = P(x, y, z)\,\overrightarrow{\imath} + Q(x, y, z)\,\overrightarrow{\jmath} + R(x, y, z)\,\overrightarrow{k}\,.$$

When we take the dot product of these two determinant vectors, we simply multiply the minors corresponding to each of the three components. It will be useful for us in computing $d\overrightarrow{\mathcal{S}}$ to adapt the old-fashioned notation for the determinant of partials that came up on p. 434:

$$\left|\frac{\partial\left(f_1,f_2\right)}{\partial\left(x_1,x_2\right)}\right| := \det\begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 \end{bmatrix}$$
$$= \frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_1}.$$

A quick calculation shows that

$$(\overrightarrow{\nabla}\times\overrightarrow{F})\cdot d\overrightarrow{\mathcal{S}} =$$
$$\left\{\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)\left|\frac{\partial\left(y,z\right)}{\partial\left(u,v\right)}\right| + \left(\frac{\partial R}{\partial x}-\frac{\partial P}{\partial z}\right)\left|\frac{\partial\left(x,z\right)}{\partial\left(u,v\right)}\right| + \left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\left|\frac{\partial\left(x,y\right)}{\partial\left(u,v\right)}\right|\right\} du\,dv.$$

This mess is best handled by separating out the terms involving each of the components of $\overrightarrow{F}$ and initially ignoring the "$du\,dv$" at the end. Consider the terms involving the first component, P: they are

$$-\frac{\partial P}{\partial z}\left|\frac{\partial\left(x,z\right)}{\partial\left(u,v\right)}\right| - \frac{\partial P}{\partial y}\left|\frac{\partial\left(x,y\right)}{\partial\left(u,v\right)}\right|;$$

we can add to this the term $-\frac{\partial P}{\partial x}\left|\frac{\partial(x,x)}{\partial(u,v)}\right|$ which equals zero (right?) and expand to get

$$-\frac{\partial P}{\partial z}\left(\frac{\partial x}{\partial u}\frac{\partial z}{\partial v}-\frac{\partial z}{\partial u}\frac{\partial x}{\partial v}\right)-\frac{\partial P}{\partial y}\left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}-\frac{\partial y}{\partial u}\frac{\partial x}{\partial v}\right)-\frac{\partial P}{\partial x}\left(\frac{\partial x}{\partial u}\frac{\partial x}{\partial v}-\frac{\partial x}{\partial u}\frac{\partial x}{\partial v}\right)$$
$$=\frac{\partial x}{\partial v}\left(\frac{\partial P}{\partial z}\frac{\partial z}{\partial u}+\frac{\partial P}{\partial y}\frac{\partial y}{\partial u}+\frac{\partial P}{\partial x}\frac{\partial x}{\partial u}\right)$$
$$-\frac{\partial x}{\partial u}\left(\frac{\partial P}{\partial z}\frac{\partial z}{\partial v}+\frac{\partial P}{\partial y}\frac{\partial y}{\partial v}+\frac{\partial P}{\partial x}\frac{\partial x}{\partial v}\right);$$

applying the Chain Rule to the terms inside each set of parentheses, this can be rewritten

$$=\frac{\partial x}{\partial v}\frac{\partial P}{\partial u}-\frac{\partial x}{\partial u}\frac{\partial P}{\partial v}.$$

Now, using the fact that since the parametrization is $\mathcal{C}^2$ the mixed cross-partials are equal

$$\frac{\partial^2 x}{\partial u\partial v}=\frac{\partial^2 x}{\partial v\partial u}$$

we can interpret this as a planar curl (in terms of the uv-plane)

$$\frac{\partial x}{\partial v}\frac{\partial P}{\partial u}-\frac{\partial x}{\partial u}\frac{\partial P}{\partial v}=\frac{\partial}{\partial u}\left[\frac{\partial x}{\partial v}P\right]-\frac{\partial}{\partial v}\left[\frac{\partial x}{\partial u}P\right]$$

whose integral can be calculated using Theorem 6.3.4 (Green's Theorem)

$$\begin{aligned}\iint_D\left\{\frac{\partial}{\partial u}\left[\frac{\partial x}{\partial v}P\right]-\frac{\partial}{\partial v}\left[\frac{\partial x}{\partial u}P\right]\right\}du\,dv&=\int_{\partial D}\frac{\partial x}{\partial u}P\,du+\frac{\partial x}{\partial v}P\,dv\\&=\oint_{\partial D}P\,dx.\end{aligned}$$

In a similar way, adding $\frac{\partial Q}{\partial y}\left|\frac{\partial(y,y)}{\partial(u,v)}\right|$ (*resp.* $\frac{\partial R}{\partial z}\left|\frac{\partial(z,z)}{\partial(u,v)}\right|$) to the sum of the terms involving Q (*resp.* R) (Exercise 7) we can calculate integrals of those terms using Green's Theorem:

$$\begin{aligned}\iint_D\left\{-\frac{\partial Q}{\partial z}\left|\frac{\partial(y,z)}{\partial(u,v)}\right|+\frac{\partial Q}{\partial x}\left|\frac{\partial(x,y)}{\partial(u,v)}\right|\right\}du\,dv&=\iint_D\left\{\frac{\partial}{\partial u}\left[\frac{\partial y}{\partial v}Q\right]-\frac{\partial}{\partial v}\left[\frac{\partial y}{\partial u}Q\right]\right\}du\,dv\\&=\oint_{\partial D}Q\,dy\\\iint_D\left\{\frac{\partial R}{\partial y}\left|\frac{\partial(y,z)}{\partial(u,v)}\right|+\frac{\partial R}{\partial x}\left|\frac{\partial(x,z)}{\partial(u,v)}\right|\right\}du\,dv&=\iint_D\left\{\frac{\partial}{\partial u}\left[\frac{\partial z}{\partial v}R\right]-\frac{\partial}{\partial v}\left[\frac{\partial z}{\partial u}R\right]\right\}du\,dv\\&=\oint_{\partial D}R\,dz.\end{aligned}$$

Adding these three equations yields the desired equality. □

Stokes' Theorem, like Green's Theorem, allows us to choose between integrating a vector field along a curve and integrating its curl over a surface bounded by that curve. Let us compare the two approaches in a few examples.

First, we consider the vector field

$$\overrightarrow{F}(x,y,z)=(x-y)\overrightarrow{\imath}+(x+y)\overrightarrow{\jmath}+z\overrightarrow{k}$$

and the surface $\mathfrak{S}$ given by the graph

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

inside the cylinder

$$x^2+y^2\leq 1.$$

We take the orientation of $\mathfrak{S}$ to be *upward*. To integrate the vector field over the boundary $x^2 + y^2 = 1$, $z = x^2 - y^2$, we parametrize the boundary curve $\partial\mathfrak{S}$ as

$$\begin{cases} x &= \cos\theta \\ y &= \sin\theta \\ z &= \cos^2\theta - \sin^2\theta \end{cases}$$

with differentials

$$\begin{cases} dx &= -\sin\theta\, d\theta \\ dy &= \cos\theta\, d\theta \\ dz &= (-2\cos\theta\sin\theta - 2\sin\theta\cos\theta)\, d\theta \\ &= -4\sin\theta\cos\theta\, d\theta \end{cases}$$

so the element of arclength is

$$d\vec{\mathfrak{s}} = \left\{(-\sin\theta)\vec{\imath} + (\cos\theta)\vec{\jmath} - (4\sin\theta\cos\theta)\vec{k}\right\} d\theta.$$

Along this curve, the vector field is

$$\begin{aligned}\vec{F}(\theta) &= \vec{F}\big(\cos\theta, \sin\theta, \cos^2\theta - \sin^2\theta\big) \\ &= (\cos\theta - \sin\theta)\vec{\imath} + (\cos\theta + \sin\theta)\vec{\jmath} + (\cos^2\theta - \sin^2\theta)\vec{k}.\end{aligned}$$

Their dot product is

$$\begin{aligned}\vec{F}\cdot d\vec{\mathfrak{s}} &= \{(\cos\theta - \sin\theta)(-\sin\theta) + (\cos\theta + \sin\theta)(\cos\theta) \\ &\qquad +(\cos^2\theta - \sin^2\theta)(-4\sin\theta\cos\theta)\big\}\, d\theta \\ &= \big\{-\cos\theta\sin\theta + \sin^2\theta + \cos^2\theta + \sin\theta\cos\theta \\ &\qquad -4\sin\theta\cos^3\theta - 4\sin^3\theta\cos\theta\big\}\, d\theta \\ &= \{1 - 4\sin\theta\cos\theta\}\, d\theta\end{aligned}$$

and the line integral of $\vec{F}$ over $\partial\mathfrak{S}$ is

$$\begin{aligned}\oint_{\partial\mathfrak{S}} \vec{F}\cdot d\vec{\mathfrak{s}} &= \int_0^{2\pi} (1 - 4\sin\theta\cos\theta)\, d\theta \\ &= (\theta - 2\sin^2\theta)_0^{2\pi} \\ &= 2\pi.\end{aligned}$$

Now let us consider the alternative calculation, as a flux integral. From Remark 6.5.4 we know that the natural parametrization of the surface

$$\begin{cases} x &= s \\ y &= t \\ z &= s^2 - t^2 \end{cases} \qquad s^2 + t^2 \leq 1$$

has element of surface area (with upward orientation)

$$d\vec{\mathcal{S}} = [-(2s)\vec{\imath} - (-2t)\vec{\jmath} + \vec{k}]\, ds\, dt.$$

The curl of our vector field is

$$\begin{aligned} \vec{\nabla} \times \vec{F} &= \begin{vmatrix} \vec{\imath} & \vec{\jmath} & \vec{k} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ x - y & x + y & z \end{vmatrix} \\ &= 0\vec{\imath} - 0\vec{\jmath} + 2\vec{k} \\ &= 2\vec{k}. \end{aligned}$$

Thus, the flux integral of the curl is

$$\begin{aligned} \iint_{\mathfrak{S}} (\vec{\nabla} \times \vec{F}) \cdot d\vec{\mathcal{S}} &= \iint_{\mathfrak{S}} 2\vec{k} \cdot d\vec{\mathcal{S}} \\ &= \iint_{s^2+t^2\leq 1} 2\, ds\, dt \end{aligned}$$

which we recognize as the area of the unit disc, or 2π.

As a second example, we consider the line integral

$$\oint_{\mathcal{C}} -y^3\, dx + x^3\, dy - z^3\, dz$$

where the curve $\mathcal{C}$ is given by the intersection of the cylinder $x^2 + y^2 = 1$ with the plane $x + y + z = 1$, circumvented counterclockwise when seen from above.

If we attack this directly, we parametrize $\mathcal{C}$ by

$$\begin{cases} x &= \cos\theta \\ y &= \sin\theta \\ z &= 1 - \cos\theta - \sin\theta \end{cases} \qquad 0 \leq \theta \leq 2\pi$$

with differentials

$$\begin{cases} dx &= -\sin\theta\,d\theta \\ dy &= \cos\theta\,d\theta \\ dz &= (\sin\theta - \cos\theta)\,d\theta \end{cases}$$

and the form becomes

$$\begin{aligned} -y^3\,dx + x^3\,dy - z^3\,dz = (-\sin^3\theta)[-\sin\theta\,d\theta] + (\cos^3\theta)[\cos\theta\,d\theta] \\ + (1-\cos\theta-\sin\theta)^3[(\sin\theta-\cos\theta)\,d\theta] \end{aligned}$$

leading to the integral

$$\int_0^{2\pi} \left(\sin^4\theta + \cos^4\theta - (1-\cos\theta-\sin\theta)^3(\sin\theta-\cos\theta)\right)\,d\theta$$

which is not impossible to do, but clearly a mess to try.

Note that this line integral corresponds to the circulation integral $\oint_{\mathcal{C}} \overrightarrow{F} \cdot d\overrightarrow{\mathfrak{s}}$ where

$$\overrightarrow{F}(x,y,z) = -y^3\overrightarrow{\imath} + x^3\overrightarrow{\jmath} - z^3\overrightarrow{k}.$$

If instead we formulate this as a flux integral, we take the curl of the vector field

$$\overrightarrow{\nabla} \times \overrightarrow{F} = \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ -y^3 & x^3 & -z^3 \end{vmatrix} = 0\overrightarrow{\imath} + 0\overrightarrow{\jmath} + (3x^2+3y^2)\overrightarrow{k}.$$

Note that $\mathcal{C}$ is the boundary of the part of the plane $x+y+z=1$ over the disc $x^2+y^2 \leq 1$; to make the given orientation on $\mathcal{C}$ the boundary orientation, we need to make sure that the disc is oriented *up*. It can be parametrized using polar coordinates as

$$\overrightarrow{p}(r,\theta) = (r\cos\theta)\overrightarrow{\imath} + (r\sin\theta)\overrightarrow{\jmath} + (1 - r\cos\theta - r\sin\theta)\overrightarrow{k}$$

with partials

$$\begin{aligned} \frac{\partial\overrightarrow{p}}{\partial r} &= (\cos\theta)\overrightarrow{\imath} + (\sin\theta)\overrightarrow{\jmath} - (\cos\theta+\sin\theta)\overrightarrow{k} \\ \frac{\partial\overrightarrow{p}}{\partial\theta} &= (-r\sin\theta)\overrightarrow{\imath} + (r\cos\theta)\overrightarrow{\jmath} + (r\sin\theta - r\cos\theta)\overrightarrow{k}. \end{aligned}$$

We calculate the element of oriented surface area[12] in terms of the cross product

$$\begin{aligned}
\frac{\partial \overrightarrow{p}}{\partial r} \times \frac{\partial \overrightarrow{p}}{\partial \theta} &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ \cos\theta & \sin\theta & -\cos\theta - \sin\theta \\ -r\sin\theta & r\cos\theta & r\sin\theta - r\cos\theta \end{vmatrix} \\
&= \overrightarrow{\imath}(r\sin^2\theta - r\sin\theta\cos\theta + r\cos^2\theta + r\sin\theta\cos\theta) \\
&\quad - \overrightarrow{\jmath}(r\cos\theta - r\cos^2\theta - r\sin\theta\cos\theta - r\sin^2\theta) \\
&\quad + \overrightarrow{k}(r\cos^2\theta + r\sin^2\theta) \\
&= r\overrightarrow{\imath} + r\overrightarrow{\jmath} + r\overrightarrow{k};
\end{aligned}$$

in particular, the element of oriented surface area is

$$d\overrightarrow{\mathcal{S}} = r(\overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k})\, dr\, d\theta$$

which, we note, has an *upward* vertical component, as desired. Since $\overrightarrow{\nabla} \times \overrightarrow{F}$ has only a $\overrightarrow{k}$ component,

$$\begin{aligned}
(\overrightarrow{\nabla} \times \overrightarrow{F}) \cdot d\overrightarrow{\mathcal{S}} &= (3x^2 + 3y^2)(r)\, dr\, d\theta \\
&= 3r^3\, dr\, d\theta
\end{aligned}$$

so the flux integral is given by

$$\iint_{\mathfrak{S}} (\overrightarrow{\nabla} \times \overrightarrow{F}) \cdot d\overrightarrow{\mathcal{S}} = \int_0^{2\pi} \int_0^1 3r^3\, dr\, d\theta = \int_0^{2\pi} \frac{3}{4}\, d\theta = \frac{3\pi}{2}.$$

We note that a consequence of Stokes' Theorem, like the Fundamental Theorem for Line Integrals, is that the flux integral is the same for any two surfaces that have the same boundary. However, in practice, this is only useful if we can recognize the integrand as a curl, an issue we will delay until we have the Divergence Theorem and Proposition 6.7.4 in § 6.7.

Exercises for § 6.5

Practice problems:

1. Evaluate each flux integral $\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}}$ below:

[12] Note that we can't apply Remark 6.5.4 directly here, because our input is not given in rectangular coordinates

(a) $\overrightarrow{F}(x,y,z) = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + 2z\overrightarrow{k}$, $\mathfrak{S}$ is the graph of $z = 3x + 2y$ over $[0,1] \times [0,1]$, oriented up.

(b) $\overrightarrow{F}(x,y,z) = yz\overrightarrow{\imath} + x\overrightarrow{\jmath} + xy\overrightarrow{k}$, $\mathfrak{S}$ is the graph of $z = x^2 + y^2$ over $[0,1] \times [0,1]$, oriented up.

(c) $\overrightarrow{F}(x,y,z) = x\overrightarrow{\imath} - y\overrightarrow{\jmath} + z\overrightarrow{k}$, $\mathfrak{S}$ is the part of the plane $x+y+z = 1$ in the first octant, oriented up.

(d) $\overrightarrow{F}(x,y,z) = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$, $\mathfrak{S}$ is the upper hemisphere $x^2 + y^2 + z^2 = 1$, $z \geq 0$, oriented up.

(e) $\overrightarrow{F}(x,y,z) = z\overrightarrow{k}$, $\mathfrak{S}$ is the part of the sphere $x^2 + y^2 + z^2 = 1$ between the xy-plane and the plane $z = \frac{1}{2}$, oriented outward.

(f) $\overrightarrow{F}(x,y,z) = x\overrightarrow{\imath} - y\overrightarrow{\jmath} + z\overrightarrow{k}$, $\mathfrak{S}$ is the unit sphere $x^2+y^2+z^2 = 1$, oriented outward.

(g) $\overrightarrow{F}(x,y,z) = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$, $\mathfrak{S}$ is the surface parametrized by

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= 1 - r^2, \\ 0 \leq &r \leq 1 \\ 0 \leq &\theta \leq 2\pi, \end{aligned}$$

oriented up.

(h) $\overrightarrow{F}(x,y,z) = (y+z)\overrightarrow{\imath} + (x+y)\overrightarrow{\jmath} + x + z)\overrightarrow{k}$, $\mathfrak{S}$ is the surface parametrized by

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= \theta, \\ 0 \leq &r \leq 1 \\ 0 \leq &\theta \leq 4\pi, \end{aligned}$$

oriented up.

2. Find the curl of each vector field below:

(a) $\overrightarrow{F}(x,y,z) = (xy)\overrightarrow{\imath} + (yz)\overrightarrow{\jmath} + (xz)\overrightarrow{k}$

(b) $\overrightarrow{F}(x,y,z) = (y^2 + z^2)\overrightarrow{\imath} + (x^2 + z^2)\overrightarrow{\jmath} + (x^2 + y^2)\overrightarrow{k}$

(c) $\overrightarrow{F}(x,y,z) = (e^y \cos z)\overrightarrow{\imath} + (x^2 z)\overrightarrow{\jmath} + (x^2 y^2)\overrightarrow{k}$

(d) $\overrightarrow{F}(x,y,z) = (y)\overrightarrow{\imath} + (-x)\overrightarrow{\jmath} + (z)\overrightarrow{k}$

(e) $\overrightarrow{F}(x,y,z) = (z)\overrightarrow{\imath} + (y)\overrightarrow{\jmath} + (x)\overrightarrow{k}$

(f) $\overrightarrow{F}(x,y,z) = (e^y \cos x)\overrightarrow{\imath} + (e^y \sin z)\overrightarrow{\jmath} + (e^y \cos z)\overrightarrow{k}$

3. Evaluate each circulation integral $\oint_{\mathcal{C}} \overrightarrow{F} \cdot \overrightarrow{T}\, d\mathfrak{s}$ two different ways: (i) directly, and (ii) using Stokes' Theorem and the fact that $\mathcal{C}$ is the boundary of $\mathfrak{S}$:

   (a) $\overrightarrow{F}(x,y,z) = (-y, x, z)$, $\mathcal{C}$ is given by $\overrightarrow{p}(t) = (\cos t, \sin t, 1 - \cos t - \sin t)$, and $\mathfrak{S}$ is given by $\overrightarrow{p}(s,t) = (s, t, 1-s-t)$, $s^2 + t^2 \leq 1$.

   (b) $\overrightarrow{F}(x,y,z) = y^2\overrightarrow{\imath} + z^2\overrightarrow{\jmath} + x^2\overrightarrow{k}$, $\mathcal{C}$ is given by $\overrightarrow{p}(t) = (\cos\theta, \sin\theta, \cos 2\theta)$, $0 \leq \theta \leq 2\pi$ and $\mathfrak{S}$ is given by $\overrightarrow{p}(s,t) = (s, t, s^2 - t^2)$, $s^2 + t^2 \leq 1$.

   (c) $\overrightarrow{F}(x,y,z) = (z, xz, y)$, $\mathfrak{S}$ is the part of the plane $x + y + z = 1$ over the rectangle $[0,1] \times [0,1]$, oriented up.

Theory problems:

4. (a) Evaluate the flux integral $\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}}$, where $\overrightarrow{F}(x,y,z) = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$ and $\mathfrak{S}$ is the plane $z = ax + by$ over $[0,1] \times [0,1]$, oriented up.

   (b) Give a geometric explanation for your result.

   (c) What happens if we replace this plane by the parallel plane $z = ax + by + c$?

5. Suppose $\mathfrak{S}$ is a regular surface and $\overrightarrow{n}$ is a continuous choice of unit normal vectors (*i.e.*, an orientation of $\mathfrak{S}$). Explain how we can cover $\mathfrak{S}$ with overlapping coordinate patches for which the induced local orientations are coherent. (Note that by definition, we are given a family of overlapping coordinate patches covering $\mathfrak{S}$. The issue is how to modify them so that their induced local orientations are coherent.)

6. Prove Remark 6.5.4.

7. Fill in the details of the calculation yielding the terms involving Q and R in the proof of Theorem 6.5.6.

6.6 2-forms in $\mathbb{R}^3$

The formalism introduced in § 6.4 can be extended to $\mathbb{R}^3$, giving a new language for formulating Stokes' Theorem as well as many other results.

Bilinear Functions and 2-forms on $\mathbb{R}^3$

The notion of a bilinear function given in Definition 6.4.1 extends naturally to $\mathbb{R}^3$ (in fact, to $\mathbb{R}^n$):

Definition 6.6.1. *A **bilinear function** on $\mathbb{R}^n$ is a function of two vector variables $B(\overrightarrow{v}, \overrightarrow{w})$ such that fixing one of the inputs results in a linear function of the other input:*

$$\begin{aligned} B(a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2, \overrightarrow{w}) &= a_1B(\overrightarrow{v}_1, \overrightarrow{w}) + a_2B(\overrightarrow{v}_2, \overrightarrow{w}) \\ B(\overrightarrow{v}, b_1\overrightarrow{w}_1 + b_2\overrightarrow{w}_2) &= b_1B(\overrightarrow{v}, \overrightarrow{w}_1) + b_2B(\overrightarrow{v}, \overrightarrow{w}_2) \end{aligned} \tag{6.27}$$

for arbitrary vectors in $\mathbb{R}^n$ and real scalars.

We shall explore this notion only for $\mathbb{R}^3$ in this section. As in $\mathbb{R}^2$, the dot product is one example of a bilinear function on $\mathbb{R}^3$. Using Equation (6.27) we can see that just as in the case of the plane, a general bilinear function $B(\overrightarrow{v}, \overrightarrow{w})$ on $\mathbb{R}^3$ can be expressed as a polynomial in the coordinates of its entries, with coefficients coming from the values of the bilinear function on the standard basis elements: if

$$\overrightarrow{v} = (x, y, z) = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$$

and

$$\overrightarrow{w} = (x', y', z') = x'\overrightarrow{\imath} + y'\overrightarrow{\jmath} + z'\overrightarrow{k}$$

then

$$\begin{aligned} B(\overrightarrow{v}, \overrightarrow{w}) &= B\left(x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}, x'\overrightarrow{\imath} + y'\overrightarrow{\jmath} + z'\overrightarrow{k}\right) \\ &= B(\overrightarrow{\imath}, \overrightarrow{w})\,x + B(\overrightarrow{\jmath}, \overrightarrow{w})\,y + B\left(\overrightarrow{k}, \overrightarrow{w}\right)z \\ &= B(\overrightarrow{\imath}, \overrightarrow{\imath})\,xx' + B(\overrightarrow{\imath}, \overrightarrow{\jmath})\,xy' + B\left(\overrightarrow{\imath}, \overrightarrow{k}\right)xz' \\ &\qquad + B(\overrightarrow{\jmath}, \overrightarrow{\imath})\,yx' + B(\overrightarrow{\jmath}, \overrightarrow{\jmath})\,yy' + B\left(\overrightarrow{\jmath}, \overrightarrow{k}\right)zz' \\ &\qquad + B\left(\overrightarrow{k}, \overrightarrow{\imath}\right)zx' + B\left(\overrightarrow{k}, \overrightarrow{\jmath}\right)zy' + B\left(\overrightarrow{k}, \overrightarrow{k}\right)zz'. \end{aligned}$$

This is rather hard on the eyes; to make patterns clearer, we will adopt a different notation, using indices and subscripts instead of different letters to denote components, etc. Let us first change our notation for the standard basis, writing

$$\begin{aligned} \overrightarrow{\imath} &= \overrightarrow{e}_1 \\ \overrightarrow{\jmath} &= \overrightarrow{e}_2 \\ \overrightarrow{k} &= \overrightarrow{e}_3 \end{aligned}$$

and also use subscripts for the components of a vector: instead of writing

$$\overrightarrow{v} = (x, y, z)$$

we will write

$$\overrightarrow{v} = (v_1, v_2, v_3).$$

Finally, if we use a double-indexed notation for the coefficients above

$$B(\overrightarrow{e}_i, \overrightarrow{e}_j) = b_{ij}$$

we can write the formula above in summation form

$$B(\overrightarrow{v}, \overrightarrow{w}) = \sum_{i=1}^{3} \sum_{j=1}^{3} b_{ij} v_i w_j.$$

There is another useful way to represent a bilinear function, with matrix notation. If we write

$$[B] = \begin{bmatrix} b_{11} & b_{12} & b_{12} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

then much in the same way as we wrote a quadratic form, we can write the formula above as

$$B(\overrightarrow{v}, \overrightarrow{w}) = [\overrightarrow{v}]^T \, [B] \, [\overrightarrow{w}]$$

where $[\overrightarrow{v}]$ is the column of coordinates of $\overrightarrow{v}$ and $[\overrightarrow{v}]^T$ is its transpose

$$[\overrightarrow{v}] = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$
$$[\overrightarrow{v}]^T = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}.$$

In particular, the dot product has as its matrix representative the **identity matrix** , which has 1 on the diagonal ($b_{ii} = 1$) and 0 off it ($b_{ij} = 0$ for $i \neq j$). As in the two-dimensional case, the fact that the matrix representative of this bilinear function is symmetric reflects the fact that the function is **commutative**: $B(\overrightarrow{v}, \overrightarrow{w}) = B(\overrightarrow{w}, \overrightarrow{v})$ for any pair of vectors in $\mathbb{R}^3$.

Again as in the two-dimensional case, we require *anti*-commutativity for a 2-form (in this context, this property is often called skew-symmetry):

Definition 6.6.2. *A* ***2-form*** *on* $\mathbb{R}^3$ *is an* ***anti-commutative*** *bilinear function: a function* $\omega(\overrightarrow{v}, \overrightarrow{w})$ *of two vector variables satisfying*

1. ***bilinearity:***

$$\omega\big(\alpha\overrightarrow{v} + \beta\overrightarrow{v}', \overrightarrow{w}\big) = \alpha\omega(\overrightarrow{v}, \overrightarrow{w}) + \beta\omega\big(\overrightarrow{v}', \overrightarrow{w}\big)$$

2. ***anti-commutativity=skew-symmetry:***

$$\omega(\overrightarrow{v}, \overrightarrow{w}) = -\omega(\overrightarrow{w}, \overrightarrow{v}).$$

The skew-symmetry of a 2-form is reflected in its matrix representative: it is easy to see that this property requires (and is equivalent to) the fact that $b_{ij} = -b_{ji}$ for every pair of indices, and in particular $b_{ii} = 0$ for every index i.

However, 2-forms in $\mathbb{R}^3$ differ from those on $\mathbb{R}^2$ in one very important respect: we saw in § 6.4 that every 2-form on $\mathbb{R}^2$ is a constant multiple of the 2×2 determinant, which we denoted using the wedge product. This **wedge product** can be easily extended to 1-forms on $\mathbb{R}^3$: if α and β are two 1-forms on $\mathbb{R}^3$, their wedge product is the 2-form defined by

$$(\alpha \wedge \beta)(\overrightarrow{v}, \overrightarrow{w}) := \det \begin{pmatrix} \alpha(\overrightarrow{v}) & \beta(\overrightarrow{v}) \\ \alpha(\overrightarrow{w}) & \beta(\overrightarrow{w}) \end{pmatrix}.$$

Now, all 1-forms in the plane are linear combinations of the two coordinate forms dx and dy; thus since the wedge product of any form with itself is zero and the wedge product is anti-commutative, every 2-form in the plane is a multiple of $dx \wedge dy$. However, there are *three* coordinate forms in $\mathbb{R}^3$: dx, dy, and dz, and these can be paired in three different ways (up to order); $dx \wedge dy$, $dx \wedge dz$, and $dy \wedge dz$. This means that instead of all being multiples of a single one, 2-forms on $\mathbb{R}^3$ are in general linear combinations of these three **basic 2-forms**:

$$\omega(\vec{v}, \vec{w}) = a(dx \wedge dy)(\vec{v}, \vec{w}) + b(dx \wedge dz)(\vec{v}, \vec{w}) + c(dy \wedge dz)(\vec{v}, \vec{w}) . \tag{6.28}$$

There is another way to think of this. If we investigate the action of a basic 2-form on a typical pair of vectors, we see that each of the forms $dx \wedge dy$, $dx \wedge dz$, and $dy \wedge dz$ acts as a 2×2 determinant on certain coordinates of the two vectors:

$$\begin{aligned}
(dx \wedge dy)(\vec{v}, \vec{w}) &= \det \begin{pmatrix} v_1 & v_2 \\ w_1 & w_2 \end{pmatrix} \\
(dx \wedge dz)(\vec{v}, \vec{w}) &= \det \begin{pmatrix} v_1 & v_3 \\ w_1 & w_3 \end{pmatrix} \\
(dy \wedge dz)(\vec{v}, \vec{w}) &= \det \begin{pmatrix} v_2 & v_3 \\ w_2 & w_3 \end{pmatrix}
\end{aligned}$$

which we might recognize as the minors in the definition of the cross product $\vec{v} \times \vec{w}$. Note that the "middle" minor, corresponding to $dx \wedge dz$, gets multiplied by -1 when we calculate the cross-product determinant; we can incorporate this into the form by replacing alphabetical order $dx \wedge dz$ with "circular" order $dz \wedge dx$. If we recall the motivation for the cross-product in the first place (§ 1.6), we see that these three basic forms represent the projections onto the coordinate planes of the oriented area of the parallelepiped spanned by the input vectors. In any case, we can write

$$\vec{v} \times \vec{w} = \vec{\imath}\big((dy \wedge dz)(\vec{v}, \vec{w})\big) + \vec{\jmath}\big((dz \wedge dx)(\vec{v}, \vec{w})\big) + \vec{k}\big((dx \wedge dy)(\vec{v}, \vec{w})\big).$$

But then the 2-form given by Equation (6.28) can be expressed as the dot product of $\vec{v} \times \vec{w}$ with a vector determined by the coefficients in that equation: you should check that for the expression as given in Equation (6.28), this vector is $c\vec{\imath} - b\vec{\jmath} + a\vec{k}$. Again, it is probably better to use a notation via subscripts: we rewrite the basic 1-forms as

$$\begin{aligned}
dx &= dx_1 \\
dy &= dx_2 \\
dz &= dx_3;
\end{aligned}$$

then, incorporating the modifications noted above, we rewrite Equation (6.28) as

$$\omega = a_1\,dx_2 \wedge dx_3 + a_2\,dx_3 \wedge dx_1 + a_3\,dx_1 \wedge dx_2.$$

With this notation, we can state the following equivalent representations of an arbitrary 2-form on $\mathbb{R}^3$:

Lemma 6.6.3. *Associated to every 2-form ω on $\mathbb{R}^3$ is a vector $\overrightarrow{a}$, defined by*

$$\omega(\overrightarrow{v}, \overrightarrow{w}) = \overrightarrow{a} \cdot \overrightarrow{v} \times \overrightarrow{w} \tag{6.29}$$

where

$$\begin{aligned} \omega &= a_1\,dx_2 \wedge dx_3 + a_2\,dx_3 \wedge dx_1 + a_3\,dx_1 \wedge dx_2 \\ \overrightarrow{a} &= a_1\overrightarrow{e}_1 + a_2\overrightarrow{e}_2 + a_3\overrightarrow{e}_3. \end{aligned}$$

The action of this 2-form on an arbitrary pair of vectors is given by the determinant formula

$$\omega(\overrightarrow{v}, \overrightarrow{w}) = \det\begin{pmatrix} a_1 & a_2 & a_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}. \tag{6.30}$$

Pay attention to the numbering here: the coefficient a_i with index i is paired with the basic form $dx_j \wedge dx_k$ corresponding to the *other two* indices, and these appear in an order such that i, j, k constitutes a *cyclic* permutation of $1, 2, 3$. In practice, we shall often revert to the non-subscripted notation, but this version is the best one to help us remember which vectors correspond to which 1-forms. The representation given by Equation (6.29) can be viewed as a kind of analogue of the gradient vector as a representation of the 1-form given by the derivative $d_{\overrightarrow{p}}f$ of a function $f\colon\mathbb{R}^3 \to \mathbb{R}$ at the point $\overrightarrow{p}$.

We saw in § 3.2 that the action of every linear function on $\mathbb{R}^3$ can be represented as the dot product with a fixed vector, and in § 6.1 we saw that this gives a natural correspondence between differential 1-forms and differentiable vector fields on $\mathbb{R}^3$

$$\omega = P\,dx + Q\,dy + R\,dz \leftrightarrow \overrightarrow{F} = P\overrightarrow{\imath} + Q\overrightarrow{\jmath} + R\overrightarrow{k}. \tag{6.31}$$

Now we have a correspondence between 2-forms Ω and vectors $\overrightarrow{F}$ on $\mathbb{R}^3$, defined by viewing the action of Ω on a pair of vectors as the dot product of a fixed vector with their cross product, leading to the correspondence between differential 2-forms and differential vector fields on $\mathbb{R}^3$

$$\Omega = A_1\,dx_2 \wedge dx_3 + A_2\,dx_3 \wedge dx_1 + A_3\,dx_1 \wedge dx_2 \leftrightarrow \overrightarrow{F} = a_1\overrightarrow{\imath} + a_2\overrightarrow{\jmath} + a_3\overrightarrow{k}. \tag{6.32}$$

The wedge product now assigns a 2-form to each ordered pair of 1-forms, and it is natural to ask how this can be represented as an operation on the corresponding vectors. The answer is perhaps only a little bit surprizing:

Remark 6.6.4. *Suppose α and β are two 1-forms, corresponding to the vectors $\overrightarrow{a}$ and $\overrightarrow{b}$*

$$\begin{aligned}\alpha &= a_1\,dx + a_2\,dy + a_3\,dz\\ \beta &= b_1\,dx + b_2\,dy + b_3\,dz.\end{aligned}$$

then their wedge product corresponds to the cross product $\overrightarrow{a} \times \overrightarrow{b}$:

$$(\alpha \wedge \beta)(\overrightarrow{v}, \overrightarrow{w}) = (\overrightarrow{a} \times \overrightarrow{b}) \cdot (\overrightarrow{v} \times \overrightarrow{w}):$$

$$\alpha \wedge \beta = (a_2b_3 - a_3b_2)\,dy \wedge dz + (a_3b_1 - a_1b_3)\,dz \wedge dx + (a_1b_2 - a_2b_1)\,dx \wedge dy.$$

The proof of this is a straightforward calculation (Exercise 5).

Orientation and Integration of Differential 2-forms on $\mathbb{R}^3$

Again by analogy with the case of 2-forms on the plane, we define a **differential 2-form** on a region $D \subset \mathbb{R}^3$ to be a mapping ω which assigns to each point $p \in D$ a 2-form ω_p on the tangent space $T_p\mathbb{R}^3$. From Lemma 6.6.3 we can write ω_p as a linear combination of the basic 2-forms

$$\omega_p = a_1(p)\ dx_2 \wedge dx_3 + a_2(p)\ dx_3 \wedge dx_1 + a_3(p)\ dx_1 \wedge dx_2$$

or represent it via the associated vectorfield

$$\overrightarrow{F}(p) = a_1(p)\,\overrightarrow{\imath} + a_2(p)\,\overrightarrow{\jmath} + a_3(p)\,\overrightarrow{k}.$$

We shall call the form $\boldsymbol{\mathcal{C}^r}$ if each of the three functions $a_i(p)$ is $\mathcal{C}^r$ on D.

The integration of 2-forms in $\mathbb{R}^3$ is carried out over surfaces in a manner analogous to the integration of 1-forms over curves described in § 6.1. There, we saw that the integral of a 1-form over a curve $\mathcal{C}$ depends on a choice of orientation for $\mathcal{C}$; reversing the orientation also reverses the sign of the integral. The same issue arises here, but in a more subtle way.

Suppose the orientation of $\mathfrak{S}$ is given by the unit normal vector field $\overrightarrow{n}$, and $\overrightarrow{p}(s,t)$ is a regular parametrization of $\mathfrak{S}$. We can define the **pullback** of a form ω by $\overrightarrow{p}$ as the 2-form on the domain $D \subset \mathbb{R}^2$ of $\overrightarrow{p}$ defined for $(s,t) \in D$ and $\overrightarrow{v}, \overrightarrow{w} \in T_{(s,t)}\mathbb{R}^2$ by

$$[\overrightarrow{p}^*(\omega)]_{(s,t)}(\overrightarrow{v}, \overrightarrow{w}) = \omega_{\overrightarrow{p}(s,t)}\left(T_{(s,t)}\overrightarrow{p}(\overrightarrow{v}), T_{(s,t)}\overrightarrow{p}(\overrightarrow{w})\right). \tag{6.33}$$

This pullback will at each point be a multiple of the basic form $ds \wedge dt$, say $[\overrightarrow{p}^*(\omega)]_{(s,t)} = f(s,t)\ ds \wedge dt$, and we define the integral of ω over $\mathfrak{S}$ as the (usual double) integral of f over D:

$$\int_{\mathfrak{S}} \omega := \iint_D f(s,t)\ ds\, dt. \tag{6.34}$$

So where does the orientation come in? This is a subtle and rather confusing point, going back to the distinction between area and *signed* area in the plane.

When we initially talked about "positive" orientation of an oriented triangle in the plane, we had a "natural" point of view on the standard xy-plane: a positive rotation was a counterclockwise one, which meant the direction from the positive x-axis toward the positive y-axis. Thus, we implicitly thought of the xy-plane as being the plane $z = 0$ in $\mathbb{R}^3$, and viewed it from the direction of the positive z-axis: in other words, we gave the xy-plane the orientation determined by the unit normal $\overrightarrow{k}$. Another way to say this is that our orientation amounted to choosing x as the *first* parameter and y as the *second*. With this orientation, the *signed* area of a positively oriented triangle $[A, B, C]$, coming from a determinant, agrees with the ordinary area of $\triangle ABC$, coming from the double integral $\iint_{\triangle ABC} dx\, dy$ (which is always non-negative). If we had followed Alice through the looking-glass and seen the xy-plane from *below* (that is, with the orientation reversed), then the same oriented triangle would have had *negative* signed area. Recall that this actually happens in a different plane—the xz-plane—where the orientation coming from alphabetical order (x before z) corresponds to viewing the plane from the *negative* y-axis, which is why, when we calculated the cross-product, we preceded the minor involving x and z with a minus sign.

But what is the orientation of the domain of a parametrization $\overrightarrow{p}(s,t)$ of $\mathfrak{S}$? You might say that counterclockwise, or positive, rotation is from the positive s-axis toward the positive t-axis, but this means we are automatically adopting alphabetical order, which is an artifact of our purely arbitrary choice of names for the parameters. We need to have a more "natural"—which is to say geometric—choice of orientation for our parameters. It stands to reason that this choice should be related to the orientation we have chosen for $\mathfrak{S}$. So here's the deal: we start with the orientation on $\mathfrak{S}$ given by the unit normal vector field $\overrightarrow{n}$ on $\mathfrak{S}$. This vector field can be viewed as the vector representative of a 2-form acting on pairs $\overrightarrow{v}, \overrightarrow{w}$ of vectors tangent to $\mathfrak{S}$ (at a common point: $\overrightarrow{v}, \overrightarrow{w} \in T_p\mathfrak{S}$) defined, following Equation (6.29), by

$$\Omega_p(\overrightarrow{v}, \overrightarrow{w}) = \overrightarrow{n} \cdot (\overrightarrow{v} \times \overrightarrow{w}).$$

When we pull this back by $\overrightarrow{p}$, we have a form $\overrightarrow{p}^*(\Omega)$ on the parameter space, so it is a nonzero multiple of $ds \wedge dt$, and of course the opposite multiple of $dt \wedge ds$. The orientation of parameter space corresponding to the order of the parameters *for which this multiple is positive* is the orientation **induced** by the parametrization $\overrightarrow{p}$. In other words, the "basic" 2-form on parameter space is the wedge product of ds and dt *in the order specified by the induced orientation*: when we chose the function $f(s,t)$ in Definition 6.34 which we integrate over the domain D of our parametrization (in the ordinary double-integral sense) to calculate $\iint_{\mathfrak{S}} \omega$, we should have defined it as $[\overrightarrow{p}^*(\omega)]_{(s,t)} = f(s,t)\ dt \wedge ds$ if the order given by the induced parametrization corresponded to t before s.

How does this work in practice? Given the parametrization $\overrightarrow{p}(s,t)$ of $\mathfrak{S}$, let us denote the unit vector along the positive s-axis (*resp.* positive t-axis) in parameter space by $\overrightarrow{e}_s$ (*resp.* $\overrightarrow{e}_t$). On one hand, $ds \wedge dt$ can be characterized as the unique 2-form on parameter space such that $(ds \wedge dt)(\overrightarrow{e}_s, \overrightarrow{e}_t) = 1$ (while $dt \wedge ds$ is characterized by $(dt \wedge ds)(\overrightarrow{e}_t, \overrightarrow{e}_s) = 1$). On the other hand, the pullback $\overrightarrow{p}^*(\Omega)$ acts on these vectors via

$$\overrightarrow{p}^*\Omega(\overrightarrow{e}_s, \overrightarrow{e}_t) = \Omega(T\overrightarrow{p}\,(\overrightarrow{e}_s), T\overrightarrow{p}\,(\overrightarrow{e}_t))\,.$$

Note that the two vectors in the last expression are by definition the partials of the parametrization

$$\begin{aligned} T\overrightarrow{p}\,(\overrightarrow{e}_s) &= \frac{\partial \overrightarrow{p}}{\partial s} \\ T\overrightarrow{p}\,(\overrightarrow{e}_t) &= \frac{\partial \overrightarrow{p}}{\partial t} \end{aligned}$$

and substituting this into the calculation above yields

$$\begin{aligned} \overrightarrow{p}^*\Omega(\overrightarrow{e}_s, \overrightarrow{e}_t) &= \Omega\left(\frac{\partial \overrightarrow{p}}{\partial s}, \frac{\partial \overrightarrow{p}}{\partial t}\right) \\ &= \overrightarrow{n} \cdot (\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}). \end{aligned}$$

If this is positive, then our orientation puts s before t, while if it is negative, we should put t before s.

Let's formalize this in a definition.

Definition 6.6.5. *Suppose $\overrightarrow{p}(s,t)$ is a regular parametrization of the surface $\mathfrak{S}$ oriented by the unit normal vector field $\overrightarrow{n}$.*

1. *The **basic form** on parameter space induced by $\overrightarrow{p}$ is the choice $dA = ds \wedge dt$ or $dA = dt \wedge ds$, where the order of ds and dt is chosen so that the cross product of the partials of $\overrightarrow{p}$ in the same order has a positive dot product with $\overrightarrow{n}$.*

2. *Suppose ω is a 2-form defined on $\mathfrak{S}$. Then its pullback via $\overrightarrow{p}$ is a function multiple of the basic form induced by $\overrightarrow{p}$:*

$$\overrightarrow{p}^*(\omega) = f(s,t)\ dA.$$

3. *We define the integral of ω over the surface $\mathfrak{S}$ with orientation given by $\overrightarrow{n}$ as the (ordinary) integral of f over the domain D of $\overrightarrow{p}$:*

$$\iint_{\mathfrak{S}} \omega := \iint_D f\,dA = \iint_D f(s,t)\ ds\,dt.$$

Let's see how this works in a couple of examples.

First, let $\mathfrak{S}$ be the part of the plane $x+y+z=1$ in the first quadrant, oriented *up*, and take $\omega = dx \wedge dz$. The natural parametrization of $\mathfrak{S}$ comes from regarding it as the graph of $z = 1-x-y$:

$$\left\{\begin{array}{ccc} x & = & s \\ y & = & t \\ z & = & 1-s-t \end{array}\right. .$$

The part of this in the first quadrant $x \geq 0$, $y \geq 0$, $z \geq 0$ is the image of the domain D in parameter space specified by the inequalities

$$\left\{\begin{array}{ccccc} 0 & \leq & t & \leq & 1-s \\ 0 & \leq & s & \leq & 1 \end{array}\right. .$$

The standard normal to the plane $x+y+z=1$ is $\overrightarrow{N} = \overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}$, which clearly has a positive vertical component. This is not a unit vector (we would have to divide by $\sqrt{3}$) but this is immaterial; it is only the direction that matters. The partials of the parametrization are

$$\begin{aligned}\frac{\partial\overrightarrow{p}}{\partial s} &= \overrightarrow{\imath} - \overrightarrow{k} \\ \frac{\partial\overrightarrow{p}}{\partial t} &= \overrightarrow{\jmath} - \overrightarrow{k}\end{aligned}$$

with cross product

$$\begin{aligned}\frac{\partial\overrightarrow{p}}{\partial s} \times \frac{\partial\overrightarrow{p}}{\partial t} &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{vmatrix} \\ &= \overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}\end{aligned}$$

so of course its dot product with $\overrightarrow{N}$ is positive; thus our basic 2-form on parameter space is

$$dA = ds \wedge dt.$$

Now, the pullback of ω is simply a matter of substitution: the differentials of the components of the parametrization are

$$\begin{aligned}dx &= ds \\ dy &= dt \\ dz &= -\,ds -\,dt\end{aligned}$$

so the pullback of ω, which is simply the expression for ω in terms of our parameters and their differentials, is

$$\begin{aligned}\omega &= dx \wedge dz \\ &= (\,ds) \wedge (-\,ds -\,dt) \\ &= -\,ds \wedge ds -\,ds \wedge dt \\ &= -\,dA\end{aligned}$$

so

$$f(s,t) = -1$$

and

$$\begin{aligned}\iint_{\mathfrak{S}} \omega &= \iint_D -1\,dA\\ &= -\iint_D ds\,dt\\ &= -\int_0^1 \int_0^{(1-s)} dt\,ds\\ &= -\int_0^1 (1-s)\,ds\\ &= -\left(s-\frac{s^2}{2}\right)_0^1\\ &= -\frac{1}{2}.\end{aligned}$$

As a second example, we take $\mathfrak{S}$ to be the part of the sphere $x^2+y^2+z^2=1$ cut out by the horizontal planes $z=-\frac{1}{\sqrt{2}}$ and $z=\frac{1}{2}$, the xz-plane, and the vertical half-plane containing the z-axis together with the vector $\overrightarrow{\imath}+\overrightarrow{\jmath}$. We orient $\mathfrak{S}$ *inward* (that is, toward the origin) and let $\omega = z\,dx\wedge dy$. The natural way to parametrize this is using spherical coordinates (with $\rho=1$):

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

The domain of this parametrization is

$$\begin{aligned}\frac{\pi}{3} \leq &\phi \leq \frac{3\pi}{4}\\ 0 \leq &\theta \leq \frac{\pi}{4}.\end{aligned}$$

The partials of the parametrization are

$$\begin{aligned}\frac{\partial\overrightarrow{p}}{\partial\phi} &= (\cos\phi\cos\theta)\overrightarrow{\imath} + (\cos\phi\sin\theta)\overrightarrow{\jmath} - (\sin\phi)\overrightarrow{k}\\ \frac{\partial\overrightarrow{p}}{\partial\theta} &= (-\sin\phi\sin\theta)\overrightarrow{\imath} + (\sin\phi\cos\theta)\overrightarrow{\jmath};\end{aligned}$$

their cross product is

$$\begin{aligned}\frac{\partial \overrightarrow{p}}{\partial \phi} \times \frac{\partial \overrightarrow{p}}{\partial \theta} &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ \cos\phi\cos\theta & \cos\phi\sin\theta & -\sin\phi \\ -\sin\phi\sin\theta & \sin\phi\cos\theta & 0 \end{vmatrix} \\ &= (\sin^2\phi\cos\theta)\overrightarrow{\imath} + (\sin^2\phi\sin\theta)\overrightarrow{\jmath} + (\sin\phi\cos\phi)\overrightarrow{k}.\end{aligned}$$

It is hard to see how this relates to the inward normal from this formula; however, we need only check the sign of the dot product at one point. At $(1,0,0)$, where $\phi = \frac{\pi}{2}$ and $\theta = 0$, the cross product is $\overrightarrow{\imath}$, while the *inward* pointing normal is $-\overrightarrow{\imath}$. Therefore, the basic form is

$$dA = d\theta \wedge d\phi.$$

To calculate the pullback of ω, we first find the differentials of the components of $\overrightarrow{p}$:

$$\begin{aligned}dx &= \cos\phi\cos\theta\,d\phi - \sin\phi\sin\theta\,d\theta \\ dy &= \cos\phi\sin\theta\,d\phi + \sin\phi\cos\theta\,d\theta \\ dz &= -\sin\phi\,d\phi.\end{aligned}$$

Then

$$\begin{aligned}\omega &= z\,dx \wedge dy \\ &= (\cos\phi)\{(\cos\phi\cos\theta\,d\phi - \sin\phi\sin\theta\,d\theta) \wedge (\cos\phi\sin\theta\,d\phi + \sin\phi\cos\theta\,d\theta)\} \\ &= (\cos\phi)\{(\cos\phi\cos\theta\sin\phi\cos\theta)\,d\phi \wedge d\theta - (\sin\phi\sin\theta\cos\phi\cos\theta)\,d\theta \wedge d\phi\} \\ &= (\cos\phi)\{(\cos\phi\sin\phi\cos^2\theta + \sin\phi\cos\phi\sin^2\theta)\,d\phi \wedge d\theta \\ &= (\cos^2\phi\sin\phi)\,d\phi \wedge d\theta \\ &= -\cos^2\phi\sin\phi\,dA.\end{aligned}$$

Thus,

$$\begin{aligned}
\iint_{\mathfrak{S}} \omega &= \iint_D -\cos^2\phi \sin\phi \, dA \\
&= \int_0^{\pi/4} \int_{\pi/3}^{3\pi/4} -\cos^2\phi \sin\phi \, d\phi \, d\theta \\
&= \int_0^{\pi/4} \left(\frac{1}{3}\cos^3\phi\right)_{\pi/3}^{3\pi/4} d\theta \\
&= \frac{1}{3}\int_0^{\pi/4} \left(-\frac{1}{2\sqrt{2}-\frac{1}{8}}\right) d\theta \\
&= -\frac{1}{3}\left(\frac{1}{2\sqrt{2}}+\frac{1}{8}\right)\frac{\pi}{4} \\
&= -\frac{\pi}{12}\left(\frac{1}{2\sqrt{2}}+\frac{1}{8}\right) \\
&= -\frac{\pi(4+\sqrt{2})}{96\sqrt{2}}.
\end{aligned}$$

Stokes' Theorem in the Language of Forms

To translate between flux integrals of vector fields and integrals of forms over oriented surfaces, we first look more closely at the "basic form" dA induced by a parametrization $\overrightarrow{p}(s,t)$ of the oriented surface $\mathfrak{S}$. This was defined in terms of the pullback of the form Ω which acted on a pair of vectors tangent to $\mathfrak{S}$ at the same point by dotting their cross product with the unit normal $\overrightarrow{n}$ defining the orientation of $\mathfrak{S}$. To calculate this pullback, let us take two vectors in parameter space and express them in terms of the unit vectors $\overrightarrow{e}_s$ and $\overrightarrow{e}_t$ in the direction of the s-axis and t-axis, respectively:

$$\begin{aligned}
\overrightarrow{v} &= v_s\overrightarrow{e}_s + v_t\overrightarrow{e}_t \\
\overrightarrow{w} &= w_s\overrightarrow{e}_s + w_t\overrightarrow{e}_t.
\end{aligned}$$

We note for future reference that these coordinates can be regarded as the values of the coordinate forms ds and dt on the respective vectors:

$$\begin{aligned}
v_s &= ds(\overrightarrow{v}) \\
v_t &= dt(\overrightarrow{v}).
\end{aligned}$$

Now, the pullback of Ω acts on $\overrightarrow{v}$ and $\overrightarrow{w}$ as follows:

$$\overrightarrow{p}^*\Omega(\overrightarrow{v},\overrightarrow{w}) = \overrightarrow{p}^*\Omega(v_s\overrightarrow{e}_s + v_t\overrightarrow{e}_t, w_s\overrightarrow{e}_s + w_t\overrightarrow{e}_t)$$

and using the linearity and antisymmetry of the form (or just Equation (6.20)) we can write this as

$$= \overrightarrow{p}^*\Omega(\overrightarrow{e}_s, \overrightarrow{e}_t)\det\begin{pmatrix} v_s & v_t \\ w_s & w_t \end{pmatrix}.$$

By definition of the pullback, the first factor is given by the action of Ω on the images of $\overrightarrow{e}_s$ and $\overrightarrow{e}_t$ under the linearization of the parametrization, which are just the partials of the parametrization. Also, using our earlier observation concerning the coordinate forms together with the definition of the wedge product, we see that the second factor is simply the action of $ds \wedge dt$ on $\overrightarrow{v}$ and $\overrightarrow{w}$:

$$\begin{aligned} &= \Omega\left(\frac{\partial \overrightarrow{p}}{\partial s}, \frac{\partial \overrightarrow{p}}{\partial t}\right)(ds \wedge dt)(\overrightarrow{v}, \overrightarrow{w}) \\ &= \left\{\overrightarrow{n} \cdot \left(\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right)\right\}\{(ds \wedge dt)(\overrightarrow{v}, \overrightarrow{w})\}. \end{aligned}$$

Note that if we reverse the roles of s and t in both factors, we introduce two changes of sign, so we can summarize the calculation above as

$$\overrightarrow{p}^*(\Omega) = \left\{\overrightarrow{n} \cdot \left(\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right)\right\} ds \wedge dt = \left\{\overrightarrow{n} \cdot \left(\frac{\partial \overrightarrow{p}}{\partial t} \times \frac{\partial \overrightarrow{p}}{\partial s}\right)\right\} dt \wedge ds.$$

This says that

$$\overrightarrow{p}^*(\Omega) = \left\|\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right\| dA$$

where dA is the "basic form" on parameter space determined by the orientation of $\mathfrak{S}$. This looks suspiciously like the element of surface area which we use to calculate surface integrals:

$$d\mathcal{S} = \left\|\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right\| ds\,dt;$$

in fact, the latter is precisely the expression we would put inside a double integral to calculate $\iint_{\mathfrak{S}} \Omega$:

$$\iint_{\mathfrak{S}} \Omega = \int_D \left\|\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right\| dA = \iint_D \left\|\frac{\partial \overrightarrow{p}}{\partial s} \times \frac{\partial \overrightarrow{p}}{\partial t}\right\| ds\,dt = \iint_{\mathfrak{S}} 1\, d\mathcal{S}.$$

So Ω is the 2-form version of the element of surface area; we will refer to it as the **area form** of the oriented surface $\mathfrak{S}$.

The following is a simple matter of chasing definitions (Exercise 7):

Remark 6.6.6. *If the 2-form ω corresponds, according to Equation (6.32), to the vector field $\overrightarrow{F}$, then the integral of ω over an oriented surface equals the flux integral of $\overrightarrow{F}$ over the same surface:*

$$\omega \leftrightarrow \overrightarrow{F} \Rightarrow \int_{\mathfrak{S}} \omega = \iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}}. \tag{6.35}$$

We also need to extend the notion of exterior derivatives to differential 1-forms in $\mathbb{R}^3$. Formally, we do just what we did in § 6.4 for differential 1-forms in $\mathbb{R}^2$: a differential 1-form on $\mathbb{R}^3$ can be written

$$\omega = P(x,y,z)\; dx + Q(x,y,z)\; dy + R(x,y,z)\; dz$$

and we define its exterior derivative by wedging the differential of each coefficient function with the coordinate form it is associated to:

$$\begin{aligned} d\omega &= (\,dP) \wedge dx + (\,dQ) \wedge dy + (\,dR) \wedge dz \\ &= \left(\frac{\partial P}{\partial x}\,dx + \frac{\partial P}{\partial y}\,dy + \frac{\partial P}{\partial z}\,dz\right) \wedge dx + \left(\frac{\partial Q}{\partial x}\,dx + \frac{\partial Q}{\partial y}\,dy + \frac{\partial Q}{\partial z}\,dz\right) \wedge dy \\ &\quad + \left(\frac{\partial R}{\partial x}\,dx + \frac{\partial R}{\partial y}\,dy + \frac{\partial R}{\partial z}\,dz\right) \wedge dz \\ &= \frac{\partial P}{\partial y}\,dy \wedge dx + \frac{\partial P}{\partial z}\,dz \wedge dx + \frac{\partial Q}{\partial x}\,dx \wedge dy + \frac{\partial Q}{\partial z}\,dz \wedge dy \\ &\quad + \frac{\partial R}{\partial x}\,dx \wedge dz + \frac{\partial R}{\partial y}\,dy \wedge dz \\ &= \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy \wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz \wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy. \end{aligned}$$

As with the wedge product, it is straightforward to show that this corresponds in our dictionary for translating between vector fields and forms to the curl (Exercise 6):

Remark 6.6.7. *If the 1-form ω corresponds, according to Equation (6.31), to the vector field $\overrightarrow{F}$, then its exterior derivative corresponds, according to Equation (6.32), to the curl of $\overrightarrow{F}$:*

$$\omega \leftrightarrow \overrightarrow{F} \quad \Leftrightarrow \quad d\omega \leftrightarrow \overrightarrow{\nabla} \times \overrightarrow{F}.$$

Using this dictionary, we can now state Stokes' Theorem in terms of forms:

Theorem 6.6.8 (Stokes' Theorem, Differential Form)**.** *Suppose ω is a differential 1-form defined on an open set in $\mathbb{R}^3$ containing the surface $\mathfrak{S}$ with boundary $\partial\mathfrak{S}$.*

Then

$$\oint_{\partial\mathfrak{S}} \omega = \iint_{\mathfrak{S}} d\omega.$$

Exercises for § 6.6

Practice problems:

1. Which of the following polynomials give bilinear functions on $\mathbb{R}^3$? For each one that does, give the matrix representative and decide whether it is commutative, anti-commutative, or neither.

   (a) $x_1x_2 + y_1y_2 - z_1z_2$

   (b) $x_1y_1 + x_2y_2 - y_1z_1 + y_2z_2$

   (c) $x_1y_2 - y_1z_2 + x_2z_1 + z_1y_2 + y_1x_2 - z_2x_1$

   (d) $(x_1 + 2y_1 + 3z_1)(x_2 - y_2 + 2z_2)$

   (e) $(x_1 + y_1 + z_1)(2x_2 + y_2 + z_2) - (2x_1 + y_1 + z_1)(x_2 + y_2 + z_2)$

   (f) $(x_1 - 2y_1 + 3z_1)(x_2 - y_2 - z_2) - (x_1 - y_1 - z_1)(2y_2 - x_2 - 3z_2)$

2. For each vector field $\overrightarrow{F}$ below, write the 2-form ω associated to it via Equation (6.29) as $\omega = A\,dx \wedge dy + B\,dy \wedge dz + C\,dz \wedge dx$.

   (a) $\overrightarrow{F} = \overrightarrow{\imath}$

   (b) $\overrightarrow{F} = \overrightarrow{\jmath}$

   (c) $\overrightarrow{F} = \overrightarrow{k}$

   (d) $\overrightarrow{F} = \overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}$

   (e) $\overrightarrow{F} = 2\overrightarrow{\imath} - 3\overrightarrow{\jmath} + 4\overrightarrow{k}$

   (f) $\overrightarrow{F} = (x + y)\overrightarrow{\imath} + (x - y)\overrightarrow{\jmath} + (y + z)\overrightarrow{k}$

   (g) $\overrightarrow{F} = y\overrightarrow{\imath} + z\overrightarrow{\jmath} + x\overrightarrow{k}$

   (h) $\overrightarrow{F} = x^2\overrightarrow{\imath} + z^2\overrightarrow{\jmath} + (x + y)\overrightarrow{k}$

   (i) $\overrightarrow{F} = \overrightarrow{\nabla} f$, where $f(x, y, z)$ is a $\mathcal{C}^2$ function.

3. For each differential 2-form ω below, find the vector field $\overrightarrow{F}$ corresponding to it via Equation (6.29).

(a) $\omega = dx \wedge dy$

(b) $\omega = dx \wedge dz$

(c) $\omega = dy \wedge dz$

(d) $\omega = x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy$

(e) $\omega = df \wedge (dx + dy + dz)$, where df is the differential of the $\mathcal{C}^2$ function $f(x,y,z)$. (Write the answer in terms of the partial derivatives of f.)

4. Evaluate $\iint_{\mathfrak{S}} \omega$.

(a) $\omega = x\,dy \wedge dz$, $\mathfrak{S}$ is the plane $x + y + z = 1$ in the first octant, oriented up.

(b) $\omega = z\,dx \wedge dy$, $\mathfrak{S}$ is the graph $z = x^2 + y^2$ over $[0,1] \times [0,1]$, oriented up.

(c) $\omega = x\,dy \wedge dz$, $\mathfrak{S}$ is the graph $z = x^2 + y^2$ over $[0,1] \times [0,1]$, oriented up.

(d) $\omega = x^2\,dx \wedge dz$, $\mathfrak{S}$ is the graph $z = x^2 + y^2$ over $[0,1] \times [0,1]$, oriented down.

(e) $\omega = dx \wedge dy$, $\mathfrak{S}$ is the part of the sphere $x^2 + y^2 + z^2 = 1$ in the first octant, oriented outward.

(f) $\omega = dx \wedge dz$, $\mathfrak{S}$ is the part of the sphere $x^2 + y^2 + z^2 = 1$ in the first octant, oriented outward.

(g) $\omega = x\,dy \wedge dz$, $\mathfrak{S}$ is the part of the sphere $x^2 + y^2 + z^2 = 1$ in the first octant, oriented inward.

(h) $\omega = x\,dy \wedge dz - y\,dx \wedge dz + z\,dx \wedge dy$, $\mathfrak{S}$ is given by the parametrization

$$\left\{ \begin{array}{rr} x = & r\cos\theta \\ y = & r\sin\theta \\ z = & \theta \end{array} \right. , \quad \left\{ \begin{array}{lcr} 0 & \leq r \leq & 1 \\ 0 & \leq \theta \leq & 2\pi \end{array} \right. ,$$

with the orientation induced by the parametrization.

(i) $\omega = z\,dx \wedge dy$, $\mathfrak{S}$ is parametrized by

$$\left\{ \begin{array}{rr} x = & \cos^3 t \\ y = & \sin^3 t \\ z = & s \end{array} \right. , \quad \left\{ \begin{array}{lcr} 0 & \leq t \leq & 1 \\ 0 & \leq s \leq & 2\pi \end{array} \right. ,$$

with the orientation induced by the parametrization.

(j) $\omega = x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy$, $\mathfrak{S}$ is the surface of the cube $[0,1] \times [0,1] \times [0,1]$, oriented outward.

Theory problems:

5. Prove Remark 6.6.4. (*Hint:* Carry out the formal wedge product, paying careful attention to order, and compare with the correspondence on 2-forms.)

6. Prove Remark 6.6.7.

7. Prove Remark 6.6.6

Challenge problem:

8. Show that every 2-form on $\mathbb{R}^3$ can be expressed as the wedge product of two 1-forms. This shows that the notion of a "basic" 2-form on p. 640 depends on the coordinate system we use.

6.7 The Divergence Theorem

So far, we have seen how the Fundamental Theorem for Line Integrals (Theorem 6.2.1) relates the line integral of a gradient vector field $\overrightarrow{F} = \overrightarrow{\nabla} f$ over a directed curve $\mathcal{C}$ to the values of the potential function f at the ends of $\mathcal{C}$, and how Green's Theorem (Theorem 6.3.4) and its generalization, Stokes' Theorem (Theorem 6.5.6) relate the flux integral of the curl $\overrightarrow{\nabla} \times \overrightarrow{F}$ of a vector field $\overrightarrow{F}$ on a domain D in $\mathbb{R}^2$ or a surface $\mathfrak{S}$ in $\mathbb{R}^3$ to its circulation integral around the boundary ∂D (*resp.* $\partial\mathfrak{S}$) of D (*resp.* $\mathfrak{S}$). In both cases, we have a relation between the integral in a *domain* of something obtained via an operation involving derivatives (or **differential operator**) applied to a function (in the case of the Fundamental Theorem for Line Integrals) or vector field (in the case of Green's and Stokes' Theorems) and the "integral" of that object on the *boundary* of the domain. In this section, we consider the third great theorem of integral calculus for vector fields, relating the flux integral of a vector field $\overrightarrow{F}$ over the boundary of a three-dimensional region to the integral of a related object, obtained via another differential operator from $\overrightarrow{F}$, over the region. This is known variously as the *Divergence Theorem*, *Gauss's Theorem*, or the *Ostrogradsky Theorem*; the differential operator in this case is the *divergence* of the vector field.

Green's Theorem Revisited: Divergence of a Planar Vector Field

A two-dimensional version of the Divergence Theorem is outlined in Exercise 7 in § 6.3. We recall it here:

Theorem 6.7.1 (Green's Theorem, Divergence Form)**.** *Suppose $D \subset \mathbb{R}^2$ is a regular planar region bounded by a simple, closed regular curve $\mathcal{C} = \partial D$ with positive orientation, and $\overrightarrow{F}(x,y) = P(x,y)\,\overrightarrow{\imath} + Q(x,y)\,\overrightarrow{\jmath}$ is a $\mathcal{C}^1$ vector field on D. Let $\overrightarrow{N}_-$ denote the* outward *pointing unit normal vector field to $\mathcal{C}$.*

Then

$$\oint_{\mathcal{C}} \overrightarrow{F} \cdot \overrightarrow{N}_-\, d\mathfrak{s} = \iint_D \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dA. \tag{6.36}$$

We note that the left side of Equation (6.36), the line integral around $\mathcal{C}$ of the *normal* component of $\overrightarrow{F}$ (by contrast with the *tangential* component which appears in Theorem 6.3.4), is the analogue in one lower dimension of the flux integral of $\overrightarrow{F}$; if we imagine a simplified two-dimensional model of fluid flow, with $\overrightarrow{F}$ the velocity (or momentum) field, then this measures the amount of "stuff" leaving D per unit time. The right side of Equation (6.36) differs from Theorem 6.3.4 in that instead of the *difference* of *cross*-derivatives of the components of $\overrightarrow{F}$ we have the *sum* of the "pure" derivatives—the x-partial of the x-component of $\overrightarrow{F}$ plus the y-partial of the y-component of $\overrightarrow{F}$. This is called the **divergence** of $\overrightarrow{F}$:

$$\operatorname{div}(P\overrightarrow{\imath} + Q\overrightarrow{\jmath}) = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}.$$

To gain some intuition about the divergence, we again think of $\overrightarrow{F}$ as the velocity field of a fluid flow, and consider the effect of this flow on the area of a small square with sides parallel to the coordinate axes (Figure 6.19). As in our intuitive discussion of curl on p. 605, a constant vector field will not affect the area; it will be the *change* in $\overrightarrow{F}$ which affects the area. In the previous discussion, we saw that the *vertical* change $\frac{\partial P}{\partial y}$ in the *horizontal* component of $\overrightarrow{F}$ (*resp.* the *horizontal* change $\frac{\partial Q}{\partial x}$ in the *vertical* component of $\overrightarrow{F}$) tends to a *shear* effect, and these effects will not change the area (by Cavalieri's principle). However, the the *horizontal* change $\frac{\partial P}{\partial x}$ in the *horizontal* component of $\overrightarrow{F}$ will tend to "stretch" the projection of the base of the square onto the x-axis, and similarly the the *vertical* change $\frac{\partial Q}{\partial y}$ in the *vertical* component of $\overrightarrow{F}$ will "stretch" the height, which is to say the vertical dimension of the square. A stretch in either of these directions increases the area of the square. Thus, we see, at least on a purely heuristic level, that div $\overrightarrow{F}$ measures *the tendency of the velocity field to increase areas.* As before, this argument comes with a disclaimer: rigorously speaking, this interpretation of divergence is a *consequence* of Theorem 6.7.1 (Exercise 15 gives a proof based on the Change-of-Variables formula, Theorem 5.3.11).

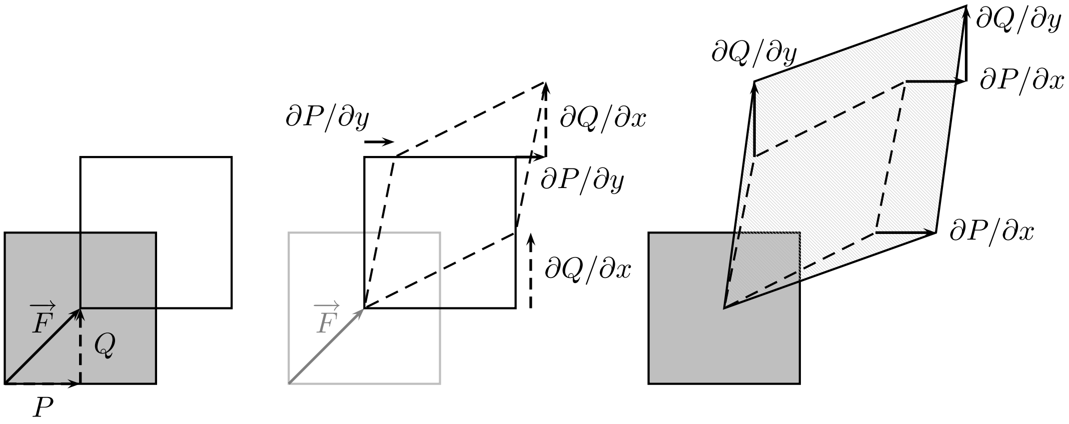


Figure 6.19: Interpretation of planar divergence

Divergence of a Vector Field in $\mathbb{R}^3$

For a vector field in space, there are three components, and it is formally reasonable that the appropriate extension of divergence to this case involves adding the partial of the *third* component.

Definition 6.7.2. *The* ***divergence*** *of a vector field*

$$\overrightarrow{F}(x,y,z) = P(x,y,z)\,\overrightarrow{\imath} + Q(x,y,z)\,\overrightarrow{\jmath} + R(x,y,z)\,\overrightarrow{k}$$

is

$$\operatorname{div}\overrightarrow{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

The heuristic argument we gave in the planar case can be extended, with a little more work, to an interpretation of this version of divergence as measuring the tendency of a fluid flow in $\mathbb{R}^3$ to increase *volumes* (Exercise 9). Note that the divergence of a vector field is a *scalar*, by contrast with its curl, which is a *vector*. If one accepts the heuristic argument that this reflects change in volume, then this seems natural: rotation has a direction (given by the axis of rotation), but volume is itself a scalar, and so its rate of change should also be a scalar. Another, deeper reason for this difference will become clearer when we consider the version of this theory using differential forms.

We can use the "del" operator $\overrightarrow{\nabla} = \overrightarrow{\imath}\frac{\partial}{\partial x} + \overrightarrow{\jmath}\frac{\partial}{\partial y} + \overrightarrow{k}\frac{\partial}{\partial z}$ to fit divergence into the formal scheme we used to denote the calculation of the differential of a function and the curl of a vector field: the divergence of $\overrightarrow{F}$ is the *dot* product of $\overrightarrow{\nabla}$ with $\overrightarrow{F}$:

$$\operatorname{div}\overrightarrow{F} = \overrightarrow{\nabla}\cdot\overrightarrow{F}.$$

Just to solidify our sense of this new operator, let us compute a few examples: if

$$\overrightarrow{F}(x,y,z) = ax\overrightarrow{\imath} + bx\overrightarrow{\jmath} + cx\overrightarrow{k}$$

then

$$\operatorname{div}\overrightarrow{F} = a + b + c.$$

This makes sense in terms of our heuristic: a fluid flow with this velocity increases the scale of each of the coordinates by a constant increment per unit time, and so we expect volume to be increased by $a + b + c$.

By contrast, the vector field

$$\overrightarrow{F}(x,y,z) = -y\overrightarrow{\imath} + x\overrightarrow{\jmath}$$

has divergence

$$\begin{aligned}\operatorname{div}\overrightarrow{F} &= -1 + 1\\ &= 0.\end{aligned}$$

A heuristic explanation for this comes from the geometry of the flow: this vectorfield represents a "pure" rotation about the z-axis, and rotating a body does not change its volume. In fact, the same is true of the "screwlike" (technically, *helical*) motion associated to the vector field obtained by adding a constant field to the vector field above. In fact, the vector fields which we use to represent the "infinitesimal" rotation induced by a flow—in other words, the vector fields which are themselves the curl of some other vector field—all have zero divergence. This is an easy if somewhat cumbersome calculation which we leave to you (Exercise 7):

Remark 6.7.3. *Every curl is **divergence-free**: if*

$$\vec{F} = \vec{\nabla} \times \vec{G}$$

for some $\mathcal{C}^2$ vector field $\vec{G}$, then

$$\operatorname{div} \vec{F} = 0.$$

Using our heuristic above, if the velocity vector field of a fluid is divergence-free, this means that the fluid has a kind of rigidity: the volume of a moving "blob" of the fluid neither increases nor decreases with the flow: such a fluid flow is called **incompressible**. A physical example is water, by contrast with a gas, which is highly compressible. [13]

This result has a converse:

Proposition 6.7.4. *A $\mathcal{C}^1$ vector field whose divergence vanishes on a simply-connected region $D \subset \mathbb{R}^3$ is the curl of some other vector field in D. That is, if*

$$\vec{F}(x,y,z) = P(x,y,z)\,\vec{\imath} + Q(x,y,z)\,\vec{\jmath} + R(x,y,z)\,\vec{k}$$

satisfies

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} = 0$$

then there exists a $\mathcal{C}^2$ vector field

$$\vec{G}(x,y,z) = g_1(x,y,z)\,\vec{\imath} + g_2(x,y,z)\,\vec{\jmath} + g_3(x,y,z)\,\vec{k}$$

such that $\vec{F} = \vec{\nabla} \times \vec{G}$—that is,

$$\frac{\partial g_3}{\partial y} - \frac{\partial g_2}{\partial z} = P \tag{6.37}$$

$$\frac{\partial g_1}{\partial z} - \frac{\partial g_3}{\partial x} = Q \tag{6.38}$$

$$\frac{\partial g_2}{\partial x} - \frac{\partial g_1}{\partial y} = R. \tag{6.39}$$

[13] A divergence-free vectorfield is also sometimes referred to as a **solenoidal** vector field.

A proof of this converse statement is outlined in Exercises 11-13. We call the vector field $\overrightarrow{G}$ a **vector potential** for $\overrightarrow{F}$ if $\overrightarrow{F} = \overrightarrow{\nabla} \times \overrightarrow{G}$. There is also a theorem, attributed to Hermann von Helmholtz (1821-1894), which states that every vector field can be written as the sum of a conservative vector field and a divergence-free one: this is called the **Helmholtz decomposition**. A proof of this is beyond the scope of this book.

The Divergence Theorem

Recall that a region $\mathfrak{D} \subset \mathbb{R}^3$ is z-regular if we can express it as the region between two graphs of z as a continuous function of x and y, in other words if we can specify $\mathfrak{D}$ by an inequality of the form

$$\varphi(x, y) \leq z \leq \psi(x, y)\,, \quad (x, y) \in \mathcal{D}$$

where $\mathcal{D}$ is some elementary region in $\mathbb{R}^2$; the analogous notions of y-regular and x-regular regions $\mathfrak{D}$ are fairly clear. We shall call $\mathfrak{D} \subset \mathbb{R}^3$ **fully regular** if it is simultaneously regular in all three directions, with the further proviso that the graphs $z = \varphi(x, y)$ and $z = \psi(x, y)$ (and their analogues for the conditions of x- and y-regularity) are both regular surfaces. This insures that we can take flux integrals across the faces of the region. We shall always assume that our region is regular, so that the boundary is piecewise regular; for this theorem we orient the boundary *outward*.

Theorem 6.7.5 (Divergence Theorem)**.** *Suppose*

$$\overrightarrow{F}(x, y, z) = P(x, y, z)\,\overrightarrow{\imath} + Q(x, y, z)\,\overrightarrow{\jmath} + R(x, y, z)\,\overrightarrow{k}$$

is a $\mathcal{C}^1$ vector field on the regular region $\mathfrak{D} \subset \mathbb{R}^3$.

Then the flux integral of $\overrightarrow{F}$ over the boundary $\partial\mathfrak{D}$, oriented outward, equals the (triple) integral of its divergence over the interior of $\mathfrak{D}$:

$$\iint_{\partial\mathfrak{D}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} = \iiint_{\mathfrak{D}} \operatorname{div} \overrightarrow{F}\, dV. \tag{6.40}$$

Proof. Equation (6.40) can be written in terms of coordinates:

$$\iint_{\partial\mathfrak{D}} (P\,\overrightarrow{\imath} + Q\,\overrightarrow{\jmath} + R\,\overrightarrow{k}) \cdot d\overrightarrow{\mathcal{S}} = \iiint_{\mathfrak{D}} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dV$$

and this in turn can be broken into three separate statements:

$$\begin{aligned}\iint_{\partial\mathfrak{D}} P\overrightarrow{\imath}\cdot d\overrightarrow{\mathcal{S}} &= \iiint_{\mathfrak{D}} \frac{\partial P}{\partial x}\,dV\\ \iint_{\partial\mathfrak{D}} Q\overrightarrow{\jmath}\cdot d\overrightarrow{\mathcal{S}} &= \iiint_{\mathfrak{D}} \frac{\partial Q}{\partial y}\,dV\\ \iint_{\partial\mathfrak{D}} R\overrightarrow{k}\cdot d\overrightarrow{\mathcal{S}} &= \iiint_{\mathfrak{D}} \frac{\partial R}{\partial z}\,dV.\end{aligned}$$

We shall prove the third of these; the other two are proved in essentially the same way (Exercise 8). For this statement, we view $\mathfrak{D}$ as a z-regular region, which means that we can specify it by a set of inequalities of the form

$$\begin{aligned}\varphi(x,y) \leq& x \leq \psi(x,y)\\ c(x) \leq& y \leq d(x)\\ a \leq& x \leq b.\end{aligned}$$

Of course the last two inequalites might also be written as limits on x in terms of functions of y, but the assumption that $\mathfrak{D}$ is simultaneously y-regular means that an expression as above is possible; we shall not dwell on this point further. In addition, the regularity assumption on $\mathfrak{D}$ means that we can assume the functions $\varphi(x,y)$ and $\psi(x,y)$ as well as the functions $c(x)$ and $d(x)$ are all $\mathcal{C}^2$.

With this in mind, let us calculate the flux integral of $R(x,y,z)\overrightarrow{k}$ across $\partial\mathfrak{D}$. The boundary of a z-regular region consists of the graphs $z=\psi(x,y)$ and $z=\varphi(x,y)$ forming the top and bottom boundary of the region and the vertical cylinder built on the boundary of the region $\mathcal{D}$ in the xy-plane determined by the second and third inequalities above. Note that the normal vector at points on this cylinder is horizontal, since the cylinder is made up of vertical line segments. This means that the dot product $R\overrightarrow{k}\cdot\overrightarrow{n}$ is zero at every point of the cylinder, so that this part of the boundary contributes nothing to the flux integral $\iint_{\partial\mathfrak{D}} R\overrightarrow{k}\cdot d\overrightarrow{\mathcal{S}}$. On the top graph $z=\psi(x,y)$ the outward normal has a positive vertical component, while on the bottom graph $z=\varphi(x,y)$ the outward normal has a *negative* vertical component. In particular, the element of oriented surface area on the top has the form

$$d\overrightarrow{\mathcal{S}} = \left(-\psi_x\overrightarrow{\imath} - \psi_y\overrightarrow{\jmath} + \overrightarrow{k}\right)\,dA$$

while on the bottom it has the form

$$d\overrightarrow{\mathcal{S}} = \left(\varphi_x\overrightarrow{\imath} + \varphi_y\overrightarrow{\jmath} - \overrightarrow{k}\right)\,dA.$$

Pulling this together with our earlier observation, we see that

$$\begin{aligned}\iint_{\partial\mathfrak{D}} R\overrightarrow{k}\cdot d\overrightarrow{\mathcal{S}} &= \iint_{z=\psi(x,y)} R\overrightarrow{k}\cdot d\overrightarrow{\mathcal{S}} + \iint_{z=\varphi(x,y)} R\overrightarrow{k}\cdot d\overrightarrow{\mathcal{S}}\\ &= \iint_{\mathcal{D}} \left(R(x,y,\psi(x,y))\,\overrightarrow{k}\right)\cdot\left(-\psi_x\overrightarrow{\imath} - \psi_y\overrightarrow{\jmath} + \overrightarrow{k}\right)\,dA\\ &\quad + \iint_{\mathcal{D}} \left(R(x,y,\varphi(x,y))\,\overrightarrow{k}\right)\cdot\left(\varphi_x\overrightarrow{\imath} + \varphi_y\overrightarrow{\jmath} - \overrightarrow{k}\right)\,dA\\ &= \iint_{\mathcal{D}} \left(R(x,y,\varphi(x,y)) - R(x,y,\psi(x,y))\right)\,dA.\end{aligned}$$

The quantity in parentheses can be interpreted as follows: given a vertical "stick" through $(x,y)\in\mathcal{D}$, we take the difference between the values of R at the ends of its intersection with $\mathfrak{D}$. Fixing (x,y), we can apply the Fundamental Theorem of Calculus to the function $f(z) = R(x,y,z)$ and conclude that for each $(x,y)\in\mathcal{D}$,

$$R(x,y,\varphi(x,y)) - R(x,y,\psi(x,y)) = \int_{\varphi(x,y)}^{\psi(x,y)} \frac{\partial R}{\partial z}(x,y,z)\,dz$$

so that

$$\begin{aligned}\iint_{\partial\mathfrak{D}} R\overrightarrow{k}\cdot d\overrightarrow{\mathcal{S}} &= \iint_{\mathcal{D}} \left(R(x,y,\varphi(x,y)) - R(x,y,\psi(x,y))\right)\,dA\\ &= \iint_{\mathcal{D}} \left(\int_{\varphi(x,y)}^{\psi(x,y)} \frac{\partial R}{\partial z}(x,y,z)\,dz\right)\,dA\\ &= \iiint_{\mathfrak{D}} \frac{\partial R}{\partial z}(x,y,z)\,dV\end{aligned}$$

as required. The other two statements are proved by a similar argument, which we leave to you (Exercise 8). □

We note that, as in the case of Green's Theorem, we can extend the Divergence Theorem to any region which can be partitioned into regular regions.

It can also be extended to regular regions with "holes" that are themselves regular regions. For example, suppose $\mathfrak{D}$ is a regular region and $B_\varepsilon(\overrightarrow{x}_0)$ is a ball of radius $\varepsilon > 0$ centered at $\overrightarrow{x}_0$ and contained in the interior of $\mathfrak{D}$ (that is, it is inside $\mathfrak{D}$ and disjoint from its boundary). Then the region $\mathfrak{D}\setminus B_\varepsilon(\overrightarrow{x}_0)$ consisting of points in $\mathfrak{D}$ but at distance at least ε from $\overrightarrow{x}_0$ is "$\mathfrak{D}$ with a hole at $\overrightarrow{x}_0$"; it has two boundary components: one is $\partial\mathfrak{D}$, oriented outward, and the other is the sphere of radius ε centered at $\overrightarrow{x}_0$, and oriented *into* the ball. Suppose for a moment that F is defined inside the ball, as well. Then the flux integral over the boundary of $\mathfrak{D} \setminus B_\varepsilon(\overrightarrow{x}_0)$ is the flux integral over the boundary of $\mathfrak{D}$, oriented outward, *minus* the flux integral over the boundary of the ball (also oriented outward). The latter is the integral of div $\overrightarrow{F}$ over the ball, so it follows that the flux integral over the boundary of $\mathfrak{D} \setminus B_\varepsilon(\overrightarrow{x}_0)$ is the integral of div $\overrightarrow{F}$ over its interior. Now, this last integral is independent of what F does inside the hole, provided it is $\mathcal{C}^1$ and agrees with the given value along the boundary. Any $\mathcal{C}^1$ vector field F defined on and outside the sphere can be extended to its interior (Exercise 14), so we have

Corollary 6.7.6. *If the ball $B_\varepsilon(\overrightarrow{x}_0)$ is interior to the regular region $\mathfrak{D}$, then the flux integral of a $\mathcal{C}^1$ vector field $\overrightarrow{F}$ over the* boundary *of $\mathfrak{D}$ with a hole $\iint_{\partial(\mathfrak{D}\setminus B_\varepsilon(\overrightarrow{x}_0))} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}}$ equals the integral of* div $\overrightarrow{F}$ *over the interior of $\mathfrak{D} \setminus B_\varepsilon(\overrightarrow{x}_0)$:*

$$\iint_{\partial(\mathfrak{D}\setminus B_\varepsilon(\overrightarrow{x}_0))} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} = \iiint_{\mathfrak{D}\setminus B_\varepsilon(\overrightarrow{x}_0)} \operatorname{div} \overrightarrow{F}\, dV. \tag{6.41}$$

In particular, if $\overrightarrow{F}$ s divergence-free in $\mathfrak{D} \setminus B_\varepsilon(\overrightarrow{x}_0)$ then the outward flux of $\overrightarrow{F}$ over $\partial(\mathfrak{D} \setminus B_\varepsilon(\overrightarrow{x}_0))$ equals the outward flux of $\overrightarrow{F}$ over the sphere of radius ε centered at $\overrightarrow{x}_0$.

Like Stokes' Theorem, the Divergence Theorem allows us to compute the same integral two different ways. We consider a few examples.

First, let us calculate directly the flux of the vector field

$$\overrightarrow{F}(x,y,z) = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$$

out of the sphere $\mathfrak{S}$ of radius R about the origin.

The natural parametrization of this sphere uses spherical coordinates:

$$\begin{cases} x &= R\sin\phi\cos\theta \\ y &= R\sin\phi\sin\theta \\ z &= R\cos\phi, \end{cases}, \qquad \begin{cases} 0 \leq & \phi \leq \pi \\ 0 \leq & \theta \leq 2\pi. \end{cases}$$

The partials are

$$\frac{\partial\overrightarrow{p}}{\partial\phi} = (R\cos\phi\cos\theta)\overrightarrow{\imath} + (R\cos\phi\sin\theta)\overrightarrow{\jmath} - (R\sin\phi)\overrightarrow{k}$$
$$\frac{\partial\overrightarrow{p}}{\partial\theta} = (-R\sin\phi\sin\theta)\overrightarrow{\imath} + (R\sin\phi\cos\theta)\overrightarrow{\jmath}$$

with cross product

$$\begin{aligned}\frac{\partial\overrightarrow{p}}{\partial\phi}\times\frac{\partial\overrightarrow{p}}{\partial\theta} &= \begin{vmatrix} \overrightarrow{\imath} & \overrightarrow{\jmath} & \overrightarrow{k} \\ R\cos\phi\cos\theta & R\cos\phi\sin\theta & -R\sin\phi \\ -R\sin\phi\sin\theta & R\sin\phi\cos\theta & 0 \end{vmatrix} \\ &= (R^2\sin^2\phi\cos\theta)\overrightarrow{\imath} + (R^2\sin^2\phi\sin\theta)\overrightarrow{\jmath} + (R^2\sin\phi\cos\phi)\overrightarrow{k}.\end{aligned}$$

To check whether this gives the outward orientation, we compute the direction of this vector at a point where it is easy to find, for example at $(1,0,0) = \overrightarrow{p}\left(\frac{\pi}{2},0\right)$:

$$\left(\frac{\partial\overrightarrow{p}}{\partial\phi}\times\frac{\partial\overrightarrow{p}}{\partial\theta}\right)\left(\frac{\pi}{2},0\right) = R^2\overrightarrow{\imath}$$

Which points out of the sphere at $(1,0,0)$. Thus, the element of outward oriented surface area is

$$d\overrightarrow{\mathcal{S}} = \left((R^2\sin^2\phi\cos\theta)\overrightarrow{\imath} + (R^2\sin^2\sin\theta)\overrightarrow{\jmath} + (R^2\sin\phi\cos\phi)\overrightarrow{k}\right)d\phi\,d\theta.$$

On the surface, the vector field is

$$\overrightarrow{F}(\phi,\theta) = (R\sin\phi\cos\theta)\overrightarrow{\imath} + (R\sin\phi\sin\theta)\overrightarrow{\jmath} + (R\cos\phi)\overrightarrow{k}$$

so

$$\begin{aligned}\overrightarrow{F}\cdot d\overrightarrow{\mathcal{S}} &= \Big((R\sin\phi\cos\theta)(R^2\sin^2\phi\cos\theta) + (R\sin\phi\sin\theta)(R^2\sin^2\phi\sin\theta) \\ &\quad + (R\cos\phi)(R^2\sin\phi\cos\phi)\Big)\,d\phi\,d\theta \\ &= R^3(\sin^3\phi\cos^2\theta + \sin^3\phi\sin^2\theta + \sin\phi\cos^2\phi)\,d\phi\,d\theta \\ &= R^3\sin\phi(\sin^2\phi + \cos^2\phi)\,d\phi\,d\theta \\ &= R^3\sin\phi\,d\phi\,d\theta.\end{aligned}$$

The flux integral is therefore

$$\begin{aligned}\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} &= \int_0^{2\pi}\int_0^{\pi} R^3 \sin\phi \, d\phi \, d\theta \\ &= \int_0^{2\pi} \left(-R^3\cos\phi\right)_{\phi=0}^{\pi} \, d\theta \\ &= 2R^3 \int_0^{2\pi} \, d\theta \\ &= 2R^3(2\pi) \\ &= 4\pi R^3.\end{aligned}$$

Now let us see how the same calculation looks using the Divergence Theorem. The divergence of our vector field is

$$\begin{aligned}\operatorname{div}\overrightarrow{F} &= 1+1+1 \\ &= 3\end{aligned}$$

so the Divergence Theorem tells us that

$$\begin{aligned}\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} &= \iiint_{\mathfrak{D}} \operatorname{div}\overrightarrow{F} \, dV \\ &= \iiint_{\mathfrak{D}} 3 \, dV \\ &= 3\mathcal{V}(\mathfrak{D})\end{aligned}$$

where $\mathfrak{D}$ is the sphere of radius R, with volume $\frac{4\pi R^3}{3}$, and our integral is

$$\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} = 4\pi R^3.$$

As another example, let us calculate the flux of the vector field

$$\overrightarrow{F}(x,y,z) = x^2\overrightarrow{\imath} + y^2\overrightarrow{\jmath} + z^2\overrightarrow{k}$$

over the same surface. We have already calculated that the element of outward oriented surface area is

$$d\overrightarrow{\mathcal{S}} = \left((R^2\sin^2\phi\cos\theta)\overrightarrow{\imath} + (R^2\sin^2\sin\theta)\overrightarrow{\jmath} + (R^2\sin\phi\cos\phi)\overrightarrow{k}\right) d\phi \, d\theta.$$

This time, our vector field on the surface is

$$\overrightarrow{F}(\phi,\theta) = (R^2\sin^2\phi\cos^2\theta)\overrightarrow{\imath} + (R^2\sin^2\phi\sin^2\theta)\overrightarrow{\jmath} + (R^2\cos^2\phi)\overrightarrow{k}$$

and its dot product with $d\overrightarrow{\mathcal{S}}$ is

$$\begin{aligned}\overrightarrow{F}\cdot d\overrightarrow{\mathcal{S}} &= \Bigl((R^2\sin^2\phi\cos^2\theta)(R^2\sin^2\phi\cos\theta) + (R^2\sin^2\phi\sin^2\theta)(R^2\sin^2\sin\theta) \\ &\quad + (R^2\cos^2\phi)(R^2\sin\phi\cos\phi)\bigr)\,d\phi\,d\theta \\ &= R^4\bigl(\sin^4\phi\cos^3\theta + \sin^4\phi\sin^3\theta + \sin\phi\cos^3\phi\bigr)\,d\phi\,d\theta \\ &= R^4\bigl(\frac{1}{4}\left(1-2\cos 2\phi+\cos^2\phi\right)\left(\cos^3\theta+\sin^3\theta\right) + \sin\phi\cos^3\phi\bigr)\,d\phi\,d\theta \\ &= R^4\bigl(\Bigl(\frac{3}{8}-\frac{1}{2}\cos 2\phi+\frac{1}{8}\cos 4\phi\Bigr)\bigr)(\cos^3\theta+\sin^3\theta) + \sin\phi\cos^3\phi\,d\phi\,d\theta\end{aligned}$$

and our flux integral is

$$\begin{aligned}\iint_{\mathfrak{S}}\overrightarrow{F}\cdot d\overrightarrow{\mathcal{S}} &= \int_0^{2\pi}\int_0^{\pi} R^4\bigl(\Bigl(\frac{3}{8}-\frac{1}{2}\cos 2\phi+\frac{1}{8}\cos 4\phi\Bigr)\bigr)(\cos^3\theta+\sin^3\theta) + \sin\phi\cos^{3\phi}\,d\phi\,d\theta \\ &= R^4\int_0^{2\pi}\bigl(\Bigl(\frac{3}{8}\theta-\frac{1}{4}\sin 2\phi+\frac{1}{32}\sin 4\phi\Bigr)\left(\cos^3\theta+\sin^3\theta\right) - \frac{1}{4}\cos^4\phi\bigr)_{\phi=0}^{\pi}\,d\phi\,d\theta \\ &= R^4\int_0^{2\pi}\left(\frac{3\pi}{8}\right)\left(\cos^3\theta+\sin^3\theta\right)\cos^3+\sin^3)\,d\theta \\ &= \frac{3\pi}{8}R^4\int_0^{2\pi}\bigl((1-\sin^2\theta)\cos\theta+(1-\sin^2\theta)\cos\theta\bigr)\,d\theta \\ &= \frac{3\pi}{8}R^4\bigl(\sin\theta-\frac{1}{3}\sin^3\theta-\cos\theta+\frac{1}{3}\cos^3\theta\bigr)_0^{2\pi} \\ &= 0.\end{aligned}$$

Now if we use the Divergence Theorem instead, we see that the divergence of our vector field is

$$\operatorname{div}\overrightarrow{F} = 2x+2y+2z$$

so by the Divergence Theorem

$$\begin{aligned}\iint_{\mathfrak{S}}\overrightarrow{F}\cdot d\overrightarrow{\mathcal{S}} &= \iiint_{\mathfrak{D}}\operatorname{div}\overrightarrow{F}\,dV \\ &= \iiint_{\mathfrak{D}} 2(x+y+z)\,dV\end{aligned}$$

which is easier to do in spherical coordinates:

$$
\begin{aligned}
&= \int_0^{2\pi}\int_0^{\pi}\int_0^{R} 2(\rho\sin\phi\cos\theta + \rho\sin\phi\sin\theta + \rho\cos\phi)\rho^2\sin\phi\,d\rho\,d\phi\,d\theta \\
&= \int_0^{2\pi}\int_0^{\pi}\int_0^{R} 2\rho^3\sin\phi(\sin\phi\cos\theta + \sin\phi\sin\theta + \cos\phi)\,d\rho\,d\phi\,d\theta \\
&= \int_0^{2\pi}\int_0^{\pi}\left(\frac{\rho^4}{4}\right)_0^R \sin\phi(\sin\phi\cos\theta + \sin\phi\sin\theta + \cos\phi)\,d\rho\,d\phi\,d\theta \\
&= \frac{R^4}{4}\int_0^{2\pi}\int_0^{\pi}\big(\sin^2\phi(\cos\theta + \sin\theta) + \sin\phi\cos\phi\big)\,d\phi\,d\theta \\
&= \frac{R^4}{4}\int_0^{2\pi}\int_0^{\pi}\big(\frac{1}{2}(1-\cos 2\phi)(\cos\theta + \sin\theta) + \sin\phi\cos\phi\big)\,d\theta \\
&= \frac{R^4}{4}\int_0^{2\pi}\big(\left(\frac{\phi}{2} - \frac{1}{2}\sin 2\phi\right)(\cos\theta + \sin\theta) + \frac{1}{2}\sin^2\phi\big)_{\phi=0}^{\pi}\,d\theta \\
&= \frac{R^4}{4}\int_0^{2\pi}\big(\frac{\pi}{2}(\cos\theta + \sin\theta) + 0\big)\,d\theta \\
&= \frac{\pi R^4}{8}\big(\sin\theta - \cos\theta\big)_0^{2\pi} \\
&= 0.
\end{aligned}
$$

We note in passing that this triple integral could have been predicted to equal zero on the basis of symmetry considerations. Recall that the integral of an odd function of one real variable $f(t)$ (*i.e.*, if $f(-t) = -f(t)$) over a symmetric interval $[-a, a]$ is zero. We call a region $\mathfrak{D} \subset \mathbb{R}^3$ **symmetric in z** if it is unchanged by reflection across the xy-plane, that is, if whenever the point (x, y, z) belongs to $\mathfrak{D}$, so does $(x, y, -z)$. (The adaptation of this definition to symmetry in x or in y is left to you in Exercise 10.) We say that a function $f(x, y, z)$ is **odd in z** (*resp.* **even in z**) if reversing the sign of z but leaving x and y unchanged reverses the sign of f (*resp.* does not change f): for odd, this means

$$f(x, y, -z) = -f(x, y, z)$$

while for even it means

$$f(x, y, -z) = f(x, y, z)\,.$$

Remark 6.7.7. *If $f(x,y,z)$ is odd in z and $\mathfrak{D}$ is z-regular and symmetric in z, then*

$$\iiint_{\mathfrak{D}} f(x,y,z)\;dV = 0.$$

(To see this, just set up the triple integral and look at the innermost integral.)

Recall that one of the useful consequences of the Fundamental Theorem for Line Integrals was that the line integral of a conservative vector field depends only on the endpoints of the curve, not on the curve itself; more generally, if the curl of a vector field is zero in a region, then the line integral of the field over a curve is not changed if we deform it within that region, holding the endpoints fixed. A similar use can be made of the Divergence Test. We illustrate with an example.

Let us find the flux integral over $\mathfrak{S}$ the upper hemisphere $z = \sqrt{1-x^2-y^2}$, oriented up, of the vector field

$$\overrightarrow{F}(x,y,z) = (1+z)e^{y^2}\overrightarrow{\imath} - (z+1)e^{x^2}\overrightarrow{\jmath} + (x^2+y^2)\overrightarrow{k}.$$

Whether we think of the hemisphere as the graph of a function or parametrize it using spherical coordinates, the terms involving exponentials of squares are serious trouble. However, note that this vector field is divergence-free:

$$\operatorname{div}\overrightarrow{F} = \frac{\partial}{\partial x}\left[(1+z)e^{y^2}\right] + \frac{\partial}{\partial y}\left[(z+1)e^{x^2}\right] + \frac{\partial}{\partial z}\left[(x^2+y^2)\right] = 0.$$

Thus, if we consider the half-ball $\mathfrak{D}$ bounded above by the hemisphere and below by the unit disc in the xy-plane, the Divergence Theorem tells us that

$$\iint_{\partial\mathfrak{D}} \overrightarrow{F}\cdot d\overrightarrow{\mathcal{S}} = \iiint_{\mathcal{D}} \operatorname{div}\overrightarrow{F}\,dV = 0.$$

Now, the boundary of $\mathfrak{D}$ consists of two parts: the hemisphere, $\mathfrak{S}$, and the disc, $\mathcal{D}$. The outward orientation on $\partial\mathfrak{D}$ means an upward orientation on the hemisphere $\mathfrak{S}$, but a *downward* orientation on the disc $\mathcal{D}$. Thus, the flux integral over the whole boundary equals the flux integral over the upward-oriented hemisphere, plus the flux integral over the downward-oriented disc—which is to say, *minus* the flux over the *upward*-oriented disc. Since the difference of the two upward-oriented discs equals zero, they are equal. Thus

$$\iint_{\mathfrak{S}} \overrightarrow{F}\cdot d\overrightarrow{\mathcal{S}} = \iint_{\mathcal{D}} \overrightarrow{F}\cdot d\overrightarrow{\mathcal{S}}.$$

But on $\mathcal{D}$,

$$d\overrightarrow{\mathcal{S}} = \overrightarrow{k}\, dA$$

so, substituting $z = 0$ we see that the vector field on the disc is

$$\overrightarrow{F}(x, y) = e^{y^2}\overrightarrow{\imath} - e^{x^2}\overrightarrow{\jmath} + (x^2 + y^2)\overrightarrow{k}$$

and

$$\overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} = (x^2 + y^2)\, dA.$$

this is easy to integrate, especially when we use polar coordinates:

$$\begin{aligned}\iint_{\mathcal{D}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}} &= \int_0^{2\pi}\int_0^1 (r^2)(r\, dr\, d\theta)\\ &= \int_0^{2\pi} (\frac{r^4}{4})\, d\theta\\ &= \int_0^{2\pi} \frac{1}{4}\, d\theta\\ &= \frac{\pi}{2}.\end{aligned}$$

Exercises for § 6.7

Practice problems:

1. Use Green's Theorem to calculate the integral $\int_{\mathcal{C}} \overrightarrow{F} \cdot \overrightarrow{N}\, d\mathfrak{s}$, where $\overrightarrow{N}$ is the outward unit normal and $\mathcal{C}$ is the ellipse $x^2 + 4y^2 = 4$, traversed counterclockwise:

   (a) $\overrightarrow{F}(x, y) = x\overrightarrow{\imath} + y\overrightarrow{\jmath}$

   (b) $\overrightarrow{F}(x, y) = y\overrightarrow{\imath} + x\overrightarrow{\jmath}$

   (c) $\overrightarrow{F}(x, y) = x^2\overrightarrow{\imath} + y^2\overrightarrow{\jmath}$

   (d) $\overrightarrow{F}(x, y) = x^3\overrightarrow{\imath} + y^3\overrightarrow{\jmath}$

2. Find the flux integral $\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}}$, where $\mathfrak{S}$ is the unit sphere oriented outward, for each vector field below:

(a) $\overrightarrow{F}(x,y,z) = (x+y^2)\overrightarrow{\imath} + (y-z^2)\overrightarrow{\jmath} + (x+z)\overrightarrow{k}$

(b) $\overrightarrow{F}(x,y,z) = (x^3+y^3)\overrightarrow{\imath} + (y^3+z^3)\overrightarrow{\jmath} + (z^3-x^3)\overrightarrow{k}$

(c) $\overrightarrow{F}(x,y,z) = 2xz\overrightarrow{\imath} + y^2\overrightarrow{\jmath} + xz\overrightarrow{k}$

(d) $\overrightarrow{F}(x,y,z) = 3xz^2\overrightarrow{\imath} + y^3\overrightarrow{\jmath} + 3x^2z\overrightarrow{k}$

3. For each vector field $\overrightarrow{F}$ below, find the flux integral $\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}}$, where $\mathfrak{S}$ is the boundary of the unit cube $[0,1]\times[0,1]\times[0,1]$, oriented outward, in two different ways: (i) directly (you will need to integrate over each face separately and then add up the results) and (ii) using the Divergence Theorem.

   (a) $\overrightarrow{F}(x,y,z) = x\overrightarrow{\imath} + y\overrightarrow{\jmath} + z\overrightarrow{k}$

   (b) $\overrightarrow{F}(x,y,z) = \overrightarrow{\imath} + \overrightarrow{\jmath} + \overrightarrow{k}$

   (c) $\overrightarrow{F}(x,y,z) = x^2\overrightarrow{\imath} + y^2\overrightarrow{\jmath} + z^2\overrightarrow{k}$

4. Find the flux of the vector field

$$\overrightarrow{F}(x,y,z) = 10x^3y^2\overrightarrow{\imath} + 3y^5\overrightarrow{\jmath} + 15x^4z\overrightarrow{k}$$

   over the outward-oriented boundary of the solid cylinder $x^2+y^2 \leq 1$, $0 \leq z \leq 1$.

5. Find the flux integral $\iint_{\mathfrak{S}} \overrightarrow{F} \cdot d\overrightarrow{\mathcal{S}}$ for the vector field $\overrightarrow{F}(x,y,z) = yz\overrightarrow{\imath} + x\overrightarrow{\jmath} + xz\overrightarrow{k}$ over the boundary of each region below:

   (a) $x^2+y^2 \leq z \leq 1$

   (b) $x^2+y^2 \leq z \leq 1$ and $x \geq 0$

   (c) $x^2+y^2 \leq z \leq 1$ and $x \leq 0$.

6. Calculate the flux of the vector field $\overrightarrow{F}(x,y,z) = 5yz\overrightarrow{\imath} + 12xz\overrightarrow{\jmath} + 16x^2y^2\overrightarrow{k}$ over the surface of the cone $z^2 = x^2+y^2$ above the xy-plane and below the plane $z = 1$.

Theory problems:

7. Prove Remark 6.7.3. (*Hint:* Start with an expression for $\overrightarrow{G}$, calculate its curl, then take the divergence of that.)

8. Fill in the details of the argument for P and Q needed to complete the proof of Theorem 6.7.5.

9. Extend the heuristic argument given on p. 655 to argue that the divergence of a vector field in $\mathbb{R}^3$ reflects the tendency of a fluid flow to increase volumes.

10. (a) Formulate a definition of what it means for a region $\mathfrak{D} \subset \mathbb{R}^3$ to be symmetric in x (*resp.* in y).

    (b) Formulate a definition of what it means for a function $f(x, y, z)$ to be even, or odd, in x (*resp.* in y).

    (c) Prove that if a function $f(x, y, z)$ is odd in x then its integral over a region which is x-regular and symmetric in x is zero.

    (d) What can you say about $\iiint_{\mathfrak{D}} f(x, y, z)\, dV$ if f is *even* in x and $\mathfrak{D}$ is x-regular and symmetric in x?

In Exercises 11-13, you will prove Proposition 6.7.4, that every divergence-free vector field $\overrightarrow{F}$ is the curl of some vector field $\overrightarrow{G}$, by a direct construction based on [55] and [36, p. 560]. Each step will be illustrated by the example $\overrightarrow{F}(x, y, z) = yz\,\overrightarrow{\imath} + xz\,\overrightarrow{\jmath} + xy\,\overrightarrow{k}$.

11. (a) Given a continuous function $\phi(x, y, z)$, show how to construct a vector field whose divergence is ϕ. (*Hint:* This can even be done with a vector field parallel to a predetermined coordinate axis.)

    (b) Given a continuous function $\varphi(x, y)$, show how to construct a *planar* vector field $\overrightarrow{G}(x, y) = g_1(x, y)\,\overrightarrow{\imath} + g_2(x, y)\,\overrightarrow{\jmath}$ whose planar curl equals φ. (*Hint:* Consider the divergence of the related vector field $\overrightarrow{x, y}(=)\, g_2(x, y)\,\overrightarrow{\imath} - g_1(x, y)\,\overrightarrow{\jmath}$.)

    (c) Construct a planar vector field $\overrightarrow{G}(x, y) = g_1(x, y)\,\overrightarrow{\imath} + g_2(x, y)\,\overrightarrow{\jmath}$ with planar curl

$$\frac{\partial g_2}{\partial x}(x, y) - \frac{\partial g_1}{\partial y}(x, y) = xy.$$

12. Note that in this problem, we deal with horizontal vector fields.

    (a) Show that the curl of a horizontal vector field

$$\overrightarrow{G}(x, y, z) = g_1(x, y, z)\,\overrightarrow{\imath} + g_2(x, y, z)\,\overrightarrow{\jmath}$$

is determined by the planar curl of its restriction to each horizontal plane together with the derivatives of its components with respect to z:

$$\overrightarrow{\nabla} \times \overrightarrow{G} = -\frac{\partial g_2}{\partial z}\overrightarrow{\imath} + \frac{\partial g_2}{\partial z}\overrightarrow{\jmath} + \left(\frac{\partial g_2}{\partial x} - \frac{\partial g_1}{\partial x}\right)\overrightarrow{k}\,.$$

(b) Construct a horizontal vector field whose restriction to the xy-plane agrees with your solution to Exercise 11c—that is, such that

$$\frac{\partial g_2}{\partial x}(x,y,0) - \frac{\partial g_1}{\partial y}(x,y,0) = xy$$

which also satisfies

$$\begin{aligned}\frac{\partial g_1}{\partial z}(x,y,z) &= xz\\ \frac{\partial g_2}{\partial z}(x,y,z) &= yz.\end{aligned}$$

at all points (x,y,z). Verify that the resulting vector field $\vec{G}(x,y,z)$ satisfies

$$\vec{\nabla} \times \vec{G} = yz\vec{\imath} + xz\vec{\jmath} + xy\vec{k}.$$

13. Now suppose that

$$\vec{F}(x,y,z) = P(x,y,z)\,\vec{\imath} + Q(x,y,z)\,\vec{\jmath} + R(x,y,z)\,\vec{k}$$

is any $\mathcal{C}^1$ vector field satisfying

$$\operatorname{div}\vec{F} = 0.$$

Show that if

$$\vec{G}(x,y,z) = g_1(x,y,z)\,\vec{\imath} + g_2(x,y,z)\,\vec{\jmath}$$

is a $\mathcal{C}^2$ horizontal vector field satisfying

$$\begin{aligned}\frac{\partial g_1}{\partial z}(x,y,z) &= Q(x,y,z)\\ \frac{\partial g_2}{\partial z}(x,y,z) &= -P(x,y,z)\\ \frac{\partial g_2}{\partial x}(x,y,0) - \frac{\partial g_1}{\partial y}(x,y,0) &= R(x,y,0)\end{aligned}$$

then

$$\vec{\nabla} \times \vec{G} = \vec{F}$$

by showing that the extension of the third condition off the xy-plane

$$\frac{\partial g_2}{\partial x}(x,y,z) - \frac{\partial g_1}{\partial y}(x,y,z) = R(x,y,z)$$

holds for all z.

Challenge problems:

14. **Filling holes:** In this problem, you will show that given a vector field $\overrightarrow{F}$ defined and $\mathcal{C}^1$ on a neighborhood of a sphere, there exists a new vector field $\overrightarrow{G}$, defined and $\mathcal{C}^1$ on the neighborhood "filled in" to include the ball bounded by the sphere, such that $\overrightarrow{F} = \overrightarrow{G}$ on the sphere and its exterior. Thus, we can replace $\overrightarrow{F}$ with $\overrightarrow{G}$ on both sides of Equation (6.41), justifying our argument extending the Divergence Theorem to regions with holes (Corollary 6.7.6).

    (a) Suppose $\phi(t)$ is a $\mathcal{C}^1$ function defined on an open interval containing $[a, b]$ satisfying

    $$\begin{aligned} \phi(a) &= 0 \\ \phi'(a) &= 0 \\ \phi(b) &= 1 \\ \phi'(b) &= 0. \end{aligned}$$

    Show that the function defined for all t by

    $$\psi(t) = \begin{cases} 0 & \text{for } t \leq a, \\ \phi(t) & \text{for } a \leq t \leq b, \\ 1 & \text{for } t \geq b \end{cases}$$

    is $\mathcal{C}^1$ on the whole real line.

    (b) Given $a < b$, find values of α and β such that

    $$\phi(t) = \frac{1}{2}\left(1 - \cos\left(\alpha t + \beta\right)\right)$$

    satisfies the conditions above.

    (c) Given $a < b$ and $\phi(t)$ as above, as well as $f(t)$ defined and $\mathcal{C}^1$ on a neighborhood $(b - \varepsilon, b + \varepsilon)$ of b, show that

    $$g(t) = \begin{cases} 0 & \text{for } t < a, \\ \psi(t)\, f(t) & \text{for } a \leq t < b + \varepsilon \end{cases}$$

    is $\mathcal{C}^1$ on $(-\infty, b + \varepsilon)$.

(d) Given a $\mathcal{C}^1$ vector field $\overrightarrow{F}$ on a neighborhood $\mathcal{N}$ of the sphere $\mathcal{S}$ of radius R centered at $\overrightarrow{c}$

$$\begin{aligned}\mathcal{S} &= \{\overrightarrow{x} \mid (\overrightarrow{x} - \overrightarrow{c})^2 = R^2\}\\ \mathcal{N} &= \{\overrightarrow{x} \mid R^2 - \varepsilon \leq (\overrightarrow{x} - \overrightarrow{c})^2 \leq R^2 + \varepsilon\}\end{aligned}$$

(where $(\overrightarrow{x} - \overrightarrow{c})^2 := (\overrightarrow{x} - \overrightarrow{c}) \cdot (\overrightarrow{x} - \overrightarrow{c})$) show that the vector field $\overrightarrow{G}$ defined by

$$\overrightarrow{G}(\overrightarrow{x}) = \begin{cases} \overrightarrow{0} & \text{for } (\overrightarrow{x} - \overrightarrow{c})^2 \leq R^2 - \varepsilon, \\ \phi\big((\overrightarrow{x} - \overrightarrow{c})^2\big)\,\overrightarrow{F}(\overrightarrow{x}) & \text{for } \overrightarrow{x} \in \mathcal{N} \end{cases}$$

is $\mathcal{C}^1$ on

$$B_\varepsilon(\overrightarrow{c}) \cup \mathcal{N} = \{\overrightarrow{x} \mid (\overrightarrow{x} - \overrightarrow{c})^2 \leq R^2 + \varepsilon\}.$$

(e) Sketch how to use this to show that a $\mathcal{C}^1$ vector field defined on a region $\mathfrak{D}$ with holes can be extended to a $\mathcal{C}^1$ vector field on the region with the holes filled in. (You may assume that the vector field is actually defined on a neighborhood of each internal boundary sphere.)

15. In this problem (based on [30, pp. 362-3]), you will use the Change-of-Variables Formula (Theorem 5.3.11) to show that the divergence of a planar vector field gives the rate of change of area under the associated flow. The analogous three-dimensional proof is slightly more involved; it is given in the work cited above.

We imagine a fluid flow in space: the position at time t of a point whose position at time $t = 0$ was (x, y) is given by

$$\begin{aligned}u &= u(x, y, t)\\ v &= v(x, y, t)\end{aligned}$$

or, combining these into a mapping $F\colon \mathbb{R}^3 \to \mathbb{R}^2$,

$$\begin{aligned}(u, v) &= F(x, y, t)\\ &= F_t(x, y)\end{aligned}$$

where $F_t(x, y)$is the transformation $F_t\colon \mathbb{R}^2 \to \mathbb{R}^2$ taking a point located at (x, y) when $t = 0$ to its position at time t; that is, it is the mapping F with t fixed.

The velocity of this flow is the vector field

$$\begin{aligned} V(u,v) &= \left(\frac{\partial u}{\partial t}, \frac{\partial v}{\partial t}\right) \\ &= (u', v'). \end{aligned}$$

V may also vary with time, but we will suppress this in our notation. Let $\mathcal{D} \subset \mathbb{R}^2$ be a regular planar region; we denote the area of its image under F_t as

$$\mathcal{A}\,(t) = \mathcal{A}\,(F_t(D))\,;$$

by Theorem 5.3.11, this is

$$= \iint_{\mathcal{D}} |J_t| \; dx\,dy$$

where

$$JF_t = \begin{bmatrix} \partial u/\partial x & \partial u/\partial x \\ \partial v/\partial x & \partial v/\partial y \end{bmatrix}$$

is the Jacobian matrix of F_t, and

$$\begin{aligned} |J_t| &= \det\ JF_t \\ &= \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x} \end{aligned}$$

is its determinant. (Strictly speaking, we should take the absolute value, but it can be shown that for a continuous flow, this determinant is always positive.)

(a) Show that

$$\frac{d}{dt}\left[|J_t|\right] = \frac{\partial u'}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u'}{\partial y}\frac{\partial v}{\partial x} + \frac{\partial v'}{\partial y}\frac{\partial u}{\partial x} - \frac{\partial v'}{\partial x}\frac{\partial u}{\partial y}.$$

(b) Show that

$$\begin{aligned} \operatorname{div} V &:= \frac{\partial u'}{\partial u} + \frac{\partial v'}{\partial v} \\ &= \frac{\partial u'}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial u'}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial v'}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial v'}{\partial y}\frac{\partial y}{\partial v}. \end{aligned}$$

(c) Show that the inverse of JF_t is

$$\begin{aligned} JF_t^{-1} &:= \begin{bmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{bmatrix} \\ &= \frac{1}{|J_t|} \begin{bmatrix} \partial v/\partial y & -\partial u/\partial y \\ -\partial v/\partial x & \partial v/\partial y \end{bmatrix}. \end{aligned}$$

(*Hint:* Use the Chain Rule, and show that the product of this with JF_t is the identity matrix.)

(d) Regarding this matrix equation as four equations (between corresponding entries of the two matrices), substitute into the previous formulas to show that

$$\operatorname{div} V = \frac{1}{|J_t|} \frac{d}{dt} [|J_t|] .$$

(e) Use this to show that

$$\frac{d}{dt} [\mathcal{A}(t)] = \iint_{F_t(\mathcal{D})} \operatorname{div} V \, du \, dv.$$

6.8 3-forms and the Generalized Stokes Theorem

Multilinear Algebra

In §§6.4 and 6.6 we encountered the notion of a *bilinear function*: a function of two vector variables which is "linear in each slot": it is linear as a function of one of the variable when the other is held fixed. This has a natural extension to more vector variables:

Definition 6.8.1. *A **trilinear function** on $\mathbb{R}^3$ is a function $f(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z})$ of three vector variables $\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z} \in \mathbb{R}^3$ such that fixing the values of two of the variables results in a linear function of the third: given $\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{v}, \overrightarrow{w} \in \mathbb{R}^3$ and $\alpha, \beta \in \mathbb{R}$,*

$$\begin{aligned} f\left(\alpha\overrightarrow{v} + \beta\overrightarrow{w}, \overrightarrow{a}, \overrightarrow{b}\right) &= \alpha F\left(\overrightarrow{v}, \overrightarrow{a}, \overrightarrow{b}\right) + \beta F\left(\overrightarrow{w}, \overrightarrow{a}, \overrightarrow{b}\right) \\ f\left(\overrightarrow{a}, \alpha\overrightarrow{v} + \beta\overrightarrow{w}, \overrightarrow{b}\right) &= \alpha F\left(\overrightarrow{a}, \overrightarrow{v}, \overrightarrow{b}\right) + \beta F\left(\overrightarrow{a}, \overrightarrow{w}, \overrightarrow{b}\right) \\ f\left(\overrightarrow{a}, \overrightarrow{b}, \alpha\overrightarrow{v} + \beta\overrightarrow{w}\right) &= \alpha F\left(\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{v}\right) + \beta F\left(\overrightarrow{a}, \overrightarrow{b}, \overrightarrow{w}\right). \end{aligned}$$

As is the case for linear and bilinear functions, knowing what a trilinear function does when all the inputs are basis vectors lets us determine what it does to any inputs. This is most easily expressed using indexed notation: Let us write

$$\begin{aligned}\overrightarrow{\imath} &= \overrightarrow{e}_1\\ \overrightarrow{\jmath} &= \overrightarrow{e}_2\\ \overrightarrow{k} &= \overrightarrow{e}_3\end{aligned}$$

and for each triple of indices $i_1, i_2, i_3 \in \mathbb{R}^3$

$$c_{i_1,i_2,i_3} := f(e_{i_1}, e_{i_2}, e_{i_3})\,.$$

Then the function f can be expressed as a homogeneous degree three polynomial in the components of its inputs as follows: for

$$\begin{aligned}\overrightarrow{x} &= x_1\overrightarrow{\imath} + x_2\overrightarrow{\jmath} + x_3\overrightarrow{k} = \sum_{i=1}^{3} x_i\overrightarrow{e}_i\\ \overrightarrow{y} &= y_1\overrightarrow{\imath} + y_2\overrightarrow{\jmath} + y_3\overrightarrow{k} = \sum_{j=1}^{3} y_j\overrightarrow{e}_j\\ \overrightarrow{z} &= z_1\overrightarrow{\imath} + z_2\overrightarrow{\jmath} + z_3\overrightarrow{k} = \sum_{k=1}^{3} z_k\overrightarrow{e}_k\end{aligned}$$

we have

$$f(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z}) = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} c_{ijk}x_iy_jz_k. \tag{6.42}$$

This can be proved by a tedious but straightforward calculation (Exercise 6).

Unfortunately, there is no nice trilinear analogue to the matrix representation of a bilinear function. However, we are not interested in *arbitrary* trilinear functions, only the ones satisfying the following additional condition, the appropriate extension of anti-commutativity:

Definition 6.8.2. *A trilinear function $f(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z})$ is **alternating** if interchanging any pair of inputs reverses the sign of the function:*

$$\begin{aligned}f(\overrightarrow{y}, \overrightarrow{x}, \overrightarrow{z}) &= -f(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z})\\ f(\overrightarrow{x}, \overrightarrow{z}, \overrightarrow{y}) &= -f(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z})\\ f(\overrightarrow{z}, \overrightarrow{y}, \overrightarrow{x}) &= -f(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z})\,.\end{aligned}$$

*A **3-form** on $\mathbb{R}^3$ is an alternating trilinear function on $\mathbb{R}^3$.*

Several properties follow immediately from these definitions (Exercise 7):

Remark 6.8.3. *If the trilinear function $f(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z})$ is alternating, the coefficients c_{ijk} in Equation (6.42) satisfy:*

1. *If any pair of indices is equal, then $c_{ijk} = 0$;*
2. *The six coefficients with distinct indices are equal up to sign; more precisely,*

$$\begin{aligned} c_{123} &= c_{231} = c_{312} \\ c_{132} &= c_{321} = c_{213} \end{aligned}$$

   and the coefficients in the first list are the negatives of those in the second list.

I particular, every 3-form on $\mathbb{R}^3$ is a constant multiple of the determinant

$$f(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z}) = c\Delta\left(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z}\right) = c\det\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}$$

where c is the common value of the coefficients in the first list above.

Regarded as a 3-form, the determinant in this remark assigns to three vectors in $\mathbb{R}^3$ the oriented volume of the parallelepiped they determine; we will refer to this as the **volume form** on $\mathbb{R}^3$, and denote it by

$$dx \wedge dy \wedge dz := \Delta.$$

We then consider any such formal "triple wedge product" of dx, dy, and dz in another order to be plus or minus the volume form, according to the alternating rule: that is, we posit that swapping neighboring entries in this product reverses its sign, giving us the following list of the six possible wedge products of all three coordinate forms:

$$\begin{aligned} dx \wedge dy \wedge dz &= -\,dy \wedge dx \wedge dz \\ &= \ dy \wedge dz \wedge dx \\ &= -\,dz \wedge dy \wedge dx \\ &= \ dz \wedge dx \wedge dy \\ &= -\,dz \wedge dx \wedge dy. \end{aligned}$$

Now, we can define the wedge product of a basic 1-form and a basic 2-form by removing the parentheses and comparing with the list above: for example,

$$dx \wedge (\,dy \wedge dz) = (\,dx \wedge dy) \wedge dz = \,dx \wedge dy \wedge dz$$

and, in keeping with the alternating rule, a product in which the same coordinate form appears twice is automatically zero. Finally, we extend this product to an arbitrary 1-form and an arbitrary 2-form on $\mathbb{R}^3$ by making the product distribute over linear combinations. As an example, if

$$\alpha = 3\,dx + \,dy + \,dz$$

and

$$\beta = \,dx \wedge dy + 2\,dx \wedge dz + \,dy \wedge dz$$

then

$$\begin{aligned}\alpha \wedge \beta &= (3\,dx + \,dy + \,dz) \wedge (\,dx \wedge dy + 2\,dx \wedge dz + \,dy \wedge dz)\\ &= 3\,dx \wedge (\,dx \wedge dy) + 6\,dx \wedge (\,dx \wedge dz) + 3\,dx \wedge (\,dy \wedge dz)\\ &\quad + \,dy \wedge (\,dx \wedge dy) + 2\,dy \wedge (\,dx \wedge dz) + \,dy \wedge (\,dy \wedge dz)\\ &\quad + \,dz \wedge (\,dx \wedge dy) + 2\,dz \wedge (\,dx \wedge dz) + \,dz \wedge (\,dy \wedge dz)\\ &= 0 + 0 + \,dx \wedge dy \wedge dz + 0 + 2\,dy \wedge dx \wedge dz + 0 + \,dz \wedge dx \wedge dy + 0 + 0\\ &= 3\,dx \wedge dy \wedge dz - 2\,dx \wedge dy \wedge dz + \,dx \wedge dy \wedge dz\\ &= 2\,dx \wedge dy \wedge dz.\end{aligned}$$

Calculus of Differential Forms

Now, in the spirit of § 6.6, we can define a **differential 3-form** on a region $\mathfrak{D} \subset \mathbb{R}^3$ to be a mapping Ω which assigns to each point $p \in \mathcal{D}$ a 3-form Ω_p on the tangent space $T_p\mathbb{R}^3$ to $\mathbb{R}^3$ at p. By the discussion above, any such mapping can be expressed as

$$\Omega_p = F(p)\ dx \wedge dy \wedge dz.$$

We can also extend the idea of an exterior derivative to 2-forms: if

$$\omega = f(x, y, z)\ dx_1 \wedge dx_2$$

(where each of x_i, $i = 1, 2$ is x, y or z), then its **exterior derivative** is the 3-form

$$d\omega = d(f(x,y,z)\ dx_1 \wedge dx_2) = \ df \wedge\ dx_1 \wedge dx_2.$$

The differential df of f involves three terms, corresponding to the three partial derivatives of f, but two of these lead to triple wedge products in which some coordinate form is repeated, so only one nonzero term emerges. We then extend the definition to general 2-forms using a distributive rule. For example, if

$$\omega = (x^2 + xyz)\, dy \wedge dz + (y^2 + 2xyz)\, dz \wedge dx + (z^2 + xyz)\, dx \wedge dy$$

then

$$\begin{aligned} d\omega &= \big((2x + yz)\, dx + xz\, dy + xy\, dz\big) \wedge\ dy \wedge dz \\ &\quad + \big(2yz\, dx + (2y + 2xz)\, dy + 2xy\, dz\big) \wedge\ dz \wedge dx \\ &\quad + \big(yz\, dx + xz\, dy + (2z + xy)\, dz\big) \wedge\ dx \wedge dy \\ &= (2x + yz)\, dx \wedge dy \wedge dz + 0 + 0 \\ &\quad + 0 + (2y + 2xz)\, dy \wedge dzx + 0 \\ &\quad + 0 + 0 + (2z + xy_{)}\, dz \wedge dxy \\ &= (2x + yz)\, dx \wedge dy \wedge dz + (2y + 2xz)\, dx \wedge dy \wedge dz + (2z + xy)\, dx \wedge dy \wedge dz \\ &= (2x + 2y + 2z + yz + 2xz + xy)\, dx \wedge dy \wedge dz. \end{aligned}$$

It is a straightforward calculation to check the following

Remark 6.8.4. *If the* 2*-form*

$$\omega_{x,y,z} = a(x,y,z)\ dy \wedge dz + b(x,y,z)\ dz \wedge dx + c(x,y,z)\ dx \wedge dy$$

corresponds to the vector field

$$\overrightarrow{F}(x,y,z) = a(x,y,z)\, \overrightarrow{\imath} + b(x,y,z)\, \overrightarrow{\jmath} + c(x,y,z)\, \overrightarrow{k}$$

then its exterior derivative corresponds to the divergence of $\overrightarrow{F}$*:*

$$d\omega = (\operatorname{div} \overrightarrow{F})\, dx \wedge dy \wedge dz.$$

Finally, we define the integral of a 3-form over a region $\mathcal{D} \subset \mathbb{R}^3$ by formally identifying the basic volume form with dV: if

$$\Omega_p = f(p)\ dx \wedge dy \wedge dz$$

then

$$\int_{\mathcal{D}} \Omega = \iiint_{\mathfrak{D}} f\, dV.$$

Pay attention to the distinction between the 3-*form* $dx \wedge dy \wedge dz$ and the element of volume $dV = dx\, dy\, dz$: changing the order of dx, dy and dz in the 3-form affects the sign of the integral, while changing the order of integration in a triple integral does not. The form is associated to the standard **right-handed orientation** of $\mathbb{R}^3$; the 3-forms obtained by transposing an odd number of the coordinate forms, like $dy \wedge dx \wedge dz$, are associated to the *opposite*, **left-handed orientation** of $\mathbb{R}^3$.

As an example, consider the 3-form

$$\Omega_{(x,y,z)} = xyz\, dx \wedge dy \wedge dz;$$

its integral over the "rectangle" $[0,1] \times [0,2] \times [1,2]$ is

$$\int_{[0,1]\times[0,2]\times[1,2]} \Omega = \iiint_{[0,1]\times[0,2]\times[1,2]} xyz\ dV$$

which is given by the triple integral

$$\begin{aligned}
\int_0^1 \int_0^2 \int_1^2 xyz\, dz\, dy\, dx &= \int_0^1 \int_0^2 \left(\frac{xyz^2}{2}\right)_{z=1}^{2} dy\, dx \\
&= \int_0^1 \int_0^2 \left(\frac{3xy}{2}\right) dy\, dx \\
&= \int_0^1 \left(\frac{3xy^2}{4}\right)_{y=0}^{2} dx \\
&= \int_0^1 3x\, dx \\
&= \frac{3x^2}{2}\Big|_0^1 \\
&= \frac{3}{2}.
\end{aligned}$$

Finally, with all these definitions, we can reformulate the Divergence Theorem in the language of forms:

Theorem 6.8.5 (Divergence Theorem, Differential Form)**.** *If ω is a $\mathcal{C}^2$ 2-form defined on an open set containing the regular region $\mathcal{D} \subset \mathbb{R}^3$ with boundary surface(s) $\partial\mathcal{D}$, then the integral of ω over the boundary $\partial\mathcal{D}$ of $\mathcal{D}$ (with boundary orientation) equals the integral of its exterior derivative over $\mathcal{D}$:*

$$\int_{\partial\mathcal{D}} \omega = \int_{\mathcal{D}} d\omega.$$

Generalized Stokes Theorem

Looking back at Theorem 6.4.3, Theorem 6.6.8 and Theorem 6.8.5, we see that Green's Theorem, Stokes' Theorem and the Divergence Theorem, which look so different from each other in the language of vector fields (Theorem 6.3.4, Theorem 6.5.6, and Theorem 6.7.5), can all be stated as one unified result in the language of differential forms. To smooth the statement, we will abuse terminology and refer to a region $\mathcal{D} \subset \mathbb{R}^n$ ($n = 2$ or 3) as an "n-dimensional surface in $\mathbb{R}^n$":

Theorem 6.8.6 (Generalized Stokes Theorem)**.** *If $\mathfrak{S}$ is an oriented k-dimensional surface in $\mathbb{R}^n$ ($k \leq n$) with boundary $\partial\mathfrak{S}$ (given the boundary orientation) and ω is a $\mathcal{C}^2$ $(k-1)$-form on $\mathbb{R}^n$ defined on $\mathfrak{S}$, then*

$$\int_{\partial\mathfrak{S}} \omega = \int_{\mathfrak{S}} d\omega.$$

So far we have understood k to be 2 or 3 in the above, but we can also include $k = 1$ by regarding a directed curve as an oriented "1-dimensional surface", and defining a "0-form" to be a function $f: \mathbb{R}^n \to \mathbb{R}$; a "0-dimensional surface" in $\mathbb{R}^n$ to be a point or finite set of points, and an orientation of a point to be simply a sign ±: the "integral" of the 0-form associated to the function f is simply the value of the function at that point, preceded with the sign given by its orientation. Then the boundary of a directed curve in $\mathbb{R}^n$ ($n = 2$ or 3) is its pair of endpoints, oriented as $p_{end} - p_{start}$, and the statement above becomes the Fundamental Theorem for Line Integrals; furthermore, the same formalism gives us the Fundamental Theorem of Calculus when $n = 1$, given that we regard an interval as a "1-dimensional surface" in $\mathbb{R}^1$.

In fact, this statement has a natural interpretation in abstract n-space $\mathbb{R}^n$ (where cross products, and hence the language of vector calculus does not have a natural extension), and gives a powerful tool for the study of functions and differential equations, as well as the topology of manifolds.

Exercises for § 6.8

Practice problems:

1. Calculate the exterior product $d\alpha \wedge d\beta$:

   (a) $\alpha = 3\,dx + 2x\,dy$, $\beta = 2\,dx \wedge dy - \,dy \wedge dz + x\,dx \wedge dz$

   (b) $\alpha = 3\,dx \wedge dy + 2x\,dy \wedge dz$, $\beta = 2x\,dx - \,dy + z\,dz$

   (c) $\alpha = x\,dx + y\,dy + z\,dz$, $\beta = \,dx \wedge dy - 2x\,dy \wedge dz$

   (d) $\alpha = x\,dx \wedge dy + xy\,dy \wedge dz + xyz\,dx \wedge dz$, $\beta = x\,dx - yz\,dy + xy\,dz$

2. Express the given form as $f(x, y, z)\ dx \wedge dy \wedge dz$:

   (a) $(\,dx + \,dy + \,dz) \wedge (2\,dx - \,dy + \,dz) \wedge (\,dx + \,dy)$

   (b) $(\,dx - \,dy) \wedge (2\,dx + \,dz) \wedge (\,dx + \,dy + \,dz)$

   (c) $(x\,dy + y\,dz) \wedge\ d(x^2y\,dy - xz\,dx)$

   (d) $d((x\,dy + y\,dz) \wedge\ dg)$, where $g(x, y, z) = xyz$.

3. Calculate the exterior derivative $d\omega$:

   (a) $\omega = \,dx \wedge dy + x\,dy \wedge dz$

   (b) $\omega = xy\,dx \wedge dy + xz\,dy \wedge dz$

   (c) $\omega = xyz(\,dx \wedge dy + \,dx \wedge dz + \,dy \wedge dz)$

   (d) $\omega = (xz - 2y)\,dx \wedge dy + (xy - z^2)\,dx \wedge dz$

4. Calculate the integral $\int_{\mathfrak{D}} \omega$:

   (a) $\omega = (xy + yz)\,dx \wedge dy \wedge dz$, $\mathfrak{D} = [0, 1] \times [0, 1] \times [0, 1]$

   (b) $\omega = (x - y)\,dx \wedge dy \wedge dz$, $\mathfrak{D}$ is the region cut out of the first octant by the plane $x + y + z = 1$.

   (c) $\omega = (x^2 + y^2 + z^2)\,dx \wedge dy \wedge dz$, $\mathfrak{D}$ is the unit ball $x^2 + y^2 + z^2 \leq 1$.

5. Calculate $\int_{\mathfrak{S}} \omega$ two ways: (i) directly, and (ii) using the Generalized Stokes Theorem.

   (a) $\omega = z\,dx \wedge dy$, $\mathfrak{S}$ is the cube with vertices $(0, 0, 0)$, $(1, 0, 0,)$, $(1, 1, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(1, 0, 1,)$, $(1, 1, 1)$, and$(0, 1, 1)$, oriented outward.

   (b) $\omega = x\,dy \wedge dz - y\,dx \wedge dz + z\,dx \wedge dy$, $\mathfrak{S}$ is the sphere $x^2 + y^2 + z^2 = 1$, oriented outward.

Theory problems:

6. Verify Equation (6.42).

7. Prove Remark 6.8.3.

8. Show that the only alternating trilinear function on $\mathbb{R}^2$ is the constant zero function.

9. Show that if
$$\alpha = P\,dx + Q\,dy + R\,dz$$
is the 1-form corresponding to the vector $\overrightarrow{v} = P\overrightarrow{\imath} + Q\overrightarrow{\jmath} + R\overrightarrow{\jmath}$ and
$$\beta = a\,dy \wedge dz + b\,dz \wedge dx + c\,dx \wedge dy$$
is the 2-form corresponding to the vector $\overrightarrow{w} = a\overrightarrow{\imath} + b\overrightarrow{\jmath} + c\overrightarrow{k}$, then
$$\alpha \wedge \beta == (\overrightarrow{v} \cdot \overrightarrow{w})\,dx \wedge dy \wedge dz = \beta \wedge \alpha.$$
Note that, unlike the product of two 1-forms, the wedge product of a 1-form and a 2-form is commutative.

10. Prove Remark 6.8.4.

11. Show that if $\omega = d\alpha$ is the exterior derivative of a 1-form α, then
$$d\omega = 0.$$

A

Conic Sections: Apollonius' approach

Here we give some more details of the argument for certain assertions in § 2.1. Our exposition loosely follows [25, pp. 355-9].

Recall the setup (Figure A.1):

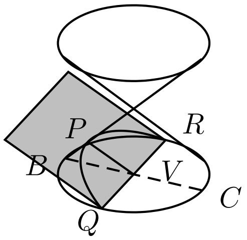


Figure A.1: Conic Section

We wish to investigate the conic section $\gamma = \mathcal{P} \cap \mathcal{K}$, where the plane $\mathcal{P}$ does not contain the origin, and intersects any horizontal plane in a line parallel to the x-axis. The yz-plane (which is perpendicular to any such line) intersects γ in a point P, and possibly in a second point P'—the vertices of γ. Given a point Q on γ distinct from the vertices (*i.e.*, not in the yz-plane), the horizontal plane $\mathcal{H}$ through Q intersects $\mathcal{K}$ in a circle containing Q. The intersection of $\mathcal{H}$ with the yz-plane is a line through the center of this circle, and so contains a diameter BC of the circle. We draw the chord of the circle through Q perpendicular to this diameter, denoting by R the other end of the chord, and by V its intersection with BC. The line segments QV and PV are, respectively, the **ordinate** and **abcissa**.

We know that V bisects QR, and also, by Prop. 13, Book VI of the *Elements*, that

$$|QV|^2 = |QV| \cdot |VR| = |BV| \cdot |VC| . \tag{A.1}$$

Parabolas

Suppose first that PV is parallel to AC, so that P is the only vertex of γ. Consider the triangle $\triangle ABC$ in the yz-plane, noting that P lies on AB and V lies on BC (Figure A.2). Since AC is parallel to PV, the triangles $\triangle ABC$ and $\triangle PBV$ are similar and (since $AD = AC$) isosceles. In particular,

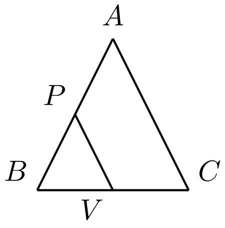


Figure A.2: Equation (A.2)

$$\frac{|BV|}{|BP|} = \frac{|BC|}{|BA|}$$

and (again since AC and PV are parallel)

$$\frac{|VC|}{|BC|} = \frac{|PA|}{|BA|}$$

or equivalently

$$\frac{|VC|}{|PA|} = \frac{|BC|}{|BA|}.$$

Multiplication of these two equations yields

$$\frac{|BV|\,|VC|}{|BP|\,|PA|} = \left(\frac{|BC|}{|BA|}\right)^2 .$$

Since $|BP| = |PV|$, we conclude that

$$|BV|\,|VC| = \left[\left(\frac{|BC|}{|BA|}\right)^2 |PA|\right] |PV| . \qquad (A.2)$$

Note that replacing Q with another point Q' on γ replaces $\mathcal{H}$ with a parallel plane $\mathcal{H}'$, and gives a picture similar to Figure A.2 (Figure A.3). In particular, the quantity in brackets in Equation (A.2) depends only on γ: it is called the **parameter of ordinates** for γ—we will denote it by p. Apollonius represents it [1] by a line segment PL perpendicular to the abcissa PV (indicated in Figure A.3).

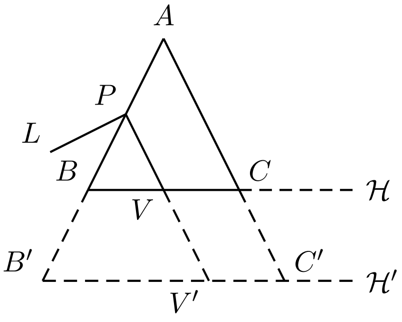


Figure A.3: Independence of Q

As noted in § 2.1, rectangular coordianates in $\mathcal{P}$ with the origin at P and axes parallel to QV ($y = |QV|$) and PV ($x = |PV|$) lead to the equation

$$y^2 = px \qquad (A.3)$$

where p is the parameter of ordinates defined above.

For the other two cases, when PV is not parallel to AC, then the line PV (extended) meets the line AB (extended) at the second vertex P'. If ϕ denotes the (acute) angle between $\mathcal{P}$ and a horizontal plane $\mathcal{H}$, then V

[1] Fried and Unguru argue that for Apollonius, the orthia is a specific line segment, not the representation of an independent quantity.

lies between P and P' if $0 \leq \phi < \frac{\pi}{2} - \alpha$ and P lies between V and P' if $\frac{\pi}{2} - \alpha < \phi \leq \frac{\pi}{2}$.

Ellipses: In the first case (see Figure A.4), let J be the point at which the line through A parallel to PV (and hence to PP') meets BC (extended). As in the case of the parabola (but with C replaced by J), the triangles $\triangle ABJ$ and $\triangle PBV$ are similar (but no longer isosceles), so

$$\frac{|BV|}{|PV|} = \frac{|BJ|}{|AJ|}.$$

Also, since the lines AP' and VJ are transversals to the two parallels AJ and PP' meeting at C, the triangles $\triangle AJC$ and $\triangle P'VC$ are similar, so

$$\frac{|VC|}{|VP'|} = \frac{|JC|}{|AJ|}.$$

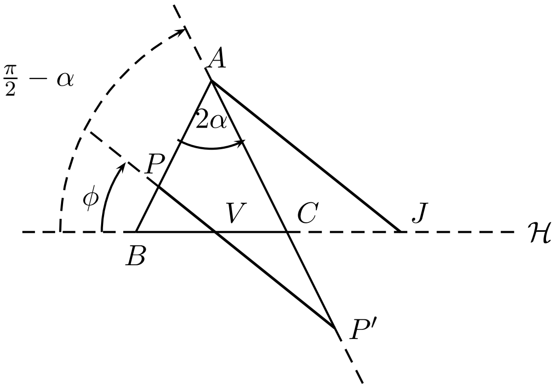


Figure A.4: $0 \leq \phi \leq \frac{\pi}{2} - \alpha$

Multiplying these equalities and invoking Equation (2.2), we have

$$\frac{|QV|^2}{|PV|\,|VP'|} = \frac{|BV|\,|VC|}{|PV|\,|VP'|} = \frac{|BJ|\,|JC|}{|AJ|^2}$$

or, as the analogue of Equation (A.2),

$$|QV|^2 = \left[\left(\frac{|BJ|\,|JC|}{|AJ|^2}\right)|VP'|\right]|PV|\,. \tag{A.4}$$

Again as in the case of the parabola, the fraction in parentheses in Equation (A.4) depends only on the curve γ. We again form the "orthia" of γ, a line segment PL perpendicular to PV with length[2]

$$p = |PL| = \left(\frac{|BJ|\,|JC|}{|AJ|^2}\right)\left|PP'\right|. \tag{A.5}$$

Now let S be the intersection of LP' with the line through V parallel to PL (Figure A.5). Note that the triangles $\triangle LP'P$ and $\triangle SP'V$ are similar, so

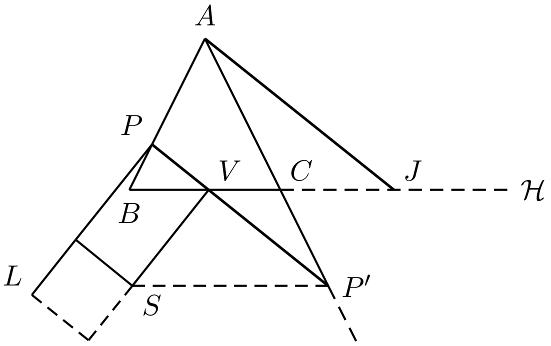


Figure A.5: Definition of S

$$\frac{|PL|}{|PP'|} = \frac{|VS|}{|VP'|}$$

[2] This defintion of the orthia is on the face of it quite different from the definition in the case of the parabola. I have not been able to find a reasonable explanation of why the two definitions yield analogous line segments. It can be shown (M. N. Fried, private correspondence) that if one considers the orthia of hyperbolic or elliptic sections whose diameter has inclination approaching that of a generator (*i.e.*, approaching a parabolic section), then these orthia tend toward the parabolic orthia of the limit. This, however, is clearly an anachronistic point of view, and Fried has pointed out that Apollonius never discusses varying the section. In fact, in his view, Apollonius did not intend the properties of the conic sections with respect to application of areas—known as the **symptomata**—to unify the different types; he viewed them as separate objects, and only noted these somewhat analogous properties as incidental observations. Fried's point of view (but not a specific commentary on this issue) is given at length in [15, pp. 74-90].

and substituting this (via Equation (A.5)) into Equation (A.4), we have

$$\begin{aligned}|QV|^2 &= \left[\left(\frac{|PL|}{|PP'|}\right)|VP'|\right]|PV| \\ &= \left[\left(\frac{|VS|}{|VP'|}\right)|VP'|\right]|PV|\end{aligned}$$

or

$$|QV|^2 = |VS| \cdot |PV|\,. \tag{A.6}$$

This is like Equation (A.3), but $|PL|$ is replaced by the shorter length $|VS|$.

To obtain the rectangular equation of the ellipse, we set

$$d = |PP'|$$

(the **diameter**): by similarity of $\triangle LP'P$ and $\triangle SP'V$,

$$\frac{|VS|}{|PL|} = \frac{|VP'|}{|PP'|} = 1 - \frac{|PV|}{|PP'|}$$

so (again setting $x = |PV|$ and $y = |QV|$) we have as the equation of the ellipse

$$y^2 = |VS|\,x = p\left(1 - \frac{x}{d}\right)x = px - \frac{p}{d}x^2. \tag{A.7}$$

Hyperbolas: In the final case, when $\frac{\pi}{2} - \alpha < \phi \leq \frac{\pi}{2}$, P lies between V and P' (Figure A.6). Formally, our constructions in this case are the same as in the case of the ellipse: as before, the two similarities $\triangle ABJ \sim \triangle PBV$ and $\triangle AJC \sim \triangle P'VC$ lead to Equation (A.4); we form the orthia PL perpendicular to PV satisfying Equation (A.5) and let S be the intersection of $P'L$ (extended) with the line through V parallel to PL (Figure A.7). The same arguments as in the ellipse case yield Equation (A.6), but this time the segment VS *exceeds* PL.

A verbatim repetition of the calculation leading to Equation (A.7) leads to its hyperbolic analogue,

$$y^2 = px + \frac{p}{d}x^2. \tag{A.8}$$

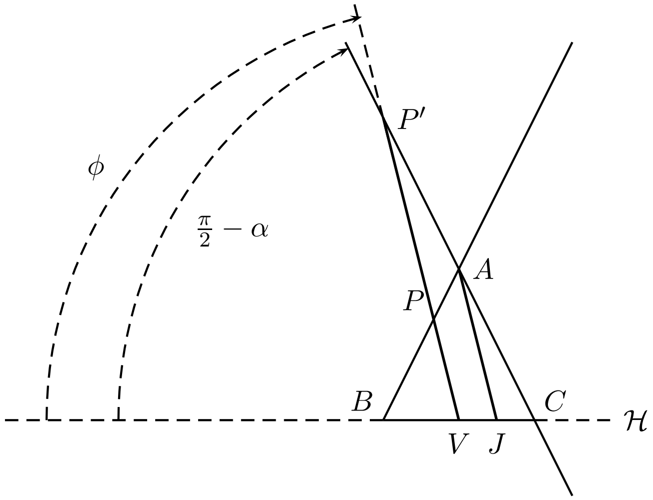


Figure A.6: $\frac{\pi}{2} - \alpha < \phi \le \frac{\pi}{2}$

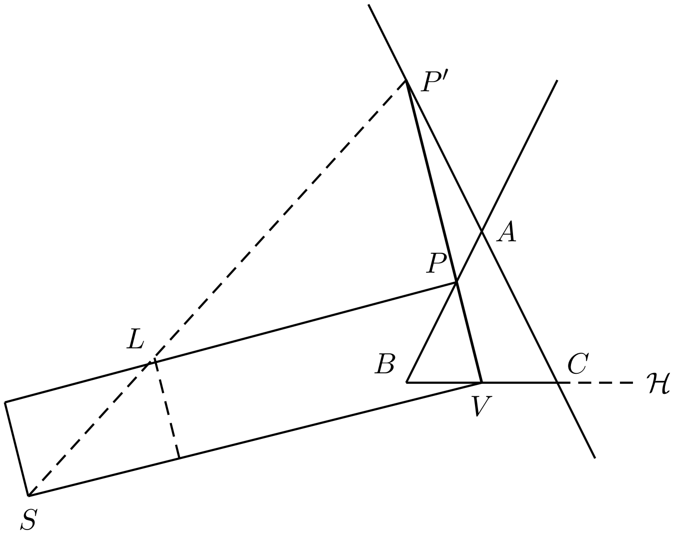


Figure A.7: Definition of S

B
Conic Sections: The Focus-Directrix Property

Here we give a proof of Pappus' lemma (Lemma 2.1.1):

> *If the distance of a point from a fixed point be in a given ratio to its distance from a fixed straight line, the locus of the point is a conic section, which is an ellipse, a parabola, or a hyperbola according as the ratio is less than, equal to, or greater than, unity.*

Our *pictures* will illustrate the elliptic case, but the *arguments* are general.

Proof. We assume as before that γ is the intersection of $\mathcal{K}$ with a plane $\mathcal{P}$ that intersects any horizontal plane $\mathcal{H}$ in a line parallel to the x-axis, making a dihedral angle ϕ with $\mathcal{H}$. Denote the intersection of $\mathcal{P}$ with the axis of $\mathcal{K}$ by K, and as before, let P be a vertex of γ (an intersection of γ with the yz-plane). Let E be the intersection of the axis of $\mathcal{K}$ with the bisector of the angle $\angle APK$ (Figure B.1) and draw perpendiculars EB to AP and EF to PK. Since $\angle EFP$ and $\angle EBP$ are both right angles and $\angle EPB = \angle EPF$, the (right) triangles $\triangle EFP$ and $\triangle EBP$, which share a common hypotenuse EP, are congruent, so $|EB| = |EF|$. Let $\mathcal{H}$ be the horizontal plane through B, and construct the sphere $\mathcal{S}$ with center at E and radius $|EF|$. Then $\mathcal{S}$ is tangent to the plane $\mathcal{P}$ at F, and to the cone $\mathcal{K}$ at all points on the circle

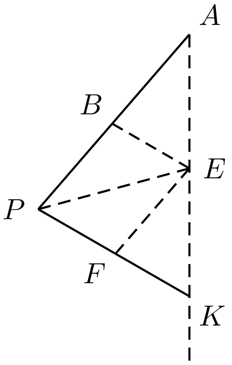


Figure B.1: Definition of E, B and F

of intersection of $\mathcal{H}$ with $\mathcal{K}$. Let ℓ be the line of intersection of $\mathcal{H}$ with $\mathcal{P}$ (by assumption, ℓ is parallel to the x-axis) and let D be the intersection of ℓ with the yz-plane (Figure B.2). Note that

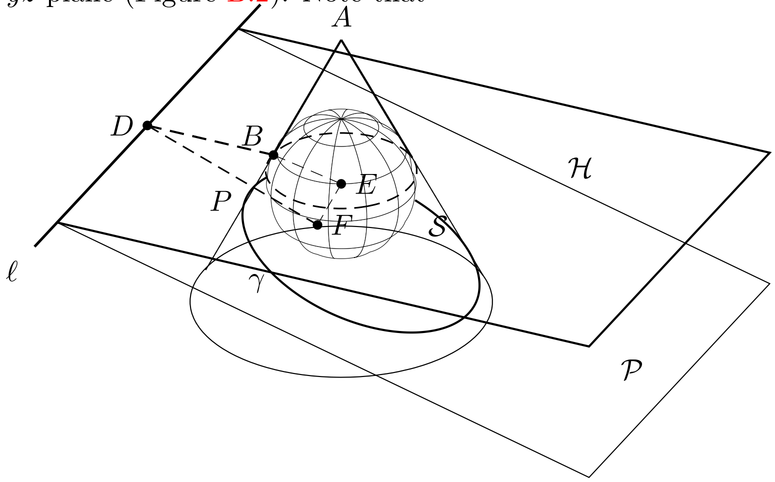


Figure B.2: Definition of $\mathcal{S}$, ℓ, and D

$$\angle BDP = \phi$$

and

$$\angle DBP = \frac{\pi}{4}.$$

Given a point Q on γ, let

- $H =$ the intersection of $\mathcal{H}$ with the generator of $\mathcal{K}$ through Q;
- $G =$ the point on ℓ nearest to Q;
- $T =$ the point of $\mathcal{H}$ directly above (or below) Q.

(Figure B.3).

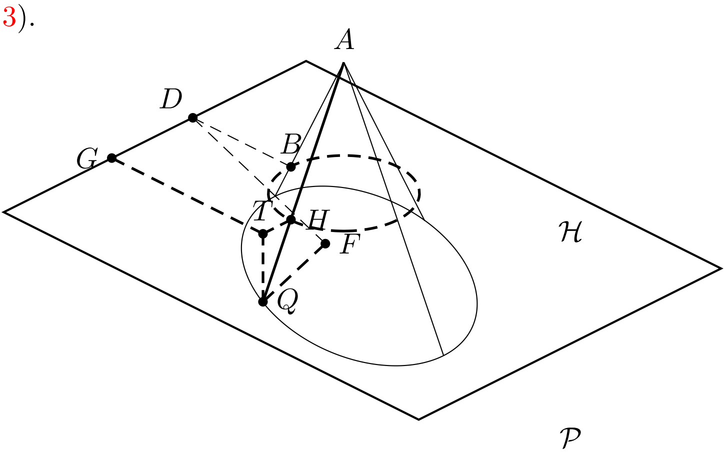


Figure B.3: Definition of H, G, and T

We wish to compare the ratio $|PF| \, / \, |PD|$ with $|QF| \, / \, |QG|$. Note that

- since QF and QH are both tangents from Q to $\mathcal{S}$,
$$|QF| = |QH|\,;$$
- $\triangle TQH$ is a right triangle in the plane containing the axis of $\mathcal{K}$ and Q, so
$$|QH| = \frac{|QT|}{\sin(\angle THQ)};$$
- rotating this plane into the plane containing the axis and P, since all generators make the same angle α with the axis,
$$\angle THQ = \angle DBP = \frac{\pi}{2} - \alpha$$
so
$$|QH| = \frac{|QT|}{\sin \alpha};$$

- the plane containing Q, T and G is parallel to the yz-plane, so

$$\angle TGQ = \angle BDP = \phi$$

  and

$$|QG| = \frac{|QT|}{\sin(\angle tGQ)} = \frac{|QT|}{\sin\phi};$$

- it follows that

$$\frac{|QF|}{|QG|} = \frac{\sin\phi}{\sin(\frac{\pi}{2}-\alpha)}.$$

But the right-hand side of this equality is independent of the choice of Q on γ, and the quantities on the left are the distances from Q to, respectively, the focus and the directrix.

This proves that the ratio is constant, and is less than, equal to, or greater than one as ϕ is less than, equal to, or greater than $\frac{\pi}{2}-\alpha$, establishing the lemma.

□

C
Kepler and Newton

The empirical starting point of Newton's argument for universal gravitation in the *Principia* was a trio of observations known as **Kepler's Laws of Planetary Motion**. Johannes Kepler (1571-1630) spent decades analyzing the data from extensive astronomical observations by the Imperial Astronomer in Prague, Tycho Brahe (1546-1601), whom he succeeded in the post upon the latter's death. He finished compiling the tables begun by Brahe: the *Rudolphine Tables*, published in 1627, were by far the most accurate and extensive astronomical tables of the time. Kepler was also a deep theorist, and distilled from his analysis of this data the following general observations concerning the orbits of the known planets:

1. *The area swept out by a line from the sun to each planet is proportional to the time elapsed.*

2. *The orbit of each planet is an ellipse with the sun at its focus.*

3. *The square of the period of a planet is proportional to the cube of its mean distance from the sun.*

The evolution of these Laws is an involved story [50]; the first observation was initially introduced—as a computational approximation—in Kepler's *Mysterium Cosmographicum* (1596); later, after he observed that the orbit of Mars was an ellipse with equal areas swept in equal times, he put forth the first two observations—as general laws—in *Astronomia nova* (1609); the third observation was published in *Harmonie Mundi* (1618).

Newton begins *Principia* with a series of definitions, followed by three "Axioms, or Laws of Motion":[1]

Law 1. *Every body perseveres in its state of being at rest or of moving "uniformly straight forward", except insofar as it is compelled to change its state by forces impressed.*

Law 2. *A change in motion is proportional to the motive force impressed and takes place along the straight line in which that force is impressed.*

Law 3. *To any action there is always an opposite and equal reaction; in other words, the actions of two bodies upon each other are always equal and opposite in direction.*

Then come some corollaries of these Laws. Book I, "The Motion of Bodies", begins with a section called "The method of first and ultimate ratios", containing eleven preliminary mathematical lemmas setting forth the basic mathematical methods of calculus as used in *Principia* (for example, Lemma 11 is essentially the limit $\lim_{\theta\to 0} \sin\theta/\theta$). Section 2, "To find centripetal forces", begins the analysis of motion; we will concentrate on two of Newton's results here. First, *Proposition 1* asserts that Kepler's Law of Areas (not named as such) is a consequence of having a central force, and *Proposition 2* asserts the converse. Newton's elegant, geometric proof of this is given in Exercise 1. Second, at the start of Section 3, *Propositions 11-13* assert that motion along a conic section (ellipse, hyperbola, and parabola, respectively) which also satisfies Kepler's Law of Areas involves a central force directed toward the focus and of length inversely proportional to the square of the distance from the focus. Again, Newton gives an elegant geometric proof, based in part on the properties of conic sections which we considered in § 2.1; this proof however uses more detailed results about conic sections that would take us too far afield to reproduce here. In this section we will see how Proposition 1 as well as Proposition 11 can be obtained using vector methods.

To work with Kepler's Law of Areas and Newton's Propositions 1-2, we need to formulate the notion of "area swept out" by a line from the center of force (the sun). To this end, it will be useful to put the center at the origin, so the position vector $\vec{\mathcal{O}P} = \vec{p}(t)$ goes from the sun to the planet.

[1] All quotes are from [9].

To define the "area swept out" by $\overrightarrow{p}(t)$ over a time period $a \leq t \leq b$, we partition $[a,b]$ via

$$\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n = b\};$$

this induces a partition of the curve itself via the succesive points

$$P_j = \overrightarrow{p}(t_j) \quad j = 0, \ldots, n.$$

Now for each time interval

$$I_j = [t_{j-1}, t_j]$$

form the oriented triangle

$$\triangle_j = [\mathcal{O}, P_{j-1}, P_j]\,;$$

its oriented area can be calculated as

$$\begin{aligned} \triangle\vec{\mathcal{A}}(\triangle_j) &= \frac{1}{2}\overrightarrow{\mathcal{O}P_{j-1}} \times \overrightarrow{\mathcal{O}P_j} \\ &= \frac{1}{2}\overrightarrow{p}(t_{j-1}) \times \overrightarrow{p}(t_j) \\ &= \frac{1}{2}\overrightarrow{p}(t_{j-1}) \times \triangle\overrightarrow{p}_j \end{aligned}$$

where

$$\begin{aligned} \triangle\overrightarrow{p}_j &= \overrightarrow{p}_{t_j} - \overrightarrow{p}_{t_{j-1}} \\ &= \overrightarrow{P_{j-1}P_j} \end{aligned}$$

is the net displacement over the time interval I_j. Then the net (oriented) area over $[a,b]$ is approximated by the "Riemann sum"

$$\mathcal{R}(\mathcal{P}, \vec{\mathcal{A}}) = \frac{1}{2}\sum_{j=1}^{n} \overrightarrow{p}(t_{j-1}) \times \triangle\overrightarrow{p}_j.$$

We want to take a limit of such sums as the mesh size goes to zero; this is made a bit easier by the further approximation that, for sufficiently small mesh size, the velocity is nearly constant over each time interval I_j, so we can write

$$\triangle\overrightarrow{p}_j \approx \overrightarrow{v}_j \triangle t_j$$

where

$$\overrightarrow{v}_j = \overrightarrow{v}(t_j) = \dot{\overrightarrow{p}}(t_j)$$

and

$$\triangle t_j = \|I_j\| = t_j - t_{j-1}.$$

We can then see that the limiting value of our Riemann sums as mesh size goes to zero, or the **net oriented area** swept out by the line $\mathcal{O}P$ for $a \leq t \leq b$, is the definite integral

$$\vec{\mathcal{A}}(a,b) = \frac{1}{2}\int_a^b \overrightarrow{p}(t) \times \overrightarrow{v}(t)\ dt.$$

Note that if $\overrightarrow{p}(t)$ is a curve in the xy-plane, then the same process replacing oriented areas with their lengths (or unsigned areas) gives a similar definition of the (unsigned) area swept out by the line $\mathcal{O}P$ in the plane:

$$\mathcal{A}(a,b) = \frac{1}{2}\int_a^b \|\overrightarrow{p} \times \overrightarrow{v}\|\ dt.$$

In taking the cross product above, we can replace $\overrightarrow{v}$ with its component $\overrightarrow{v}^{\perp}$ perpendicular to $\overrightarrow{p}$; then noting that, in terms of polar coordinates,

$$\|\overrightarrow{p}\| = r$$

and

$$\left\|\overrightarrow{v}^{\perp}\right\| = r\dot{\theta}$$

we see that the formula for area in polar coordinates is a special case:

$$\mathcal{A}(a,b) = \int_a^b \frac{r^2}{2}\dot{\theta}\, dt.$$

If we differentiate $\vec{\mathcal{A}}(a,t)$ with respect to t (using the Fundamental Theorem of Calculus on the components of the integrand) we obtain

Remark C.0.7. *The rate of change of the (oriented) area swept out by the line $\mathcal{O}P$, where $P = \overrightarrow{p}(t)$, is*

$$\frac{d\vec{\mathcal{A}}}{dt} = \frac{1}{2}\overrightarrow{p}(t) \times \overrightarrow{v}(t)\,. \tag{C.1}$$

From this we easily conclude

Proposition C.0.8 (*Principia*, Props. I.1 and I.2)**.** *For a regular motion $\overrightarrow{p}(t)$, the following are equivalent:*

1. *The acceleration $\overrightarrow{a}(t)$ is a scalar multiple of $\overrightarrow{p}(t)$ (i.e., the force is central, and directed toward or away from the origin).*
2. *The curve traced out by $\overrightarrow{p}(t)$ is contained in a plane through the origin, and the rate of change of the area swept out by $\mathcal{O}P$ is constant.*

Proof. If we differentiate Equation (C.1), using the product rule, we obtain

$$\begin{aligned}\frac{d^2\vec{\mathcal{A}}}{dt^2} &= \frac{1}{2}\frac{d}{dt}\left[\overrightarrow{p} \times \overrightarrow{v}\right] \\ &= \frac{1}{2}(\dot{\overrightarrow{p}} \times \overrightarrow{v} + \overrightarrow{p} \times \dot{\overrightarrow{v}}) \\ &= \frac{1}{2}(\overrightarrow{v} \times \overrightarrow{v} + \overrightarrow{p} \times \overrightarrow{a})\end{aligned}$$

or

$$\frac{d^2\vec{\mathcal{A}}}{dt^2} = \frac{1}{2}(\overrightarrow{p} \times \overrightarrow{a}) \tag{C.2}$$

(since $\overrightarrow{v} \times \overrightarrow{v} = \vec{0}$).

If $\overrightarrow{a}$ and $\overrightarrow{p}$ are linearly dependent then this quantity is the zero vector, which means that $\frac{d\vec{\mathcal{A}}}{dt}$ is a constant vector, say

$$\frac{d\vec{\mathcal{A}}}{dt} = \overrightarrow{K}.$$

Substituting this into Equation (C.1), we see that both the initial position and the velocity (for all time) are perpendicular to $\overrightarrow{K}$, so contained in the plane through the origin perpendicular to $\overrightarrow{K}$, and hence the whole motion is contained in the plane. Furthermore, the rate of change of the (plane, unsigned) area swept out by $\mathcal{O}P$ is the length of $\overrightarrow{K}$, which is constant by Equation (C.2).

Conversely, if $\overrightarrow{p}(t)$ lies in a plane through the origin, let $\overrightarrow{u}$ be a unit vector normal to this plane, so

$$\overrightarrow{u} \cdot \overrightarrow{p}(t) = 0$$

and also

$$\frac{d}{dt}[\overrightarrow{u} \cdot \overrightarrow{p}] = \overrightarrow{u} \cdot \overrightarrow{v} = 0.$$

Thus

$$\frac{d\vec{\mathcal{A}}}{dt} = \frac{1}{2}\overrightarrow{p} \times \overrightarrow{v}$$

is a scalar multiple of $\overrightarrow{u}$:

$$\frac{d\vec{\mathcal{A}}}{dt} = \pm \left\| \frac{d\vec{\mathcal{A}}}{dt} \right\| \overrightarrow{u}.$$

But if the rate of change of planar area $\left\| \frac{d\vec{\mathcal{A}}}{dt} \right\|$ is constant, then $\frac{d\vec{\mathcal{A}}}{dt}$ is a constant vector, and using Equation (C.2) we have

$$\overrightarrow{p} \times \overrightarrow{a} = \vec{0}$$

which means that $\overrightarrow{p}$ and $\overrightarrow{a}$ are linearly dependent, as required. □

Now, we will show that Kepler's first two laws imply that the planet is subject to an inverse-square force directed at the sun. The mathematical abstraction of Newton's Proposition 11 in Book I of *Principia* is:

Proposition C.0.9 (*Principia*, Prop. I.11)**.** *If $\overrightarrow{p}(t)$ traces out an ellipse with focus at the origin, such that the vector $\mathcal{O}P = \overrightarrow{p}(t)$ sweeps out equal areas in equal times, then*

$$\overrightarrow{a}(t) = -\frac{k}{\|\overrightarrow{p}(t)\|^3}\overrightarrow{p}(t)$$

for some constant k.

Proof. We saw in § 2.1 that the locus of the equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{C.3}$$

where $a>b>0$, is an ellipse with eccentricity

$$e=\sqrt{1-\frac{b^2}{a^2}}$$

and one focus at

$$F(ae,0).$$

Note that Equation (C.3) can be rewritten as

$$(1-e^2)x^2+y^2=a^2(1-e^2).$$

Since varying a in this equation amounts to equal scaling in all directions, without affecting the eccentricity, we shall set $a=1$ and consider as a "model" ellipse of eccentricity[2] $0\leq e<1$ the locus of

$$x^2+\frac{y^2}{1-e^2}=1$$

which can be parametrized by

$$\begin{aligned} x &= \cos\theta(t) \\ y &= \sqrt{1-e^2}\sin\theta(t) \end{aligned}$$

where $\theta(t)$ is a monotone function of t. However, this has its focus at $F(e,0)$, and we want to move this to the origin. This is accomplished by subtracting the position vector of F, so a regular parametrization of a "model" ellipse of eccentricity $e<1$ with focus at the origin is

$$\overrightarrow{p}(t)=(\cos\theta-e)\overrightarrow{\imath}+(\sqrt{1-e^2}\sin\theta)\overrightarrow{\jmath}$$

where

$$\theta=\theta(t)$$

[2] Note that $e=0$ yields the equation of a circle.

is a differentiable function of t with nonvanishing derivative.

We note for later reference that

$$\begin{aligned}\|\overrightarrow{p}(t)\|^2 &= (\cos\theta - e)^2 + (1-e^2)\sin^2\theta \\ &= \cos^2\theta - 2e\cos\theta + e^2 + \sin^2\theta - e^2\sin^2\theta \\ &= 1 - 2e\cos\theta + e^2\cos^2\theta \\ &= (1 - e\cos\theta)^2\end{aligned}$$

so

$$\|\overrightarrow{p}(t)\| = 1 - e\cos\theta(t)\,.$$

the velocity is given by

$$\begin{aligned}\overrightarrow{v}(t) &= \frac{d}{dt}\overrightarrow{p}(t) \\ &= \dot{\theta}(-\sin\theta\overrightarrow{\imath} + \sqrt{1-e^2}\cos\theta\overrightarrow{\jmath})\end{aligned}$$

so a direct calculation yields

$$\overrightarrow{p} \times \overrightarrow{v} = \dot{\theta}\sqrt{1-e^2}(1 - e\cos\theta)\overrightarrow{k}\,.$$

That the *direction* of this vector is constant follows from the fact that our ellipse lies in the xy-plane. To sweep out equal areas in equal time, we must also make its *length* constant, which means we must make $\dot{\theta}$ inversely proportional to $(1 - e\cos\theta)$. We shall assume that[3]

$$\dot{\theta} = \frac{d\theta}{dt} = (1 - e\cos\theta)^{-1}$$

and work with it implicitly. First, we calculate

$$\begin{aligned}\ddot{\theta} &= -\dot{\theta}\frac{e\sin\theta}{(1-e\cos\theta)^2} \\ &= -(\dot{\theta})^3 e\sin\theta.\end{aligned}$$

[3] The existence of a function satisfying this condition is proved using the theory of ordinary differential equations.

Now, the acceleration vector is

$$\begin{aligned}\overrightarrow{a}(t) &= \frac{d}{dt}\left[\overrightarrow{v}(t)\right] \\ &= \ddot{\theta}(-\sin\theta, \sqrt{1-e^2}\cos\theta) + (\dot{\theta})^2(-\cos\theta, -\sqrt{1-e^2}\sin\theta) \\ &= (\dot{\theta})^3\left[-(e\sin\theta)(-\sin\theta, \sqrt{1-e^2}\cos\theta) + (1-ecos\theta)(-\cos\theta, -\sqrt{1-e^2}\sin\theta)\right]\end{aligned}$$

We calculate the vector in brackets: its first component is

$$\begin{aligned}e\sin^2\theta - (1-e\cos\theta)\cos\theta &= e\sin^2\theta - \cos\theta + e\cos^2\theta \\ &= e - \cos\theta\end{aligned}$$

while its second component is

$$\begin{aligned}&-(e\sin\theta)\sqrt{1-e^2}\cos\theta - (1-e\cos\theta)\sqrt{1-e^2}\sin\theta \\ &\quad = -e\sqrt{1-e^2}\sin\theta\cos\theta - \sqrt{1-e^2}\sin\theta + e\sqrt{1-e^2}\sin\theta\cos\theta \\ &\quad\quad = -\sqrt{1-e^2}\sin\theta;\end{aligned}$$

thus we have

$$\overrightarrow{a}(t) = (\dot{\theta})^3(e-\cos\theta, -\sqrt{1-e^2}\sin\theta).$$

But now notice that

$$\begin{aligned}\dot{\theta} &= (1-e\cos\theta) \\ &= \frac{1}{\|\overrightarrow{p}(t)\|}\end{aligned}$$

and the vector above is $-\overrightarrow{p}(t)$, so

$$\overrightarrow{a}(t) = -\frac{1}{\|\overrightarrow{p}(t)\|^3}\overrightarrow{p}(t)$$

as required. □

Exercises for Appendix C

History note:

1. Proposition 1 in Book I of the *Principia* reads

> The areas which bodies made to move in orbits describe by radii drawn to an unmoving center of force lie in unmoving planes and are proportional to their times.

The proof of this relies on a diagram of which we sketch only a part (Figure C.1)

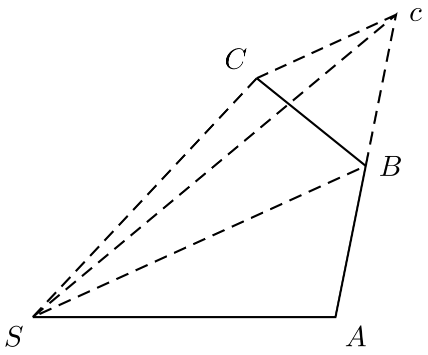


Figure C.1: Proposition 1, Book I of *Principia*

Consider first a body moving, with no force acting, for a time interval $\triangle t$, describing the line segment AB; in the continued absence of forces, it would continue, in the next time interval $\triangle t$ to travel describing the line segment Bc, of the same length as AB and parallel to it (*i.e.*, ABc forms a straight line segment). Suppose, however, that at the moment it reaches B, the body is subjected to an instantaneous force directed toward S, and as a result it is diverted and describes the line segment BC, of the same length as AB but such that the diversion cC (the effect of the force) is parallel to SB.

(a) *Show* that the triangles $\triangle SAB$ and $\triangle SBc$ have equal areas. (*Hint:* Use AB (*resp.* Bc) as the base.)

(b) *Show* that the triangles $\triangle SBC$ and $\triangle SBc$ have equal areas. (*Hint:* Use SB as the base for both.)

(c) These two assertions show two things: if we divide the motion into equal time segments and replace the (continuously acting) force over each time interval by an impulse at the end of the time interval, then

> - The area swept out over each time interval will be the same, and
> - The whole path will lie in a single plane (why?).
>
> Now, as we subdivide the path into an increasing number of (successively shorter) time segments, the polygonal paths so obtained will converge to the actual path, and these two properties will persist in the limit.

This proof has been criticized [43]: is it legitimate to replace the constantly acting acceleration due to the central force over a time period with an impulse *at the end of that period*, which is parallel to the line from the position *at the start of the period*, and in particular, how can we argue that the motion takes place in a single plane? Consider these questions, and see if you have any ideas on how they might be answered.

D

Intrinsic Geometry of Curves

As we have seen, any given curve $\mathcal{C}$ in the plane or in space can be expressed by many different vector-valued functions; it would be useful to find a "standard" parametrization that reflects as directly as possible the geometric properties of $\mathcal{C}$. In this section, we shall see that such a parametrization is given by the arclength function, and this allows us to characterize the "shape" of a curve by a very small number of measurements.

Arclength Parametrization

Suppose $\mathcal{C}$ is a regular curve, that is, it is given by a parametrization $\overrightarrow{\gamma}(t)$ defined on the interval $[a, b]$ whose velocity vector $\overrightarrow{v}(t) = \dot{\overrightarrow{\gamma}}(t)$ never vanishes. As noted before, this means that the *speed* never vanishes, and therefore the arclength function

$$\mathfrak{s}\,(t) = \int_a^t \left\| \dot{\overrightarrow{\gamma}}(t) \right\| \, dt$$

has strictly positive derivative

$$\frac{d}{dt}\left[\mathfrak{s}\,(t)\right] = \frac{d\mathfrak{s}}{dt} = \left\| \dot{\overrightarrow{\gamma}}(t) \right\| > 0$$

and hence it is a strictly increasing function of t; it follows that it is an invertible function, so that we can write

$$t = \mathfrak{t}(\mathfrak{s})$$

where $\mathfrak{t}(\mathfrak{s})$ is differentiable with derivative

$$\begin{aligned}\frac{d\mathfrak{t}(\mathfrak{s})}{d\mathfrak{s}} &= \frac{1}{d\mathfrak{s}/dt}\\ &= \frac{1}{\|\overrightarrow{v}(t)\|},\end{aligned}$$

and

$$\overrightarrow{\gamma}(s) = \overrightarrow{\gamma}(\mathfrak{t}(s))$$

is a reparametrization of $\mathcal{C}$. It follows from the Chain Rule that

$$\begin{aligned}\dot{\overrightarrow{\gamma}}(s) &= \frac{d\overrightarrow{\gamma}(\mathfrak{t}(s))}{ds}\\ &= \frac{d}{dt}\left[\overrightarrow{\gamma}(t)\right] \cdot \frac{d\mathfrak{t}(s)}{ds}\\ &= \overrightarrow{v}(t) \cdot \left(\frac{1}{\|\overrightarrow{v}(t)\|}\right)\end{aligned}$$

is a unit vector, which means it is the unit tangent vector to $\mathcal{C}$ at $\overrightarrow{\gamma}(t)$:

$$\dot{\overrightarrow{\gamma}}(s) = \overrightarrow{T}(\mathfrak{t}(s)).$$

This can be summarized in the observation

Remark D.0.10. *Suppose $\overrightarrow{\gamma}(t)$, $t \in [a,b]$ is a continuously differentiable parametrization of the curve $\mathcal{C}$. Then the following are equivalent:*

1. *The velocity vector is a unit vector for all $t \in [a,b]$:*

$$\left\|\dot{\overrightarrow{\gamma}}(t)\right\| = 1$$

or

$$\dot{\overrightarrow{\gamma}}(t) = \overrightarrow{T}(\overrightarrow{\gamma}(t)) \text{ for all } t \in [a,b];$$

2. *The parametrization has unit speed:*

$$\frac{d\mathfrak{s}}{dt} = 1 \text{ for all } t \in [a, b]\,;$$

3. *The parameter t equals the arclength $\mathfrak{s}\,(t)$ along $\mathcal{C}$ from the starting point $\overrightarrow{\gamma}(a)$ to $\overrightarrow{\gamma}(t)$.*

In the rest of this section, we shall distinguish between an arbitrary regular parametrization $\overrightarrow{p}(t)$ of a curve $\mathcal{C}$ and an **arclength parametrization** of $\mathcal{C}$, which we will denote $\overrightarrow{\gamma}(s)$.

Very often it is difficult to find an explicit formula for the arclength reparametrization function $t = \mathfrak{t}(s)$ or its inverse; in this sense arclength parametrization is of more theoretical than practical interest. However, it is possible to find it in some cases, and then use it to define geometrically significant functions of the arclength.

A particularly important case is the circle. The standard parametrization of a circle with center at the origin and radius R

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

has constant speed

$$\frac{d\mathfrak{s}}{d\theta} = R$$

so the arclength is just a multiple of the original parameter

$$\mathfrak{s}\,(\theta) = R\theta$$

or

$$\theta = \mathfrak{t}(s) = \frac{s}{R}$$

and the arclength parametrization is easily seen to be

$$\overrightarrow{\gamma}(s) = (R\cos\frac{s}{R}, R\sin\frac{s}{R})$$

with unit velocity

$$\overrightarrow{T}(s) = \dot{\overrightarrow{\gamma}}(s) = (-\sin\frac{s}{R}, \cos\frac{s}{R}).$$

Curvature

When $\overrightarrow{\gamma}(s)$ is an arclength parametrization of $\mathcal{C}$, then as we saw, the velocity has constant length 1:

$$\left\|\dot{\overrightarrow{\gamma}}(s)\right\| = \left\|\overrightarrow{T}(s)\right\| = 1$$

and it follows from Corollary 2.3.16 that the acceleration is perpendicular to the velocity:

$$\overrightarrow{T}(s) \cdot \dot{\overrightarrow{T}}(s) = 0.$$

We can then define a new vector, the **principal normal** vector to be the unit vector parallel to the acceleration

$$\overrightarrow{N}(s) = \frac{1}{\left\|\dot{\overrightarrow{T}}(s)\right\|}\dot{\overrightarrow{T}}(s)$$

and so can write the acceleration as

$$\dot{\overrightarrow{T}}(s) = \kappa(s)\,\overrightarrow{N}(s) \tag{D.1}$$

where $\kappa(s)$ is a nonnegative function of s, called the **curvature** of $\mathcal{C}$ at $\overrightarrow{\gamma}(s)$.

The name is motivated by its meaning in the case of a circle. Recall that the arclength parametrization of a circle of radius R about the origin

$$\overrightarrow{\gamma}(s) = (R\cos\frac{s}{R}, R\sin\frac{s}{R})$$

has unit velocity

$$\overrightarrow{T}(s) = \dot{\overrightarrow{\gamma}}(s) = (-\sin\frac{s}{R}, \cos\frac{s}{R})$$

and hence acceleration

$$\begin{aligned}\dot{\overrightarrow{T}}(s) &= (-\frac{1}{R}\cos\frac{s}{R}, -\frac{1}{R}\sin\frac{s}{R}) \\ &= -\frac{1}{R}(\cos\frac{s}{R}, \sin\frac{s}{R})\end{aligned}$$

so

$$\vec{N}(s) = -(\cos\frac{s}{R}, \sin\frac{s}{R})$$

and

$$\kappa(s) = \frac{1}{R}.$$

That is, *the curvature of a circle is the reciprocal of its radius.* This is semantically the right idea, since we think of a circle with smaller radius as *more* (tightly) *curved.*

To understand the geometric significance of curvature for a general curve, recall that two functions $f(x)$ and $g(x)$ have **second-order contact** at $x = x_0$ if

$$|f(x) - g(x)| = \mathfrak{o}(|x - x_0|^2)$$

which is to say

$$\frac{|f(x) - g(x)|}{|x - x_0|^2} \to 0 \text{ as } x \to x_0;$$

for sufficiently differentiable functions, this condition is equivalent to the equality at $x = x_0$ of the values and first two derivatives of the two functions. The obvious analogue of this condition for vector-valued functions $\vec{f}(t)$ and $\vec{g}(t)$ replaces $|f(x) - g(x)|$ with $\left\|\vec{f}(t) - \vec{g}(t)\right\|$: we say that $\vec{f}, \vec{g}: \mathbb{R} \to \mathbb{R}^3$ have **second-order contact** at $t = t_0$ if

$$\left\|\vec{f}(t) - \vec{g}(t)\right\| = \mathfrak{o}(|t - t_0|^2) \qquad \text{as } t \to t_0.$$

When these are arclength parametrizations of two curves—let us write $\vec{\gamma}(s)$ and $\vec{\varphi}(s)$—the derivative condition becomes

$$\vec{\gamma}(s_0) = \vec{\varphi}(s_0)$$

(they go through the same point at $s = s_0$)

$$\dot{\vec{\gamma}}(s_0) = \dot{\vec{\varphi}}(s_0)$$

(their unit tangent vectors at $s = s_0$ agree), or

$$\overrightarrow{T}_{\gamma}(s_0) = \overrightarrow{T}_{\varphi}(s_0)$$

and, using Equation (D.1),

$$\begin{aligned}\kappa_{\gamma}(s_0)\,\overrightarrow{N}_{\gamma}(s_0) &= \ddot{\overrightarrow{\gamma}}(s_0)\\ &= \ddot{\overrightarrow{\varphi}}(s_0)\\ &= \kappa_{\varphi}(s_0)\,\overrightarrow{N}_{\varphi}(s_0)\,.\end{aligned}$$

In particular, curves whose arclength parametrizations have second-order contact have the same curvature. When $\overrightarrow{\varphi}(s)$ is an arclength parametrization of a circle of radius $R = \frac{1}{\kappa_{\gamma}(s_0)}$ and center at $\overrightarrow{c} = \overrightarrow{\gamma}(s_0) + \frac{1}{R}\overrightarrow{N}_{\gamma}(s_0)$, this means

Lemma D.0.11. *Suppose $\overrightarrow{\gamma}(s)$ is an arclength parametrization of the curve $\mathcal{C}$; for a point $P = \overrightarrow{\gamma}(s_0)$ on $\mathcal{C}$, let $\kappa = \kappa(s_0)$ be the curvature of $\mathcal{C}$ at P. Then the circle of radius $R = \frac{1}{\kappa}$ and center $\overrightarrow{c} = \overrightarrow{\gamma}(s_0) + \kappa\overrightarrow{N}(s_0)$ is the (unique) circle which has second-order contact with $\mathcal{C}$ at P.*

This circle is called the **osculating circle** or **circle of curvature** for $\mathcal{C}$ at P. Note that it lies in the plane determined by $\overrightarrow{T}$ and vN, which is called the **osculating plane** of $\mathcal{C}$ at P. The vector $\frac{d}{ds}\left[\overrightarrow{T}\right] = \kappa\overrightarrow{N}$ is called the **curvature vector**, and the radius of the osculating circle $R = \frac{1}{\kappa}$ is called the **radius of curvature** of $\mathcal{C}$ at P. For a straight line, $\overrightarrow{T}$ is a constant vector, so that $\kappa\overrightarrow{N}$ is the zero vector; in this case the curvature is zero $\kappa = 0$ and the principal normal is undefined. This agrees with the fact that as $R \to \infty$, the arc of a circle becomes more and more "straight", and $\kappa = \frac{1}{R} \to 0$. In general, a curve can have zero curvature at a point without containing a straight-line segment, but in this case the principal normal vector $\overrightarrow{N}$ is still undefined.

Given an arbitrary regular parametrization of a curve, it can be difficult to explicitly find its arclength parametrization; however, by use of implicit differentiation we can still calculate the curvature and the principal normal at a given point.

Lemma D.0.12. *Suppose $\overrightarrow{p}(t)$ is a regular parametrization of the curve $\mathcal{C}$; then the curvature at $P = \overrightarrow{p}(t_0)$ is given by*

$$\kappa = \frac{\|\overrightarrow{a} \times \overrightarrow{v}\|}{\|\overrightarrow{v}\|^3} = \frac{\left\|\ddot{\overrightarrow{p}}(t_0) \times \dot{\overrightarrow{p}}(t_0)\right\|}{\left\|\dot{\overrightarrow{p}}(t_0)\right\|^3}. \tag{D.2}$$

Proof. Write $\overrightarrow{v}(t) = \dot{\overrightarrow{p}}(t)$ for the velocity, and note that by definition

$$\overrightarrow{v} = \|\overrightarrow{v}\| \overrightarrow{T}$$

so

$$\frac{d}{dt}[\overrightarrow{v}] = \frac{d}{dt}[\|\overrightarrow{v}\|] \overrightarrow{T} + \|\overrightarrow{v}\| \frac{d}{dt}\left[\overrightarrow{T}\right].$$

Let s be the arclength parameter for $\overrightarrow{p}$, so

$$\frac{ds}{dt} = \frac{d\mathfrak{s}}{dt}.$$

Using Equation (2.23) in Corollary 2.3.16 together with the definition of $\overrightarrow{T}$ we can rewrite this as

$$= \left(\frac{\dot{\overrightarrow{v}} \cdot \overrightarrow{v}}{\|\overrightarrow{v}\|}\right) \frac{\overrightarrow{v}}{\|\overrightarrow{v}\|} + \|\overrightarrow{v}\| \frac{d}{ds}\left[\overrightarrow{T}\right] \frac{d\mathfrak{s}}{dt}$$

or, using the formula for vector projection (Proposition 1.4.3) and the definition of the curvature vector and speed,

$$\frac{d}{dt}[\overrightarrow{v}] = \operatorname{proj}_{\overrightarrow{v}} \dot{\overrightarrow{v}} + \|\overrightarrow{v}\| \left(\kappa \overrightarrow{N}\right)(\|\overrightarrow{v}\|)$$

which we can solve for $\kappa \overrightarrow{N}$

$$\kappa \overrightarrow{N} = \frac{1}{\|\overrightarrow{v}\|^2}\left(\dot{\overrightarrow{v}} - \operatorname{proj}_{\overrightarrow{v}} \dot{\overrightarrow{v}}\right).$$

We recognize the vector in parentheses as the component of $\dot{\overrightarrow{v}} = \overrightarrow{a}$ normal to $\overrightarrow{v}$,

$$\overrightarrow{a} - \operatorname{proj}_{\overrightarrow{v}} \overrightarrow{a} = \overrightarrow{a}^{\perp},$$

whose length is

$$\left\|\overrightarrow{a}^{\perp}\right\| = \|\overrightarrow{a}\| \sin\theta$$

where θ is the angle between $\overrightarrow{v}$ and $\overrightarrow{a}$. We recognize this as the formula for the length of the cross product between $\overrightarrow{a}$ and the unit vector parallel to $\overrightarrow{v}$

$$\left\|\overrightarrow{a}^{\perp}\right\| = \frac{\|\overrightarrow{a} \times \overrightarrow{v}\|}{\|\overrightarrow{v}\|}$$

so that

$$\begin{aligned}\kappa &= \frac{1}{\|\overrightarrow{v}\|^2}\left\|\overrightarrow{a}^{\perp}\right\| \\ &= \frac{\|\overrightarrow{a} \times \overrightarrow{v}\|}{\|\overrightarrow{v}\|^3} \\ &= \frac{\left\|\dot{\overrightarrow{v}} \times \overrightarrow{v}\right\|}{\|\overrightarrow{v}\|^3}\end{aligned}$$

as required. □

Using this, we can calculate the curvature of some explicit curves.

The helix parametrized by

$$\overrightarrow{p}(t) = (\cos 2\pi t, \sin 2\pi t, t)$$

has velocity

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

and acceleration

$$\overrightarrow{a}(t) = (-4\pi^2 \cos 2\pi t, -4\pi^2 \sin 2\pi t, 0)$$

and we can calculate that

$$\overrightarrow{a} \times \overrightarrow{v} = (4\pi^2 \sin 2\pi t)\overrightarrow{\imath} - (4\pi^2 \cos 2\pi t)\overrightarrow{\imath} + 8\pi^3\overrightarrow{k}$$

with length

$$\begin{aligned}\|\overrightarrow{a} \times \overrightarrow{v}\| &= \sqrt{(4\pi^2)^2 + (8\pi^3)^2} \\ &= 4\pi^2\sqrt{1 + 4\pi^2}\end{aligned}$$

while the speed is

$$\|\overrightarrow{v}\| = \sqrt{1+4\pi^2}$$

so the helix has constant curvature

$$\begin{aligned}\kappa &= \frac{\|\overrightarrow{a}\times\overrightarrow{v}\|}{\|\overrightarrow{v}\|^3}\\ &= \frac{4\pi^2}{1+4\pi^2}.\end{aligned}$$

As a second example, we consider the ellipse

$$\overrightarrow{p}(\theta) = (a\cos\theta, b\sin\theta)$$

with velocity

$$\overrightarrow{v}(\theta) = (-a\sin\theta, b\cos\theta)$$

speed

$$\|\overrightarrow{v}(\theta)\| = \sqrt{a^2\sin^2\theta + b^2\cos^2\theta}$$

and velocity

$$\overrightarrow{a}(\theta) = (-a\cos\theta, -b\sin\theta);$$

we see that

$$\|\overrightarrow{a}\times\overrightarrow{v}\| = ab$$

so

$$\kappa(\theta) = \frac{ab}{(a^2\sin^2\theta + b^2\cos^2\theta)^{3/2}}.$$

Intrinsic Geometry of Plane Curves

In this subsection we shall see that the curvature, as a function of arclength, essentially determines the “shape” of a curve in the plane. We consider two curves to have the same shape if they are **congruent**: if we can move one onto the other by means of **rigid motions**: translations, rotations and reflections. Note first that the elements in the formula Equation (D.2) for curvature are unchanged by rigid motion, so we immediately have the invariance of curvature under rigid motion (even for curves in space):

Remark D.0.13. *For any pair of congruent curves, the curvature κ at corresponding points is equal.*

We would like to obtain a kind of converse for this statement, but there is a slight difficulty here. To see the nature of this difficulty, consider the graph of $y = x^3$ (Figure D.1), parametrized by

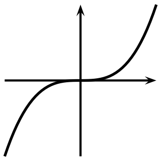


Figure D.1: The curve $y = x^3$

$$\overrightarrow{p}(x) = (x, x^3)$$

with velocity

$$\overrightarrow{v}(x) = (1, 3x^2)$$

speed

$$\|\overrightarrow{v}(x)\| = \sqrt{1 + 9x^4}$$

and acceleration

$$\overrightarrow{a}(x) = (0, 6x).$$

For $x \neq 0$, the curvature is given by

$$\begin{aligned}\kappa(x) &= \frac{\|\overrightarrow{a} \times \overrightarrow{v}\|}{\|\overrightarrow{v}\|^3} \\ &= \frac{|6x|}{(1 + 9x^4)^{3/2}}\end{aligned}$$

which also works at $x = 0$, since there $\overrightarrow{a} = \vec{0}$ so $\kappa = 0$.

Now, consider another curve (Figure D.2), the graph of

$$y = \left|x^3\right| = \begin{cases} -x^3 & \text{for } x < 0, \\ x^3 & \text{for } x \geq 0. \end{cases}$$

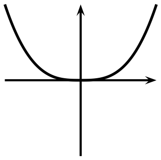


Figure D.2: The curve $y = \left|x^3\right|$

This is parametrized by

$$\overrightarrow{q}(t) = (x, \left|x^3\right|)$$

with velocity

$$\overrightarrow{v}(x) = \begin{cases} (1, -3x^2) & \text{for } x < 0, \\ (1, 3x^2) & \text{for } x \geq 0 \end{cases}$$

and acceleration

$$\overrightarrow{a}(x) = \begin{cases} (0, -6x) & \text{for } x < 0, \\ (0, 6x) & \text{for } x \geq 0. \end{cases}$$

Its curvature is given by

$$\begin{aligned} \kappa(x) &= \frac{\|\overrightarrow{a} \times \overrightarrow{v}\|}{\|\overrightarrow{v}\|^3} \\ &= \frac{|6x|}{(1+9x^4)^{3/2}} \end{aligned}$$

which agrees with the formula for $\overrightarrow{p}(x)$; furthermore, since the speed of both parametrizations is the same, even if we write κ as a function of arclength, the two curvature functions will agree. Nonetheless, the curves are not congruent: the part of $\overrightarrow{p}(x)$ for $x < 0$ is the *reflection* about the x-axis of $\overrightarrow{q}(t)$, while they are the *same* for $x \geq 0$. One can see from the geometry that for $\overrightarrow{q}(x)$ the principal normal always has an *upward* component (except when $x = 0$), while for $\overrightarrow{p}(x)$ it is *downward* when $x < 0$ and *upward* when $x > 0$.

To overcome this difficulty, it will be useful to find another interpretation of the curvature. Since the unit tangent vector $\overrightarrow{T}$ has length one, it can be written in the form

$$\overrightarrow{T}(s) = (\cos\theta(s)\,, \sin\theta(s))$$

where $\theta(s)$ is the angle (measured counterclockwise) between $\overrightarrow{T}$ and the positive x-axis. Differentiating with respect to arclength, we obtain

$$\begin{aligned}\kappa(s)\,\overrightarrow{N}(s) &= \frac{d\overrightarrow{T}(s)}{ds} \\ &= \theta'(s)\,(-\sin\theta(s)\,, \cos\theta(s))\end{aligned}$$

from which we see that

$$\begin{aligned}\kappa(s) &= \left\| \frac{d\overrightarrow{T}(s)}{ds} \right\| \\ &= \left|\theta'(s)\right|\end{aligned}$$

and

$$\overrightarrow{N}(s) = \pm(-\sin\theta(s)\,, \cos\theta(s))$$

where the sign agrees with the sign of $\theta'(s)$. Note that if we replaced the angle $\theta(s)$ with the angle between $\overrightarrow{T}(s)$ and any other (fixed) line in the plane, we would only add a constant to the function $\theta(s)$, and its derivative would be unchanged.

Remark D.0.14. *The curvature function $\kappa(s)$ of a plane curve is the (unsigned) rate of change, with respect to arclength, of the angle between the unit tangent vector $\overrightarrow{T}(s)$ and any fixed line in the plane.*

In particular, we can define the **signed curvature** to be the *signed* rate of change of $\theta(s)$

$$\kappa_{\pm}(s) = \theta'(s)\,. \tag{D.3}$$

We note briefly that a curve $\mathcal{C}$ has two distinct arclength parametrizations: the second corresponds to beginning at the endpoint of the first and "going backward"; this reverses the unit tangent vector at any given point of $\mathcal{C}$, which means the new angle $\theta(s)$ is the negative of the old one, but we are also going "backward" along the curve, so its derivative is also reversed—the two sign reversals cancel out, and so *the signed curvature is independent of which direction we go along the curve.*

Our main observation, then, is

Proposition D.0.15. *Suppose $\overrightarrow{\gamma}_i\colon \mathbb{R}\to\mathbb{R}^2$, $i = 1, 2$ are two arclength parametrizations of plane curves with the same domain $[0, b]$ with the same initial velocity ($\overrightarrow{T}_1(0) = \overrightarrow{T}_2(0)$) and the same signed curvature functions (with respect to the same line).*

Then $\overrightarrow{\gamma}_1(s) = \overrightarrow{\gamma}_2(s)$ for all $s \in [0, b]$; that is, they trace out the same plane curve.

Proof. By assumption,

$$\begin{aligned}\theta_1(0) &= \kappa_{1,\pm}(s)\\ &= \kappa_{2,\pm}(s)\\ &= \theta_2(0)\end{aligned}$$

and

$$\theta_1'(s) = \theta_2'(s) \text{ for } 0 \leq s \leq b.$$

The second condition insures that the two functions differ by a constant, which by the first condition is zero. Thus, for every $s \in [0, b]$ the two velocity vectors make the same angle with our common line, and hence (since both are unit vectors)

$$\overrightarrow{T}_1(s) = \overrightarrow{T}_2(s) \quad 0 \leq s \leq b.$$

Thus the *derivatives* of the two vector-valued functions $\overrightarrow{\gamma}_i(s)$, $i = 1, 2$ agree; but since they also have the same starting point, they are the same for all s, and we are done. □

Corollary D.0.16. *Any two regular plane curves with the same non-vanishing curvature function are congruent.*

Proof. Recall that the definition of the signed curvature does not depend upon which line we use to measure the angle $\theta(s)$; let us use, for each of the two curves, the tangent line at the starting point, so $\theta_i(s)$ is the (counterclockwise) angle between $\overrightarrow{T}_i(0)$ and $\overrightarrow{T}_i(s)$, for $i = 1, 2$. By assumption, $\theta_1'(0)$ and $\theta_2'(0)$ are equal or negatives of each other (since they both have the same absolute value); if they have opposite signs, we replace the second curve by its reflection across its initial tangent line, to make the initial sign of $\theta_i'(s)$ the same. Since neither is ever zero and both have the same absolute value, they are the same function. But then if we rotate the second curve so that its tangent line becomes parallel to the initial tangent line of the first curve, we have that they always have the same angle with parallel lines, which means their unit tangent vectors are always parallel (that is, as free vectors they agree). This means the vector-valued functions giving arclength parametrizations of the two curves differ by a constant vector; translation of the second curve so that its initial point agrees with that of the first makes this constant zero, which is to say the two curves are the same after applying a possible reflection, rotation and translation; they are congruent. □

Space Curves

For space curves, the curvature alone is not enough to determine the curve: for example, we saw earlier that a helix has constant curvature, just like a circle, but the two are never congruent. Thus we need to refine our analysis. This was done independently by Fréderic-Jean Frénet (1816-1900) and Joseph Alfred Serret (1819-1885); Frénet did it in his thesis (at Toulouse) in 1847, but only published an abstract in 1852 [14]; in the meantime Serret published his version in the same journal a year earlier [47].

Recall that the unit tangent vector $\overrightarrow{T}$ and principal normal $\overrightarrow{N}$, if nonzero, are unit vectors perpendicular to each other; therefore the osculating plane, which contains both, has as a normal vector their cross product, which is called the **binormal**

$$\overrightarrow{B}(s) = \overrightarrow{T}(s) \times \overrightarrow{N}(s)\,.$$

Now, the unit normal $\overrightarrow{N}$ is a unit vector, and hence its derivative $\dot{\overrightarrow{N}}$ is perpendicular to $\overrightarrow{N}$. This means it lies in the plane spanned by the unit tangent and binormal:

$$\dot{\overrightarrow{N}}(s) = a(s)\,\overrightarrow{T}(s) + b(s)\,\overrightarrow{B}(s)\,.$$

Furthermore, since $\overrightarrow{T}(s)$ and $\overrightarrow{B}(s)$ are perpendicular to each other, we can find the two functions $a(s)$ and $b(s)$ by taking the dot product of $\dot{\overrightarrow{N}}(s)$ with, respectively, $\overrightarrow{T}(s)$ and $\overrightarrow{B}(s)$:

$$\begin{aligned}\dot{\overrightarrow{N}}(s)\cdot\overrightarrow{T}(s) &= (a(s)\,\overrightarrow{T}(s)+b(s)\,\overrightarrow{B}(s))\cdot\overrightarrow{T}(s)\\ &= a(s)\,\overrightarrow{T}(s)\cdot\overrightarrow{T}(s)+b(s)\,\overrightarrow{B}(s)\cdot\overrightarrow{T}(s)\\ &= a(s)\,(1)+b(s)\,(0)\\ &= a(s)\end{aligned}$$

and similarly

$$b(s)=\dot{\overrightarrow{N}}(s)\cdot\overrightarrow{B}(s)\,.$$

But since $\overrightarrow{N}(s)$ and $\overrightarrow{T}(s)$ are perpendicular to each other, their dot product is zero, from which it follows by the product rule that

$$\dot{\overrightarrow{N}}(s)\cdot\overrightarrow{T}(s)+\overrightarrow{N}(s)\cdot\dot{\overrightarrow{T}}(s)=0$$

or

$$a(s)+\overrightarrow{N}(s)\cdot\kappa(s)\,\overrightarrow{N}(s)=0$$

in other words, (since $\overrightarrow{N}\cdot\overrightarrow{N}=\left\|\overrightarrow{N}\right\|^2=1$)

$$a(s)=-\kappa(s)\,.$$

We define the **torsion** to be the coefficient $b(s)$, which is usually denoted $\tau(s)$: this leads to the equation

$$\dot{\overrightarrow{N}}(s)=-\kappa(s)\,\overrightarrow{T}(s)+\tau(s)\,\overrightarrow{B}(s)\,.$$

Finally, we can calculate the derivative of $\overrightarrow{B}(s)$:

$$\begin{aligned}\dot{\overrightarrow{B}}(s) &= \frac{d}{ds}\left[\overrightarrow{T}(s)\times\overrightarrow{N}(s)\right]\\ &= \dot{\overrightarrow{T}}(s)\times\overrightarrow{N}(s)+\overrightarrow{T}(s)\times\dot{\overrightarrow{N}}(s)\\ &= \kappa(s)\,\overrightarrow{N}(s)\times\overrightarrow{N}(s)+\overrightarrow{T}(s)\times-\kappa(s)\,\overrightarrow{T}(s)+\tau(s)\,\overrightarrow{B}(s)\\ &= \vec{0}+\tau(s)\,\overrightarrow{T}(s)\times\overrightarrow{B}(s)\,.\end{aligned}$$

It is easy to check that

$$\overrightarrow{T}(s) \times \overrightarrow{B}(s) = -\overrightarrow{N}(s)$$

so we are led to the system of three equations known as the **Frenet-Serret formulas**:

$$\begin{array}{rclcl} \overrightarrow{T}' & = & & \kappa\overrightarrow{N} & \\ \overrightarrow{N}' & = & -\kappa\overrightarrow{T} & & +\tau\overrightarrow{B} \\ \overrightarrow{B}' & = & & -\tau\overrightarrow{N}. & \end{array} \tag{D.4}$$

Assuming nonvanishing curvature $\kappa(s) \neq 0$, the three vectors $\overrightarrow{T}(s)$, $\overrightarrow{N}(s)$ and $\overrightarrow{B}(s)$ are mutually perpendicular unit vectors for each value of s. A set of three mutually perpendicular unit vectors in space is called a **frame** (a standard example is $\{\overrightarrow{\imath}, \overrightarrow{\jmath}, \overrightarrow{k}\}$). A frame $\{\overrightarrow{e}_1, \overrightarrow{e}_2, \overrightarrow{e}_3\}$ has the property that every vector $\overrightarrow{v} \in \mathbb{R}^3$ has a unique expression as a combination of $\overrightarrow{e}_1$, $\overrightarrow{e}_2$ and $\overrightarrow{e}_3$. It turns out that this, together with Equation (D.4), means that we have an analogue of Proposition D.0.15 for curves in space; however, the proof of this is beyond our present means: it requires either the theory of linear systems of ordinary differential equations or the differentiation of matrix-valued functions. However, we state without proof

Theorem D.0.17 (Frenet-Serret Theorem)**.** *Two regular curves with the same curvature and torsion functions, with nowhere vanishing curvature, are congruent.*

To understand these constructions, let us carry them out for the family of curves generalizing the helix of § 2.2

$$\overrightarrow{p}(t) = (a\cos t, a\sin t, bt)$$

where a and b are constants. The velocity of this parametrization is

$$\overrightarrow{v} = (-a\sin t, a\cos t, b)$$

with length (*i.e.*, speed)

$$\frac{d\mathfrak{s}}{dt} = \sqrt{a^2 + b^2};$$

since this is constant, the reparametrization by arclength is

$$\mathfrak{t}(s) = \frac{s}{\sqrt{a^2 + b^2}}$$

or

$$\overrightarrow{\gamma}(s) = \left(a\cos\frac{s}{\sqrt{a^2+b^2}}, a\sin\frac{s}{\sqrt{a^2+b^2}}, \frac{bs}{\sqrt{a^2+b^2}}\right).$$

The unit tangent is

$$\overrightarrow{T}(s) = \left(-\frac{a}{\sqrt{a^2+b^2}}\left[\sin\frac{s}{\sqrt{a^2+b^2}}\right], \frac{a}{\sqrt{a^2+b^2}}\left[\cos\frac{s}{\sqrt{a^2+b^2}}\right], \frac{b}{\sqrt{a^2+b^2}}\right).$$

The derivative of $\overrightarrow{T}$s with respect to s is

$$\kappa(s)\,\overrightarrow{N}(s) = \left(-\frac{a}{a^2+b^2}\left[\cos\frac{s}{\sqrt{a^2+b^2}}\right], -\frac{a}{a^2+b^2}\left[\sin\frac{s}{\sqrt{a^2+b^2}}\right], 0\right);$$

in particular, the curvature is constant

$$\kappa(s) = \kappa = \frac{a}{a^2+b^2}$$

and the normal vector is

$$\overrightarrow{N}(s) = \left(-\cos\frac{s}{\sqrt{a^2+b^2}}, \sin\frac{s}{\sqrt{a^2+b^2}}, 0\right).$$

Now the binormal vector is

$$\begin{aligned}\overrightarrow{B}(s) &= \overrightarrow{T}(s)\times\overrightarrow{N}(s)\\ &= \left(\frac{b}{\sqrt{a^2+b^2}}\right)\left(\sin\frac{s}{\sqrt{a^2+b^2}}\overrightarrow{\imath} - \cos\frac{s}{\sqrt{a^2+b^2}}\overrightarrow{\jmath}\right) + \left(\frac{a}{\sqrt{a^2+b^2}}\right)\overrightarrow{k}\end{aligned}$$

and the derivative of $\overrightarrow{N}(s)$ with respect to s is

$$\dot{\overrightarrow{N}}(s) = \left(\frac{1}{\sqrt{a^2+b^2}}\sin\frac{s}{\sqrt{a^2+b^2}}, -\frac{1}{\sqrt{a^2+b^2}}\cos\frac{s}{\sqrt{a^2+b^2}}, 0\right)$$

so we can calculate the torsion as

$$\begin{aligned}\tau(s) &= \dot{\overrightarrow{N}}(s)\cdot\overrightarrow{B}(s)\\ &= \frac{b}{a^2+b^2}\sin^2\frac{s}{\sqrt{a^2+b^2}} + \frac{b}{a^2+b^2}\cos^2\frac{s}{\sqrt{a^2+b^2}} + 0\\ &= \frac{b}{a^2+b^2}.\end{aligned}$$

We leave it to you to verify the other components of $\dot{\overrightarrow{N}}(s)$ and $\dot{\overrightarrow{B}}(s)$. From Theorem D.0.17, we see that helices can be characterized as those curves with constant curvature and torsion; in particular, a "helix" with zero torsion is a circle. In general, it is easy to see that a curve has zero torsion precisely if it is contained in a plane; in a sense, the torsion measures the "speed" with which the curve leaves its osculating plane. For more on the intrinsic geometry of curves (and surfaces) see any book on differential geometry, for example [51].

E
Matrix Basics

Matrices and matrix operations offer an efficient, systematic way to handle several numbers at once. In this appendix, we review the basics of matrix algebra. Determinants are considered separately, in Appendix F.

An $\boldsymbol{m \times n}$ **matrix** is an array consisting of m **rows** with n entries each, aligned vertically in n **columns**. We shall deal primarily with matrices whose dimensions m and n are at most 3. An example is the **coordinate column** $[\overrightarrow{x}]$ of a vector $\overrightarrow{x} = x_1 \overrightarrow{\imath} + x_2 \overrightarrow{\jmath} + x_3 \overrightarrow{k}$:

$$[\overrightarrow{x}] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

which is a 1×3 matrix. A 3×3 matrix has the form

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

which has three rows

$$\begin{aligned} \text{row}_1(A) &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix} \\ \text{row}_2(A) &= \begin{bmatrix} a_{21} & a_{22} & a_{23} \end{bmatrix} \\ \text{row}_3(A) &= \begin{bmatrix} a_{31} & a_{32} & a_{33} \end{bmatrix} \end{aligned}$$

(which are 1×3 matrices) and three columns

$$\mathrm{col}_1(A) = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} \quad \mathrm{col}_2(A) = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \quad \mathrm{col}_3(A) = \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}$$

(which are 3×1). Note that the entry a_{ij} is in $\mathrm{row}_i(A)$ and $\mathrm{col}_j(A)$.

E.1 Matrix Algebra

Matrix sums and scaling

Matrices add and scale much the way vectors do: addition is componentwise, and scaling consists of multiplying all the entries by the same number. Thus, if A is the matrix above with entries a_{ij} and B is 3×3 with entries b_{ij} then their sum has entries $a_{ij} + b_{ij}$, and cA has entries ca_{ij}:

$$A+B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} \end{bmatrix}, \quad cA = \begin{bmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \\ ca_{31} & ca_{32} & ca_{33} \end{bmatrix}.$$

As an example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 \\ 1 & -2 & 0 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 4 \\ 5 & 3 & 6 \\ 8 & 7 & 8 \end{bmatrix}$$

and

$$2\begin{bmatrix} 1 & 0 & 1 \\ 1 & -2 & 0 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 2 \\ 2 & -4 & 0 \\ 2 & -2 & -2 \end{bmatrix}.$$

These operations obey the same rules as vector sums and scaling:

- matrix addition is **commutative**: $A + B = B + A$;
- matrix addition is **associative**: $A + (B + C) = (A + B) + C$;
- scaling **distributes** over scalar sums and matrix sums: $(c + d)A = cA + dA$ and $c(A + B) = cA + cB$.
- The matrix $\mathcal{O}$ all of whose entries are zero acts as an **additive identity**: $\mathcal{O} + A = A + \mathcal{O} = A$ for every matrix A of the same size as $\mathcal{O}$.

Matrix Products

We saw in § 3.2 that a homogeneous polynomial of degree one (*a.k.a.* linear function) can be viewed as multiplying the coordinate column of each input vector $\overrightarrow{x} \in \mathbb{R}^3$ by a row of coefficients; this can also be interpreted as a dot product:

$$\ell(\overrightarrow{x}) = a_1x_1 + a_2x_2 + a_3x_3 = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \overrightarrow{a} \cdot \overrightarrow{x}$$

where $\overrightarrow{a} = a_1\overrightarrow{\imath} + a_2\overrightarrow{\jmath} + a_3\overrightarrow{k}$: we can regard the row on the left as the **transpose** $[\overrightarrow{a}]^T$ of the coordinate column $[\overrightarrow{a}]$ of $\overrightarrow{a}$. Based on this we can define the product of the 3×3 matrix A with a column vector (3×1 matrix) $[\overrightarrow{x}]$ as the 3-column (3×1 matrix) that results from multiplying each row of A by $[\overrightarrow{x}]$:

$$A[\overrightarrow{x}] = \begin{bmatrix} [\mathrm{row}_1(A)]\,[\overrightarrow{x}] \\ [\mathrm{row}_2(A)]\,[\overrightarrow{x}] \\ [\mathrm{row}_3(A)]\,[\overrightarrow{x}] \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{bmatrix};$$

for example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1(2) + 2(-1) + 3(1) \\ 2(2) + 3(-1) + 1(1) \\ 3(2) + 1(-1) + 2(1) \end{bmatrix} = \begin{bmatrix} 2 - 2 + 3 = 3 \\ 4 - 3 + 1 = 2 \\ 6 - 1 + 2 = 7 \end{bmatrix}.$$

Similarly, the product of a row $[\overrightarrow{x}]^T$ with a 3×3 matrix results from multiplying $[\overrightarrow{x}]^T$ by each column of A:

$$\begin{aligned}[\overrightarrow{x}]^T A &= \begin{bmatrix} [\overrightarrow{x}]^T [\mathrm{col}_1(A)] & [\overrightarrow{x}]^T [\mathrm{col}_2(A)] & [\overrightarrow{x}]^T [\mathrm{col}_3(A)] \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + a_{21}x_2 + a_{31}x_3 & a_{12}x_1 + a_{22}x_2 + a_{32}x_3 & a_{13}x_1 + a_{23}x_2 + a_{33}x_3 \end{bmatrix};\end{aligned}$$

for example,

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

Note that the products $[\overrightarrow{x}]\,A$ and $A\,[\overrightarrow{x}]^T$ are undefined.

Finally, we can define the product AB of two matrices A and B by multiplying each *row* of A by each *column* of B: the entry $(AB)_{ij}$ in $\mathrm{row}_i(AB)$ and $\mathrm{col}_j(AB)$ is

$$(AB)_{ij} = \mathrm{row}_i(A)\,\mathrm{col}_j(B).$$

This only makes sense if the *width* of A matches the *height* of B: if A is $m \times n$ and B is $n \times p$ then AB is $m \times p$.

For example,

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

Proposition E.1.1. *Matrix multiplication satisfies the following:*

- *It is* ***associative****:*

$$A(BC) = (AB)C$$

  whenever the product makes sense;

- *It* ***distributes*** *over matrix sums:*

$$A(B+C) = AB + AC \text{ and } (A+B)C = AC + BC.$$

- *It is in general* **not commutative***: unless A and B are both* square *matrices ($n \times n$ for some n), the two products AB and BA, if defined, are of different sizes, and even for two $n \times n$ matrices the two products can be different (Exercise 1).*

Recall that for the operation of matrix *addition*, the zero matrix $\mathcal{O}$ consisting of all zeroes was an *additive identity*, which meant that adding it to any matrix did not change the matrix—a role analogous to that of the number 0 for addition in $\mathbb{R}$. Of course, there are actually many different zero matrices: given any pair of dimensions m and n, the $m \times n$ zero matrix $\mathcal{O}_{m\times n}$ acts as the additive identity for addition of $m \times n$ matrices. What about matrix mutliplication? Is there a matrix that plays a role for matrix multiplication analogous to that played by the number 1 for multiplication in $\mathbb{R}$, namely that *multiplying* something by it changes nothing (called a **multiplicative identity**)? We see first that since the height (*resp.* width) of a matrix product matches the height of the second (*resp.* width of the first) factor, a multiplicative identity for matrix multiplication must be square (Exercise 2): if A is $m \times n$ then any matrix I for which $IA = A$ (*resp.* $AI = A$) must be $m \times m$ (*resp.* $n \times n$). It turns out that the matrix playing this role is the square matrix with all diagonal entries 1 and all other entries 0:

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & \vdots \\ \vdots & 0 & \dots & \vdots \\ \vdots & \vdots & \dots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

When we want to specify the size of this matrix, we use a subscript: I_n denotes the $n \times n$ **identity matrix**: for any $m \times n$ matrix A, $I_m A = A = AI_n$.

Transpose and Symmetric Matrices

The **transpose** A^T of an $m \times n$ matrix A is the $n \times m$ matrix obtained by changing each row of A into a column of A^T: for $i = 1, \dots, n$,

$$\mathrm{col}_i(A^T) = \mathrm{row}_i(A)$$

and also

$$\mathrm{row}_i(A^T) = \mathrm{col}_i(A).$$

This is the same as reflecting A across its diagonal[1]:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix};$$

for example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}.$$

The relation of transposition to matrix sums and scaling is easy to check (Exercise 3):

$$(cA + B)^T = cA^T + B^T.$$

[1] The **diagonal** of a matrix is the set of entries whose row number and column number match.

A more subtle but very useful fact is the relation of transposition to matrix products (Exercise 4):

$$(AB)^T = B^T A^T;$$

the transpose of a matrix product is the product of the factors, transposed *and in the opposite order.*

A matrix is **symmetric** if it equals its transpose:

$$A^T = A;$$

for example, the matrix

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

is symmetric. Note that symmetry is possible only for a square matrix.

E.2 Matrices and Systems of Equations: Row Reduction

A system of three linear equations in three unknowns

$$\begin{array}{llll} a_{11}x_1 & +a_{12}x_2 & +a_{13}x_3 & = b_1 \\ a_{21}x_1 & +a_{22}x_2 & +a_{23}x_3 & = b_2 \\ a_{31}x_1 & +a_{32}x_2 & +a_{33}x_3 & = b_3 \end{array} \tag{E.1}$$

can be solved systematically via **elimination** . [2] The idea is that we try to rewrite the system in a different form in which the first variable, x_1, appears only in the first equation; then we work on the second and third equations, trying to eliminate the second variable x_2 from the third equation; with

[2] The method of elimination was effectively present in Chinese mathematical texts of the third century AD. In Western European literature, it was presented by Newton in the notes of his Lucasian lectures (deposited in the University archives at Cambridge in 1683) , a projected book titled *AlgebræUniversalis* [53, vol. V, esp. pp. 535-621], later published over his objections—see [53, vol. V, pp.10-15]—in 1707 and in English translation in 1728. Subsequently, versions of this method were mentioned in textbooks by Nathaniel Hammond (1742) and Lacroix (1800). The method became standard in certain circles as a result of its use in connection with least-squares calculations by C. F. Gauss (1777-1855). As a result of the latter, the method is commonly referred to as **Gaussian elimination**. These comments are based on [18], which gives a detailed history of the development of this method.

luck, this leads to a system of equations that looks like

$$\begin{array}{llll} a'_{11}x_1 & +a'_{12}x_2 & +a'_{13}x_3 & = b'_1 \\ & a'_{22}x_2 & +a'_{23}x_3 & = b'_2 \\ & & a'_{33}x_3 & = b'_3 \end{array} ;$$

we can then work our way *up*: we solve the *last* equation for x_3, then substitute the value obtained for x_3 into the *second* equation and solve for x_2, and finally substitute both of the values obtained for x_2 and x_3 into the first equation, and solve for x_1.

By "rewriting the system in a different form" we mean that we replace our system of three equations with a new system of three equations in such a way that we can be sure the *solutions* of the two systems are identical. This involves three basic operations on the system. Two of these are quite simple:

- if we *multiply both sides of one equation by a nonzero number*, we don't change the solutions,
- nor do we change the solutions by *shuffling the order of the equations.*
- The third and most important operation is: replace exactly one of the equations–say the i^{th}–with the sum of it and (a multiple of) one of the other equations–say the j^{th}–leaving the other two equations unchanged. We shall refer to this as *adding [a multiple of] the j^{th} equation to the i^{th}.*

To see how this works, consider the system

$$\begin{array}{rrrl} x_1 & +x_2 & +x_3 & = 2 \\ x_1 & +2x_2 & +2x_3 & = 3 \\ 2x_1 & +3x_2 & +x_3 & = 1 \end{array} ; \tag{E.2}$$

We can eliminate x_1 from the *second* equation by subtracting the *first* from it

$$\begin{array}{rrrl} x_1 & +x_2 & +x_3 & = 2 \\ & x_2 & +x_3 & = 1 \\ 2x_1 & +3x_2 & +x_3 & = 1 \end{array}$$

and then we can eliminate x_1 from the *third* equation by subtracting *twice* the first equation from the third:

$$\begin{array}{rrrl} x_1 & +x_2 & +x_3 & = 2 \\ & x_2 & +x_3 & = 1 \\ & x_2 & -x_3 & = -3 \end{array} .$$

Now, we subtract the *second* equation from the *third* to eliminate x_2 from the latter:

$$\begin{array}{llll} x_1 & +x_2 & +x_3 & = 2 \\ & x_2 & +x_3 & = 1 \\ & & -2x_3 & = -4 \end{array} .$$

The last equation can be solved for $x_3 = -2$; substituting this into the first and second equations leads to

$$\begin{array}{llll} x_1 & +x_2 & -2 & = 2 \\ & x_2 & -2 & = 1 \\ & & x_3 & = -2 \end{array}$$

which lets us solve the *second* equation for $x_2 = 3$, and substituting this into the first equation leads us to

$$\begin{array}{llll} x_1 & +3 & -2 & = 2 \\ & x_2 & & = 3 \\ & & x_3 & = -2 \end{array}$$

and we see that $x_1 = 1$. Thus the solution to our system is

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 3 \\ x_3 &= -2. \end{aligned}$$

Writing out the steps above involved a lot of uninformative notation: the fact that the unknowns are called x_1, x_2 and x_3 is irrelevant—we could as well have called them x, y and z. Also, many of the "plus signs" are redundant. We can use matrix notation to represent the essential information about our system—which consists of the coefficients of the unknowns and the numbers on the right hand side—in a 3×4 matrix: for the system above, this would be

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 1 & 2 & 2 & 3 \\ 2 & 3 & 1 & 1 \end{array}\right].$$

This matrix is called the **augmented matrix** of the system. We often separate the last column, which represents the right side of the system, from the three columns on the left, which contain the coefficients of the unknowns in the system: think of the vertical line as a stand-in for the "equals" signs in the various equations.

The operations we performed on the equations are represented by **row operations**:

- Multiply all the entries of one row by a nonzero number;
- Interchange two rows;
- Add (a multiple of) row j to row i.

In the matrix representation, eliminating the first variable from all but the first equation amounts to making sure that the only nonzero entry in the first column is the one in the first row. Note that this was accomplished by subtracting (multiples of) the first row from each of the rows below it, so we can combine the two instances of the third row operation into one operation: *use the first row to clear the first column below the first row.* In matrix terms, this step of our example can be written

$$\left[\begin{array}{ccc|c} \mathbf{1} & 1 & 1 & 2 \\ 1 & 2 & 2 & 3 \\ 2 & 3 & 1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} \mathbf{1} & 1 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & -3 \end{array}\right].$$

The major player in this operation was the first entry of the first row; we have set it in boldface to highlight this.

The next step in our example was to use the second row to clear the second column below the second row. Note however that our final goal is to *also* eliminate x_2 from the *first* equation—that is, to make sure that the only nonzero entry in the second *column* is in the second *row*: we accomplish this in our example by subtracting the second row from the first as well as the third: notice that since we have made sure that the first entry of the second row is zero, this doesn't affect the first column. The major player in this operation is the first nonzero entry in the second row, which we have also highlighted with boldface:

$$\left[\begin{array}{ccc|c} \mathbf{1} & 1 & 1 & 2 \\ 0 & \mathbf{1} & 1 & 1 \\ 0 & 1 & -1 & -3 \end{array}\right] \to \left[\begin{array}{ccc|c} \mathbf{1} & 0 & 0 & 1 \\ 0 & \mathbf{1} & 1 & 1 \\ 0 & 0 & -2 & -4 \end{array}\right].$$

Finally, solving the last equation for x_3 amounts to dividing the last row by the coefficient of x_3, that is, by -2; then back-substitution into the preceding equations amounts to using the last row to clear the rest of the third column. For this, the major player is the first nonzero entry in the third row: using it to clear the third column

$$\left[\begin{array}{ccc|c} \mathbf{1} & 0 & 0 & 1 \\ 0 & \mathbf{1} & 1 & 1 \\ 0 & 0 & \mathbf{-2} & -4 \end{array}\right] \to \left[\begin{array}{ccc|c} \mathbf{1} & 0 & 0 & 1 \\ 0 & \mathbf{1} & 0 & 3 \\ 0 & 0 & \mathbf{1} & -2 \end{array}\right]$$

results in the augmented matrix of our solution—that is, this last matrix is the augmented matrix of the system

$$\begin{array}{cccl} x_1 & & & = 1 \\ & x_2 & & = 3 \\ & & x_3 & = -2 \end{array}$$

which exhibits the solution of the system explicitly.

This technique is called **row reduction**: we say that our original matrix *reduces* to the one above. The full story of row reduction is more complicated than suggested by our example: row reduction doesn't always lead to a nice solution like the one we found to the system (E.2). For instance, the system

$$\begin{array}{cccl} x_1 & +x_2 & +x_3 & = 2 \\ x_1 & +x_2 & +2x_3 & = 3 \\ 2x_1 & +2x_2 & +3x_3 & = 5 \end{array} \tag{E.3}$$

looks like three equations in three unknowns, but this is a bit bogus, since (as you might notice) the third equation is just the sum of the other two: in effect, the third equation doesn't give us any information beyond that in the first two, so we expect the set of points in $\mathbb{R}^3$ which satisfy all three equations to be the same as those that satisfy just the first two.

Let us see how this plays out if we try to use row reduction on the augmented matrix of the system: using the first row to clear the first column leads to

$$\left[\begin{array}{ccc|c} \mathbf{1} & 2 & 1 & 2 \\ 1 & 2 & 2 & 3 \\ 2 & 4 & 3 & 5 \end{array}\right] \to \left[\begin{array}{ccc|c} \mathbf{1} & 2 & 1 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right].$$

Now, we can't use the second row to clear the second column, since the $(2, 2)$ entry is zero. In some cases this is not a problem: if at this stage the *third* row had a nonzero entry in the second column, we could perform a row interchange to obtain a matrix in which such a clearing operation would be possible. In this case, though, the third row is just as bad, so we try for the next best thing: we use the *second row* to clear the *third column*:

$$\left[\begin{array}{ccc|c} \mathbf{1} & 2 & 1 & 2 \\ 0 & 0 & \mathbf{1} & 1 \\ 0 & 0 & 1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} \mathbf{1} & 2 & 0 & 1 \\ 0 & 0 & \mathbf{1} & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

At this point, we have cleared all we can. The system represented by this matrix

$$\begin{array}{cccl} x_1 & +2x_2 & & = 1 \\ & & x_3 & = 1 \\ & & 0 & = 0 \end{array}$$

clearly displays the fact that the third equation provides no new information. The second equation tells us that all the points satisfying the system must have third coordinate $x_3 = 1$, and the first equation tells us that the first two coordinates are related by

$$x_1 = 1 - 2x_2.$$

There is nothing in the system that limits the value of x_2; it is a **free variable**. Any particular choice of value for the free variable x_2 determines the value for x_1 (via the first equation) and thus (since the value for x_3 is determined by the second equation) a particular solution to the whole system.

We can formulate our general process of reduction as follows:

- Start with the first column: using a row interchange if necessary, make sure the entry in the first row and first column is nonzero, then use the first row to clear the rest of the first column.

- Next consider the entries in the second column *below the first row*: if necessary, use a row interchange to insure that the first entry *below the first row* in the second column is nonzero, and then use it to clear that column. If this is impossible (because the second column has only zeroes below the first row), give up and go on to the next *column*.

- A given row can be used to clear a column at most once; the "used" rows at any stage are the highest ones in the matrix, and any subsequent clearing uses a lower row. Continue until you run out of columns to clear, or rows to clear them with.

This will result in a matrix with the following properties:

- The first nonzero entry in each row (called its **leading entry**) is the *only* nonzero entry in its *column*.

- As one moves down row-by-row, the leading entries move to the right.

The alert reader will note that this allows the possibility that some row(s) consist only of zeroes; the second property requires that all of these entries are at the bottom of the matrix. We add one final touch to this process: if we divide each nonzero row by its (nonzero) leading entry, we insure that

- The leading entries are all 1.

A matrix having all three properties is said to be *row-reduced*.

Formally stated,

Definition E.2.1. *Any matrix satisfying the conditions*

- *The leading entry in any (nonzero) row is a* 1*;*
- *the leading entries move right as one goes down the rows, with any rows consisting entirely of zeroes appearing at the bottom;*
- *a leading entry is the only nonzero entry in its* column*;*

*is a **reduced matrix** or a matrix in **reduced row-echelon form**.*

A sketch of reduced row-echelon form is

$$\begin{bmatrix} \mathbf{1} & * & 0 & *\ldots & 0 & * \\ 0 & 0 & \mathbf{1} & *\ldots & 0 & * \\ \vdots & \vdots & 0 & 0\ldots & \mathbf{1} & \vdots \\ 0 & 0 & 0 & 0\ldots & 0 & 0 \end{bmatrix}$$

where the asterisks indicate that in a column not containing a leading entry, the entries in all rows whose leading entry is in an *earlier* column (which for an augmented matrix are the coefficients of the free variables) can take on *arbitrary* values.

The process of row-reduction can be applied to a matrix of any size to obtain an equivalent *reduced* matrix of the same size. [3] We shall see below how it can be helpful to reduce a matrix which is not necessarily given as the augmented matrix of some system of equations.

We saw above that a system in which some equation is redundant—in the sense that it can be obtained from the other equations—will reduce to a matrix with a row of zeroes. There is another possibility: that the equations contradict each other: for example, if we change only the right hand side of the third equation in (E.3),

$$\begin{array}{rrrl} x_1 & +x_2 & +x_3 & = 2 \\ x_1 & +x_2 & +2x_3 & = 3 \\ 2x_1 & +2x_2 & +3x_3 & = 6 \end{array} \tag{E.4}$$

[3] It is a fact, which we will not need to use, and which we shall not prove, that this reduced matrix is unique: any two sequences of row operations which start from the same matrix and end up with a reduced one will yield the same result.

then the third equation contradicts the other two, since its left side is the sum of the others, while its right side is not. You should confirm that the row reduction here proceeds as follows:

$$\left[\begin{array}{ccc|c} \mathbf{1} & 2 & 1 & 2 \\ 1 & 2 & 2 & 3 \\ 2 & 4 & 3 & 6 \end{array}\right] \to \left[\begin{array}{ccc|c} \mathbf{1} & 2 & 1 & 2 \\ 0 & 0 & \mathbf{1} & 1 \\ 0 & 0 & 1 & 2 \end{array}\right] \to \left[\begin{array}{ccc|c} \mathbf{1} & 2 & 0 & 1 \\ 0 & 0 & \mathbf{1} & 1 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

Interpreted as a system of equations, this reads

$$\begin{array}{cccc} x_1 & +2x_2 & & = 1 \\ & & x_3 & = 1 \\ & & 0 & = 1; \end{array}$$

the last equation is clearly nonsensical—that is, no values of x_1, x_2 and x_3 can make this equation hold, so no such values can represent a solution of our system. This "standard nonsense equation" always shows up in the form of *a leading entry in the last column* of the (reduced) augmented matrix.

From the preceding, we see that the reduction of the augmented matrix of a system of three equations in three unknowns results in one of the scenarios described in the following remark:

Remark E.2.2.

- *If the last column (to the right of the vertical bar) in the reduced matrix contains a leading entry of some row, the corresponding equation is* $0 = 1$ *and the system is* ***inconsistent****: there are* ***no solutions****.*
- *If the reduced matrix has three leading entries, located in the* (i, i) *position for* $i = 1, 2, 3$*, the three equations have the form* $x_i = b_i$*, exhibiting the* ***unique*** *solution of the system.*
- *If there are fewer than three leading entries in the reduced matrix, none of which is in the last column, then the system is consistent, but there are free variables, whose value can be assigned at will; each such assignment determines one of the* ***infinitely many*** *solutions of the system.*

E.3 Matrices as Transformations

A 3×3 matrix A generates a transformation that moves points around in 3-space. For any 3-vector $\overrightarrow{x} \in \mathbb{R}^3$, the coordinate column $[\overrightarrow{x}]$ can be multiplied by A to get a new 3×1 matrix $A[\overrightarrow{x}]$. This in turn is the coordinate column of a new vector in $\mathbb{R}^3$, which (by abuse of notation) we denote $A\overrightarrow{x}$. The operation that moves the point with position vector $\overrightarrow{x}$ to the point at $A\overrightarrow{x}$, which we will refer to as the transformation $\overrightarrow{x} \mapsto A\overrightarrow{x}$, has several basic properties (Exercise 5):

Proposition E.3.1.

- *The origin stays put:* $A\mathcal{O} = \mathcal{O}$.
- *The transformation is* ***linear****: it respects vector sums and scaling:*

$$\begin{aligned} A(\overrightarrow{v} + \overrightarrow{w}) &= A\overrightarrow{v} + A\overrightarrow{w} \\ A(c\overrightarrow{v}) &= c(A\overrightarrow{v}). \end{aligned}$$

  Geometrically, this means that straight lines go to straight lines.

- *The columns of A tell us where the standard basis vectors go:*

$$\begin{aligned} [A\overrightarrow{\imath}] &= \mathrm{col}_1(A) \\ [A\overrightarrow{\jmath}] &= \mathrm{col}_2(A) \\ \left[A\overrightarrow{k}\right] &= \mathrm{col}_3(A) \end{aligned}$$

  so this information determines where every other vector goes.

The matrix product has a natural interpretation in terms of transformations. Suppose A and B are 3×3 matrices. Multiplying a vector $\overrightarrow{x}$ by B moves it to a new position, that is, the result is a new column $B\overrightarrow{x}$. Now, we can multiply this product by A, moving the new column to another position $A(B\overrightarrow{x})$. We have performed the **composition** of the two transformations: first, move by B, then move by A. But by the associative property of matrix products (Proposition E.1.1), $A(B\overrightarrow{x}) = (AB)\overrightarrow{x}$. This says

Remark E.3.2. *The transformation given by the product matrix AB is the composition of the transformations given by A and B respectively: multiplying $\overrightarrow{x}$ by AB has the same effect as first multiplying $\overrightarrow{x}$ by B and then multiplying the resulting column by A.*

Nonsingular Matrices and Invertible Transformations

A 3×3 matrix is **singular** if there is some *nonzero* vector $\overrightarrow{x} \neq \overrightarrow{0}$ satisfying

$$A\overrightarrow{x} = \overrightarrow{0};$$

otherwise, it is **nonsingular**.

Nonsingular matrices yield transformations with special properties:

Proposition E.3.3. *1. If A is nonsingular, then the transformation is* ***one-to-one****: no two points land in the same place (if $\overrightarrow{x} \neq \overrightarrow{x}'$ then $A\overrightarrow{x} \neq A\overrightarrow{x}'$). This condition means that any equation of the form $A\overrightarrow{x} = \overrightarrow{y}$ can have* ***at most*** *one solution.*

2. If A is nonsingular, then the transformation is ***onto****: every point of $\mathbb{R}^3$ gets hit; equivalently, for every $\overrightarrow{y} \in \mathbb{R}^3$, the equation $A\overrightarrow{x} = \overrightarrow{y}$ has* ***at least*** *one solution.*

By contrast, if A is singular, then there is a plane that contains $A\overrightarrow{x}$ for every $\overrightarrow{x} \in \mathbb{R}^3$.

The first property follows from the calculation

$$A\overrightarrow{x} - A\overrightarrow{x}' = A(\overrightarrow{x} - \overrightarrow{x}');$$

if $\overrightarrow{x} \neq \overrightarrow{x}'$ then the vector on the right side is nonzero, and the fact that the left side is nonzero means $A\overrightarrow{x} \neq A\overrightarrow{x}'$.

The second property is more subtle; we outline a proof in Exercise 6. The proof of the last statement is outlined in Exercise 7.

This point of view gives us a second way to think about our system (E.1) in matrix terms. Let us separate the augmented matrix into the 3×3 **matrix of coefficients**[4] and the last column. A quick calculation shows that if we multiply this matrix by the column whose entries are the unknowns, the product is the column of expressions obtained by substituting our unknowns into the left sides of our three equations:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{bmatrix}.$$

Thus, if we denote the coefficient matrix by A, the column of unknowns by X, and the column of numbers on the right side of (E.1) by B

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix},$$

then the system (E.1) of three equations in three unknowns becomes the matrix equation

$$AX = B.$$

[4] or informally, the coefficient matrix

Now, the analogous equation in *one* unknown, $ax = b$, can be solved (provided $a \neq 0$) via division of both sides by a. We haven't talked about division by a matrix, because in general this is not possible. However, when a 3×3 matrix is nonsingular, then it has a kind of "reciprocal" A^{-1}. This is called the **inverse** of A.

Definition E.3.4. *An $n \times n$ matrix A is* ***invertible*** *if there exists another $n \times n$ matrix A^{-1} such that $AA^{-1} = I$ and $A^{-1}A = I$, where I is the $n \times n$ identity matrix.*

When A is invertible, A^{-1} is called the ***inverse*** *of A.*

When A is invertible, then we can multiply both sides of the equation above by its inverse to solve for X:

$$\begin{aligned} A^{-1}(AX) &= A^{-1}B \\ (A^{-1}A)X &= A^{-1}B \\ IX &= A^{-1}B \\ X &= A^{-1}B. \end{aligned}$$

In particular, using $B = [\overrightarrow{y}]$, we have

Remark E.3.5. *If A is invertible, then A^{-1} represents the transformation that takes $A\overrightarrow{x}$ back to $\overrightarrow{x}$; in other words, for any vector $\overrightarrow{y}$, the unique solution of the equation $A\overrightarrow{x} = \overrightarrow{y}$ is $\overrightarrow{x} = A^{-1}\overrightarrow{y}$.*

In Exercise 15, we outline a proof of

Proposition E.3.6. *Suppose A is a 3×3 matrix. Then the following are equivalent:*

1. *A is nonsingular.*
2. *The transformation $\overrightarrow{x} \mapsto A\overrightarrow{x}$ is one-to-one and onto.*
3. *A is invertible.*

How do we find the inverse matrix if it exists? This involves extending our notion of an augmented matrix.

Given A a 3×3 matrix, we look for a 3×3 matrix A^{-1} satisfying [5]

$$AA^{-1} = I. \tag{E.5}$$

[5] We are abusing notation, since the inverse matrix is actually required to satisfy also the second equation $A^{-1}A = I$. We will see (Exercise 16) that any solution of Equation (E.5) automatically also satisfies the second equation.

This can be regarded as a single matrix equation involving the three 3×3 matrices A, A^{-1} and I, with unknown A^{-1}, but it can also be regarded as three separate systems of equations, resulting from matching corresponding columns on each side of the equation. Recall, however, that each column of a matrix product is the product of the left factor with the corresponding column of the right factor. Thus, we can write three "column equations"

$$\mathrm{col}_j(AA^{-1}) = A\,\mathrm{col}_j(A^{-1}) = \mathrm{col}_j(I), \text{ for } j = 1, 2, 3.$$

Each column equation is really a system of three equations in three unknowns, and all of them have the same coefficient matrix, A. If we try to solve one of these systems by row reduction, we end up reducing an augmented matrix whose first three columns are those of A, and whose fourth column is the appropriate column of the identity matrix I. Since reduction proceeds left to right, this means we apply the same sequence of row operations for each value of j, and so the first three columns of the augmented matrix always come out the same. Thus we can save some time by reducing a *single* 3×6 **super-augmented matrix**, consisting of a copy of A followed by a copy of I, which we denote $[A|I]$. The solutions of the column equations will be the columns of A^{-1}. This sketch explains how the following works.

Proposition E.3.7. *If A is an invertible matrix, then reduction of the super-augmented matrix $[A|I]$ leads to a row-reduced 3×6 matrix whose left half is the identity matrix, and whose right half is the solution A^{-1} of Equation (E.5):*

$$[A|I] \to [I|A^{-1}].$$

If A is singular, then the reduction process ends up with a matrix for which some row has its leading entry on the right side. This means that at least one of the column equations is an inconsistent system of equations, with no solutions—so the inverse of A can't exist.

To see how this works, let us find the inverse of the matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 3 & 1 \end{bmatrix}.$$

You should confirm the following reduction of the superaugmented matrix:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 & 1 & 0 \\ 2 & 3 & 1 & 0 & 0 & 1 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -2 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & -2 & -1 & -1 & 1 \end{array}\right] \to$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 1 & 0 & -\frac{3}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{array}\right].$$

The right half of this is the inverse:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 3 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -\frac{3}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

You should check this by multiplying the two matrices to see if you get the identity.

E.4 Rank

When a matrix A is not invertible, the behavior of a system of equations with coefficient matrix A can be clarified using the *rank* of A. This notion is not limited to square matrices.

Recall that a collection of vectors $\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3$ is **linearly dependent** if there is a linear combination $a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2 + a_3\overrightarrow{v}_3$ which equals the zero vector, and with at least one of the coefficients a_i not zero; we will refer to such a relation $a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2 + a_3\overrightarrow{v}_3 = \overrightarrow{0}$ as a **dependency relation** among the vectors. A particular kind of dependency relation is one in which one of the coefficients is -1, which means that one of the vectors is a linear combination of the rest: for example, if $a_3 = -1$, the relation $a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2 - \overrightarrow{v}_3 = \overrightarrow{0}$ is the same as $\overrightarrow{v}_3 = a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2$. It is easy to see (Exercise 8) that any dependency relation can be rewritten as one of this type. A collection of vectors is **linearly independent** if there is no dependency relation among them: the only way to combine them to get the zero vector is by multiplying them all by zero. Note that the zero vector cannot be included in an independent set of vectors (Exercise 9).

This idea naturally extends to a collection of any number of vectors, and we can also apply it to the rows of a matrix.

Definition E.4.1. *The* ***rank*** *of a matrix A is the maximum number of linearly independent rows in A: that is, A has rank r if some set of r of its rows is linearly independent, but every set of more than r rows is linearly dependent.*

An important basic property of rank is that it is not changed by row operations: this requires a little bit of thought (Exercise 12):

Remark E.4.2. *If A' is obtained from A via a sequence of row operations, then they have the same rank.*

From this observation, we can draw several useful conclusions. Note that for a matrix in *row-reduced form*, the nonzero rows are independent (Exercise 10), and this is clearly the largest possible independent set of rows. Since each nonzero row starts with a leading entry, we see immediately

Remark E.4.3. *The rank of A equals the number of nonzero rows (or equivalently, the number of leading entries) in the reduced matrix equivalent to A.*

Note, however (Exercise 11a), that the *columns* which contain leading entries of a reduced matrix are also linearly independent, and every other column is a combination of them; furthermore, row operations don't change dependency relations among the columns (Exercise 11b). Thus, we can also say that

Remark E.4.4. *The rank of A equals the maximum number of linearly independent columns in A.*

.

Finally, we can use the rank to characterize the solution sets of systems of equations with a given coefficient matrix:

Proposition E.4.5. *Suppose the $m \times n$ matrix A has rank r, and consider the system of equations with augmented matrix $[A|b]$ (where b is the column of constants on the right-hand side). Then*

1. *The system is consistent precisely if A and $[A|b]$ have the same rank.*

2. *If the system is consistent, then the solution set is determined by k free variables, where $r + k = n$ (n is the width of A).*

3. *In particular, if the rank equals the height m of A, then the system is consistent for every right-hand side b, and if the rank is less than m there are right-hand sides making the system inconsistent.*

The proof of this is outlined in Exercise 13.

Exercises for Appendix E

1. Let
$$A = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}.$$
Show that $AB \neq BA$.

2. Suppose A is an $m \times n$ matrix.

   (a) Explain why, if B is a matrix satisfying $BA = A$, then B must be a square matrix of size $m \times m$.

   (b) Explain why, if C satisfies $AC = A$, then C must be $n \times n$.

   (c) Suppose B and C are both $n \times n$ matrices such that for every $n \times n$ matrix A, $BA = A$ and $AC = C$. Show that $B = C$.

3. Show that if A and B are $m \times n$ matrices and c is a scalar, then
$$(cA + B)^T = cA^T + B^T.$$

4. (a) Show that for any two 2×2 matrices A and B, the transpose of AB is
$$(AB)^T = B^T A^T.$$

   (b) Can you prove the analogue for 3×3 matrices? (*Hint:* Think of the formula for the ij entry of AB as the product of a row of A and a column of B; reinterpret this as a dot product of vectors.)

5. Prove Proposition E.3.1.

6. Suppose the 3×3 matrix A has the property that whenever $\overrightarrow{v}$ is a nonzero vector, then $A\overrightarrow{v} \neq \overrightarrow{0}$.

   (a) Show that the vectors $A\overrightarrow{\imath}$, $A\overrightarrow{\jmath}$ and $A\overrightarrow{k}$ are linearly independent. (*Hint:* If $aA\overrightarrow{\imath} + bA\overrightarrow{\jmath} + cA\overrightarrow{k} = \overrightarrow{0}$, show that $a\overrightarrow{\imath} + b\overrightarrow{\jmath} + c\overrightarrow{k}$ is a vector whose image under multiplication by A is zero. Conclude that $a = b = c = 0$.)

   (b) Show that the columns of a nonsingular matrix are linearly independent.

(c) Use this to show that the vector equation $A\overrightarrow{x} = \overrightarrow{y}$ has a solution $\overrightarrow{x}$ for every $\overrightarrow{y}$.

7. Suppose A is a 3×3 singular matrix; specifically, suppose $\overrightarrow{x}_0 = a\overrightarrow{\imath} + b\overrightarrow{\jmath} + c\overrightarrow{k}$ is a nonzero vector with $A\overrightarrow{x}_0 = \overrightarrow{0}$.

   (a) Show that
   $$aA_1 + bA_2 + cA_3 = \overrightarrow{0}$$
   where A_j is the j^{th} column of A.

   (b) Show that, if $c \neq 0$, then $A\overrightarrow{k}$ can be expressed as a linear combination of $A\overrightarrow{\imath}$ and $A\overrightarrow{\jmath}$; in general, at least one of the vectors $A\overrightarrow{\imath}$, $A\overrightarrow{\jmath}$ and $A\overrightarrow{k}$ can be expressed as a linear combination of the other two.

   (c) In particular, if $c \neq 0$, show that the image $A\overrightarrow{x}$ of *any* vector $\overrightarrow{x} \in \mathbb{R}^3$ lies in the plane spanned by $A\overrightarrow{\imath}$ and $A\overrightarrow{\jmath}$.

8. Suppose that
   $$a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2 + a_3\overrightarrow{v}_3 = \overrightarrow{0}$$
   is a dependency relation among the vectors $\overrightarrow{v}_i$. Show that if $a_i \neq 0$ then $\overrightarrow{v}_i$ can be expressed as a linear combination of the other two vectors.

9. Show that any collection of vectors which includes the zero vector is linearly dependent.

10. Show that the nonzero rows of a matrix in reduced row-echelon form are linearly independent. (*Hint:* Show that, in any linear combination of those rows, the entry in the space corresponding to the leading entry of a given row equals the coefficient of that row in the combination.)

11. (a) Suppose A is a matrix in reduced row-echelon form. Show that the columns containing the leading entries in A are linearly independent, and that every other column of A is a linear combination of these.

    (b) Suppose
    $$a_1 \operatorname{col}_1(A) + a_2 \operatorname{col}_2(A) + a_3 \operatorname{col}_3(A) = \overrightarrow{0}$$
    is a dependency relation among the columns of A, and A' is obtained from A via a row operation. Show that the same relation holds among the columns of A'.

12. In this problem, you will show that the rank of a matrix is not changed by row operations.

    (a) Suppose A and A' are matrices whose rows are the same, but in different order (that is, A' is obtained from A by shuffling the rows). Show that A and A' have the same rank.

    (b) Suppose the vectors $\overrightarrow{v}_1$, $\overrightarrow{v}_2$ and $\overrightarrow{v}_3$ are linearly independent (*resp.* linearly dependent). Show that the same is true of the vectors $c_1\overrightarrow{v}_1$, $c_2\overrightarrow{v}_2$, and $c_3\overrightarrow{v}_3$, where c_i, $i = 1, 2, 3$, are any nonzero real numbers. This shows that multiplying a row of A by a nonzero number does not change the rank.

    (c) Suppose
    $$a_1\overrightarrow{v}_1 + a_2\overrightarrow{v}_2 + a_3\overrightarrow{v}_3 = \overrightarrow{0}$$
    is a dependency among the rows of A. Suppose $\overrightarrow{v}'_1 = \overrightarrow{v}_1 + b\overrightarrow{v}_2$ (*i.e.*, the first row is replaced by itself plus a multiple of another row). Assuming $\overrightarrow{v}_2 \neq \overrightarrow{0}$, find a nontrivial dependency relation among $\overrightarrow{v}'_1$, $\overrightarrow{v}_2$ and $\overrightarrow{v}_3$. This shows that the rank cannot *increase* under such a row operation.

    (d) Using the fact that the third row operation is reversible, show that it does not change the rank of A.

Challenge problems:

13. In this exercise, you will prove Proposition E.4.5. Suppose that A is $m \times n$ with rank r, and fix an m-column b.

    (a) Show that $\text{rank}(A) \leq \min m, n$.

    (b) Show that $\text{rank}(A) \leq \text{rank}([A|b])$.

    (c) Show that a leading entry appears in the last column of the reduced form of $[A|b]$ if and only if $\text{rank}([A|b]) = \text{rank}(A) + 1$. Conclude that the system with augmented matrix $[A|b]$ is consistent if and only if $\text{rank}([A|b]) = \text{rank}(A)$.

    (d) Show that the number of columns of $[A|b]$ not containing a leading entry equals $n - r$.

    (e) Show that if $r = m$ then $\text{rank}([A|b]) = \text{rank}(A)$.

    (f) Show that if $r < m$ then there is a choice of b for which $\text{rank}([A|b]) > \text{rank}(A)$. (*Hint:* First prove this assuming $[A|b]$ is in reduced row-echelon form. Then, reverse the row operations that went from $[A|b]$ to its reduced form to prove it in general.)

14. Suppose A is a 3×3 matrix.

    (a) Show that if $A'A = I$ for some 3×3 matrix A', then the transformation $\overrightarrow{x} \mapsto A\overrightarrow{x}$ is onto.

    (b) Show that if $AA'' = I$ for some 3×3 matrix A'', then the transformation $\overrightarrow{x} \mapsto A\overrightarrow{x}$ is one-to-one.

    (c) Show that if A' and A'' both exist, then they must be equal. (*Hint:* Parse the product $A'AA''$ two different ways.)

15. Prove Proposition E.3.6 as follows:

    (a) By Proposition E.3.3, if A is nonsingular, then $\overrightarrow{x} \mapsto A\overrightarrow{x}$ is both one-to-one and onto. Show that if A is singular, then A is neither one-to-one nor onto. (*Hint:* Use the definition for one property, and Exercise 7 for the other.)

    (b) Use Exercise 14 to show that if A is invertible then $\overrightarrow{x} \mapsto A\overrightarrow{x}$ is one-to-one and onto.

    (c) Suppose $\overrightarrow{x} \mapsto A\overrightarrow{x}$ is one-to-one and onto. Then show that there is a unique, well-defined transformation which takes any vector $\overrightarrow{y}$ to the unique solution $\overrightarrow{x}$ of the equation $A\overrightarrow{x} = \overrightarrow{y}$. Show that this transformation is linear—that is, that if $A\overrightarrow{x}_1 = \overrightarrow{y}_1$ and $A\overrightarrow{x}_2 = \overrightarrow{y}_2$ then for $\overrightarrow{y} = \alpha_1\overrightarrow{x}_1 + \alpha_2\overrightarrow{x}_2$, the unique solution of $A\overrightarrow{x} = \overrightarrow{y}$ is $\overrightarrow{x} = \alpha_1\overrightarrow{x}_1 + \alpha_2\overrightarrow{x}_2$.

    (d) Suppose $\overrightarrow{x} = \overrightarrow{a}_i$ is the unique solution of $A\overrightarrow{x} = \overrightarrow{e}_i$, where $\overrightarrow{e}_i$, $i = 1, 2, 3$ are the standard basis for $\mathbb{R}^3$. Let B be the 3×3 matrix with columns $[\overrightarrow{a}_i]$. Show that the transformation $\overrightarrow{y} \mapsto B\overrightarrow{y}$ is the same as the transformation defined in the preceding item, and conclude that B is the matrix inverse of A. It follows that if the transformation $\overrightarrow{x} \mapsto A\overrightarrow{x}$ is one-to-one and onto, then A is invertible.

16. Show that a 3×3 matrix A for which the transformation $\overrightarrow{x} \mapsto A\overrightarrow{x}$ is *either* one-to-one *or* onto is *both* one-to-one *and* onto. Conclude that if the equation

$$AB = I$$

has a solution, then A is invertible, and $B = A^{-1}$.

F

Determinants

In § 1.6, we defined a 2×2 determinant as shorthand for the calculation of the signed area of a parallelogram, and then a 3×3 determinant as a formal calculation for the cross product (which in turn determines an oriented area); subsequently, we saw that a numerical determinant can be interpreted as the signed volume of a parallelipiped. In this appendix, we see how these interpretations as well as others can be used to establish some basic properties of 2×2 and 3×3 determinants which are very useful (and easy to use), but whose proofs are not so obvious. Our approach to these will go back and forth between formal calculation and geometric interpretation. We begin by warming up with the 2×2 case.

F.1 2×2 Determinants

Recall that the determinant of a 2×2 matrix is given by the formula

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{F.1}$$

In words, we multiply each of the two entries in the first row of the matrix by the (one) entry which is *neither* in the same *row*, *nor* in the same *column* as the entry, and then we subtract. We shall see later that this way of phrasing it extends to two different ways of seeing 3×3 determinants.

Our original definition of a 2×2 determinant really views it as a function $\Delta\left(\overrightarrow{v},\overrightarrow{w}\right)$ of its rows, regarded as vectors, and from the formulation above we immediately get the first basic algebraic properties of this function (as detailed in Proposition 1.6.2):

Remark F.1.1. *The function $\Delta\left(\overrightarrow{v},\overrightarrow{w}\right)$ has the following properties:*

additivity: $\Delta\left(\overrightarrow{v}_1+\overrightarrow{v}_2,\overrightarrow{w}\right)=\Delta\left(\overrightarrow{v}_1,\overrightarrow{w}\right)+\Delta\left(\overrightarrow{v}_2,\overrightarrow{w}\right)$*, and* $\Delta\left(\overrightarrow{v},\overrightarrow{w}_1+\overrightarrow{w}_2\right)=\Delta\left(\overrightarrow{v},\overrightarrow{w}_1\right)+\Delta\left(\overrightarrow{v},\overrightarrow{w}_2\right)$*;*

scaling: $\Delta\left(\alpha\overrightarrow{v},\overrightarrow{w}\right)=\alpha\Delta\left(\overrightarrow{v},\overrightarrow{w}\right)=\Delta\left(\overrightarrow{v},\alpha\overrightarrow{w}\right)$*;*

antisymmetry: $\Delta\left(\overrightarrow{w},\overrightarrow{v}\right)=-\Delta\left(\overrightarrow{v},\overrightarrow{v}\right)$*.*

There is another way to write the formula (F.1):

$$a_{11}a_{22}-a_{12}a_{21}=(-1)^{1+1}a_{11}a_{22}+(-1)^{1+2}a_{12}a_{21};$$

that is, we attach to each entry of the first row a sign (depending on whether the sum of its indices is even or odd), then multiply this entry by this sign times the (unique) entry in the other row and other column. One way to remember the sign attached to a given position in the matrix is to note that the signs form a checkerboard pattern, starting with "+" in the upper left corner:

$$\begin{vmatrix} + & - \\ - & + \end{vmatrix}.$$

The advantage of this notation becomes apparent when we note that replacing the exponent for each product by the sum of indices in the *second* factor, nothing changes:

$$(-1)^{1+1}a_{11}a_{22}+(-1)^{1+2}a_{12}a_{21}=(-1)^{2+2}a_{11}a_{22}+(-1)^{2+1}a_{12}a_{21};$$

this corresponds to taking each entry a_{2i} of the *second* row and multiplying it by the sign $(-1)^{2+i}$ times the entry in the other row and column, and summing over *this* row. Similarly, if we followed this process not along a *row* but along a *column*—say the first–we would be looking at a_{11}, the first entry of this column, times the sign $(-1)^{1+1}$ attached to it, times the entry a_{22} in the other row and column, plus the *second* entry of this *column*, a_{21}, times the sign $(-1)^{2+1}$ attached to *it*, times the entry a_{12} in the other row and column.

To pull this together, we extend the idea of a **cofactor** to each entry of the matrix (not just in the first row): the cofactor of the entry a_{ij} in a matrix A is the sign $(-1)^{i+j}$ attached to its position in A, times its minor, which we denote A_{ij}, but which here is just the entry in the other row and column from a_{ij}. This leads us to several variations on how to calculate a 2×2 determinant:

Remark F.1.2 (Expansion by cofactors for a 2×2 determinant)**.** *For the entry a_{ij} in row i and column j of the 2×2 matrix A, define its **cofactor** to be $(-1)^{i+j}A_{ij}$, where A_{ij} is the (unique) entry of A which is* not *in row i and* not *in column j.*

Then det A *is the sum of the entries in any one row or column of A, each multiplied by its cofactor.*

As an immediate consequence of this, we also see that taking a transpose does not change the determinant:

Corollary F.1.3 (Determinant of transpose)**.** *The determinant of any 2×2 matrix and that of its transpose are equal:* $\det A^T = \det A$.

This is because expanding along a *row* of A^T is the same as expanding along the corresponding *column* of A.

F.2 3×3 Determinants

In § 1.6, we defined the determinant of a 3×3 matrix in terms of expansion by cofactors of entries in the first row: if

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{F.2}$$

then

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} \\ &= \sum_{j=1}^{3} (-1)^{1+j} a_{1j} \det A_{1j}. \end{aligned}$$

where

$$A_{11} = \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & a_{22} & a_{23} \\ \cdot & a_{32} & a_{33} \end{pmatrix}$$
$$A_{12} = \begin{pmatrix} \cdot & \cdot & \cdot \\ a_{21} & \cdot & a_{23} \\ a_{31} & \cdot & a_{33} \end{pmatrix}$$
$$A_{13} = \begin{pmatrix} \cdot & \cdot & \cdot \\ a_{21} & a_{22} & \cdot \\ a_{31} & a_{32} & \cdot \end{pmatrix}.$$

We shall see that there are several other ways to think of this calculation, but to do so we will first go through a rather baroque exercise. The full formula for det A in terms of the entries is a sum of six triple products: each product involves an entry from each row (and at the same time one from each column), preceded by a sign. Let us write this out, putting the factors in each triple product in the order of their rows, that is, each triple product is written in the form $\pm a_{1j_1}a_{2j_2}a_{3j_3}$:

$$\begin{aligned}\det A &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} \\ &\quad - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ &\qquad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.\end{aligned}$$

We will refer to this formula in our discussion as the *grand sum*. Now in each triple product, the *second* indices include one each of $1, 2, 3$, in a different order for each product. Let us look at just these orders and the signs associated to them: writing $\pm(j_1j_2j_3)$ in place of $\pm a_{1j_1}a_{2j_2}a_{3j_3}$, we have the pattern

$$+(123) - (132) - (213) + (231) + (312) - (321).$$

What is the pattern here? One way to get at it is to consider the *order reversals* in each list—that is, we ask of each list

1. Does 1 come earlier than 2?

2. Does 1 come earlier than 3?

3. Does 2 come earlier than 3?

Each "no" answer counts as an order reversal. Then you should check that the signs above are determined by the rule: *each sign reversal contributes a factor of* -1 *to the sign in front of the list.* This means a list has a "plus" (*resp.* "minus") sign if it involves an even (*resp.* odd) number of order reversals. For example, in the list (213), the *first* question is answered "*no*" but the other two are answered "yes": there is *one* order reversal, and so this list is preceded by a *minus.* By contrast, in the list (312) the first question is answered "yes" but the other *two* are answered "*no*": there are *two* order reversals, and this list is preceded by a *plus.*

Armed with this pattern, we can justify a variety of other ways to calculate a 3 × 3 determinant.

Each entry a_{ij} of the matrix appears in two triple products, formed by picking one entry from each of the rows different from i, in such a way that they come from distinct *columns* (both different from j). In other words, these two triple products can be written as a_{ij} times the product of either the *diagonal* or the *antidiagonal* entries of the submatrix A_{ij} of A, formed by eliminating row i and column j We call the A_{ij} **minor** of a_{ij}, extending our terminology and notation on p. 89. The relative order of the column numbers other than j in these two products will be correct in one case and reversed in the other. In both cases, their order *relative to* j will be *the same* if j comes first ($i = 1$) or last ($i = 3$). If j is in the middle ($i = 2$), then we can compare the two patterns (ajb) and (bja): whatever the order of a (*resp.* b) relative to j in the first pattern, it will be reversed in the second; thus the number of order reversals relative to j will have the *same parity* in both cases. From this we see that the combined contribution to the grand sum of the two triple products that include a_{ij} will be, up to sign, the product $a_{ij} \det A_{ij}$.

To determine the sign attached to this product, we need to deduce the number of order reversals in the pattern with j in the i^{th} position and the other two in their correct relative order. First, any of the patterns ($1ab$), ($a2b$) and ($ab3$) with $a < b$ is (123) (*no* order reversals), while ($3ab$) (*resp.* ($ab1$)) is (312) (*resp.* (231)) (*two* reversals): thus, if i and j have the *same* parity, the sign is *plus.* Second, ($2ab$) (*resp.* ($ab2$), ($a1b$), ($a3b$)) is (213) (*resp.* (132), (213), (132)) (one reversal), so if i and j have *opposite* parity, the sign is *minus.* This can be summarized as saying

Remark F.2.1. *The two terms in the grand sum which include* a_{ij} *combine as* $(-1)^{i+j} a_{ij} \det A_{ij}$.

The pattern of signs associated to various positions in the matrix can be visualized as a checkerboard of + and − signs with + in the upper right corner:

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}.$$

We define the **cofactor** of a_{ij} in A to be

$$\text{cofactor}(ij) = (-1)^{i+j} \det A_{ij}.$$

If we choose any row or column of A, the grand sum can be interpreted as the sum of entries in that row or column times their cofactors:

$$\begin{aligned} \det A &= \sum_{j=1}^{3} (-1)^{i+j} a_{ij} \det A_{ij} \\ &= \sum_{j=1}^{3} a_{ij} \cdot \text{cofactor}(ij) \qquad \text{for row } i \\ \det A &= \sum_{i=1}^{3} (-1)^{i+j} a_{ij} \det A_{ij} \\ &= \sum_{i=1}^{3} a_{ij} \cdot \text{cofactor}(ij) \qquad \text{for column } j. \end{aligned}$$

This is called the **expansion by cofactors** of det A along that row or column.

As in the 2×2 case, expansion of the determinant of det A^T along its first *column* is the same as expansion of det A along its first *row*:

Corollary F.2.2 (Determinant of transpose)**.** *The determinant of any* 3×3 *matrix and that of its transpose are equal:* $\det A^T = \det A$.

There is an alternative way to visualize the grand sum, analogous to the 2×2 case (Exercise 1).

We can regard the determinant as a function of the three rows of A, by analogy with the treatment of 2×2 determinants in the preceding subsection: given vectors

$$\begin{aligned} \overrightarrow{u} &= (u_1, u_2, u_3) \\ \overrightarrow{v} &= (v_1, v_2, v_3) \\ \overrightarrow{w} &= (w_1, w_2, w_3) \end{aligned}$$

we set

$$\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right) := \det\begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_{33} \\ w_1 & w_2 & w_3 \end{pmatrix}.$$

This function of three vector variables has properties analogous to those formulated in Remark F.1.1 for 2 × 2 determinants:

Remark F.2.3. *The function $\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right)$ has the following properties:*

additivity:

$$\begin{aligned}\Delta\left(\overrightarrow{u}_1+\overrightarrow{u}_2,\overrightarrow{v},\overrightarrow{w}\right) &= \Delta\left(\overrightarrow{u}_1,\overrightarrow{v},\overrightarrow{w}\right)+\Delta\left(\overrightarrow{u}_2,\overrightarrow{v},\overrightarrow{w}\right)\\ \Delta\left(\overrightarrow{u},\overrightarrow{v}_1+\overrightarrow{v}_2,\overrightarrow{w}\right) &= \Delta\left(\overrightarrow{u},\overrightarrow{v}_1,\overrightarrow{w}\right)+\Delta\left(\overrightarrow{u},\overrightarrow{v}_2,\overrightarrow{w}\right)\\ \Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}_1+\overrightarrow{w}_2\right) &= \Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}_1\right)+\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}_2\right);\end{aligned}$$

scaling:

$$\Delta\left(\alpha\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right)=\Delta\left(\overrightarrow{u},\alpha\overrightarrow{v},\overrightarrow{w}\right)=\Delta\left(\overrightarrow{u},\overrightarrow{v},\alpha\overrightarrow{w}\right)=\alpha\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right);$$

alternating: *interchanging any pair of inputs reverses the sign of Δ:*

$$\begin{aligned}\Delta\left(\overrightarrow{v},\overrightarrow{u},\overrightarrow{w}\right) &= -\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right)\\ \Delta\left(\overrightarrow{u},\overrightarrow{w},\overrightarrow{v}\right) &= -\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right)\\ \Delta\left(\overrightarrow{w},\overrightarrow{v},\overrightarrow{u}\right) &= -\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right).\end{aligned}$$

The additivity and scaling properties in a given row are obvious if we use expansion by cofactors along that row; a function of three variables with additivity and scaling in each variable is called a **trilinear function**. To see the alternating property, we note first that interchanging the second and third rows amounts to interchanging the rows of each minor A_{1j} of entries in the first row, which reverses the sign of each cofactor det A_{1j}. Thus, using expansion by cofactors along the first row, interchanging the second and third rows of A reverses the sign of the determinant det A. But a similar argument applied to *any* row of A shows that interchanging the *other two* rows of A reverses the sign of det A.

Invoking Remark 6.8.3, we can characterize the function $\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right)$:

Remark F.2.4. *The* 3 × 3 *determinant $\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right)$ is the unique alternating trilinear function on $\mathbb{R}^3$ satisfying*

$$\Delta\left(\overrightarrow{\imath},\overrightarrow{\jmath},\overrightarrow{k}\right)=1.$$

A different characterization of this function is given in Remark 1.7.2: $\Delta\left(\overrightarrow{u},\overrightarrow{v},\overrightarrow{w}\right)$ gives the signed volume of the parallelepiped whose sides are $\overrightarrow{u}$, $\overrightarrow{v}$ and $\overrightarrow{w}$.

F.3 Determinants and Invertibility

Using the geometric interpretation of a 2×2 determinant as a signed area (Proposition 1.6.1), we saw (Corollary 1.6.3) that a 2×2 determinant is nonzero precisely if its rows are linearly independent. Since (by an easy calculation) transposing a 2×2 matrix does not change its determinant, a 2×2 determinant is nonzero precisely if the *columns* are linearly independent, and this in turn says that the underlying matrix is nonsingular. Thus we have the observation

Remark F.3.1. *For any 2×2 matrix A, the following are equivalent:*

1. $\det A \neq 0$.

2. *The rows of A are linearly independent.*

3. *The columns of A are linearly independent.*

4. *A is nonsingular.*

The 3×3 analogue of Remark F.3.1 takes slightly more work, but we have most of the ingredients in place. Remark 4 tells us that the first two properties are equivalent. Remark E.4.4 then tells us that the second and third properties are equivalent. Finally, Proposition 2 tells us that (for a square matrix) that the third and fourth properties are equivalent. In fact, an elaboration of this argument can be applied to a square matrix of any size, although some of the details involve ideas that are beyond the scope of this book.

Exercises for Appendix F

1. Consider the following calculation: given a 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

write down A and next to it a second copy of its first two columns:

$$\left[\begin{array}{ccc|cc} a_{11} & \mathbf{a_{12}} & a_{13} & a_{11} & \boxed{a_{12}} \\ a_{21} & a_{22} & \mathbf{a_{23}} & \boxed{a_{21}} & a_{22} \\ a_{31} & a_{32} & \boxed{a_{33}} & \mathbf{a_{31}} & a_{32} \end{array}\right].$$

There are three "downward diagonals" in this picture, one starting from each entry in the first row of A; we have put the downward diagonal starting from the second entry in boldface. There are also three "upward diagonals", starting from the entries in the last row of A; we have framed the entries in the upward diagonal starting from the third entry. Then by analogy with 2×2 determinants, consider the sum of the products along the three downward diagonals minus the sum of products along the upward ones. Verify that this agrees with the grand sum giving det A in the text. [1]

2. Suppose $F(\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w})$ is an alternating trilinear function.

   (a) Show that if two of these vectors are equal then $F(\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w}) = 0$

   (b) Show that an alternating trilinear function applied to three linearly dependent vectors must equal zero. (*Hint:* Use Exercise 8 in Appendix E together with additivity and homogeneity to rewrite $F(\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w})$ as a sum of multiples of terms $F(\overrightarrow{v}_1, \overrightarrow{v}_2, \overrightarrow{v}_3)$ in each of which two entries are equal.)

[1] However, unlike expansion by cofactors, this procedure breaks down for larger determinants.

G

Surface Area: the Counterexample of Schwarz and Peano

We present here an example, due to H. A. Schwarz [45] and G. Peano [42][1] which shows that the analogue for surfaces of the usual definition of the length of a curve cannot work.

Recall that if a curve $\mathcal{C}$ was given as the path of a moving point $\overrightarrow{p}(t)$, $a \leq t \leq b$, we partitioned $[a, b]$ via $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_n\}$ and approximated $\mathcal{C}$ by a piecewise-linear path consisting of the line segments $[\overrightarrow{p}(t_{i-1}), \overrightarrow{p}(t_i)]$, $i = 1, \ldots, n$. Since a straight line is the shortest distance between two points, the distance travelled by $\overrightarrow{p}(t)$ between $t = t_{i-1}$ and $t = t_i$ is at least the length of this segment, which is $\|\overrightarrow{p}(t_i - \overrightarrow{p}(t_{i-1}))\|$. Thus, the total length of the piecewise-linear approximation is a lower bound

[1] Schwarz tells the story of this example in a note [46, pp. 369-370]. Schwarz initially wrote down his example in a letter to one Angelo Gnocchi in December 1880, with a further clarification in January 1881. In May, 1882 Gnocchi wrote to Schwarz that in a conversation with Peano, the latter had explained that Serret's definition of surface area (to which Schwarz's example is a counterexample) could not be correct, giving detailed reasons why it was wrong; Gnocchi had then told Peano of Schwarz's letters. Gnocchi reported the example in the Notices of the Turin Academy, at which point it came to the attention of Charles Hermite (1822-1901), who publihsed the correspondence in his *Cours d'analyse*. Meanwhile, Peano published his example. After seeing Peano's article, Schwarz contacted him and learned that Peano had independently come up with the same example in 1882.

for the length of the actual path: we say $\mathcal{C}$ is *rectifiable* if the supremum of the lengths of all the piecewise-linear paths arising from different partitions of $[a, b]$ is finite, and in that case define the (arc)length of the curve to be this supremum. We saw in § 2.5 that every regular arc (that is, a curve given by a one-to-one differentiable parametrization with non-vanishing velocity) the length can be calculated from any regular parametrization as

$$\mathfrak{s}\left(\mathcal{C}\right) = \int_a^b \left\| \dot{\overrightarrow{p}}(t) \right\| \, dt.$$

The analogue of this procedure could be formulated for surface area as follows. [2] Let us suppose for simplicity that a surface $\mathfrak{S}$ in $\mathbb{R}^3$ is given by the parametrization $\overrightarrow{p}(s, t)$, with domain a rectangle $[a, b] \times [c, d]$. If we partition $[a, b] \times [c, d]$ as we did in § 5.1, we would like to approximate $\mathfrak{S}$ by rectangles in space whose vertices are the images of "vertices" $p_{i,j} = \overrightarrow{p}(x_i, y_j)$ of the subrectangles R_{ij}. This presents a difficulty, since *four* points in $\mathbb{R}^3$ need not be coplanar. However, we can easily finesse this problem if we note that *three* points in $\mathbb{R}^3$ are *always* contained in some plane. Using diagonals (see Figure G.1)[3] we can divide each subrectangle R_{ij} into two triangles, say

$$\begin{aligned} T_{ij1} &= \triangle p_{i-1,j-1} p_{i,j-1} p_{i,j} \\ T_{ij2} &= \triangle p_{i-1,j-1} p_{i-1,j} p_{i,j}. \end{aligned}$$

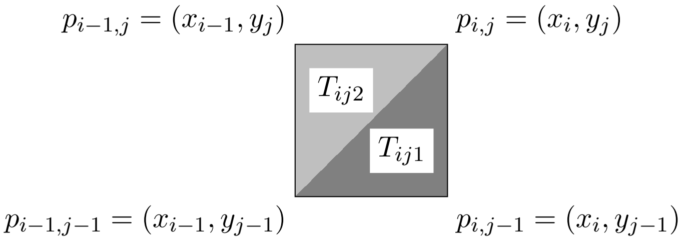

$p_{i-1,j} = (x_{i-1}, y_j)$
$p_{i,j} = (x_i, y_j)$
T_{ij2}
T_{ij1}
$p_{i-1,j-1} = (x_{i-1}, y_{j-1})$
$p_{i,j-1} = (x_i, y_{j-1})$

Figure G.1: Dividing R_{ij} into triangles

This tiling $\{T_{ijk} \,|\, i = 1, \ldots, m, \quad j = 1, \ldots, n, \quad k = 1, 2\}$ of the rectangle $[a, b] \times [c, d]$ is called a **triangulation**. Now, it would be natural to try to look at the total of the areas of the triangles whose vertices are the points

[2] This approach was in fact followed by J. M. Serret

[3] There are two ways to do divide each rectangle, but as we shall see, this will not prove to be an issue.

$\overrightarrow{p^*}_{i,j} = \overrightarrow{p}(p_{i,j})$ on the surface

$$T^*_{ij1} = \triangle \overrightarrow{p^*}_{i-1,j-1}\overrightarrow{p^*}_{i,j-1}\overrightarrow{p^*}_{i,j}$$
$$T^*_{ij2} = \triangle \overrightarrow{p^*}_{i-1,j-1}\overrightarrow{p^*}_{i-1,j}\overrightarrow{p^*}_{i,j}.$$

and define the area of $\mathfrak{S}$ to be the supremum of these over all triangulations of $[a,b] \times [c,d]$.

Unfortunately, this approach doesn't work; an example found (simultaneously) in 1892 by H. A. Schwartz and G. Peano shows

Proposition G.0.2. *There exist triangulations for the standard parametrization of the cylinder such that the total area of the triangles is arbitrarily high.*

Proof. Consider the finite cylinder surface

$$x^2 + y^2 = 1$$
$$0 \leq z \leq 1.$$

We partition the interval $[0,1]$ of z-values into m equal parts using the $m+1$ horizontal circles $z = \frac{k}{m}$, $k = 0, \dots, m$. Then we divide each circle into n equal arcs, but in such a way that the endpoints of arcs on any particular circle are directly above or below the midpoints of the arcs on the neighboring circles. One way to do this is to define the angles

$$\theta_{jk} = \begin{cases} \frac{2\pi j}{n} & \text{for } k \quad \textit{even}, \\ \frac{2\pi j}{n} - \frac{\pi}{n} & \text{for } k \ \textit{odd} \end{cases}$$

and then the points

$$p_{jk} = (\cos\theta_{jk}, \sin\theta_{jk}, \frac{k}{m}).$$

That is, the points $\{p_{jk}\}$ for k fixed and $j = 1, \dots, n$ divide the k^{th} circle into n equal arcs. Now consider the triangles whose vertices are the endpoints of an arc and the point on a neighboring circle directly above or below the midpoint of that arc (Figure G.2).

The resulting triangulation of the cylinder is illustrated in Figure G.3.

To calculate the area of a typical triangle, we note first (Figure G.4) that its base is a chord of the unit circle subtending an arc of $\triangle\theta = \frac{2\pi}{n}$ radians; it follows from general principles that the length of the chord is

$$b_n = 2\sin\frac{\triangle\theta}{2} = 2\sin\frac{\pi}{n}.$$

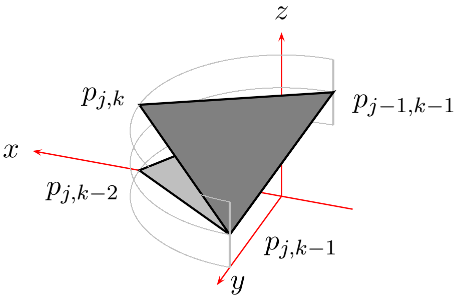


Figure G.2: Typical Triangles

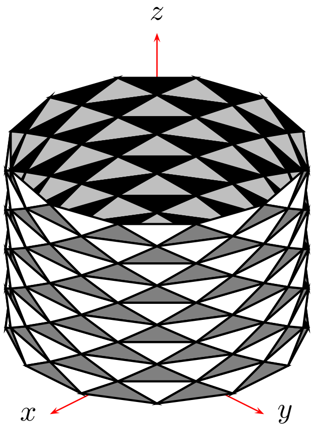


Figure G.3: Triangulation of the Cylinder

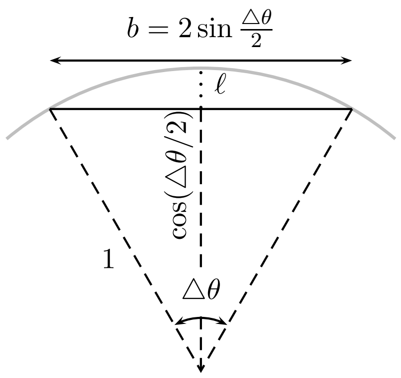

$b = 2\sin\frac{\triangle\theta}{2}$
ℓ
$\cos(\triangle\theta/2)$
1
$\triangle\theta$

Figure G.4: The Base of a Typical Triangle

We also note for future reference that the part of the perpendicular bisector of the chord from the chord to the circle has length

$$\ell = 1 - \cos\frac{\triangle\theta}{2} = 1 - \cos\frac{\pi}{n}.$$

To calculate the height of a typical triangle, we note that the vertical distance between the plane containing the base of the triangle and the other vertex is $\frac{1}{m}$, while the distance (in the plane containing the base) from the base to the point below the vertex (the dotted line in Figure G.5) is $\ell = 1 - \cos\frac{\pi}{n}$; it follows that the height of the triangle (the dashed line in Figure G.5) is itself the hypotenuse of a right triangle with sides ℓ and $\frac{1}{m}$, so

$$h_{m,n} = \sqrt{\left(\frac{1}{m}\right)^2 + \ell^2} = \sqrt{\left(\frac{1}{m}\right)^2 + \left(1 - \cos\frac{\pi}{n}\right)^2}.$$

Thus the area of a single triangle of our triangulation (for a given choice

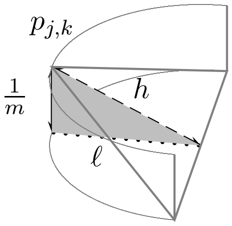


Figure G.5: Calculating the Height of a Typical Triangle

of m and n) is

$$\begin{aligned}\triangle A_{m,n} &= \frac{1}{2}b_n h_{m,n}\\ &= \frac{1}{2}\left[2\sin\frac{\pi}{n}\right]\sqrt{\left(\frac{1}{m}\right)^2+\left(1-\cos\frac{\pi}{n}\right)^2}\\ &= \left[\sin\frac{\pi}{n}\right]\sqrt{\left(\frac{1}{m}\right)^2+\left(1-\cos\frac{\pi}{n}\right)^2}.\end{aligned}$$

Now let us count the number of triangles. There are $m+1$ horizontal circles, each cut into n arcs, and the chord subtending each arc is the base of exactly two triangles of our triangulation, except for the two "end" circles, for which each chord is the base of *one* triangle. This means there are $2mn$ triangles, giving a total area of

$$\begin{aligned}A_{m,n} &= 2mn\triangle Am, n\\ &= 2mn\left[\sin\frac{\pi}{n}\right]\sqrt{\left(\frac{1}{m}\right)^2+\left(1-\cos\frac{\pi}{n}\right)^2}\\ &= 2\left[n\sin\frac{\pi}{n}\right]\sqrt{1+m^2\left(1-\cos\frac{\pi}{n}\right)^2}.\end{aligned}$$

Now, the quantity in brackets converges to π as $n \to \infty$ and, for n fixed, the square root goes to ∞ as $m \to \infty$; it follows that we can fix a sequence of pairs of values $\{(m_k, n_k)\}$ (for example, as Schwarz suggests, $m = n^3$) such that the quantity A_{m_k,n_k} diverges to infinity, establishing that the supremum of the total area of piecewise-linear approximations of a cylinder is infinite, and hence gives a bad definition for the area of the cylinder itself. □

To see what is going on here, we might note that at the vertex opposite the chord of each triangle, the plane tangent to the cylinder is vertical, while the plane of the triangle makes an angle with it of size $\theta(m,n)$, where

$$\tan\theta(m,n) = \frac{\ell}{1/m} = \frac{1-\cos\frac{\pi}{n}}{1/m} = m\left(1-\cos\frac{\pi}{n}\right).$$

If for example $m = n^3$, one can check (using, *e.g.*, L'Hôpital's rule) that the tangent goes to infinity with n, so in the limit the triangles approach being *perpendicular* to the cylinder.

It turns out that the approach of Serret (using these triangulations) can be made to work, provided we replace the *supremum* of all such total areas with the *limit*, as $\varepsilon \to 0$, of the *infimum* of all such total areas for triangulations with mesh size less than ε. In effect, this means we are looking at triangulations in which the triangles are as close as possible to being tangent to the cylinder.

Bibliography

[1] Tom Archibald. Analysis and physics in the nineteenth century: The case of boundary-value problems. In Hans Niels Jahnke, editor, *A History of Analysis*, pages 197–211. AMS, 2003. 6

[2] Archimedes. Measurement of a circle. In *The Works of Archimedes* [26], pages 91–98. Dover reprint, 2002. 6a, 2.5

[3] Archimedes. On spirals. In *The Works of Archimedes* [26], pages 151–188. Dover reprint, 2002. 2.1, 7, 2.2

[4] Carl B. Boyer. *History of Analytic Geometry.* Yeshiva University, 1956. Dover reprint, 2004. 1.4

[5] David M. Burton. *The History of Mathematics.* McGraw-Hill, 5th edition, 2003. 16

[6] Sandro Caparrini. The discovery of the vector representation of moments and angular velocity. *Archive for History of the Exact Sciences*, 56:151–181, 2002. 3

[7] Sandro Caparrini. Early theories of vectors. In M. Corradi, A. Becchi, and F. Foce, editors, *Between Mechanics and Architecture: The Work of Clifford Ambrose Truesdell and Edoardo Benvenuto*, pages 173–193. Birkhauser, 2003. Proceedings of an International Symposium 30 November-1 December 2001, Genoa, Italy. 3

[8] Sandro Caparrini. The theory of vectors at the beginning of the nineteenth century. In C. Alvarez, J. Rafael Martinez, P. Radelet de Grave, and J. Lacki, editors, *Variar para encontrar. Varier pour mieux trouver. The Lore of Variation: Finding Pathways to Scientific Knowledge*, pages 235–257. Universidad Nacional Autnoma de Mxico, Universitas Catholica Lovaniensis, Universit de Genve, 2004. 3

[9] I. Bernard Cohen and Anne Whitman. *Isaac Newton:The Principia (Mathematical Principles of Natural Philosophy), A new translation.*

Univ. of California Press, 1999. Translation from the Latin by I. Bernard Cohen and Anne Whitman of *Philosophiae Naturalis Principia Mathematica* (1687, 1713, 1726). 1, 39

[10] Michael J. Crowe. *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System.* Notre Dame University Press, 1967. Dover Reprint, 1994. 3

[11] C. H. Edwards. *Advanced Calculus of Several Variables.* Academic Press, 1973. reprinted by Dover publications. 10

[12] Steven B. Engelsman. *Families of Curves and the Origins of Partial Differentiation*, volume 93 of *North-Holland Mathematical Studies.* North-Holland, 1984. 14

[13] Howard Eves. *An Introduction to the History of Mathematics.* Saunders College Publishing, 5 edition, 1983. 15, 6, 7, 8

[14] Fréderic-Jean Frénet. Sur les courbes à double courbure. *Journal de Mathématiques*, 17:437–447, 1852. D

[15] Michael N. Fried and Sabetai Unguru. *Apollonius of Perga's Conica: Text, Context, Subtext.* Number 222 in Mnemosyne: Bibliotheca Classica Batava. Brill, 2001. 5, 2

[16] Guido Fubini. Sugli integrali multipli. *R. Acc. Lincei Rend., Roma*, 16:608–14, 1907. 7

[17] Josiah Willard Gibbs. Elements of vector analysis. Privately printed. Included in *Scientific Works of J. Willard Gibbs*, vol. 2 (pp. 17-90), 1881, 1884. 1.2, 1.4

[18] Joseph F. Grcar. How ordinary elimination became gaussian elimination. Posted as `arXiv:09907.2397v1`, 2009. 2

[19] George Green. *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism.* privately published, 1828. 6.3

[20] E. Hairer and G. Wanner. *Analysis by its History.* Springer, 1995. 17

[21] William Rowan Hamilton. *Lectures on Quaternions.* Hodges and Smith, 1853. 6.5, 11

[22] William Rowan Hamilton. *Elements of Quaternions*. Longmans, Green, 1866. Second edition (2 vols., 1899-1901) reprinted by Chelsea (1969). 1.4

[23] Thomas Hawkins. *Lebesgue's Theory of Integration: Its Origins and Development*. University of Wisconsin Press, 1970. Second Edition (1975) reprinted with corrections, AMS-Chelsea, 2002. 5.2

[24] Thomas Heath. *A History of Greek Mathematics*. Clarendon Press, 1921. Dover Reprint, 1981. 1.2, 6, 8

[25] Thomas Heath. *A Manual of Greek Mathematics*. Oxford University Press, 1931. Dover Reprint, 2003. 2.1, 2.1, 2.1, 2.1, 9c, A

[26] Thomas L. Heath. *The Works of Archimedes*. Cambridge University Press, 1897, 1912. Dover reprint, 2002. 2.2, 2, 3

[27] Thomas L. Heath. *The Thirteen Books of Euclid's Elements*. Cambridge University Press, 1908. Dover reprint, 1956. 15, 16, 8, 11, 2.1

[28] Otto Hesse. Über die Criterien des Maximums und Minimums der einfachen Integrale. *Journal für Reine und Angewandte Mathematik*, 54:227–73, 1857. 3.7

[29] Camille Jordan. *Cours d'analyse de l'École Polytechnique*. Gauthier-Villars, 1882. 1. ed. 1882-87; 2. ed. 1893-96, 3 ed. 1909-15. 6.3

[30] Wilfred Kaplan. *Advanced Calculus*. Addison-Wesley, 4 edition, 1991. 15

[31] Morris Kline. *Mathematical Thought from Ancient to Modern Times*. Oxford Univ. Press, 1972. 8, 2.1

[32] Steven G. Krantz and Harold R. Parks. *The Implicit Function Theorem: History, Theory, and Applications*. Birkhäuser, 2002. (document), 8, 6

[33] Joseph-Louis Lagrange. Recherches sur la méthode de maximis et minimis. *Misc. Taurinensia*, 1, 1759. Oeuvres, vol. 1, pp. 3-20. 17

[34] Joseph-Louis Lagrange. *Méchanique Analitique*. Desaint, 1788. 1.4

[35] Eli Maor. *The Pythagorean Theorem, A 4,000-Year History*. Princeton University Press, 2007. 15

[36] Jerrold E. Marsden and Anthony J. Tromba. *Vector Calculus*. W. H. Freeman & Co., fifth edition, 2003. 6.7

[37] James Clerk Maxwell. On the mathematical classification of physical quantities. *Proceedings of the London Mathematical Society*, 3:224–232, 1871. reprinted in [40, vol. 2, pp. 257-266]. 11

[38] James Clerk Maxwell. *A Treatise on Electricity and Magnetism*. Oxford, Clarendon Press, 1873. Dover Reprint, 1954. 11

[39] Isaac Newton. Mathematical principles of natural philosophy. In *Isaac Newton: The Principia* [9], pages 401–944. Translation from the Latin by I. Bernard Cohen and Anne Whitman of *Philosophiae Naturalis Principia Mathematica* (1687, 1713, 1726). 1.2, 1.7

[40] W. D. Niven, editor. *The Scientific Papers of James Clerk Maxwell*. Cambridge University Press, 1890. Dover reprint, 1965. 37

[41] Apollonius of Perga. *Conics, Books I-IV*. Green Lion Press, 2000. translation by R. Catesby Taliaferro (1939), in revised edition edited by Dana Densmore. Two volumes: I-III, and IV. 5

[42] Giuseppe Peano. Sulle definizione dell'area di una superficie. *Rendiconti Acc. Lincei*, 6:54–57, 1890. G

[43] Bruce Pourciau. Newton's argument for the first proposition of the *Principia*. *Archive for History of Exact Sciences*, 57:267–311, 2003. 1

[44] Daniel Reem. New proofs of basic theorems in calculus. *Math Arxiv*, 0709.4492v1, 2007. 13

[45] Herman Schwarz. Sure une definition erronée de l'aire d'une surface courbe. In *Gesammelte Math. Abhandlungen* [46], pages 309–311. Reprint, Chelsea, 1972. G

[46] Herman Schwarz. *Gesammelte Math. Abhandlungen*, volume 2. Springer, 1890. Reprint, Chelsea, 1972. 1, 45

[47] Joseph Alfred Serret. *Journal de Mathématiques*, 16:193–207, 1851. D

[48] George F. Simmons. *Calculus Gems: Brief Lives and Memorable Mathematics*. McGraw-Hill, 1992. reissued by MAA, 2007. 13

[49] Michael Spivak. *Calculus*. Publish or Perish, Inc., second edition, 1980. 6

[50] Bruce Stephenson. *Kepler's Physical Astronomy.* Princeton Univ. Press, 1987. C

[51] Dirk J. Struik. *Lectures in Classical Differential Geometry.* Addison-Wesley, 2 edition, 1961. Dover reprint, 1988. D

[52] Oswald Veblen. Theory on plane curves in non-metrical analysis situs. *Transactions, Amer. Math. Soc.*, 6(1):83–98, 1905. 6.3

[53] D. T. Whiteside. *The Mathematical Papers of Isaac Newton*. Cambridge University Press, 1967-1981. 2

[54] Edwin Bidwell Wilson. *Vector Analysis: A Textbook for the Use of Students of Mathematics, founded on the lectures of J. Willard Gibbs.* Yale Centennial Publications. Charles Scribner's Sons, 1901. Dover Reprint of second edition (1909), 1960. 1.2, 1.4, 11

[55] Shirley Llamado Yap. The poincaré lemma and an elementary construction of vector potentials. *American Mathematical Monthly*, 116(3):261–267, 2009. 6.7

Index