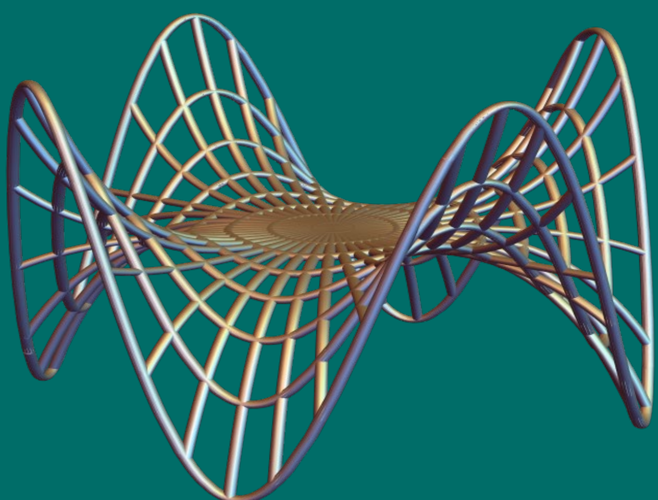
Vector Calculus

With

Differential Forms

By

Gabriel Lugo



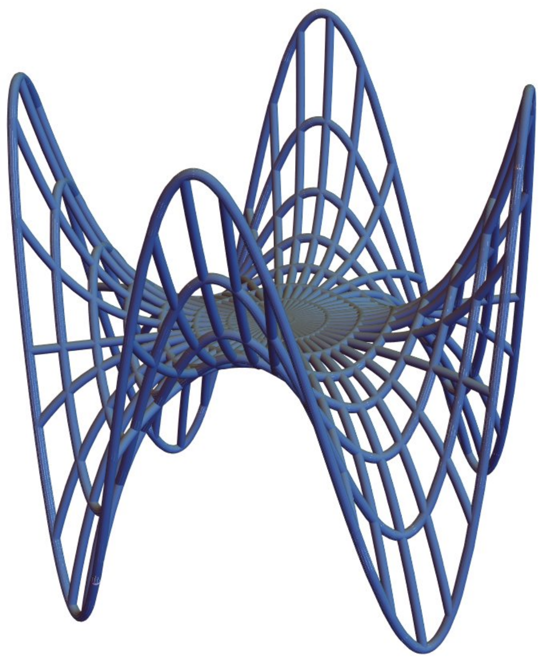
Vector Calculus

With

Differential Forms

Gabriel Lugo

UNCW 2023



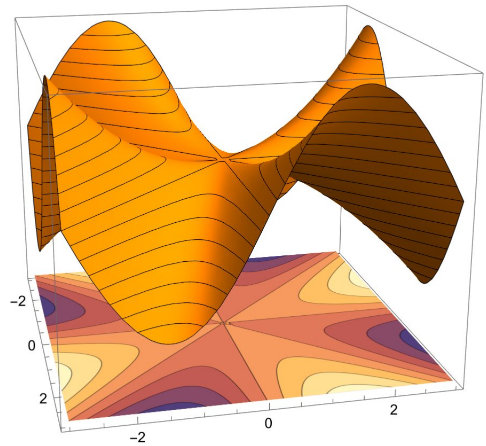
Suggested citation: Lugo, Gabriel. Vector Calculus with Differential Forms. Wilmington. University of North Carolina Wilmington, 2024.

Cover illustration: Hyperbolic Paraboloid.

UNC Wilmington, Department of Mathematics and Statitstics

Distributed by UNCW

All figures and illustrations were produced by the author or taken from photographs taken by the author and family members.



This book is dedicated to my family, for without their love and support, this work would have not been possible. I would also like thank Dr. Russell Herman for continuing inspiration and for all the contributions to the historical facts about the masters who developed this wonderful subject. Finally, I would like to include a note of appreciation to the Chair of the department, Dr. Johannes Hattingh and the able administrative assistants, Sarah Payne and Beth Casper for all their help in facilitating the acquisition of a 3D-printer, and for supporting the production of vast array of 3D-models for classroom illustrations.

G. Lugo (2024)

Contents

Preface

This book is based on typical lecture notes from a Calculus III course, which the author has taught for many years at UNCW. The course website is located at

http://people.uncw.edu/lugo/courses/M261

At this site, under the heading "Course Handouts", there is a copy of a set of actual lecture notes. The book is typeset in Latex using the Miktex editor. Mathematical graphics were created with Maple and Mathematica. The material is self-contained and includes all the standard topics usually covered in a third semester calculus course. The manuscript includes an array of examples on each topic to illustrate some variety of the types of problems students need to master. As of this edition, there is no list of exercises at the end of each section; for now, we refer the student to any fat calculus text for extra problems. As soon as time permits, problem sets will be added. After reading these notes, students should not have any difficulty working out any of the typical problems expected as background for students in science and engineering.

Calculus was co-invented by Newton for the main purpose of explaining Kepler's laws of celestial mechanics. Most of the other ideas in calculus were developed from physics as well. It is the opinion of this author, that it is a mistake to teach calculus devoid of the physical and historical context from which it was developed. Thus, Whenever it is appropriate, snippets of physics will be inserted into the narrative in a effort to enrich the meaning of the concepts. The expectation is that students in the course will read the book thoroughly and, at bare minimum, work out on their own, the examples shown in the text. Many significant concepts which in typical calculus textbooks are buried in the "advanced problems" are included in these notes as part of the content.

When it is pedagogically advantageous, we either deviate slightly from the traditional order of topics, or use an alternative presentation that students might find more efficient and comprehensible. Occasionally, we sacrifice some rigor for the sake of clarity and ease of applications. For example, we do not present a rigorous proof the chain rule, nor do we expect students to establish if a function is differentiable. Instead we rely on the intuitive notions of the differential invented by Newton and Leibnitz to motivate most topics. We also introduce elementary differential forms to augment understanding of the change of variables, surface area, and the theorems and Stokes and Gauss. We advocate that differential forms become part of the regular curriculum for all third semester calculus course. As it is the case with most standard textbooks, some optional material such as a section on the derivation of Kepler's laws is included to be covered at the discretion of the instructor. It is sad that the latter, being the reason why Newton co-invented calculus, is considered optional in the standard calculus curriculum. The chapter on Kepler's laws is not optional for this author.

Gabriel Lugo (2024)

Chapter 1

Vectors and the Geometry of Space

1.1 Euclidean Space

1.1.1 Points and Planes

1.1.1 Definition Euclidean n-space $\mathbf{R}^n$ is defined as the set of n-tuples $P(p_1, p_2, \ldots, p_n)$, where $p_i \in \mathbf{R}$, for each $i = 1, 2, \ldots, n$.

In this course we will be primarily concerned with Euclidean 3-space $\mathbf{R}^3$. For visualization purposes, let O be an arbitrary point which we will call the origin. Through the origin, we draw three mutually perpendicular copies of the real line. We call the three copies of the real line the coordinate axes. By default, we denote the axes x, y and z respectively, making sure the axes are always oriented according to the right hand rule - namely, if one places the palm of the right hand along the positive direction of the x-axis and curls the hand toward the positive direction of the y axis, then the thumb points in the direction of the positive z-axis. The ticks on the axes can be scaled as needed to adjust the "window" size, as shown in figure 1.1. In this picture, we have made the common choice to render the x-axis by perspective, sticking out of the "board." The three planes spanned by the coordinate axes are called the xy, the yz and the xz planes, respectively. We can now use the coordinate axes to assign a unique label $P(x_0, y_0, z_0)$ to any point P in $\mathbf{R}^3$, where in the entries of the triplet,

x_0 is the signed distance from the point to the yx-plane,

y_0 is the signed distance from the point to the xz-plane, and

z_0 is the signed distance from the point to the xy-plane.

For example, to locate the point $P(2, 3, 1)$, start at the origin, step 2 units in the x-direction, 3 units in the y-direction and 1 unit in the z-direction. To get a sense of the 3-dimensional location of the point, it is essential to either, draw line segments for each step, or better yet, draw a rectangular parallelepiped

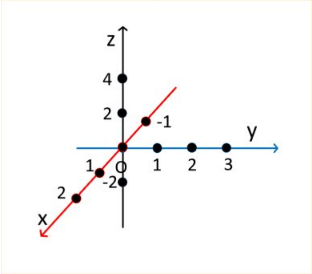


Fig. 1.1: Euclidean 3 Space

(a cuboid) to locate the point diagonally opposed to the origin. Figure 1.2 shows the points (2,3,1) and (-1,2,2). Next to points, **planes** are the simplest

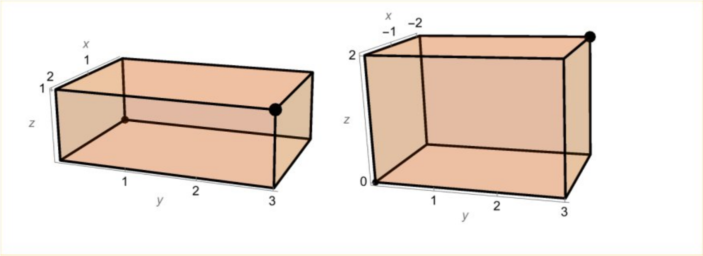


Fig. 1.2: Point Plot

geometrical objects to visualize. The general equation of a plane in $\mathbf{R}^3$ is of the form

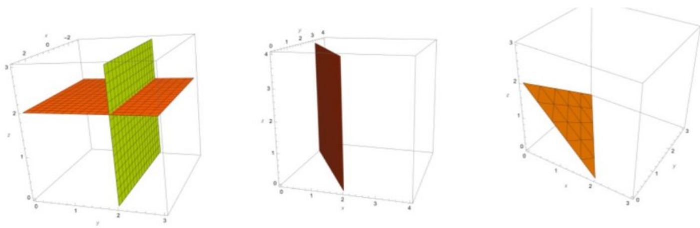
$$Ax + By + Cz = D.$$

There are three possible cases.

Case 1. Two variables are missing as in the equation $z = 2$. Since the variables x and y are unconstrained, the graph is a plane passing through the point $(0,0,2)$ which is parallel to the xy-plane. The analog $y = 2$ in $\mathbf{R}^2$ is a line passing through the point $(0,2)$, which is parallel to the x-axis.

Case 2. One variable is missing as in equation $2x + 3y = 6$. Had this been an equation in 2-dimensions, the graph would have been a straight line passing through the points $(3,0)$ and $(0,2)$. However, here we are in $\mathbf{R}^3$ and the points in the graph must include three coordinates, the z-coordinate being arbitrary. The graph is the plane one obtains by extruding the line in the direction of the missing coordinate, as shown in figure 1.3. The word of the day here is "**extrusion**". In 3d graphics, extrusion means to stretch a flat object in a direction orthogonal to the plane in which the object lies, thereby producing a 3-dimensional shape. In general, if we have the graph of an equation in two variables, in dimension 3, the graph is obtained by extruding the curve in the direction of the missing coordinate. Thus, for example,

- The 3d graph of $z = x$ is a plane containing the y-axis and the line at 45 degrees in the xz-plane
- The 3d graph of $x^2 + y^2 = 1$ is a circular cylinder symmetric with respect to the z-axis.
- The 3d graph of $x^2 + z^2 = 1$ is a circular cylinder symmetric with respect to the y-axis.
- The 3d graph of $y = x^2$ is a parabolic cylinder with generator lines parallel to the z-axis. A shape similar to bending a thin piece of cardboard.



a) z=2, y=2 b) 2 x +3 y =6 b) 2 x +3 y + 2z =6

Fig. 1.3: Graphs of Planes

Case 3. No missing coordinates. In this case, the equation of the plane is best written in the standard form $Ax + By + Cz = D$. We will treat planes formally in section 1.5, but there is no reason to wait until then to learn how to visualize them. Consider the example $2x + 3y + 2x = 6$ shown in figure 1.3. According to one of the Euclidean postulates, a plane is determined uniquely by three non-collinear points. The easiest three points to locate are the intersections of the plane with the coordinate axes. These are obtained by successively setting two of the coordinates equal to zero and solving for the remaining coordinate. In the example, the x-intercept is obtained by setting $y = z = 0$. Then $2x = 6$ and $x = 3$. The three intercepts are $P(3, 0, 0)$, $Q(0, 2, 0)$ and $R(0, 0, 3)$. We render the plane by drawing the triangle $\triangle PQR$. Of course, the plane that contains the triangle extends infinitely in all directions. Typically, we render planes by drawing a triangle or a parallelogram contained in the plane.

1.1.2 Distance Formula

Let $P(x_0, y_0, z_0)$ and $Q(x_1, y_1, z_1)$ be two points in $\mathbf{R}^3$. We make think of the points P and Q as main diagonal vertices of a cuboid as shown in 1.4. The segment $\overline{PQ}$ is the hypothenuse of a right triangle with height $(z_1 - z_0)$, and whose base is itself the hypothenuse of another right triangle with sides $(x_1 - x_0)$ and $(y_1 - y_0)$. Applying Pythagoras' theorem twice, we deduce that the distance $d(P, Q)$ between the two points is given by

$$d(P, Q) = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}. \tag{1.1}$$

The formula is a natural extension of the distance formula one learns in 2-dimensional coordinate geometry. In particular, if P is the origin O, the distance $d(O, Q)$ is given by the square root of the sum of the squares of the coordinates of Q.

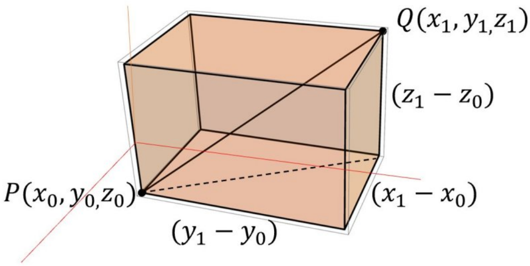


Fig. 1.4: Distance Formula

From the same diagram 1.4, it is fairly obvious that the coordinates of the midpoint M of the segment $\overline{PQ}$ are given by the average of the coordinates of P and Q, namely

$$M\left(\frac{x_0+x_1}{2}, \frac{y_0+y_1}{2}, \frac{z_0+z_1}{2}\right). \tag{1.2}$$

The distance formula can be extrapolated to $\mathbf{R}^n$ by simply taking the square root of the sum of the squares of the differences of the coordinates of the two points. In a similar way, the midpoint formula can also be generalized to $\mathbf{R}^n$.

1.1.3 Spheres

A sphere S^2 in $\mathbf{R}^3$ is completely determined by its center and its radius. If the center is located at $C(x_0, y_0, z_0))$ and the radius is R, the sphere consists of all points $X(x, y, z)$ such that distance $d(C, X) = R$. It follows immediately that the equation of the sphere is given by,

$$\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2} = R,$$

or

$$(x-x_0)^2+(y-y_0)^2+(z-z_0)^2 = R^2. \tag{1.3}$$

Any equation of the form 1.3 is a sphere and any sphere can be written in the form 1.3. One can instantly recognize the equation of a sphere because it is quadratic on x, y and z with equal coefficients (that by division can be set to 1), and there are no cross terms.

Example Find the equation of the sphere with radius 3 centered at $C(4, 3, 3)$.
Solution. Let $X(x, y, z)$ be an arbitrary point on the sphere. Setting the distance $d(C, X)^2 = 3^2$, we get

$$(x-4)^2+(y-3)^2+(z-3)^2 = 9.$$

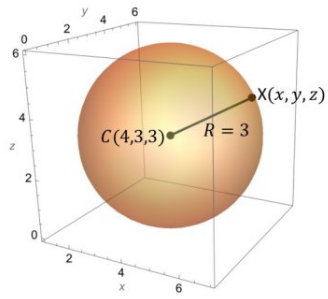


Fig. 1.5: Sphere

Example Find the equation of a sphere with center $C(1,2,3)$ and containing the point $P(4,3,-1)$.
Solution

$$\begin{aligned} R^2 =& (4-1)^2 + (3-2)^2 + (-1-3)^2 \\ =& 9+1+16 \\ =& 26 \end{aligned}$$

Hence

$$(x-1)^2 + (y-2)^2 + (z-3)^2 = 26.$$

Example Given: $x^2 + y^2 + z^2 = x + y + z$, find the center and radius of the sphere. The equation is quadratic and the coefficients of x, y and z are equal to 1 already. Thus, this represents a sphere. We complete the squares to rewrite in standard form.

$$\begin{gathered} (x^2 - x \quad) + (y^2 - y \quad) + (z^2 - z \quad) = 0, \\ (x^2 - x + \tfrac{1}{4}) + (y^2 - y + \tfrac{1}{4}) + (z^2 - z + \tfrac{1}{4}) = \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4}, \\ (x - \tfrac{1}{2}) + (y - \tfrac{1}{2})^2 + (z - \tfrac{1}{2})^2 = \tfrac{3}{4}. \end{gathered}$$

So, the sphere has center at $C(1/2, 1/2, 1/2)$ and radius $R = \sqrt{3}/2$.
Example Conceptual hints

- Find the equation of a sphere if one of its diameters has end points $P(5,4,3)$ and $Q(1,6,-1)$. Hint: Use the distance formula to find the radius $R = \frac{1}{2}d(P,Q)$, and the midpoint formula to find the center.
- Determine if three points A, B and C lie on a straight line. Hint: Compute $d(A,B)$, $d(A,C)$ and $d(B,C)$. Determine if the largest of these three numbers is equal to the sum of the other two.
- Determine if three points A, B and C are the coordinates of a right triangle. Hint: Compute $d(A,B)$, $d(A,C)$ and $d(B,C)$. Determine if the square of the largest of these three numbers is equal to the sum of the squares of the other two.

1.2 Vectors

1.2.1 Definition The space of Euclidean vectors $\mathbf{E}^3$ is defined as the set of ordered triplets $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$, where $v_i \in \mathbf{R}$, for each $i = 1, 2, 3$. The definition can be extended to any dimension n by replacing triplets to n-tuples. Given any two triplets $\mathbf{A} = (a_1, a_2, a_3)$, $\mathbf{B} = (b_1, b_2, b_3), \in \mathbf{E}^3$ and any real number c, we define two operations:

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= (a_1 + b_1, a_2 + b_2, a_3 + b_3), \\ c\,\mathbf{A} &= (c\,a_1, c\,a_2, c\,a_3). \end{aligned} \tag{1.4}$$

The triplet $\mathbf{0} = \langle 0, 0, 0 \rangle$ is called the zero vector. Real numbers such as $c \in \mathbf{R}$ are called **scalars**. In this course, all scalars are real numbers, but one could conceive of a geometry in which the scalars are complex numbers or some other entities.
The two operations of vector sum and multiplication by a scalar satisfy 8 natural properties (VS1...VS8), inherited from real numbers, as follows:

- VS1: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$,
- VS2: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$,
- VS3: $\mathbf{A} + \mathbf{0} = \mathbf{A}$,
- VS4: $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$,
- VS5: $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$
- VS6: $(c + d)\mathbf{A} = c\mathbf{A} + d\mathbf{A}$,
- VS7: $(cd)\mathbf{A} = c(d\mathbf{A})$,
- VS8: $1\mathbf{A} = \mathbf{A}$.

A space with two operations which satisfy the 8 properties listed above is called a **vector space**. The triplet that represents the vector are called the components of the vector.

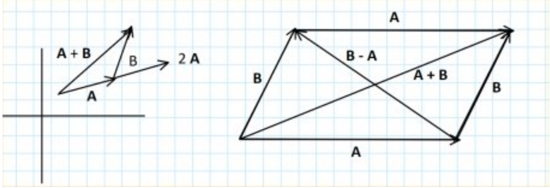


Fig. 1.6: Vector Operations

Vectors in $\mathbf{E}^3$ are not the same as points in $\mathbf{R}^3$ and they must not be confused. For a point $(a_1, a_2, a_3) \in \mathbf{R}^3$, what we mean by $P(a_1, a_2, a_3)$ is just a label to identify the location of the point. On the other hand, in calculus and elementary physics, vectors are viewed as triplets, independent of their location in space. Vectors can be added, points can't. Similarly, it makes no sense to multiply a point by a number. Vectors are usually regarded as arrows characterized by a direction and a length. Thus, a vector $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ can be parallel-transported so that the foot of the arrow can be located at any point in space. The tip of the arrow would then be located at the point obtained by starting at the foot of the arrow, followed by steps of size a_1 in the x-direction, a_2 in the y-direction and a_3 in the z-direction. This representation leads to a 1-1 correspondence between points and vectors which often causes confusion to the novice. Given a point $P(a, b, c) \in \mathbf{R}^3$ we can associate to it uniquely, a vector $\mathbf{P} = \langle a, b, c\rangle$ which corresponds to the arrow with foot at the origin and head at the point P. We then say that $\mathbf{P}$ is the **position vector** of the point P. Triplets representing coordinates of a point will always be written with parenthesis and vector components with angle brackets.

Geometry

With the representation of vectors by arrows just described, the geometry of the two vector operations can be easily visualized.

Vector addition: If $\mathbf{A}$ and $\mathbf{B}$ are vectors with corresponding representation by arrows, the vector $\mathbf{A}+\mathbf{B}$ is obtained by placing the foot of the arrow of the $\mathbf{B}$ vector at the head of the $\mathbf{A}$ vector, then drawing a new vector with foot at the foot of $\mathbf{A}$ and head at the head of $\mathbf{B}$. The geometry is illustrated in 2 dimensions in figure 1.6. Thus, for example If

$$\begin{aligned} \mathbf{A} &= \langle 3, 1\rangle, \\ \mathbf{B} &= \langle 4, 2\rangle, \quad \text{then} \\ \mathbf{A}+\mathbf{B} &= \langle 7, 3\rangle, \end{aligned}$$

Multiplication by a scalar: Let $c > 0$ be a scalar and $\mathbf{A}$ a vector. Then $c\mathbf{A}$ is represented by an arrow in the same direction, stretched by the factor c if $c > 1$ or contracted by the factor c if $0 < c < 1$. If c is negative, a similar effect is obtained, but then the resulting arrow points in the opposite direction.

Examples

- The vector $2\mathbf{A}$ is twice as long as $\mathbf{A}$ and points in the same direction as $\mathbf{A}$.

- The vector $-3\mathbf{A}$ is three times as long as $\mathbf{A}$, but points in the opposite direction.

- The vector $\frac{1}{2}\mathbf{A}$ is half as long as $\mathbf{A}$ and points in the same direction as $\mathbf{A}$.

- If the vector $\mathbf{F}$ represents a force, then the vector $-\mathbf{F}$ represents a force of the same magnitude but in the opposite direction.

- If $\mathbf{F_1}$ and $\mathbf{F_2}$ represent two forces acting on an a mass (regarded as a point-mass), then the total force acting on the mass is given by $\mathbf{F_1} + \mathbf{F_2}$.

- Parallelogram law: As shown in figure 1.6, two vectors $\mathbf{A}, \mathbf{B}$ drawn with a common foot, span a parallelogram. Then the diagonal vector with foot at the common foot of $\mathbf{A}$ and $\mathbf{B}$ represents $\mathbf{A} + \mathbf{B}$, whereas the diagonal from the head of $\mathbf{A}$ to the head of $\mathbf{B}$ represents $\mathbf{B} - \mathbf{A} = \mathbf{B} + (-\mathbf{A})$. Students should take the time to draw the appropriate parallel translations to verify latter assertion. In particular, the vector $\overrightarrow{PQ}$ from P to Q is given by
$$\overrightarrow{\mathbf{PQ}} = \mathbf{Q} - \mathbf{P}.$$

Note: When writing vectors by hand, it is not practical to use boldface notation, instead, it is standard to denote vectors with an overhead arrow $\overrightarrow{\mathbf{A}}$. There is no standard notation on whether or not use uppercase for the names of vectors. For example, force vectors $\mathbf{F}$ are typically uppercase, but velocity vector $\mathbf{v}$ are most commonly lowercase.

1.2.2 Definition Let $\mathbf{A} = \langle a, b, c\rangle$ be a vector. The **norm / length /magnitude** is given by
$$\|\mathbf{A}\| = \sqrt{a^2 + b^2 + c^2}.$$
If $\mathbf{A}$ is the position vector of the point A, then $\|\mathbf{A}\|$ is the distance from A to the origin. A vector of length one is called a **unit vector**. If a vector is divided by its length, one gets a unit vector in the same direction.

1.2.3 Definition The **standard basis** of $\mathbf{R}^3$ consists of the three unit vectors
$$\begin{aligned}\mathbf{i} =& \langle 1, 0, 0\rangle = \mathbf{e}_1,\\ \mathbf{j} =& \langle 0, 1, 0\rangle = \mathbf{e}_2,\\ \mathbf{k} =& \langle 0, 0, 1\rangle = \mathbf{e}_3.\end{aligned}$$
Any vector $\mathbf{A} \in \mathbf{E}^3$ can be written as a linear combination of the basis vectors. That is
$$\begin{aligned}\mathbf{A} =& \langle a, b, c\rangle,\\ =& a\,\mathbf{i} + b\,\mathbf{j} + c\,\mathbf{k}\end{aligned}$$
Both notations are common, but a "hybrid" notation that denotes a vector using a combination of brackets and basis vectors is incorrect.
Example Find a unit vector in the direction of $\mathbf{A} = \langle 3, 1, 1\rangle$.
Solution:
$$\begin{aligned}\|\mathbf{A}\| =& \sqrt{3^2 + 1^2 + 1^2} = \sqrt{11},\\ \mathbf{u} =& \frac{\mathbf{A}}{\|\mathbf{A}\|},\\ =& \frac{\langle 3, 1, 1\rangle}{\sqrt{11}} = \langle \tfrac{3}{\sqrt{11}}, \tfrac{1}{\sqrt{11}}, \tfrac{1}{\sqrt{11}}\rangle.\end{aligned}$$

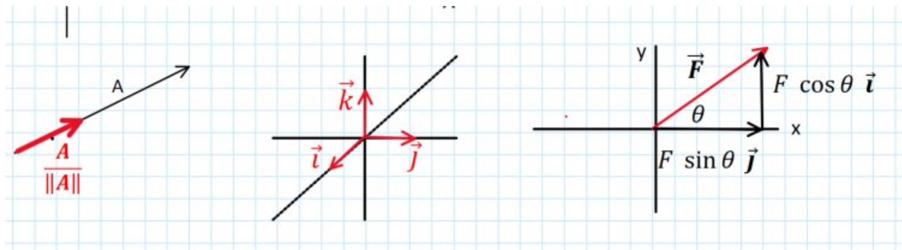


Fig. 1.7: Basis

Let $\mathbf{F} = \langle F_x, F_y \rangle \in \mathbf{E}^2$ be a vector in the plane such as a force vector. If we denote the length $\|\mathbf{F}\|$ of the vector by F and the angle the vector makes with the horizontal axis by θ, then, as shown in figure 1.7, the components of the vector in standard basis are.

$$\begin{aligned} \mathbf{F} =& F_x\,\mathbf{i} + F_y\,\mathbf{j} \quad \text{where,} \\ F_x =& F\cos\theta, \\ F_y =& F\sin\theta, \\ F =& \|\mathbf{F}\|, \\ \theta =& \tan^{-1}(F_y/F_x). \end{aligned}$$

It is understood that the direction angle θ is positive when measured counterclockwise. Unless otherwise specified, angles in calculus are measured in radians.

Definition The magnitude v of a velocity vector $\mathbf{v}$ is called the **speed**.

Example What is the angle between the vector $\mathbf{F} = \mathbf{i} + \sqrt{3}\,\mathbf{j}$ and the positive direction of the x-axis?

Solution:

$$\begin{aligned} F =& \|\mathbf{F}\| = \sqrt{1^2 + \sqrt{3}^2} = 2 \quad \text{(30-60 right triangle)}, \\ \theta =& \tan^{-1}(\sqrt{3}/1) = \pi/3. \end{aligned}$$

Example A woman walks due west on the deck of a ship at 3 mi/h. The ship is moving north at a speed of 22 mi/h. Find the speed and and direction of the woman relative to the surface of the water.

Solution: The velocity vector $\mathbf{v}$ of the woman is the sum of the two vectors.

$$\begin{aligned} \mathbf{v_1} =& -3\,\mathbf{i} + 0\,\mathbf{j}, \\ \mathbf{v_2} =& 0\,\mathbf{i} + 22\,\mathbf{j}, \\ \mathbf{v} =& \mathbf{v_1} + \mathbf{v_1} = -3\,\mathbf{i} + 22\,\mathbf{j}, \\ v =& \|\mathbf{v}\| = \sqrt{(-3)^2 + 22^2} = \sqrt{493}, \\ \theta =& \tan^1(-\tfrac{22}{3}). \end{aligned}$$

Bearing angles are usually measured in degrees. So, to compute the angle θ in a calculator one must first set the device to degree mode. The calculator result

is negative since the inverse tangent function by default has range $(-\pi/2, \pi/2)$. To get the correct answer one must add 180^o which is supplement in the third quadrant. The answer is $\theta \doteq 93^o$.
The hand-written class notes contain some typical static equilibrium problems with hanging masses. The problems on this topic are for enrichment only and will not be worked out in class. Solutions to two static equilibrium problems appear in page 5/55 of the handwritten version posted at the course web site.

`http://people.uncw.edu/lugo/courses/M261/M261_ClassNotes_OCR.pdf.`

1.3 Dot Products

The two operations of vector addition and multiplication by a scalar which are permitted under structure of a vector space, are not sufficient for our purposes. We need to equip the space with two additional vector multiplication structures. The first of these multiplications is called the dot product

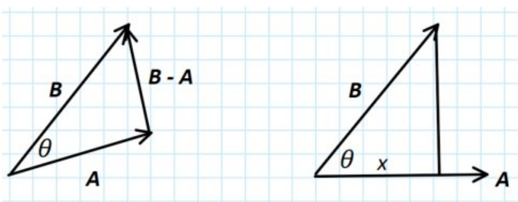


Fig. 1.8: Dot Product

1.3.1 Definition Let $\mathbf{A} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{B} = \langle b_1, b_2, b_3\rangle$ be two vectors in 3-space and denote the relative angle between the two vectors by θ, as shown in figure 1.8. The dot product of the two vectors is defined by

$$\mathbf{A}\cdot\mathbf{B} = a_1b_1 + a_2b_2 + a_3b_3 = \sum_{k=1}^{3} a_kb_k. \tag{1.5}$$

Notes:

- The dot product of two vectors is a scalar. For this reason, the dot product is also called in the literature, the scalar product or the **inner product**. To avoid possible confusion with the vector space operation of multiplication by a scalar, we will not use the term scalar product again in these notes.
- The dot product operation generalizes to n-dimensions by simply changing the sum from 1 to n for vectors in $\mathbf{E}^n$.
- By the "angle" between two vectors we mean the angle formed by the two vectors arranged to have a common foot.

- It follows immediately from the definition that

$$\mathbf{A} \cdot \mathbf{A} = \|\mathbf{A}\|^2. \tag{1.6}$$

1.3.2 Theorem The dot product of the two vectors **A** and **B** satisfies the equation

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\|\,\|\mathbf{B}\|\,\cos\theta \tag{1.7}$$

The proof follows from applying the law of cosines to the triangle shown in figure 1.8, and comparing to the length of $\mathbf{C} = \mathbf{B} - \mathbf{A}$.

$$\begin{aligned}
\|\mathbf{B} - \mathbf{A}\|^2 =& \|\mathbf{A} - \mathbf{B}\|^2 \\
=& (\mathbf{A} - \mathbf{B}) \cdot (\mathbf{A} - \mathbf{B}) \\
=& \mathbf{A} \cdot \mathbf{A} + \mathbf{B} \cdot \mathbf{A}\mathbf{B} - 2\mathbf{A} \cdot \mathbf{B}, \\
=& \|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 - 2\mathbf{A} \cdot \mathbf{B}, \\
=& \|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 - 2\|\mathbf{A}\|\,\|\mathbf{B}\|\,\cos\theta, \quad \text{(Law of Cosines)}.
\end{aligned}$$

The theorem is central because it allows for a geometry and a physics interpretation of the dot product.

1.3.1 Geometry of Dot Products

If we let x denote the (scalar) projection of the vector **B** onto **A**, we see immediately from the definition of cosines applied to figure 1.8, that

$$\begin{aligned}
x \equiv \mathit{Proj}(\mathbf{B}, \mathbf{A}) =& \|\mathbf{B}\|\,\cos\theta, \\
=& \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|}
\end{aligned} \tag{1.8}$$

The terminology $\mathit{Proj}_{\mathbf{A}}\mathbf{B}$ is also used in the literature. As the name indicates, the scalar projection gives a number also called the **component** of the the vector **B** in the direction of **A**. Clearly, the scalar projection of **B** onto **A** is not the same as the scalar projection of **A** onto **B**. Thus, it is essential that in formula 1.8, one has the norm of the correct vector in the denominator. It might be easier to remember just the formula for the dot product, and draw the triangular picture in figure 1.8.

The vector projection of **B** onto **A**, is obtained by multiplying the scalar projection by a unit vector in the direction of **A**, namely

$$\begin{aligned}
\overrightarrow{\mathbf{Proj}(\mathbf{B}, \mathbf{A})} &= \left(\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|}\right) \frac{\mathbf{A}}{\|\mathbf{A}\|}, \\
&= \left(\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|^2}\right) \mathbf{A}, \\
&= \left(\frac{\mathbf{A} \cdot \mathbf{B}}{\mathbf{A} \cdot \mathbf{A}}\right) \mathbf{A}.
\end{aligned} \tag{1.9}$$

The last step makes the formula slightly more "symmetrical" and easier to remember. It also serves as preparation for future encounters with projections of wave functions onto eigenstates in quantum mechanics, or root systems in representation theory.

The dot product formula can be "reversed engineered" to find the angle θ between two vectors. Namely, to find θ, one solves equation the equation 1.7 for $\cos\theta$, then one finds θ by taking the inverse cosine. In particular, two vectors are perpendicular if and only their dot product is zero.

$$\mathbf{A} \perp \mathbf{B} \Leftrightarrow \mathbf{A} \cdot \mathbf{B} = 0. \tag{1.10}$$

Example Given $\mathbf{A} = \langle -1, 1, 0\rangle$ and $\mathbf{B} = \langle 2, -1, 1\rangle$, find: a) $\mathbf{A} \cdot \mathbf{B}$, b) the component of $\mathbf{B}$ in the direction of $\mathbf{A}$, c) the angle between the two vectors. Solution:

$$\begin{aligned}
\mathbf{A} \cdot \mathbf{B} =& (-1)(2) + (1)(-1) + (0)(1) = -3, \\
Proj(\mathbf{B}, \mathbf{A}) =& \frac{-3}{\sqrt{(-1)^2 + 1^2 + 0^2}} = \frac{-3}{\sqrt{2}}, \\
\theta =& \cos^{-1}\left(\frac{-3}{\sqrt{2}\sqrt{6}}\right) = \cos^{-1}\left(\frac{-3}{\sqrt{12}}\right).
\end{aligned}$$

Example Determine whether or not the points $P(1, -3, -2), Q(2, 0, -4)$ and $R(6, -2, -5)$ are the vertices of a right triangle.
Solution: Let $\mathbf{A} = \overrightarrow{PQ}, \mathbf{B} = \overrightarrow{PR}, \mathbf{C} = \overrightarrow{QR}$ and compute the dot products.

$$\begin{aligned}
\mathbf{A} =& \langle 1, 3, -2\rangle, \\
\mathbf{B} =& \langle 5, 1, -3\rangle, \\
\mathbf{C} =& \langle 4, -2, -1\rangle, \\
\mathbf{A} \cdot \mathbf{B} =& 5 + 3 + 6 = 14, \\
\mathbf{A} \cdot \mathbf{C} =& 4 - 6 + 2 = 0, \quad \longrightarrow \mathbf{A} \perp \mathbf{C}, \quad \text{Right angle.} \\
\mathbf{B} \cdot \mathbf{C} =& 20 - 2 + 3 = 21.
\end{aligned}$$

1.3.2 Physics of Dot Products - Work

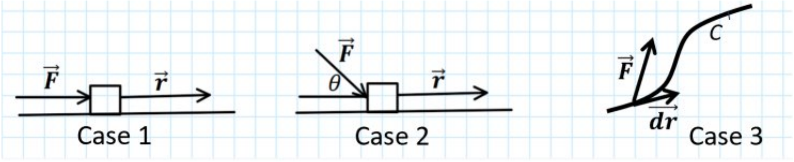


Fig. 1.9: Work

Let $\mathbf{F}$ be a force vector and $\mathbf{r}$ a position vector representing the displacement of a point mass. We consider three cases as depicted in figure 1.9.
Case 1: Here we assume the force vector $\mathbf{F}$ is constant and the force is exerted upon a mass along a constant direction vector $\mathbf{r}$. The work W done by the

force on the mass is then defined by the magnitude of the force multiplied by the displacement. The displacement is the magnitude of $\mathbf{r}$. In other words,

$$W = \|\mathbf{F}\|\,\|\mathbf{r}\| \tag{1.11}$$

This formula appears in a first introduction to physics as $W =$ force $\times$ distance. In MKS units, the force is measured in Newtons and the displacement in meters. The standard unit of work is a Joule. This is same as the standard unit of energy, the difference being that work can be negative.

1 Joule = 1 Newton-meter,

If one lifts vertically a mass m from a table to a height h, the magnitude of the force of gravity is just the weight mg. It takes energy to lift the mass, and since energy is conserved, the energy must be stored in the mass in the form of potential energy

$$PE = mgh.$$

An object with a weight of 1 Newton is about the size of one brass "Fig Newton". The work against gravity to lift such an object i meter is about 1 Joule.

Case 2: Here we assume again that the force vector is constant but the force is exerted upon a mass along a constant direction vector $\mathbf{r}$ which makes an angle θ with the force vector. Then, only the scalar projection of $\mathbf{F}$ onto $\mathbf{r}$, that is only the component of the force in the given direction, has an effect in moving the mass. Thus, The work done by the force on the mass is given by

$$\begin{aligned} W =& \|\mathbf{F}\| \cos\theta \, \|\mathbf{r}\|, \\ W =& \mathbf{F}\cdot\mathbf{r}. \end{aligned} \tag{1.12}$$

Case 3. This is the most general case in which instead of a force, we have a force field. That is, for each point in space there there is a force vector which may vary in magnitude and direction from point to point. We also assume that instead of a straight line, the mass is being pushed by the force along a curve C given by a position vector $\mathbf{r}(t)$, which depends on time t. True to the spirit of calculus, one divides the curve into infinitesimally small line segments with direction $d\mathbf{r}$, as shown in figure 1.11. The vector field will not have significant change as it moves a mass from a point on the curve to another point infinitesimally close along the curve. So, case 2 above applies and we get an infinitesimal amount of work

$$dW = \mathbf{F}\cdot d\mathbf{r}$$

The total work is then found by integrating along the curve

$$W = \int_C \mathbf{F}\cdot d\mathbf{r}. \tag{1.13}$$

The integral that appears in this general case is called a line integral. We will have to wait until section 5.2 to learn how to compute such line integrals. Computing work by line integrals is a major objective of this course. Thus, in some sense, this course is mostly about energy!

Example Find the work done by a force $\mathbf{F} = 8\,\mathbf{i} - 6\,\mathbf{j} + \mathbf{9}\,\mathbf{k}$ that moves an object along a straight line from the point $P(0, 10, 8)$ to the point $Q(6, 12, 20)$.

Solution:

$$\begin{aligned}
\mathbf{F} =& \langle 8, -6, 9\rangle, \\
\mathbf{r} =& \overrightarrow{PQ} = \mathbf{Q} - \mathbf{P}, \\
=& \langle 6, 2, 12\rangle. \\
W =& \mathbf{F}\cdot\mathbf{r} = 8(6) - 6(2) + 9(12) = 144.
\end{aligned}$$

In a physics course, it would be unthinkable to pose a problem like this one without stating the units.

1.4 Cross Products

Let

$$\begin{aligned}
\mathbf{A} &= a_1\,\mathbf{i} + a_2\,\mathbf{j} + a_3\,\mathbf{k}, \\
\mathbf{B} &= b_1\,\mathbf{i} + b_2\,\mathbf{j} + b_3\,\mathbf{k}.
\end{aligned}$$

The cross product $\mathbf{A}\times\mathbf{B}$ is a new vector defined by the determinant

$$\mathbf{A}\times\mathbf{B} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}, \tag{1.14}$$

$$= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}\mathbf{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}\mathbf{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\mathbf{k}, \tag{1.15}$$

where, recalling from pre-calculus, the determinant of a 2×2 matrix is given by the product of the diagonal terms minus the product of the off-diagonal ones,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \tag{1.16}$$

.

Example

$$\begin{aligned}
\mathbf{A} &= 3\,\mathbf{i} - 2\,\mathbf{j} + \mathbf{k}, \\
\mathbf{B} &= \mathbf{i} + 2\,\mathbf{j} + 3\,\mathbf{k}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{A}\times\mathbf{B} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 3 & -2 & 1 \\ 1 & 2 & 3 \end{vmatrix}, \\
&= \begin{vmatrix} -2 & 1 \\ 2 & 3 \end{vmatrix}\mathbf{i} - \begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix}\mathbf{j} + \begin{vmatrix} 3 & -2 \\ 1 & 2 \end{vmatrix}\mathbf{k}, \\
&= -8\,\mathbf{i} - 8\,\mathbf{j} + 8\,\mathbf{k}
\end{aligned}$$

We will be doing lots of computations of cross products. Since the computations involve only simple arithmetic, It will be shown in class how to compute cross products effectively in one line.
From elementary properties of matrices, we have the following facts:

1. Determinants are defined only for square matrices. Since we only have the three basis vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ in the definition of cross products, this concept only makes sense in dimension 3. Never write a determinant symbol for a 2×3 or other non-square arrays!.

2. If in a square matrix, one switches two rows, the determinant changes by a minus sign. Hence
$$\mathbf{B} \times \mathbf{A} = -\mathbf{A} \times \mathbf{B},$$
$$\mathbf{A} \times \mathbf{A} = -\mathbf{A} \times \mathbf{A} = 0.$$

3. If in a square matrix, one row is a multiple of another, the determinant is 0. Hence, if $\mathbf{B}$ is a non-zero multiple of $\mathbf{A}$, that is $\mathbf{B} = c\,\mathbf{A}, \quad c \neq 0$, then the cross product is 0. We conclude that if the cross product of two vectors is zero, then vectors are parallel. That is,
$$\mathbf{A} \times \mathbf{B} = 0 \Leftrightarrow \mathbf{A} \parallel \mathbf{B}.$$

4. By direct computation from the definitions one can verify that
$$(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{A} = 0,$$
$$(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{B} = 0,$$
so the vector $(\mathbf{A} \times \mathbf{B})$ is perpendicular to both, $\mathbf{A}$ and $\mathbf{B}$.

We have the following three neat theorems.

1.4.1 Triple Product

Let

$$\begin{aligned} \mathbf{A} =& \langle a_1, a_2, a_3 \rangle, \\ \mathbf{B} =& \langle b_1, b_2, b_3 \rangle, \\ \mathbf{C} =& \langle c_1, c_2, c_3 \rangle. \end{aligned}$$

Then we have the following important vector identities

1.4.1 Theorem Triple Product

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3, \end{vmatrix} \tag{1.17}$$

The quantity on the left hand side is called the **scalar triple product** or just the **triple product**. The triple product is also denoted by $(\mathbf{ABC})$.

1.4.2 Theorem Triple Cross Product

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C}, \tag{1.18}$$

If in the right hand side of the equation, one writes the vectors in front of the coefficients, one gets the common mnemonic **BAC** minus **CAB**.

1.4.3 Theorem Cross-Dot-Cross Formula

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = \begin{vmatrix} \mathbf{A} \cdot \mathbf{C} & \mathbf{A} \cdot \mathbf{D} \\ \mathbf{B} \cdot \mathbf{C} & \mathbf{B} \cdot \mathbf{D} \end{vmatrix} \tag{1.19}$$

Proof of triple product. For the the first assertion 1.17 we have,

$$\begin{aligned}
\mathbf{A} =& a_1 \,\mathbf{i} + a_2 \,\mathbf{j} + a_3 \,\mathbf{k}, \\
\mathbf{B} \times \mathbf{C} =& \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \mathbf{k}, \\
\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) =& a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}, \\
=& \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3. \end{vmatrix}
\end{aligned}$$

Again, by properties of determinants, if one swaps two rows, the determinant changes by a minus sign. By swapping two appropriate rows at a time, one can verify that

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A})$$

so there is no minus sign ambiguity in the notation (**ABC**) for any three cyclic permutation of the letters.

Proof of BAC-CAB At this stage we do not have enough fancy tools such as tensor calculus to provide an elegant proof of equation 1.18, so basically, the uninspiring approach is to compute both sides and compare to show they yield the same answer. We present a slightly more instructive computation that at least has some value in extracting the reason why the formula works. We start with the right-hand-side and bootstrap our way to the left-hand-side.

$$\begin{aligned}
(\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C} =& (a_1c_1 + a_2c_2 + a_3c_3)\langle b_1, b_2, b_3\rangle - \\
& (a_1b_1 + a_2b_2 + a_3b_3)\langle c_1, c_2, c_3\rangle, \\
=& [b_1(a_1c_1 + a_2c_2 + a_3c_3) - c_1(a_1b_1 + a_2b_2 + a_3b_3)]\,\mathbf{i} + \\
& [b_2(a_1c_1 + a_2c_2 + a_3c_3) - c_2(a_1b_1 + a_2b_2 + a_3b_3)]\,\mathbf{j} + \\
& [b_3(a_1c_1 + a_2c_2 + a_3c_3) - c_3(a_1b_1 + a_2b_2 + a_3b_3)]\,\mathbf{k}.
\end{aligned}$$

Keeping in mind our goal, as we extract the coefficients of **i** we want the expression in terms of the minors of the determinant, so we factor our a_2 and a_3. We adjust the signs of the factors to keep the indices leading to easy to read determinants. We do the same for the coefficients of **j** and **k**.

$$\begin{aligned}(\mathbf{A}\cdot\mathbf{C})\mathbf{B}-(\mathbf{A}\cdot\mathbf{B})\mathbf{C} &= [\quad a_2(b_1c_2-b_2c_1)+a_3(b_1c_3-b_3c_1)]\,\mathbf{i}+\\ &\quad [-a_1(b_1c_2-b_2c_1)+a_3(b_2c_3-b_3c_2)]\,\mathbf{j}+\\ &\quad [-a_1(b_1c_3-b_3c_1)-a_2(b_2c_3-b_3c_2)]\,\mathbf{k}+\\ &= \left(\quad a_2\begin{vmatrix} b_1 & b_2\\ c_1 & c_2\end{vmatrix}+a_3\begin{vmatrix} b_1 & b_3\\ c_1 & c_3\end{vmatrix}\right)\,\mathbf{i}\\ &\quad \left(-a_1\begin{vmatrix} b_1 & b_2\\ c_1 & c_2\end{vmatrix}+a_3\begin{vmatrix} b_2 & b_3\\ c_2 & c_3\end{vmatrix}\right)\,\mathbf{j}\\ &\quad \left(-a_1\begin{vmatrix} b_1 & b_3\\ c_1 & c_3\end{vmatrix}-a_2\begin{vmatrix} b_2 & b_3\\ c_2 & c_3\end{vmatrix}\right)\,\mathbf{k}\\ &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_1 & a_2 & a_3\\ \begin{vmatrix} b_2 & b_3\\ c_2 & c_3\end{vmatrix} & -\begin{vmatrix} b_1 & b_3\\ c_1 & c_3\end{vmatrix} & \begin{vmatrix} b_1 & b_1\\ c_1 & c_2\end{vmatrix}\end{vmatrix}\\ &= \mathbf{A}\times(\mathbf{B}\times\mathbf{C}).\end{aligned}$$

Proof of Cross-Dot-Cross formula. The proof of equation 1.19 follows from a slightly tricky application of the "BAC-CAB" formula and the preceding remark on the order of operations.

$$\begin{aligned}(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D} &=(\mathbf{A}\times\mathbf{B}\times\mathbf{C})\cdot\mathbf{D}),\\ &=[(\mathbf{A}\cdot\mathbf{C})\mathbf{B}-(\mathbf{A}\cdot\mathbf{B})\mathbf{C}]\cdot\mathbf{D},\\ &=(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D})-(\mathbf{A}\cdot\mathbf{B})(\mathbf{C}\cdot\mathbf{D}),\\ &=\begin{vmatrix}\mathbf{A}\cdot\mathbf{C} & \mathbf{A}\cdot\mathbf{D}\\ \mathbf{B}\cdot\mathbf{C} & \mathbf{B}\cdot\mathbf{D}.\end{vmatrix}\end{aligned}$$

The special case of when $\mathbf{C}=\mathbf{A}$ and $\mathbf{D}=\mathbf{B}$ is even neater. We get

$$\begin{aligned}\|\mathbf{A}\times\mathbf{B}\|^2 &=(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{A}\times\mathbf{B}),\\ &=\begin{vmatrix}\mathbf{A}\cdot\mathbf{A} & \mathbf{A}\cdot\mathbf{B}\\ \mathbf{B}\cdot\mathbf{A} & \mathbf{B}\cdot\mathbf{B}\end{vmatrix},\\ \|\mathbf{A}\times\mathbf{B}\|^2 &=\|\mathbf{A}\|^2\|\mathbf{B}\|^2-\|\mathbf{A}\cdot\mathbf{B}\|^2, \qquad (1.20)\\ &=\|\mathbf{A}\|^2\|\mathbf{B}\|^2-\|\mathbf{A}\|^2\|\mathbf{B}\|^2\cos^2\theta\\ &=\|\mathbf{A}\|^2\|\mathbf{B}\|^2(1-\cos^2\theta),\\ &=\|\mathbf{A}\|^2\|\mathbf{B}\|^2\sin^2\theta.\end{aligned}$$

Taking the square root of both sides taking into account that the angle θ between two vector is between 0 and π so that the sine is positive, we get

$$\|\mathbf{A}\times\mathbf{B}\|=\|\mathbf{A}\|\|\mathbf{B}\|\sin\theta. \qquad (1.21)$$

The intermediate result

$$\|\mathbf{A}\times\mathbf{B}\|^2=\|\mathbf{A}\|^2\|\mathbf{B}\|^2-\|\mathbf{A}\cdot\mathbf{B}\|^2, \qquad (1.22)$$

is called **Lagrange's identity**.

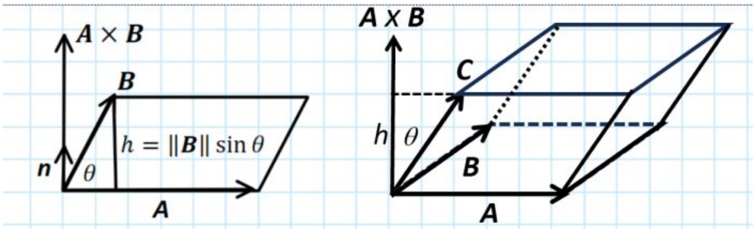


Fig. 1.10: Cross Product

1.4.2 Geometry of Cross Products

From figure 1.10 and equation 1.21 we see that the magnitude of the cross product of two vectors gives the area of the parallelogram subtended by the two vectors. This is a very important geometrical fact. If we use the notation $Area_{\square}$ for the subtended area, we have just shown that

$$Area_{\square} = \|\mathbf{A} \times \mathbf{B}\|. \tag{1.23}$$

The area $Area_{\triangle}$ of the subtended triangle would be half of this. Since the cross product vector is also perpendicular to the two vectors, we can visualize the $\mathbf{A} \times \mathbf{B}$ as the area of the subtended parallelogram times a unit vector $\mathbf{n}$ that is **normal** (perpendicular) to the plane containing the parallelogram.

$$\mathbf{A} \times \mathbf{B} = (Area_{\square})\,\mathbf{n}, \quad \text{where } \|\mathbf{n}\| = 1, \ \ \mathbf{n} \perp \mathbf{A}, \ \ \mathbf{n} \perp \mathbf{B}.$$

The orientation of the unit normal vector $\mathbf{n}$ is chosen according to the right hand rule. A plane is uniquely defined by a point and a vector perpendicular to the plane. Thus, the cross product is an essential tool to find equations of planes. By dividing a curved surface into a grid of infinitesimal parallelograms with areas given by cross products of infinitesimal vectors, we can integrate and thus find the full surface area. This will a major topic to be covered in section 4.7.

The diagram on the right of figure 1.10 shows a parallelepiped subtended by three vectors $\mathbf{A}, \mathbf{B}, \mathbf{C}$. The base of the parallelepiped is the area spanned by $\mathbf{A}$ and $\mathbf{B}$ and the height h is the projection of $\mathbf{C}$ onto the unit normal vector $\mathbf{n}$, that is, $h = \|\mathbf{C}\| \cos\theta$. Hence, the volume is

$$V = (Area_{\square})h = \|\mathbf{A} \times \mathbf{B}\|\,\|\mathbf{C}\| \cos\theta = (\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C}. \tag{1.24}$$

Thus, the volume subtended is the absolute value of the determinant of the matrix with rows given by the components of the three vectors. A parallelepiped can be divided into two congruent prisms, and as shown in figure 1.11, each prism can be divided into three tetrahedra of equal volumes. Three vectors $\mathbf{A}, \mathbf{B}, \mathbf{C}$, subtend a tetrahedron, a prism and a parallelepiped. The volumes are given by

Volume(Parallelepiped) $= |(\mathbf{ABC})|$,
Volume(Prism) $= \frac{1}{2}|(\mathbf{ABC})|$,
Volume(Tetrahedron) $= \frac{1}{6}|(\mathbf{ABC})|$.

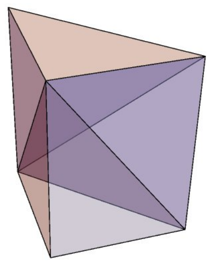


Fig. 1.11: Prism

Determinants were known to Leibnitz in the 1600's but cross products were not introduced until the 1840's by Hamilton in his treatment of quaternions. The formula for the cross product 1.15 in terms of cofactor expansion along a row, is due to Laplace. There are 6 different permutations of the vectors in the triple product $(\mathbf{ABC})$, three of which are even and three odd, thus, 3 of the permutations yield the volume and the other three, the negative of the volume. Cross product of vectors in $\mathbf{E}^2$ only makes sense if one embeds the vectors in $\mathbf{R}^3$ by taking the z-component equal to zero. One can then verify immediately, that the length of the cross product of to vectors is equal the absolute value of the corresponding 2×2 matrix, and this is equal to the area of parallelogram subtended by the two vectors. There is no cross product for vectors $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ in four dimensions, but the determinant with this row vectors does exist. In that case, absolute value of the 4×4 determinant gives the hypervolume.

Geometry Examples

Example Area of a triangle
Find the area of the triangle in $\mathbf{R}^2$ with vertices at $P(1,2), Q(3,5)$ and $R(3,8)$.
Solution:

$$\begin{aligned} \text{Let } \mathbf{A} =& \overrightarrow{PQ} = \langle 2,3,0\rangle, \\ \mathbf{B} =& \overrightarrow{PR} = \langle 2,6,0\rangle. \\ \text{Then } \mathbf{A}\times\mathbf{B} =& 6\,\mathbf{k}, \\ Area_{\triangle} =& \tfrac{1}{2}\|\mathbf{A}\times\mathbf{B}\| = 3. \end{aligned}$$

Example Find the volume of the tetrahedron spanned by the vectors $\mathbf{A} = \langle 1,2,3\rangle, \mathbf{B} = \langle -1,1,2\rangle$, and $\mathbf{C} = \langle 2,1,4\rangle$.
Solution:

$$\begin{aligned} Vol =& \tfrac{1}{6}|(\mathbf{ABC})|, \\ =& \tfrac{1}{6}\begin{vmatrix} 1 & 2 & 3 \\ -1 & 1 & 2 \\ 2 & 1 & 4 \end{vmatrix}, \\ =& \frac{9}{6} = \frac{3}{2} \end{aligned}$$

1.4.4 Example Distance from a point to a plane
Let P be a point not in a plane that passes through the points Q, R and S. Let

$$\mathbf{A} = \overrightarrow{QR}, \quad \mathbf{B} = \overrightarrow{QS}, \quad \mathbf{C} = \overrightarrow{QP},$$

The vector $\mathbf{A} \times \mathbf{B}$ is perpendicular to the plane. Divide this vector by its length to get a unit normal vector $\mathbf{N}$. Let θ be the angle between $\mathbf{C}$ and $\mathbf{N}$. The distance d from the point P to the plane is just the projection $d = \|\mathbf{C}\| \cos\theta$. We compute,

$$\mathbf{N} = \frac{\mathbf{A} \times \mathbf{B}}{\|\mathbf{A} \times \mathbf{B}\|}, \quad \|\mathbf{N}\| = 1.$$

$$d = \|\mathbf{C}\| \cos\theta = \mathbf{C} \cdot \mathbf{N},$$

$$d = \frac{\mathbf{C} \cdot (\mathbf{A} \times \mathbf{B})}{\|\mathbf{A} \times \mathbf{B}\|}, \tag{1.25}$$

$$d = \frac{|(\mathbf{ABC})|}{\|\mathbf{A} \times \mathbf{B}\|} \tag{1.26}$$

It is assumed of course, that $\|\mathbf{A} \times \mathbf{B}\| \neq 0$, else, the vectors would be parallel and they would not span a plane. In the last line we inserted an absolute value, since a distance can't be negative.

1.4.3 Physics of Cross Products

1.4.4 Torque

In elementary physics, the cross product manifests itself in many ways. Suppose we were trying to turn a bolt with a wrench with lever arm given by vector $\mathbf{r}$ with a force $\mathbf{F}$ applied at the farthest point on the wrench . If the force is at a right angle to the lever arm, the magnitude τ of the torque is given by the simple lever formula of Archimedes, $\tau = \|\mathbf{r}\| \, \|\mathbf{F}\|$. The MKS units of torque are Newton-meters. In the primitive American system, the units are foot-pounds. The units are the same as the units of energy, but torque is more like the rotational version of force, so we reserve Joules units only in the case when Newton-meters refers to energy or work. If the force is applied in the direction of the lever arm, the bolt will not rotate and the torque is zero. If the force is applied at an angle θ, only the component of the force $\|\mathbf{F}\| \sin\theta$ perpendicular to the lever arm, will make a contribution to the rotation. So the magnitude of the torque is

$$\tau = \|\mathbf{r}\| \, \|\mathbf{F}\| \sin\theta = \|\mathbf{r} \times \mathbf{F}\|$$

The torque is in fact defined as the vector

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} \tag{1.27}$$

Angular Momentum

If a particle of mass m is moving with velocity vector $\mathbf{v}$, the linear momentum $\mathbf{P}$ is defined by $\mathbf{P} = m\mathbf{v}$. Students are often introduced to Newton's second law of motion as $\mathbf{F} = m\mathbf{a}$. The actual law of motion $\mathbf{F} = d\mathbf{P}/dt$ written by Newton, is that force is the rate of change of momentum. This allows for a system in which the mass of the object is not constant, as in the case of a rocket. The rotational versions of the second law are given in relation to motion with a position vector **r** about a pivot point. The relevant quantities are,

$$\begin{aligned} \boldsymbol{\omega} &= \mathbf{r} \times \mathbf{v}, && \text{Angular velocity,} \\ \boldsymbol{\alpha} &= \mathbf{r} \times \mathbf{a}, && \text{Angular acceleration,} \\ \mathbf{L} &= \mathbf{r} \times \mathbf{P}, && \text{Angular momentum,} \\ \boldsymbol{\tau} &= \mathbf{r} \times \mathbf{F}, && \text{Torque,} \qquad (1.28) \\ \boldsymbol{\tau} &= \frac{d}{dt}\mathbf{L}, && \text{Second Law.} \qquad (1.29) \end{aligned}$$

Lorentz Force

The force exerted on a particle with charge q moving with velocity $\mathbf{v}$ in the presence of an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$ is given by the Lorentz force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

This force is the principle behind the design of cyclotrons and bubble chambers.

1.5 Lines and Planes

Starting about the 8th grade, students learn that the equation $z = mx + b$ represents a straight line with slope $m = \Delta z/\Delta x$ and z-intercept b. In $\mathbf{R}^3$ we have three variables so this definition of slope m is no longer valid. Also the natural extension of the general equation of a line $Ax + By = C$ to three variables would be $Ax + By + Cz = D$ but the latter describes a plane, not a line. To find equations of lines and planes in $\mathbf{R}^3$ we allude to the following two Euclidean postulates.

- Given a point and a direction, there is only one line passing through that point in that direction.
- Given a point and a direction, there is only one plane passing through that point and perpendicular to that direction.

We stipulate a direction by a vector. We think at this time it is appropriate to issue an early warning not to confuse lines with planes! It might sound odd to issue this warming since clearly lines are different than planes. Yet, year after year, there are students who respond with the equation of a plane when prompted for the equation of a line, and viceversa. The confusion is a fatal error. The source of the confusion is that equations of lines and planes require exactly the same data, namely, a point and a vector, as shown in figure 1.12.

1.5.1 Lines

Let $P(x_0, y_0, z_0) \in \mathbf{R}^3$ be a given point and $\mathbf{v} = \langle a, b, c\rangle$ a constant vector. Suppose that $\mathbf{X} = \langle x, y, z\rangle$ the position vector for an arbitrary point on the line $\mathcal{L}$ determined by P and $\mathbf{v}$. The notation $\mathbf{r} = \langle x, y, z\rangle$ is also commonly used.

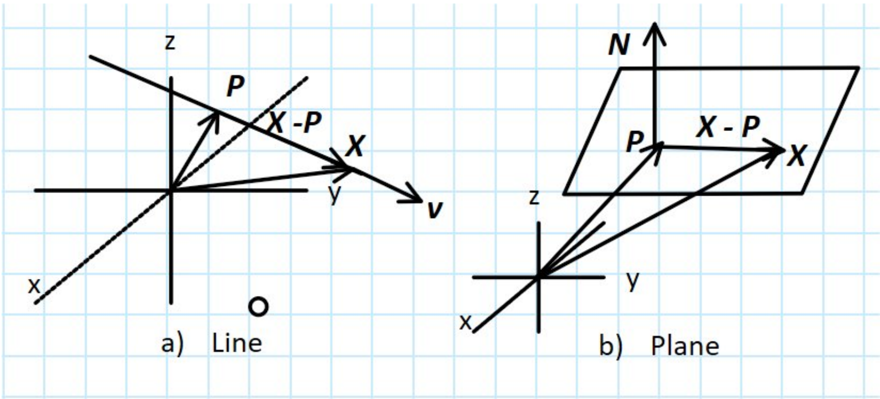


Fig. 1.12: Lines and Planes

Then, the vector $\overrightarrow{PX} = \mathbf{X} - \mathbf{P}$ must be a multiple of $\mathbf{v}$ as shown in figure 1.26; that is, $\mathbf{X} - \mathbf{P} = t\mathbf{v}$, where t is a real number. Therefore, the vector equation of the line must be

$$\mathbf{X}(t) = \mathbf{r}(t) = \mathbf{P} + t\mathbf{v}, \quad \text{where, } t \in \mathbf{R}. \tag{1.30}$$

A remark on notation is in order. Naming the position vector of an arbitrary point in $\mathbf{R}^3$ by $\mathbf{X}$ (or $\mathbf{x}$) is logistically a natural choice. Using $\mathbf{r}$ instead is also a good choice because the length r of the vector would represent the distance of the point to the origin, which would be historically consistent with Newton's $1/r^2$ law. Equation 1.30 says that to get to a point on $\mathcal{L}$ with position vector $\mathbf{r}$, one starts at the initial position vector $\mathbf{P}$, then flows by t units along the vector $\mathbf{v}$. The variable t is called a **parameter**, which we may interpret as being time in some unit. An equation for a position vector in $\mathbf{R}^3$ depending on one parameter describes a 1-dimensional continuum, as it should be for a line. In practice, equation 1.30 is most useful when written in terms of its components. We have

$$\langle x, y, z\rangle = \langle x_0, y_0, z_0\rangle + t\langle a, b, c\rangle, \quad \text{hence,}$$

$$\begin{aligned} x =& x_0 + at,\\ y =& y_0 + bt,\\ z =& z_0 + ct. \end{aligned} \tag{1.31}$$

So, the single vector equation 1.30 of the line is equivalent to three **scalar parametric equations** 1.31. Unless otherwise specified in the problem, all equations of lines in this course should be answered in parametric form.
We can eliminate the parameter and reduce to a system of two equations

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}. \tag{1.32}$$

This is called the **symmetric** form of the equation of the line. Parametric equations are the preferred form for at least two reasons. First, the parametric form has built-in dynamics. It is interpreted not just as a graph of a line, but rather, as the trajectory of a particle moving along a straight line. In fact, getting a bit ahead of ourselves and take derivative of equation 1.30 with respect to t, we see that ($\mathbf{v}$) is the velocity vector.

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}.$$

Since $\mathbf{v}$ is constant, the second derivative is zero, which means the acceleration is zero and so is the force. This is consistent with Newton's first law. A particle with constant velocity moves along a straight line unless acted upon by an external force. The parametric equation is not unique. For example, we could replace $\mathbf{v}$ by $2\mathbf{v}$, and the particle would still be moving along the same straight line, but now with twice the velocity.

A second reason why parametric equations are more suitable for physics and engineering is that the symmetric form is not quite right in the case when one or two of the components of the direction vector is zero. For example, if $a = 0$, then the the first term of equation 1.32 must be separated to read $x = 0$. It is common to abuse notation and leave zero components in the denominator of equation 1.32 with the understanding this just means that the variable in the numerator is constant. Of course, the equations must be consistent with $y = mb + b$ in two variables. This is easily seen from the symmetric equation

$$\begin{aligned} \frac{x - x_0}{a} &= \frac{y - y_0}{b}, \\ \frac{y - y_0}{x - x_0} &= \frac{b}{a} = m. \end{aligned}$$

This is just the point-slope formula for a line passing through the point (x_0, y_0) and slope $m = b/a$ as it should be since that is the slope of the direction vector $\langle a, b \rangle$.

1.5.2 Planes

Equations of planes are just a easy to construct. Let $P(x_0, y_0, z_0) \in \mathbf{R}^3$ be a given point and $\mathbf{N} = \langle A, B, C \rangle$ a constant vector. Suppose that $\mathbf{X} = \langle x, y, z \rangle$ is an arbitrary point on the plane $\mathcal{P}$ determined by P and normal vector $\mathbf{N}$. Then, the vector $\overrightarrow{PX} = \mathbf{X} - \mathbf{P}$ must be perpendicular to $\mathbf{N}$ as shown in figure 1.26; that is,

$$(\mathbf{X} - \mathbf{P}) \cdot \mathbf{N} = 0. \tag{1.33}$$

In terms of the components, we have

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0 \tag{1.34}$$

Distributing the multiplications, we get the standard **linear equation** of a plane

$$Ax + By + Cz = D, \tag{1.35}$$

where $D = Ax_0 + By_0 + Cz_0$ is just a constant. There is no particular geometric significance to D other than if $D = 0$, the plane goes through the origin, and the larger the absolute value of D, the further the plane is from the origin.

Example Line passing through two points

Find the equation of the line passing through $P(1, 3, -1)$ and $Q(2, 3, 5)$.

Solution: Let $\mathbf{v} = \overrightarrow{PQ} = \langle 1, 0, 6\rangle$. Then, equation 1.31 gives immediately

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

The symmetric equation would be

$$\frac{x-1}{1} = \frac{y-3}{0} = \frac{z+1}{6}$$

Notice that when $t = 0$ we are at the point P and when $t = 1$ we are at the point Q. This can't be a coincidence. Let P and Q are arbitrary points and choose the direction vector as $\mathbf{v} = \overrightarrow{PQ}$. Then we have,

$$\begin{aligned} \mathbf{r}(t) =& \mathbf{P} + t\mathbf{v}, \\ =& \mathbf{P} + t\mathbf{Q} - \mathbf{P}, \\ \mathbf{r}(t) =& (1-t)\mathbf{P} + t\mathbf{Q}. \quad \text{Therefore,} \qquad (1.36) \\ \mathbf{r}(0) =& \mathbf{P}, \\ \mathbf{r}(1) =& \mathbf{Q}. \end{aligned}$$

So, with the choice $\mathbf{v} = \overrightarrow{PQ}$, the values $t \in [0, 1]$ describe the line segment $\overline{PQ}$. We will need this fact often when performing line integrals later. The fact is also used in basic computer graphics like Paint, to render line segments.

Example Plane containing three points

Find the equation of the plane that passes through three points, say $P(3, 0, 0)$, $Q(0, 2, 0)$ and $Q(0, 0, 6)$.

Solution: Let $\mathbf{A} = \overrightarrow{PQ}$ and $\mathbf{B} = \overrightarrow{PR}$. We have,

$$\begin{aligned} \mathbf{P} =& \langle 2, 0, 0\rangle \\ \mathbf{A} =& \langle -3, 2, 0\rangle \\ \mathbf{B} =& \langle -3, 0, 6\rangle \\ \mathbf{A} \times \mathbf{B} =& \langle 12, 18, 6\rangle = 6\langle 2, 3, 1\rangle, \\ \mathbf{N} =& \langle 2, 3, 1\rangle, \\ 2(x-3) + 3(y-0) + 1(z-0) =& 0, \quad \text{from equation 1.35} \\ 2x + 3y + z =& 6 \end{aligned}$$

A neat way to write the equation of plane that has non-zero coordinate intercepts $(a, 0, 0), (0, b, 0), (0, 0, c)$ is

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

Example Line parallel to another line
Find the line through $P(-6, 2, 3)$ and parallel to the line $\frac{1}{2}x = \frac{1}{3}y = z + 1$.
Solution

$$\begin{aligned}
\frac{x}{2} =& \frac{y}{3} = \frac{z+1}{1},\\
\mathbf{v} =& \langle 2, 3, 1 \rangle, \quad \text{from equation } 1.32\\
x =& -6 + 3t,\\
y =& 2 + 3t,\\
z =& 3 + t
\end{aligned}$$

Example Plane parallel to another plane
Find the equation of the plane through the point $P(3, -2, 8)$ and parallel to the plane $z = x + y$.
Solution. The required plane is parallel to the given plane $x + y - z = 0$ so we can use the same normal $\mathbf{N} = \langle 1, 1, -1 \rangle$. So we have,

$$\begin{aligned}
1(x-3) + 1(y+2) - (z-8) =& 0,\\
x + y - z = -7
\end{aligned}$$

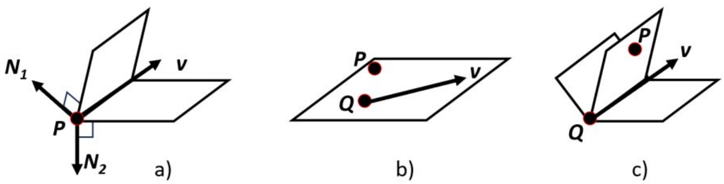


Fig. 1.13: a) Line of intersection of two planes. b) Plane containing a point and a line. c) Plane containing a point and the line of intersection of two planes

Example Line of intersection of two planes
Find the line of intersection of the plane $\mathcal{P}_1 : \; x + 2y + 3z = 1$ with the plane $\mathcal{P}_2 : \; x - y + z = 1$.
Solution: First we need a point on the intersection of the planes. Since we have two equations and three unknowns, the system is under-determined and we have infinite number of solutions - in fact, we have a whole line worth of solutions. All we need is one. We could, for example, find the solution corresponding to choosing $z = 0$. This gives two equations and two unknowns that can be solved by any algebraic method of your choice.

$$\begin{aligned}
x + 2y =& 1,\\
x - y =& 1,\\
x =& 1, \; y = 0,\\
\mathbf{P} =& \langle 1, 0, 0 \rangle
\end{aligned}$$

A line on a plane is perpendicular to the normal to the plane. Since the line here lies on both planes, the direction vector is perpendicular to the normals of both planes. We have

$$\begin{aligned}\mathbf{N}_1 =&\langle 1,2,3\rangle,\\ \mathbf{N}_2 =&\langle 1,-1,1\rangle,\\ \mathbf{v}=\mathbf{N_1}\times\mathbf{N_2} =&\langle 5,3,-3\rangle,\end{aligned}$$

Therefore, the equation of the line of intersection is

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

Example Plane containing a point and a line
Find the equation of the plane passes through the point $P(3,5,-1)$ and contains the line $x=4-t,\ y=2t-1, z=-3t$.
Solution: It is important to visualize these analytic geometry problems as in figure 1.13(b) to develop a solution strategy. The data required to find the equation of a plane is a point and a normal vector. A point P is given. For the normal vector we need to find two vectors on the plane and use the cross product. One vector on the plane is the direction vector $\mathbf{v}=\langle -1,2,-3\rangle$. For another vector on the plane we use $\mathbf{w}=\overrightarrow{QP}$, where Q is a point on the line. We have

$$\begin{aligned}\mathbf{P} =&\langle 3,5,-1\rangle,\\ \mathbf{Q} =&\langle 4,-1,0\rangle, \quad \text{Point on the line with } t=0,\\ \mathbf{w}=\overrightarrow{QP} =&\langle 1,-6,1\rangle,\\ \mathbf{v} =&\langle -1,2,-3\rangle,\\ \mathbf{N}=\mathbf{w}\times\mathbf{v} =&\langle 16,2,-4\rangle.\end{aligned}$$

So, the equation of the plane is

$$\begin{aligned}16(x-3)+2(y-5)-4(z+1) =&0,\\ 16x+2y-4z =&62.\end{aligned}$$

Example Plane containing a point and the line of intersection of two planes
Find the equation of the plane through a point P and containing the line of intersection of two other planes.
Solution. We present only the strategy for solving the problem. We need a point on the plane and a vector normal to the plane. For the normal vector we first find the equation of the line of intersection of the two planes as shown in 1.13(c). The problem then reduces to one just like the previous one.

1.6 Quadric Surfaces

Following a parallel treatment of curves in elementary algebra, we now move from linear to quadratic equations in $\mathbf{R}^3$. The key to visualization of quadric surfaces is having good familiarity with conic sections. The fundamental equations for conic sections are:

- Parabola: $z = x^2$.
- Ellipse: $\frac{x^2}{a^2} + \frac{z^2}{c^2} = 1$.
- Hyperbola: $\frac{x^2}{a^2} - \frac{z^2}{c^2} = c$.

  As shown in figure 1.14(b), if $c > 0$ the hyperbola opens left and right, if $c = 0$ one gets two straight lines $y = \pm x$, and if $c < 0$, the hyperbola opens up and down.

Here, we have intentionally described the conics in x-z coordinates in anticipation to making an analogy with the extensions to the corresponding surfaces. A good starting point is this: If in an equation of a curve in x-z coordinates, one replaces x^2 by the square of the polar coordinates distance to the origin $r^2 = x^2 + y^2$, one gets a **surface of revolution** with the same curve as generators and symmetry about the z-axis. The cross-sections of the surface with horizontal planes z =constant are circles. Thus, for example

Parabola	$z = x^2$	$\mapsto$	$z = x^2 + y^2$	Paraboloid
Circle	$x^2 + z^2 = 1$	$\mapsto$	$x^2 + y^2 + z^2 = 1$	Sphere
Hyperbola	$x^2 - z^2 = 1$	$\mapsto$	$x^2 + y^2 - z^2 = 1$	Hyperboloid of 1 sheet
Two lines	$x^2 - z^2 = 0$	$\mapsto$	$x^2 + y^2 - z^2 = 0$	Cone
Hyperbola	$x^2 - z^2 = -1$	$\mapsto$	$x^2 + y^2 - z^2 = -1$	Hyperboloid of 2 sheets

If instead one replaces x^2 by $(\frac{x^2}{a^2} + \frac{y^2}{b^2})$ the horizontal cross-sections become elliptical. Horizontal cross-sections are also called **level curves** or **contours.**

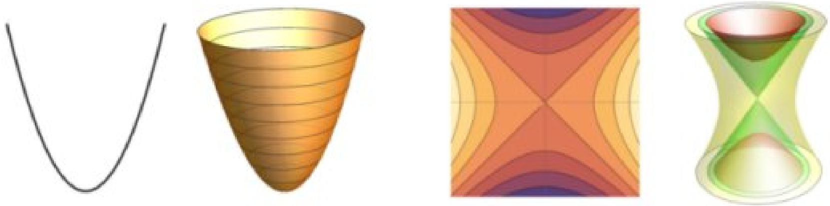


Fig. 1.14: a) Paraboloid. b) Hyperboloids - Cone

With the simple observation above, we can now identify the following quadrics as shown in figure 1.14:

- $z = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$. Elliptic paraboloid. The paraboloid opens "upwards" along the z-axis. We use quotation marks because coordinate axes can be rotated at will as long as one preserves the right-hand orientation. Also, if in the equation one interchanges two variables, say x and z, the shape of the surface would be identical but the central axis of symmetry would now be the x-axis. For example, $x = \frac{y^2}{a^2} + \frac{z^2}{b^2}$ is a paraboloid in which the x-axis is the axis of symmetry. If $a = b$, we would have a circular paraboloid. The language is inconsistent here in regards to the suffix of the first descriptor. It really should be called a "circlic paraboloid", but nobody uses that!

- $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$. Elliptic hyperboloid of one sheet. In the figure 1.14, the hyperboloid of one sheet is the one that looks like a nuclear plant cooling tower. The cross-sections with horizontal planes $z =$ constant are ellipses. The cross-sections with vertical planes $x =$ constant, or $y =$ constant are hyperbolas. For instant algebraic pattern recognition, the equation is quadratic in all three variables, the right hand side is a positive constant, and only one variable has a minus sign. One minus sign, one sheet.

- $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = -1$. Elliptic hyperboloid of two sheets. In the figure, the hyperboloid of two sheets is the one that looks like two cups, one opening "up" and one opening "down." The cross-sections (when they exist) with horizontal planes $z =$ constant are ellipses. The cross-sections with vertical planes $x =$ constant, or $y =$ constant are hyperbolas. For instant algebraic pattern recognition, the equation is quadratic in all three variables, the right hand side is a negative constant, and only one variable has a minus sign. Two minus signs, two sheets.

- $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 0$. Elliptic cone. The most common cones are circular as in the equation $z^2 = x^2 + y^2$. The cone had two funnels, one opens up and one opens down. The graph is not a function of x and y because there are two values of z for each point on the xy-plane other than the origin. In the language of pre-calculus, it fails the vertical line test. If one desires to get a function, one one must extract the square root and choose a sign. Thus, for example $z = \sqrt{x^2 + y^2}$ would be a circular cone with the funnel opening upwards. In this common example, notice that the cross-section with the plane $y = 0$ gives two straight lines $z = \pm x$ at 45 degrees.

- $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$. Ellipsoid. All the cross-sections with constant planes are ellipses. If $a = b = c = R$, the ellipsoid is a sphere of radius R. If the ellipsoid is obtained by rotating an ellipse along its minor axis then it is "pancake-shaped" like an M&M and is called an oblate spheroid. If the ellipsoid is "cigar-shaped" like a rugby ball, it is called a prolate spheroid.

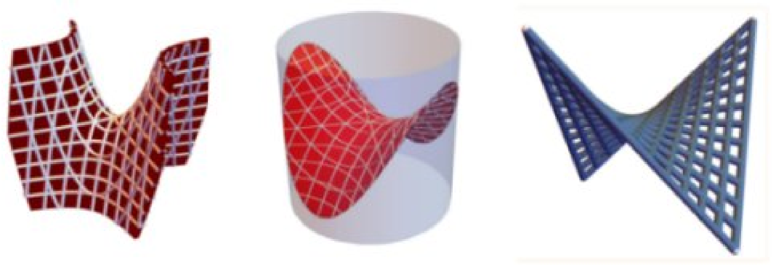


Fig. 1.15: Hyperbolic Paraboloid

Saddles: These quadric surfaces do not arise from extensions of rotating plane curves. The generic equation of a saddle is

- $z = \dfrac{x^2}{a^2} - \dfrac{y^2}{b^2}$. Hyperbolic paraboloid.

  The level curves $z = c$, where c is a constant, depend on the value of c. If $c > 0$ the level curves are hyperbolas opening left and right. If $c = 0$, the level "curve" is two straight lines $y = \pm(a/b)x$. If $c < 0$, the level curves are hyperbolas opening up and down. Thus, if z represents the height of the function, we can think of the plane curves in figure 1.14(b) as being a topographical map of the surface. The cross-section by the vertical plane $y = 0$ is a parabola that opens up and the cross-section $x = 0$ is a parabola that opens down. Because of this, the technical name of the surface is a hyperbolic paraboloid, but a saddle is a perfectly good mathematical term. The plot on the left of figure 1.15 shows the surface $z = x^2 - y^2$ with tube-enhanced level curves. The graph in the middle is given by exactly the same equation, but the domain region in the xy-plane is a circle.

- $z = 2xy$. Hyperbolic paraboloid

  The level curves $2xy = c$ of the surface $z = 2xy$ look exactly like the level curves of $z = x^2 - y^2$ but rotated 45 degrees. Thus, when one renders the surface over a standard square domain centered at the origin, the graph appears to have pointed corners as in the surface depicted on the right in figure 1.15.

Cylinders. An equation of a curve in two variables in $\mathbf{R}^3$ actually represents a surface obtained by extruding the curve in the direction of the missing coordinate. These surfaces are called **cylinders**. Again, there is an inconsistency with the "oid " suffix.

- $y = x^2$. Parabolic cylinder.

  The shape looks like a bent page. The equation is missing the z-coordinate, so the graph is just the standard parabola extruded along the z-axis.

- $x^2 + y^2 = R^2$. Circular cylinder.

  This is the shape that one commonly conjures when thinking of a cylinder. It looks like a tube.

- $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$. Elliptic cylinder.
- $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$. Hyperbolic cylinder.
- $y = f(x)$. Cylinder with level curve in the shape of the curve $y = f(x)$

The quadric surfaces listed in this section are all given in generic form. If any variable such as x in the equations is replaced by $(x - h)$, the shape does not alter in any way. The graph is just a translation of the original graph, in this case, h units to the right. It is important that the student learns to instantly identify a given quadric surface by the shapes of the conic cross sections and the algebraic structure of the equation. Quadric surfaces will appear very often in later chapter as examples to illustrate new concepts.

Chapter 2

Vector Functions

A vector function is a function that takes one or more variables as input and outputs a vector. More specifically, a vector function is a map $f : \mathbf{R}^m \to \mathbf{R}^n$. The set of all possible inputs is called the **domain** and the set of all possible outputs is called the **range** of the function. The domain could be all of $\mathbf{R}^m$ or a subset thereof. If $n = 1$, that is, when the output are real numbers, it is an overkill to think of this as a vector function since for $\mathbf{R}$, the vector operations are the same as the operations of real numbers. So, a function $f : \mathbf{R}^m \to \mathbf{R}$ is called a **real-valued function**. In the first two semesters of calculus we consider single-variable, real-valued functions $f : \mathbf{R} \to \mathbf{R}$. In this course we are interested in the following cases:

- $f : \mathbf{R} \to \mathbf{R}^n$. These represent curves in parametric form. If $n = 2$ we have parametric plane curves and if $n = 3$, we have parametric space curves. We well cover these in section 2.1.

- $f : \mathbf{R}^n \to \mathbf{R}$. These are called multivariate, real-valued functions. When $n = 2$, that is, $f : \mathbf{R}^2 \to \mathbf{R}$ the functions can be visualized as surfaces $z = f(x, y)$ in 3-space. We will study these extensively in chapter 3.

- $f : \mathbf{R}^n \to \mathbf{R}^n$. These transformations are called change of coordinates. We discuss briefly the two main examples in section 2.5.

- $f : \mathbf{R}^2 \to \mathbf{R}^3$. These represent surfaces in parametric form. We provide a short introduction to parametric surfaces in section 2.6 and treat the subject of surface integrals in chapter 5

  The main subject of the current chapter is the study of parametric curves in $\mathbf{R}^3$, culminating on a derivation of Kepler's laws.

2.1 Space Curves in $\mathbf{R}^3$

A space curve in $\mathbf{R}^3$ is a map $\mathbf{r} : \mathbf{R} \to \mathbf{R}^3$

$$\begin{aligned} t &\xrightarrow{\mathbf{r}} \langle x, y, z\rangle, \\ t &\mapsto \langle x(t), y(t), z(t)\rangle, \\ \mathbf{r}(t) &= \langle x(t), y(t), z(t)\rangle, \quad \text{or}, \\ \mathbf{r}(t) &= x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}. \end{aligned} \tag{2.1}$$

The map assigns to each value of the variable t, a position vector $\mathbf{r}(t)$. Below is a list of the most common curves you are expected to recognize when you step on one.

1. Lines

   $$\mathbf{r}(t) = \langle x_0 + at, y_0 + bt, z_0 + ct\rangle \tag{2.2}$$

   The single parametric vector equation is clearly equivalent to three parametric scalar equations

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

2. Circles

   Circles are plane curves. We present only the equation of circles parallel to one of the coordinate planes. For example, a circle of radius R on the plane $z = c$, can be represented by the vector parametric equation

   $$\mathbf{r}(t) = \langle R\cos t, R\sin t, c\rangle \tag{2.3}$$

   The vector parametric equation is equivalent to the three scalar equations

   $$\begin{aligned} x(t) =& R\cos t, \\ y(t) =& R\sin t, \\ z(t) =& c. \end{aligned}$$

   It is sometimes possible to **eliminate the parameter** to get a Cartesian equation that might be more familiar. In the case in question, the key is to recall Pythagoras's theorem in trigonometric form. Indeed, applying the Pythagoras trigonometric identity $\cos^2 t + \sin^2 t = 1$, we get immediately

   $$x^2 + y^2 = R^2, \quad z = c.$$

3. Ellipses

   Ellipses are also plane curves. We present only the equation of ellipses parallel to one of the coordinate planes. For example, an ellipse on the plane $z = c$, can be represented by the vector parametric equation

   $$\mathbf{r}(t) = \langle a\cos t, b\sin t, c\rangle \tag{2.4}$$

The vector parametric equation is equivalent to the three scalar equations

$$
\begin{aligned}
x(t) =& a\cos t,\\
y(t) =& b\sin t,\\
z(t) =& c.
\end{aligned}
$$

To verify this represents an ellipse we eliminate the parameter, again evoking the fundamental trigonometric identity $\cos^2 t + \sin^2 t = 1$. We have

$$
\begin{aligned}
\frac{x}{a} &= \cos t, \quad \frac{x^2}{a^2} = \cos^2 t,\\
\frac{y}{b} &= \sin t, \quad \frac{y^2}{b^2} = \sin^2 t, \quad \text{which gives,}\\
\frac{x^2}{a^2} + \frac{y^2}{b^2} &= 1, \quad z = c.
\end{aligned}
$$

If we want the center shifted to (x_0, y_0, c) we just write

$$
\begin{aligned}
x(t) =& x_0 + a\cos t,\\
y(t) =& y_0 + b\sin t,\\
z(t) =& c.
\end{aligned}
$$

4. Hyperbolas

As above, we present only the equation of hyperbolas parallel to one of the coordinate planes. It is often the case that students coming into this course have not been properly exposed to hyperbolic functions of have little recollection of them. To partially ameliorate this deficiency, we present a brief review. Define, the hyperbolic cosine and hyperbolic sine functions as

$$
\cosh x = \frac{e^x + e^{-x}}{2}, \tag{2.5}
$$

$$
\sinh x = \frac{e^x - e^{-x}}{2}, \tag{2.6}
$$

We probably see a hyperbolic cosine function every day, because this is the function describing the shape of a hanging cable. Now, here is a picture 2.1 of something you don't see every day. This is the most beautiful hyperbolic cosine function in the world. It is located at an idyllic place in the Andes mountains near Bogotá, designed by architect Martha Lugo. It follows immediately that

$$
\begin{aligned}
\frac{d}{dx}\cosh x = \sinh x,\\
\frac{d}{dx}\sinh x = \cosh x,
\end{aligned}
$$

Fig. 2.1: $y = \cosh x$

A short computation gives

$$\begin{aligned}\cosh^2 x - \sinh^2 &= \frac{1}{4}[(e^x + e^{-x})^2 - (e^x - e^{-x})^2],\\ &= \frac{1}{4}[e^{2x} + 2e^x e^{-x} + e^{-2x} - (e^{2x} - 2e^x e^{-x} + e^{-2x})],\\ &= \frac{1}{4}(4e^x e^{-x}),\end{aligned}$$

hence

$$\cosh^2 x - \sinh^2 x = 1$$

This is the hyperbolic analog of the trigonometric Pythagoras' theorem. For each formula in trigonometry, there is a corresponding formula for hyperbolic functions. There is a $(\tanh x)$ function, a $(\operatorname{sech} x)$ function, and other co-hyperbolic functions defined exactly the same way as in trigonometry. There are sum and difference formulas, double angle formulas, half angle formulas, and inverse hyperbolic formulas. For now, the few facts listed above is all we need.

A hyperbola on the plane $z = c$, can be represented by the vector parametric equation

$$\mathbf{r}(t) = \langle a\cosh t, b\sinh t, c\rangle \tag{2.7}$$

The vector parametric equation is equivalent to the three scalar equations

$$\begin{aligned}x(t) =& a\cosh t,\\ y(t) =& b\sinh t,\\ z(t) =& c.\end{aligned}$$

To verify this represents a hyperbola, we eliminate the parameter, but this time we use the fundamental hyperbolic identity $\cosh^2 t - \sinh^2 t = 1$. We have

$$\begin{aligned}\frac{x}{a} &= \cosh t, \quad \frac{x^2}{a^2} = \cosh^2 t,\\ \frac{y}{b} &= \sinh t, \quad \frac{y^2}{b^2} = \sinh^2 t, \quad \text{which gives,}\end{aligned}$$

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad z = c.$$

Note. Hyperbolic functions will be used extensively in this course, so this might be a good time to review that appropriate section of the fat calculus textbook. An alternative parametrization of the hyperbola above is given by

$$\mathbf{r}(t) = \langle a \sec t, b \tan t, c \rangle$$

Here one uses the fundamental identity $\sec^2 t - \tan^2 t = 1$ to eliminate the parameter.

5. Parabolas

Again, we only present a parametrization of a parabola on a plane parallel to one of the coordinate planes. On a plane, the equations of parabolas are actually functions of x or y. When a curve on a plane parallel to the xy-plane is expressed by an actual function (of x or y), there is always a trivial parametrization. For example, the parabola $x = ay^2 + by + c$ on the plane $z = d$ can be written as

$$\mathbf{r}(t) = \langle at^2 + bt + c, t, d \rangle \tag{2.8}$$

In general, if $y = f(x)$, one can embed the plane curve on $\mathbf{R}^3$ by the trivial parametrization

$$\mathbf{r}(t) = \langle t, f(t), 0 \rangle$$

6. Helix

The basic curves above are all plane curves. The present is the first example in this list that represents a true space curve. The generic equation is

$$\mathbf{r}(t) = (a \cos t, a \sin t, ct). \tag{2.9}$$

This curve is called a **circular helix**. Geometrically, we may view the curve as the path described by the hypotenuse of a triangle with slope c, which is wrapped around a circular cylinder of radius a. The projection of the helix onto the xy-plane is a circle of radius a and the curve rises at a constant rate c in the z-direction, as shown in figure 2.2. The circular helix will serve as a good example to illustrate the dynamics of a particle moving along a curve, as will be discussed later on this chapter.

The curve

$$\mathbf{r}(t) = (a \cosh t, a \sinh t, ct). \tag{2.10}$$

is called a **hyperbolic helix**. It has a similar interpretation as a curve describing a hyperbola in xy-coordinates as it climbs at a constant rate up a hyperbolic cylinder.

7. Twisted Cubic

This example illustrates the simplest cubic curve in $\mathbf{R}^3$. The vector equation is

$$\mathbf{r}(t) = t\,\mathbf{i} + t^2\,\mathbf{j} + t^3\,\mathbf{k} \tag{2.11}$$

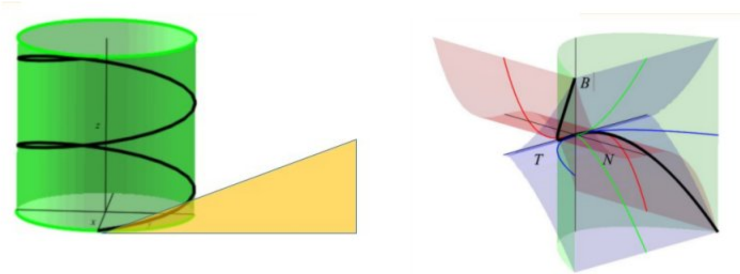


Fig. 2.2: a) Circular Helix b) Twisted Cubic

This is equivalent to the three scalar equations $x = t,\ y = t^2,\ z = t^3$. It follows that

$$y = x^2, \quad z = x^3, \quad z = y^{3/2}$$

This means that a particle moving on a trajectory along the twisted cubic is constrained to a parabolic cylinder in the xy-coordinates, a cubic cylinder on the xz-coordinates, and a half-odd power cylinder in the yz-coordinates. Here we are recalling that an an equation missing one coordinate represents a generalized cylinder obtained by extruding that curve in the direction of the missing coordinate. In other words, when projected onto the coordinate planes, the graph looks a parabola, a cubic and a half-odd power function respectively. This very neat behavior is illustrated in the fancy figure 2.2(b). In the figure we used an orthonormal frame $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ instead of the standard orthonormal basis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ for reasons that will become clear later in this chapter.

This concludes the set of basic curves students are expected to recognize and visualize on sight in this course. If a curve is not one of these, students may avail themselves of Maple, Mathematica, Matlab, or other graphing software.

2.2 Calculus of Curves

We finally start doing some calculus. I think it is more instructive to present the whole content in a "super-lecture" to get a holistic view of the theory, then we go back and do more specific examples. Since all the concepts arise from physics, it is my view that not covering the physics thoroughly would be a great disservice.

Let $\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}$.

2.2.1 Definition Let t_0 be a particular value of t. We define the limit as t approaches t_0 as

$$\lim_{t \to t_0} \mathbf{r}(t) = \lim_{t \to t_0} x(t)\,\mathbf{i} + \lim_{t \to t_0} y(t)\,\mathbf{j} + \lim_{t \to t_0} x(t)\,\mathbf{k} \tag{2.12}$$

The limit of the vector equation exists only if each of the three limits of the components exist. If the limit exists, then the computation reduces to three

single-variable calculus computations of limits. All the rules and theorems about limits for functions of one variable apply.
Example Let

$$\mathbf{r}(t) = \frac{\sin t}{t}\,\mathbf{i} + \frac{t^2-1}{t+1}\,\mathbf{j} + \frac{e^t-1}{t}\,\mathbf{k}$$

.

Then by using L'Hôpital's rule on the first and third slots, we get

$$\begin{aligned}
\lim_{t\to 0}\mathbf{r}(t) &= \lim_{t\to 0}\left[\frac{\sin t}{t}\right]\mathbf{i} + \lim_{t\to 0}\left[\frac{t^2-1}{t+1}\right]\mathbf{j} + \lim_{t\to 0}\left[\frac{e^t-1}{t}\right]\mathbf{k},\\
&= \lim_{t\to 0}\left[\frac{\cos t}{1}\right]\mathbf{i} + \lim_{t\to 0}\left[\frac{t^2-1}{t+1}\right]\mathbf{j} + \lim_{t\to 0}\left[\frac{e^t}{1}\right]\mathbf{k},\\
&= 1\,\mathbf{i} - 1\,\mathbf{j} + 1\,\mathbf{k}.
\end{aligned}$$

Definition We say that $\mathbf{r}(t)$ is continuous at t_0 if
a) $\mathbf{r}(t)$ is defined.
b) $\lim\limits_{t\to t_0}\mathbf{r}(t)$ exists.
c) $\lim\limits_{t\to t_0}\mathbf{r}(t) = \mathbf{r}(t_0)$

In other words, the vector function is continuous if each the component functions is continuous. All theorems and properties of continuous functions for a single variable apply to each to the component functions.

Definition We say that $\mathbf{r}(t)$ is differentiable at t_0 if each of the component functions is differentiable at t_0. Again, all theorems, and formulas of differentiation of a function of a single variable apply to each component function. To find the derivative of $\mathbf{r}(t)$ we just take the derivative of each component. We will interpret the parameter t as time in some appropriate unit. We have the following definitions

$$\mathbf{v}(t) = \frac{d\mathbf{r}}{dt} = \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right\rangle, \qquad \text{Velocity,} \tag{2.13}$$

$$\mathbf{a}(t) = \frac{d^2\mathbf{r}}{dt^2} = \left\langle \frac{d^2x}{dt^2}, \frac{d^2y}{dt^2}, \frac{d^2z}{dt^2} \right\rangle, \quad \text{Acceleration,} \tag{2.14}$$

$$v = \|\mathbf{v}\| = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{d^2z}{dt}\right)^2}, \quad \text{Speed.} \tag{2.15}$$

There is a clear difference between velocity which is a vector and speed which is a scalar. Conflating the two leads to fatal mistakes. At each point of the curve, the velocity vector is tangent to the curve and thus the velocity constitutes a "vector field" representing the velocity flow along that curve.
Integration of a vector function is also defined in the obvious way, namely

$$\int_a^b \mathbf{r}(t)\,dt = \left(\int_a^b x(t)\,dt\right)\mathbf{i} + \left(\int_a^b y(t)\,dt\right)\mathbf{j} + \left(\int_a^b z(t)\,dt\right)\mathbf{k}. \tag{2.16}$$

Example Projectile Motion
Suppose a projectile is launched from a position $\mathbf{r}(0) = x_0\mathbf{i} + y_0\mathbf{j}$ with a speed v_0 at an angle θ_0. Then, the initial velocity vector is

$$\mathbf{v}(0) = v_0 \cos\theta_0\, \mathbf{i} + v_0 \cos\theta_0\, \mathbf{j}.$$

Ignoring friction, the only force acting on the projectile is gravity. The acceleration due to gravity is of magnitude $a = g$ and it point downwards. We can easily integrate the equations of motion. We use the prescribed initial conditions to evaluate the constants of integration

$$\begin{aligned}
\mathbf{a}(t) &= 0\,\mathbf{i} - g\,\mathbf{j},\\
\mathbf{v}(t) &= C\,\mathbf{i} + (-gt + D)\,\mathbf{j},\\
&= v_0\cos\theta_0\mathbf{i} + (-gt + v_0\sin\theta_0)\,\mathbf{j}, \quad \text{(from the initial conditions for velocity)},\\
\mathbf{r}(t) &= (v_0\cos\theta_0\, t + E)\mathbf{i} + (-\frac{1}{2}gt^2 + v_0\sin\theta_0\, t + F)\,\mathbf{j},\\
&= (v_0\cos\theta_0\, t + x_0)\mathbf{i} + (-\frac{1}{2}gt^2 + v_0\sin\theta_0\, t + y_0)\,\mathbf{j},
\end{aligned}$$

In component notation, the motion is parabolic according to the equations

$$\begin{aligned}
x(t) &= v_0\cos\theta_0\, t + x_0,\\
y(t) &= -\frac{1}{2}gt^2 + v_0\sin\theta_0\, t + y_0.
\end{aligned} \tag{2.17}$$

This represents a complete solution to projectile motion with no friction, under a constant gravitational field. All other physics related to this model such as maximum range and height can be extracted from the equations. For example, if the particle is launched from the origin, the range is obtained by setting $y = 0$, which gives

$$t = 0, \quad t = \frac{2v_0}{g}\sin\theta_0$$

Substituting into $x(t)$, we get the range R,

$$R = \frac{2v_0^2}{g}\sin\theta_0\cos\theta_0 = \frac{v_0^2}{g}\sin 2\theta_0.$$

The maximum range (without friction) is obtained when $\sin 2\theta_0 = 1$, that is, when $\theta_0 = \pi/4$. This is a well-known result in first-year physics.
Basically, the projectile moves at a constant speed in the horizontal direction, but behaves like a free-fall particle in the vertical direction.

2.3 Arc Length and Curvature

Let $\mathbf{r}(t)$ represent a space curve. We imagine that the curve is partitioned into infinitesimal segments described by the **differential of length** vector

$$d\mathbf{r} = dx\,\mathbf{i} + dy\,\mathbf{j} + dz\,\mathbf{k}. \tag{2.18}$$

The vector $d\mathbf{r}$ represents an infinitesimal displacement in a direction tangential to the curve. By extension of the definition of the element of arc length, which in $\mathbf{R}^2$ is just an infinitesimal expression of Pythagoras' theorem, we define

$$\|d\mathbf{r}\|^2 = ds^2 = dx^2 + dy^2 + dz^2. \tag{2.19}$$

Now we perform a short but interesting manipulation.

$$\begin{aligned}\left(\frac{ds}{dt}\right)^2 &= \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2,\\ &= v^2,\end{aligned}$$

so we conclude that

$$\frac{ds}{dt} = v, \tag{2.20}$$

$$s = \int v\, dt \tag{2.21}$$

If we have initial condition $s(0) = 0$, we can write the functional equation for arc length in terms of t as,

$$s(t) = \int_0^t v(\tau) d\tau, \tag{2.22}$$

and thus take care of the constant of integration.

The physics of equation 2.20 is most intuitive. If one were travelling along some curved trajectory, no matter how curved a differentiable curve is, at the infinitesimal level it looks straight, so the rate of change ds/dt of the arc length is the instantaneous speed v. In a car, this is the number that would show in the odometer. The second equation is a cause of anxiety for some students. It says that to compute the arc length of a curve, one must integrate the speed, and that means integrating the square root of some function. Experience in the first two semesters of calculus indicates that the only functions with square roots we are equipped to integrate at this stage are those in which the radicand is: constant, linear, quadratic, a perfect square, or one from which we can extract just the right chain rule factor for a substitution. As a result, there is only a handful of neat problems that can be done analytically, and these are the very same problems that have appeared in calculus textbooks for a couple of centuries. Anxiety will disappear for the diligent student who works out all the arc length problems in a fat calculus book. We will do the most important arc length computation shortly.

If t is a function $t = t(u)$ of some other variable, we can substitute for that variable $\mathbf{r}(t(u))$. We call this a **reparametrization**. The most useful reparametrization is by arc length. In practice, it is typically difficult to find an explicit arc length parametrization of a curve since not only does one have to calculate the integral, but also one needs to be able to find the inverse function $t(s)$ from $s(t)$. On the other hand, from a theoretical point of view, arc length parameterizations are ideal, since any curve so parametrized has unit speed.

The proof of this fact is a simple application of the chain rule and the inverse function theorem. This is more evident using the chain rule in Leibnitz notation, since

$$\begin{aligned} \frac{d\mathbf{r}}{ds} &= \frac{d\mathbf{r}}{dt}\frac{dt}{ds} = \frac{\frac{d\mathbf{r}}{dt}}{\frac{ds}{dt}}, \\ &= \frac{\frac{d\mathbf{r}}{dt}}{\|\frac{d\mathbf{r}}{dt}\|} = \frac{\mathbf{v}}{v} \equiv \mathbf{T}. \end{aligned} \tag{2.23}$$

and any vector divided by its length is a unit vector. This vector is called the **unit tangent vector T** because it is equal to the velocity vector divided by its length (the speed), so the vector $\mathbf{T}$ points in the same direction as the velocity vector. The last equation above is a bit awkward to write because of the fraction, so it is more elegant to rewrite is as

$$\dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt} = v\mathbf{T}. \tag{2.24}$$

Here we have introduced the common notation from physics to indicate by a dot, the derivative with respect to time. We will reserve the prime notation for the derivative with respect to other parameters. For example, we would write $\mathbf{r}'(s)$ for the derivative with respect to s. The dot notation is also useful because it clears the superscript slot when it is needed for a square or some other power.

Let $\mathbf{r}(s)$ be a curve parametrized by arc length and let $\mathbf{T}$ be the unit tangent vector $\mathbf{T} = \mathbf{r}'(s)$. Since $\mathbf{T}$ is a unit vector, we have

$$\mathbf{T}\cdot\mathbf{T} = 1$$

Now we differentiate this equation using the product rule. We get,

$$\begin{aligned} \frac{d}{ds}(\mathbf{T}\cdot\mathbf{T}) &= 0, \\ \frac{d\mathbf{T}}{ds}\cdot\mathbf{T} + \mathbf{T}\cdot\frac{d\mathbf{T}}{ds} &= 0, \\ 2\frac{d\mathbf{T}}{ds}\cdot\mathbf{T} &= 0, \\ \frac{d\mathbf{T}}{ds}\cdot\mathbf{T} &= 0. \end{aligned}$$

We conclude that $d\mathbf{T}/ds$ is perpendicular to $\mathbf{T}$. This is a general fact for the rate of change of any vector of constant length. Now, We divide this new vector by its length and call it the **unit normal N**,

$$\mathbf{N} = \frac{\frac{d\mathbf{T}}{ds}}{\|\frac{d\mathbf{T}}{ds}\|}.$$

Once again, the fraction makes the last equation a bit awkward, so we rewrite it as

$$\frac{d\mathbf{T}}{ds} = \left\|\frac{d\mathbf{T}}{ds}\right\|\mathbf{N}.$$

The result is called the first (of three) **Frenet equations**

$$\frac{d\mathbf{T}}{ds} = \kappa\,\mathbf{N}, \quad \text{where } \kappa = \left\|\frac{d\mathbf{T}}{ds}\right\|. \tag{2.25}$$

The quantity κ is called the **curvature**. It makes sense to call κ the curvature because, if $\mathbf{T}$ is a unit vector, then $\mathbf{T}'(s)$ is not zero only if the direction of $\mathbf{T}$ is changing. The rate of change of the direction of the tangent vector is precisely what one would expect to measure how much a curve is curving.Next, we introduce a third vector

$$\mathbf{B} = \mathbf{T} \times \mathbf{N}, \tag{2.26}$$

which we will call the **binormal** vector. The triplet of vectors $(\mathbf{T}, \mathbf{N}, \mathbf{B})$ forms an orthonormal set called the **Frenet frame**; that is,

$$\begin{aligned} \mathbf{T}\cdot\mathbf{T} = \mathbf{N}\cdot\mathbf{N} = \mathbf{B}\cdot\mathbf{B} = 1, \\ \mathbf{T}\cdot\mathbf{N} = \mathbf{T}\cdot\mathbf{B} = \mathbf{N}\cdot\mathbf{B} = 0. \end{aligned} \tag{2.27}$$

2.3.1 Example Circular Motion

Before we apply the full machinery of vector calculus, we start with some elementary trigonometry. Consider a circle of radius R, and θ be the central angle measured in radians. The length s of the arc subtended by a central angle θ is given by $s = R\theta$. Now, suppose a particle moves around the circle at a constant speed v. Speed is the rate of change of arc length. Since R is constant, we get $v = R\omega$, where ω is the rate of the change θ/t of the angle. That is $\theta = \omega t$. This small result motivates the vector parametrization of the position vector of a particle moving in circular motion with constant angular speed ω

$$\mathbf{r}(t) = \langle R\cos\omega t, R\sin\omega t\rangle. \tag{2.28}$$

Now, we compute formally the dynamics of the particle using the first Frenet equation 2.25. Taking the first and second derivatives we find the velocity and the acceleration

$$\begin{aligned} \mathbf{v} = \dot{\mathbf{r}} &= \langle -R\omega\sin\omega t, R\omega\cos\omega t\rangle, \\ \dot{\mathbf{r}} &= R\omega\langle -\sin\omega t, \cos\omega t\rangle,, \\ \mathbf{a} = \ddot{\mathbf{r}} &= -R\omega^2\langle \cos\omega t, \sin\omega t\rangle, \end{aligned} \tag{2.29}$$

By inspection we see that $\mathbf{v}\cdot\mathbf{r} = 0$, so the velocity vector is perpendicular to the position vector, and hence it is tangent to the circle as shown in figure 2.3. This is consistent with an earlier statement that the velocity vector is tangential to the direction of motion. We also observe that the acceleration vector is a negative multiple of the position vector, so the acceleration vector points towards the center of the circle. By Newton's second law $\mathbf{F} = m\mathbf{a}$ we infer that the force also points toward the center. This is evidence that circular motion is produced by a central force. Using one more time Pythagoras' identity

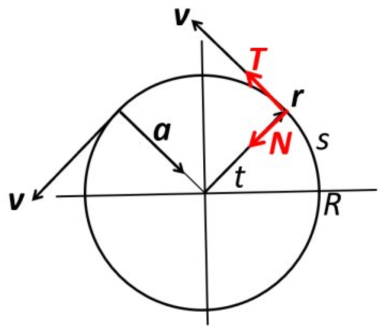


Fig. 2.3: Circular Motion

$\cos^2\phi + \sin^2\phi = 1$, we can mentally compute the speed v and the magnitude a of the acceleration

$$v = R\omega, \quad a = R\omega^2. \tag{2.30}$$

The first of these equations is consistent with our result obtained from geometry. It says that the larger the radius, the larger the speed. This should be obvious to any student who played in a marching band going around a corner. It is also the principle behind the operation of centrifuges as those use to extra honey from honeycombs, drying cycle of washing machines, and separation of biological particles as in DNA centrifuges. It follows that $\omega = v/R$. Inserting this value into the acceleration equation, we get

$$a = \frac{v^2}{R}. \tag{2.31}$$

This is a fundamental result in uniform circular motion. The acceleration is called the **centripetal acceleration**. I am in a one-man mission to have this equation taught to every high school freshman taking drivers education. The equation says that as one goes around a curve, doubling the speed quadruples the acceleration and hence the friction force on the tires required to maintain the vehicle on the road. Reducing the radius by a factor of two, doubles the acceleration. A combination of the two is likely to result in a car skidding off the road. The formula for centripetal acceleration appears on every introductory physics textbook. Unfortunately, most students at that stage have not had a course on vector calculus and it is likely they will not be exposed to the "right way" of obtaining this result.

Next, we compute the curvature. The arc length is given by

$$s = \int v\,dt = \int R\omega\,dt = R\omega t.$$

Hence $\omega t = s/R$. We substitute this into the equation for the position vector, thus resulting in the reparametrization by arc length

$$\mathbf{r}(s) = \langle R\cos(s/R), R\sin(s/R)\rangle \tag{2.32}$$

Taking the derivative with respect to s, the constant R gets cancelled by the chain rule factor, and we get the unit tangent vector

$$d\mathbf{T} = \frac{d\mathbf{r}}{ds} = \langle -\sin(s/R), \cos(s/R)\rangle$$

Differentiating again, we get

$$\frac{d\mathbf{T}}{ds} = \frac{d\mathbf{r}}{ds} = \frac{1}{R}\langle -\cos(s/R), -\sin(s/R)\rangle \tag{2.33}$$

The vector $\mathbf{N} = \langle -\cos(s/R), -\sin(s/R)\rangle$ is a unit vector, so it is fact the unit normal, and by the first Frenet equation 2.25, the multiplicative factor $1/R$ must be the curvature. Notice that $\mathbf{T}\cdot\mathbf{N} = 0$ as it should be. The neat result is

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}, \quad \text{where } \kappa = \frac{1}{R} \tag{2.34}$$

The result could not be more intuitive. The curvature of a circle is constant. The larger the radius, the smaller the curvature. If one stands on a circle with the very large radius like at a point on the Earth's equator, the curvature will be very small. This is the reason why, even today, some people think the Earth is flat. How sad. Finally, the significant result we were seeking is that the centripetal acceleration can be rewritten as

$$a = v^2\kappa. \tag{2.35}$$

2.4 Frenet Equations

If we differentiate the relation $\mathbf{B}\cdot\mathbf{B} = 1$, we find that $\mathbf{B}\cdot\mathbf{B}' = 0$, hence $\mathbf{B}'$ is orthogonal to $\mathbf{B}$. Furthermore, differentiating the equation $\mathbf{T}\cdot\mathbf{B} = 0$, we get

$$\mathbf{B}'\cdot\mathbf{T} + \mathbf{B}\cdot\mathbf{T}' = 0.$$

rewriting the last equation

$$\mathbf{B}'\cdot\mathbf{T} = -\mathbf{T}'\cdot\mathbf{B} = -\kappa\mathbf{N}\cdot\mathbf{B} = 0,$$

we also conclude that $\mathbf{B}'$ must also be orthogonal to $\mathbf{T}$. This can only happen if $\mathbf{B}'$ is orthogonal to the **TB**-plane, so $\mathbf{B}'$ must be proportional to $\mathbf{N}$. In other words, we must have

$$\mathbf{B}'(s) = -\tau\mathbf{N}(s), \tag{2.36}$$

for some quantity τ, which we will call the **torsion**. The torsion is similar to the curvature in the sense that it measures the rate of change of the binormal. Since the binormal also has unit length, the only way one can have a non-zero derivative is if $\mathbf{B}$ is changing directions. This means that if in addition $\mathbf{B}$ did not change directions, the vector would truly be a constant vector, so the curve would be a flat curve embedded into the **TN**-plane.

The quantity $\mathbf{B}'$ then measures the rate of change in the up and down direction of an observer moving with the curve always facing forward in the direction of the tangent vector. The binormal $\mathbf{B}$ is something like the flag in the back of sand dune buggy. The orthonormal Frenet frame $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ is the simplest example of what Eli Cartan called **repère mobile** (moving frame). It constitutes the starting point of the field of modern differential geometry. The advantage of this basis over the fixed $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ basis is that the Frenet frame is naturally adapted to the curve. It propagates along the curve with the tangent vector always pointing in the direction of motion, and the normal and binormal vectors pointing in the directions in which the curve is tending to curve. In particular, a complete description of how the curve is curving can be obtained by calculating the rate of change of the frame in terms of the frame itself. In aviation, the angular changes of the Frenet frame around the basis vectors are called the roll, the pitch and the yaw.

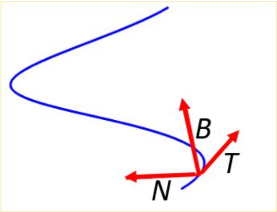


Fig. 2.4: Frenet Frame.

2.4.1 Theorem Let $\mathbf{r}(s)$ be a unit speed curve with curvature κ and torsion τ. Then

$$\begin{array}{rcl} \mathbf{T}' & = & \kappa\mathbf{N} \\ \mathbf{N}' & = & -\kappa\mathbf{T} \qquad\qquad \tau\mathbf{B} \\ \mathbf{B}' & = & -\tau\mathbf{N} \end{array}. \tag{2.37}$$

Proof We need only establish the equation for $\mathbf{N}'$. Differentiating the equation $\mathbf{N}\cdot\mathbf{N} = 1$, we get $2\mathbf{N}\cdot\mathbf{N}' = 0$, so $\mathbf{N}'$ is orthogonal to $\mathbf{N}$. Hence, $\mathbf{N}'$ must be a linear combination of $\mathbf{T}$ and $\mathbf{B}$.

$$\mathbf{N}' = a\mathbf{T} + b\mathbf{B}.$$

Taking the dot product of last equation with $\mathbf{T}$ and $\mathbf{B}$ respectively, we see that

$$a = \mathbf{N}'\cdot\mathbf{T}, \text{ and } \quad b = \mathbf{N}'\cdot\mathbf{B}.$$

On the other hand, differentiating the equations $\mathbf{N}\cdot\mathbf{T} = 0$, and $\mathbf{N}\cdot\mathbf{B} = 0$, we find that

$$\mathbf{N}'\cdot\mathbf{T} = -\mathbf{N}\cdot\mathbf{T}' = -\mathbf{N}\cdot(\kappa\mathbf{N}) = -\kappa$$
$$\mathbf{N}'\cdot\mathbf{B} = -\mathbf{N}\cdot\mathbf{B}' = -\mathbf{N}\cdot(-\tau\mathbf{N}) = \tau.$$

We conclude that $a = -\kappa$, $b = \tau$, and thus

$$\mathbf{N}' = -\kappa\mathbf{T} + \tau\mathbf{B}.$$

The Frenet frame equations (2.37) can also be written in matrix form as shown below.

$$\begin{bmatrix} \mathbf{T} \\ \mathbf{N} \\ \mathbf{B} \end{bmatrix}' = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{bmatrix} \begin{bmatrix} \mathbf{T} \\ \mathbf{N} \\ \mathbf{B} \end{bmatrix}. \tag{2.38}$$

The appearance of an antisymmetric matrix in the Frenet equations is not at all coincidental, as is deeply connected to advanced properties of the rotation group.

In the previous section we computed the curvature of circle from the first Frenet formula. In general this is completely impractical because the formula requires computing the integral for arc length and then inverting the equation to reparametrize by arc length. We need a formula to compute κ directly from the original equation of the curve, and utilizing only the parameter t. Recall equation 2.24

$$\dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt} = v\mathbf{T}.$$

Take the derivative using the product rule

$$\begin{aligned}
\ddot{\mathbf{r}} &= \frac{d}{dt}(v\mathbf{T}), \\
&= \frac{dv}{dt}\mathbf{T} + v\frac{d\mathbf{T}}{dt}, \\
&= \frac{dv}{dt}\mathbf{T} + v\frac{d\mathbf{T}}{ds}\frac{ds}{dt}, \quad \text{by the chain rule,} \\
&= \frac{dv}{dt}\mathbf{T} + v^2\frac{d\mathbf{T}}{ds}, \quad \text{since } \frac{ds}{dt} = v, \\
&= \frac{dv}{dt}\mathbf{T} + v^2\kappa\mathbf{N}, \quad \text{from the First Frenet equation 2.25.}
\end{aligned}$$

To summarize the result, we have

$$\ddot{\mathbf{r}} = \mathbf{a} = \frac{dv}{dt}\mathbf{T} + v^2\kappa\mathbf{N}. \tag{2.39}$$

Equation 2.39 is important in physics. The equation states that a particle moving along a curve in space feels a component of acceleration along the direction of motion whenever there is a change of speed, and a centripetal acceleration in the direction of the normal whenever it changes direction. The **tangential acceleration** a_T, and the **centripetal (or normal) Acceleration** a_c and any point are given by

$$a_T = \frac{dv}{dt}, \quad a_c = v^2\kappa = \frac{v^2}{r}.$$

where r is the radius of a circle called the **osculating circle**. The osculating circle has maximal tangential contact with the curve at the point in question. This is called contact of order 2. In general, a tangent to a curve has contact of order 1 with the curve in the sense that it passes through two consecutive points on the curve. The contact order is $(2-1)$. The osculating circle passes through three consecutive points on the curve in the curve; the contact order is $(3-1)$.

The osculating circle can be envisioned by a limiting process similar to that of the tangent to a curve in differential calculus. Let p be point on the curve, and let q_1 and q_2 be two nearby points. If the three points are not collinear, they uniquely determine a circle. The center of this circle is located at the intersection of the perpendicular bisectors of the segments joining two consecutive points. This circle is a "secant" approximation to the tangent circle. As the points q_1 and q_2 approach the point p, the "secant" circle approaches the osculating circle. The osculating circle, as shown in figure 2.5, always lies in the **TN**-plane, which by analogy is called the **osculating plane**. If $\mathbf{T}' = 0$, then $\kappa = 0$ and the osculating circle degenerates into a circle of infinite radius, that is, a straight line. The physics interpretation of equation 2.39 is that as a particle moves along a curve, in some sense, at an infinitesimal level, it is moving tangential to a circle, and hence, the centripetal acceleration at each point coincides with the centripetal acceleration along the osculating circle. As the points move along, the osculating circles move along with them, changing their radii appropriately.

Our focus remains on a formula for κ, since that is how we get to physics of components of the acceleration vector. For convenience, we align the two results

$$\dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt} = v\,\mathbf{T},$$
$$\ddot{\mathbf{r}} = \frac{dv}{dt}\mathbf{T} + v^2\kappa\,\mathbf{N}$$

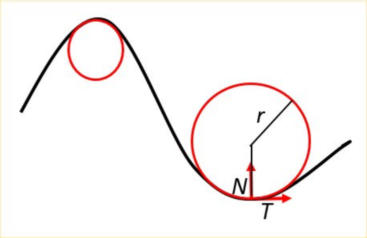


Fig. 2.5: Osculating Circle

We need to solve for κ. We need to be a bit clever since this is a vector equation and we can't use just plain algebra to isolate κ on one side of the second equation. In particular that would involve dividing by a vector, which makes no sense at all. Instead, we take the cross product of the two equations and use the facts that $\mathbf{T} \times \mathbf{T} = 0$ and $\mathbf{T} \times \mathbf{N} = \mathbf{B}$. We get

$$\dot{\mathbf{r}} \times \ddot{\mathbf{r}} = v^3\kappa\,\mathbf{B}$$

But $\mathbf{B}$ is a unit vector, that is $\|\mathbf{B}\| = 1$. Therefore, taking the length of both sides, we finally get the desired formula

$$\kappa = \frac{\|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}\|}{v^3} = \frac{\|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}\|}{\|\dot{\mathbf{r}}\|^3}. \tag{2.40}$$

The formula is entirely in terms of the derivatives with respect to t of the original curve, so after all this effort, obtaining the curvature is reduced to a step-by-step computation. The computation of the length of the cross product here is typically easier using Lagrange's identity 1.22

$$\|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}\|^2 = \begin{vmatrix} \dot{\mathbf{r}} \cdot \dot{\mathbf{r}} & \dot{\mathbf{r}} \cdot \ddot{\mathbf{r}} \\ \ddot{\mathbf{r}} \cdot \dot{\mathbf{r}} & \ddot{\mathbf{r}} \cdot \ddot{\mathbf{r}} \end{vmatrix}. \tag{2.41}$$

2.4.2 Example Circular helix

Let $\mathbf{r}(t) = \langle 3\cos t, 3\sin t, 4t\rangle$. Compute: $\dot{\mathbf{r}}, \ddot{\mathbf{r}}, v, \mathbf{T}, s, a_T, a_c$. Ordinarily we do not need to compute $\mathbf{N}$.

Solution. The curve is a helix. Work is expedited doing mental computations whenever we encounter and instance of the identity $\cos^2 t + \sin^2 t = 1$. We have

$$\begin{aligned}
\dot{\mathbf{r}} &= \langle -3\sin t, 3\cos t, 4\rangle, \\
\ddot{\mathbf{r}} &= \langle -3\cos t, -3\sin t, 0\rangle, \\
v &= \sqrt{3^2+4^2} = 5, \\
\mathbf{T} &= \frac{1}{5}\langle -3\sin t, 3\cos t, 4\rangle, \\
\mathbf{N} = \frac{\mathbf{T}'}{\|\mathbf{T}'\|} &= \langle -\cos t, -\sin t, 0\rangle, \\
\dot{\mathbf{r}} \times \ddot{\mathbf{r}} &= \langle -3(4)\sin t, -3(4)\cos t, 43\rangle, \\
\|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}\| &= 3(5) = 15, \\
\kappa &= 15/5^3 = 3/25, \\
a_c &= v^2\kappa = 3, \\
a_T &= \dot{v} = 0
\end{aligned}$$

This makes sense. The helix is curving at a uniform rate, so κ is constant. As stated earlier, the computation of the length of the cross product would have been easier using Lagrange's identity

$$\|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}\|^2 = \begin{vmatrix} 5^2 & 0 \\ 0 & 3^2 \end{vmatrix} = 15^2.$$

2.4.3 Example Hyperbolic helix

Let $\mathbf{r}(t) = \langle e^t, e^{-t}, \sqrt{2}t\rangle$. Compute: $\dot{\mathbf{r}}, \ddot{\mathbf{r}}, v, \mathbf{T}, s, a_T, a_c$.

Solution. This problem in disguise is similar to the preceding one. Note that $x = e^t, y = 1/e^t$, so $xy = 1$. This is the graph of a hyperbolic cylinder on which the curve wraps, rising at a constant rate. The curve is a hyperbolic helix. We have,

$$\begin{aligned}
\dot{\mathbf{r}} &= \langle e^t, e^{-t}, \sqrt{2}\rangle, \\
\ddot{\mathbf{r}} &= \langle e^t, -e^{-t}, 0\rangle, \\
v &= \sqrt{e^{2t} + e^{-2t} + 2} = \sqrt{e^{2t} + 2 + e^{-2t}}, \quad \text{(Looking for a perfect square)}, \\
&= \sqrt{(e^t + e^{-t})^2} = e^t + e^{-t} = 2\cosh t, \\
s &= \int 2\cosh\, dt = 2\sinh t. \quad \text{(assuming that } s(0) = 0), \\
\mathbf{T} &= \tfrac{1}{(e^t + e^{-t})}\langle e^t, e^{-t}, \sqrt{2}\rangle, \\
\dot{\mathbf{r}} \times \ddot{\mathbf{r}} &= \langle -\sqrt{2}e^{-t}, \sqrt{2}e^t, 2\rangle = \sqrt{2}\langle -e^{-t}, e^t, \sqrt{2}\rangle, \\
\|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}\| &= \sqrt{2}(e^{-t} + e^t) = 2\sqrt{2}\cosh t. \\
\kappa &= \frac{\sqrt{2}}{(e^t + e^-t)^2}, \\
a_c &= \sqrt{2}, \\
a_T &= e^t - e^{-t} = 2\sinh t.
\end{aligned}$$

A slightly fancier version of the hyperbolic helix example above is given by,

2.4.4 Example Hyperbolic helix - version 2

Let $\mathbf{r}(t) = \langle a\cosh t, a\sinh t, a\rangle$. We have

$$\begin{aligned}
\dot{\mathbf{r}} &= \langle a\sinh t, a\cosh t, a\rangle, \\
\ddot{\mathbf{r}} &= \langle a\cosh t, a\sinh t, 0\rangle, \\
v &= \sqrt{a^2\sinh^2 t + a^2\cosh^2 t + a^2}, \\
&= a\sqrt{\cosh 2t + 1}, \quad \text{(by double angle formula)}, \\
&= a\sqrt{2}\sqrt{\frac{\cosh 2t + 1}{2}}, \\
&= a\sqrt{2}\cosh t, \quad \text{(by half angle formula)}, \\
\dot{\mathbf{r}} \times \ddot{\mathbf{r}} &= \langle -a^2\sinh t, a^2\cosh t, -1\rangle, \\
\|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}\| &= a^2\cosh t, \\
a_c &= a, \\
a_T &= a\sqrt{2}\sinh t.
\end{aligned}$$

The centripetal acceleration for this curve is also constant.

Example Twisted cubic

Let $\mathbf{r} = \langle \frac{1}{3}t^3, t^2, 2t\rangle$. Compute: $\dot{\mathbf{r}}, \ddot{\mathbf{r}}, v, \mathbf{T}, s, a_T, a_c$.

Solution. This is a twisted cubic with the coefficients rigged up just right.

$$\begin{aligned}
\dot{\mathbf{r}} &= \langle t^2, 2t, 2\rangle,\\
\ddot{\mathbf{r}} &= \langle 2t, 2, 0\rangle,\\
v &= \sqrt{t^4+4t^2+4} = \sqrt{t^4+4+4t^2}, \quad \text{(Looking for a perfect square)},\\
&= \sqrt{(t^2+2)^2} = t^2+2\\
\mathbf{T} &= \frac{1}{t^2+2}\langle t^2, 2t, 2\rangle,\\
s &= \tfrac{1}{3}t^3 + 2t, \quad \text{(assuming that } s(0)=0),\\
\dot{\mathbf{r}}\times\ddot{\mathbf{r}} &= \langle -4, 4t, -2t^2\rangle = 2\langle -2, 2t, -t^2\rangle,\\
\|\dot{\mathbf{r}}\times\ddot{\mathbf{r}}\| &= 2(t^2+2),\\
\kappa &= \frac{2}{(t^2+2)^2},\\
a_c &= 2\\
a_T &= 2t.
\end{aligned}$$

This is a good example that illustrates that often, computation of the unit normal is more computationally intensive. The normal is given by $\mathbf{N} = \mathbf{T}'/\|\mathbf{T}'\|$. For the derivative of each component of $\mathbf{T}$ we can use the quotient rule. Then we have divide by the length of the vector. Nothing is hard, but it does take a bit of time and hope that the length of $\mathbf{T}'$ involves the square root of a perfect square. It does. For the final result I got.

$$\mathbf{N} = \frac{1}{2+t^2}\langle 2t, 2-t^2, -2t\rangle$$

Details are found in my hand written notes at the web site.

2.4.5 Example Plane curves

Let $\mathbf{r}(t) = (x(t), y(t))$. To find κ, we just embed the plane curve into $\mathbf{R}^3$ by writing $\mathbf{r}(t) = (x(t), y(t), 0)$ and just apply the formula. We get,

$$\begin{aligned}
\dot{\mathbf{r}} &= (\dot{x}, \dot{y}, 0),\\
\ddot{\mathbf{r}} &= (\ddot{x}, \ddot{y}, 0),\\
\kappa &= \frac{\|\dot{\mathbf{r}}\times\ddot{\mathbf{r}}\|}{\|\dot{\mathbf{r}}\|^3},\\
&= \frac{|\,\dot{x}\ddot{y} - \dot{y}\ddot{x}\,|}{(\dot{x}^2+\dot{y}^2)^{3/2}}.
\end{aligned}$$

This is the formula that appears in the section on parametric curves in Calculus II. In particular, if $y = f(x)$, we use the trivial parametrization $x = t, y = f(t)$ and we get a simper formula that also appears in Calculus II. Had it been the case that all STEM students take the three semesters of calculus, this formula would have been redundant in the curriculum. As long as we have derived the formula here, we will use it in the next example.

2.4.6 Example Let $\mathbf{r}(t) = \langle e^t \cos t, e^t \sin t\rangle$. Find the v, κ.
Solution. Without the exponentials this would have been a circle. With the factor of e^t the central distance is increasing exponentially. The graph is called a logarithmic spiral. These type of spirals are manifested all over nature as in the case of nautilus shells and spiral galaxies.

$$\begin{aligned}
\dot{\mathbf{r}} &= \langle -e^t \sin t + e^t \cos t, e^t \cos t + e^t \sin t\rangle \\
&= \langle e^t(\cos t - \sin t), e^t(\cos t + \sin t)\rangle \\
\ddot{\mathbf{r}} &= \langle -2e^t \sin t, 2e^t \cos t\rangle, \quad \text{(From the product rule)}, \\
v &= \sqrt{e^{2t}(\cos t - \sin t)^2 + e^{2t}(\cos t + \sin t)^2}, \\
&= e^t\sqrt{(\cos t - \sin t)^2 + (\cos t + \sin t)^2)}, \\
&= e^t\sqrt{(1 - 2\cos t \sin t) + (1 + 2\cos t \sin t)}, \\
v &= \sqrt{2}\, e^t, \\
\kappa &= \frac{2e^{2t}\cos t(\cos t - \sin t) + 2e^{2t}\sin t(\cos t - \sin t)}{(\sqrt{2}\, e^t)^3}, \\
&= \frac{2e^{2t}}{2\sqrt{2}e^{3t}} = \frac{1}{\sqrt{2}e^t}.
\end{aligned}$$

2.5 Change of Coordinates

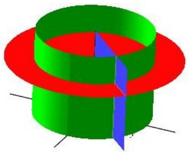
2.5.1 Cylindrical Coordinates

Cylindrical coordinates refers to the change of variables $T : \mathbf{R}^3 \to \mathbf{R}^3$ given by

$$\begin{aligned}
\langle r, \theta, z\rangle &\xrightarrow{T} \langle x, y, z\rangle, \\
\langle r, \theta, z\rangle &\mapsto \langle r\cos\theta, r\sin\theta, z\rangle.
\end{aligned}$$

Of course, this is equivalent to the three scalar equations

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



Cylindrical coordinates constitute an example of what is called a **triply orthogonal system**. This means that the surfaces obtained by setting one of the coordinates equals to a constant are mutually orthogonal at each point. For example, in cylindrical coordinates $z = k$ is a plane, $r = a$ is a cylinder of radius a, $\theta = \pi/4$ is a vertical plane at 45 degrees. The surfaces intersect at a point at which three unit vector normal to each of the three surfaces form an orthogonal frame, much like the $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ frame. It is easy to find these unit vectors. We will call the triplet $\{\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_z\}$. The position vector for a unit circle is $\mathbf{r} = \cos\theta\mathbf{i} + \sin\theta\mathbf{j}$. The unit vector $\mathbf{e}_r$ points in the same direction, so it is the same vector. The derivative of $\mathbf{e}_r$ with respect to θ is also a unit vector and it is orthogonal to $\mathbf{e}_r$, so choose this to be $\mathbf{e}_\theta$. The cylindrical basis frame is therefore given by

$$\begin{aligned} \mathbf{e}_r &= \cos\theta\,\mathbf{i} + \sin\theta\,\mathbf{j}, \\ \mathbf{e}_\theta &= -\sin\theta\,\mathbf{i} + \cos\theta\,\mathbf{j}, \\ \mathbf{e}_z &= \mathbf{k} \end{aligned} \tag{2.42}$$

Physicists usually call the adapted unit vectors $\{\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{z}}\}$. The importance of triply orthogonal systems will become self evident in the derivation of Kepler's laws in section 2.7.

Cylindrical coordinates are the natural extension of polar coordinates to three dimensions. In polar coordinates in the plane, the distance from a point with coordinates x, y to the origin is called $r = \sqrt{x^2 + y^2}$. We will not review the full array of beautiful curves in $\mathbf{R}^2$ that admit an elegant functional form in polar coordinates. For the time being, it suffices to mention the following

Basic Polar Curves

1. Circle. The equation $r = a$ represents a circle centered at the origin with radius a. The real reason why polar coordinates were introduced by Newton.

2. Shifted Circle. The equation $r = a\cos\theta$ also represents a circle. To verify this, we change to Cartesian coordinates,

$$\begin{aligned} r &= a\cos\theta, \\ r^2 &= ar\cos\theta, \quad \text{(after multiplying by } r), \\ x^2 + y^2 &= ax, \\ (x^2 - ax + \tfrac{a^2}{4}) + y^2 &= \tfrac{a^2}{4}, \quad \text{(completing the square)}, \\ (x - \tfrac{a}{2})^2 + y^2 &= \tfrac{a^2}{4}. \end{aligned}$$

   This is the equation of a circle of diameter a with center shifted to $(\frac{a}{2}, 0)$. If instead we had $r = a\sin\theta$, we would have circle of the same diameter, but with center shifted to $(0, \frac{a}{2})$.

3. Cardioids. The equations $r = a(1 \pm \cos\theta)$ and $r = a(1 \pm \sin\theta)$ are cardioids. To establish whether the cardioids open right, left, up or down, it suffices to plot the points for $\theta = \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$.

4. Polar conics. One of the major contributions to polar equations was discovered by Newton in the process of trying to integrate the equations of motion of a planet in polar coordinates. The solutions are of the form

$$r = \frac{ed}{1 \pm e\cos\theta}, \quad r = \frac{ed}{1 \pm e\sin\theta}.$$

   He quickly proved that these are indeed equations of conics, thereby proving that Kepler's law about elliptical orbits of planets could be be derived mathematically and not just empirically.

Surfaces of Revolution

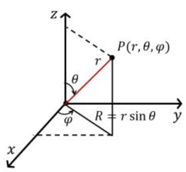
Equations of surfaces in which x and y appear in the form $(x^2 + y^2)$ can be transformed into cylindrical coordinates by replacing the expression by r^2. Equations in cylindrical coordinates in which there is no z are cylinders For example, in cylindrical coordinates, we have

$r = a,$	is a circular cylinder with axis of symmetry along z,	
$r = a\cos\theta,$	is circular cylinder with a generator along the vertical axis,	
$z = r^2,$	is the paraboloid	$z = x^2 + y^2,$
$z = r,$	is the cone	$z = \sqrt{x^2 + y^2},$
$z = \sqrt{a^2 - r^2}$	is the hemisphere	$z = \sqrt{a^2 - x^2 - y^2}$

In single variable calculus, when one makes a substitution in an integral, the integrand must be adjusted by the chain rule factor. Cylindrical coordinates are most useful when performing multiple integration over regions or surfaces that exhibit symmetry with respect to one of the coordinate axes. The change of variables analog of the chain rule factor in cylindrical coordinates often changes a multiple integral in Cartesian coordinates with nasty square roots, to a very simple integral that often can be integrated by inspection. Multiple integrals in cylindrical coordinates are treated in section 4.3.

2.5.2 Spherical Coordinates

Spherical coordinates are to spheres as polar coordinates are to circles. The motivation is the same. Even in the most fundamental problem of finding the area of a circle $x^2 + y^2 = a^2$ by integration in Cartesian coordinates, one is presented with a square root $y = \sqrt{a^2 - x^2}$ in the integrand. This requires a trigonometric substitution leading to the integral of an even power of a cosine function. In plane polar coordinates, the equation of the same circle is $r = a$, making the area integral possible in one line. The equation of a sphere of radius a centered at the origin is $x^2 + y^2 + z^2 = a^2$, so the integral for the volume in Cartesian coordinates would now involve an undesirable $z = \sqrt{a^2 - x^2 - z^2}$ integrand. To ameliorate this problem, we introduce **spherical coordinates**.



The spherical coordinates $P(r, \theta, \phi)$ of a point P in $\mathbf{R}^3$ are defined as follows. The coordinate r is the distance from P to the origin. The coordinate θ is the spherical polar angle that the line segment from P to the origin makes with z-axis. The angle ϕ is the usual plane polar coordinates angle (the azimuthal angle). To relate spherical coordinates to Cartesian coordinates we first draw the projections $z = r\cos\theta$ onto the z-axis and $R = r\sin\theta$ onto the xy-plane. The quantity R is the distance from the point with coordinates (x, y) on the xy-plane to the origin. By polar coordinates $x = R\sin\phi$, and

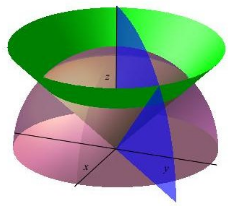
$y = R\sin\phi$. Hence, we have

$$x = r\sin\theta\cos\phi, \qquad r^2 = x^2 + y^2 + z^2,$$
$$y = r\sin\theta\sin\phi, \quad \text{and} \quad \phi = \tan^{-1}(y/x), \tag{2.43}$$
$$z = r\cos\theta \qquad \theta = \cos^{-1}(z/\sqrt{x^2+y^2+z^2}).$$

Although is is really unnecessary, the skeptics may verify algebraically that

$$x^2 + y^2 + z^2 = (r\sin\theta\cos\phi)^2 + (r\sin\theta\sin\phi)^2 + (r\cos\theta)^2 = r^2.$$

We impose the restriction $r \geq 0$, $0 \leq \theta \leq \pi$, and $0 \leq \phi \leq 2\pi$. With these restrictions we can label all the points in $\mathbf{R}^3$.



Spherical coordinates is the only other example of a triply orthogonal system that appears in Calculus III. The idea is the same as above. Surfaces obtained by setting each one of the coordinates equals to a constant are mutually orthogonal at the point of intersection. For example, in spherical coordinates $r = a$ is a sphere of radius a, $\theta = \pi/4$ is a cone at 45 degrees and $\phi = \pi/4$ is a vertical plane. The surfaces intersect at a point at which the normals to the three surfaces are mutually orthogonal; that is, if we construct unit vectors $\{\hat{\mathbf{r}}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}\}$, normal to the corresponding constant-coordinate surfaces, the triplet forms an orthonormal set adapted to the coordinate system.

Transformations in $\mathbf{R}^3$ which lead to triply orthogonal systems are not abundant. Whereas, substitutions in single variable integration are ubiquitous, we present only two "simultaneous substitutions" to help simplify certain classes of triple integrals over regions or functions that exhibit an appropriate symmetry. Triply orthogonal systems are special because in these systems the infinitesimal differential of volume is a rectangular parallelepiped (a box).

2.5.1 Remark Regrettably, we should mention that the spherical coordinates labels used by most mathematicians are confusing. Historically, since Lagrange introduced spherical coordinates in the 1700's, physicists have denoted the distance from a point in space to the origin by the coordinate r. This explains why in gravitation and in electromagnetism, one talks about the $1/r^2$ law. Most mathematicians call the distance ρ. I have never read a textbook from an author bold enough to call the gravitational force a $1/\rho^2$ law. In the standard physics notation the coordinates of a point $P(r, \theta, \phi)$ on a sphere are in this order, the distance r to the origin, the angle θ from the north pole, and the angle ϕ from the x-axis formed by the projection or r onto the xy-plane. Many calculus textbooks also change the order of the spherical coordinates to (ρ, θ, ϕ) where ϕ is the angle from the north pole. This introduces yet one more point of confusion because it makes the spherical basis vectors left-handed. Perhaps it is time to reconsider the 2002 proposal by Tevian Dray and Corinne Manogue, to universally adopt the physics convention (https://bridge.math.oregonstate.edu/papers/spherical.pdf).

We will expect students to recognize the following special quadrics surfaces in spherical coordinates

1. Sphere. The equation $r = a$ in spherical coordinates represents a sphere of radius a, centered at the origin.

2. Shifted Sphere. The equation $r = a\cos\theta$ also represents a sphere. To verify this, we transform the equation to Cartesian coordinates,

$$\begin{aligned} r &= a\cos\theta, \\ r^2 &= ar\cos\theta, \quad \text{after multiplying by } r, \\ x^2 + y^2 + z^2 &= az, \\ x^2 + y^2 + (z^2 - az + \tfrac{a^2}{4}) &= \tfrac{a^2}{4}, \quad \text{completing the square,} \\ x^2 + y^2 + (z - \tfrac{a}{2})^2 &= \tfrac{a^2}{4}. \end{aligned}$$

   This is the equation of a sphere of diameter a with center shifted to $(0, 0, \frac{a}{2})$. If instead we had $r = a\sin\theta\cos\phi$, we would have sphere of the same diameter, but with center shifted to $(\frac{a}{2}, 0, 0)$.

3. Cone. The equation $\theta = \theta_0$ represents a branch of a cone with vertex at the origin and generator that makes an angle θ_0 with the z-axis. Thus, to write the equation of a such a cone, all need to know is the angle between the z-axis and the generator of the cone. This is from the definition of the coordinate system, but again, the skeptic could corroborate the statement by a direct computation from the transformation equations. For example, for the case $\theta = \pi/4$, we have

$$\begin{aligned} \theta &= \tfrac{\pi}{4}, \\ r\cos\theta &= r\cos(\tfrac{\pi}{4}) = r\tfrac{1}{\sqrt{2}}, \\ z^2 &= \tfrac{1}{2}(x^2 + y^2 + z^2), \\ z^2 &= x^2 + y^2, \\ z &= \sqrt{x^2 + y^2} \end{aligned}$$

   If we wanted the lower branch of the cone $z^2 = x^2 + y^2$ we would have to take $\theta = 3\pi/4$.

2.5.2 **Example** Atomic Orbitals

Although in the main examples in setting up integrals in spherical coordinates in the this course will be restricted to regions consisting of spheres and cones as above, here is something neat. Define the Legendre polynomials $P_n(z)$ of degree n by the equation

$$P_n = \frac{1}{2^n n!}\frac{d^n}{dz^n}(z^2 - 1)^n. \tag{2.44}$$

Thus, for example, we compute $P_2(z)$

$$\begin{aligned} P_2(z) &= \frac{1}{2^2 2!}\frac{d^2}{dz^2}(z^2-1)^2, \\ &= \tfrac{1}{8}\frac{d^2}{dz^2}(z^4-2z^2+1), \\ &= \tfrac{1}{8}(12z^2-4), \\ &= \tfrac{1}{2}(3z^2-1) \end{aligned}$$

The idea is to compute the Legendre polynomials, follow up by making the substitution $z = \cos\theta$, and then plot the results in spherical coordinates. Here is a list of the first four Legendre polynomials. We could easily compute these by hand, but as long as we will be using a computer algebra system to graph the functions, we might as well let computer also do the calculation of the derivatives.

n	$P_n(z)$	$P_n(\cos\theta)$
1	$P_1(z) = z$	$P_1(\cos\theta) = \cos\theta$
2	$P_2(z) = \frac{1}{2}(3z^2-1)$	$P_2(\cos\theta)\frac{1}{2}(5\cos^3\theta - 3\cos\theta)$
3	$P_3(z) = \frac{1}{2}(5z^3-3z)$	$P_3(\cos\theta) = \frac{1}{2}(5\cos^3\theta - 3\cos\theta)$
4	$P_4(z) = \frac{1}{8}(35z^4-30z^2+3)$	$P_4(\cos\theta) = \frac{1}{8}(35\cos^4\theta - 30\cos^2\theta + 3)$

The table also shows the corresponding polynomials after making the substitution $z = \cos\theta$. The graphs in spherical coordinates as shown in figure 2.6.

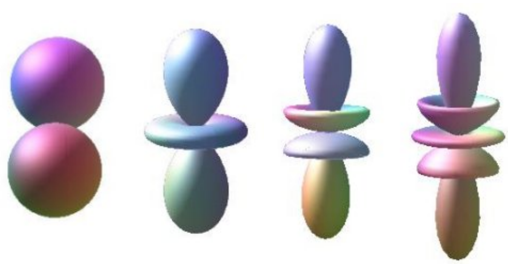


Fig. 2.6: Atomic Orbitals with quantum number $n = 1, 2, 3, 4$

Hopefully the reader recognizes these as the angular part of the atomic orbitals for principal quantum numbers $n = 1, 2, 3, 4$. Of course there must be a compelling mathematical reason for this neat result. The reason behind everything is that the Legendre functions are solutions to the angular θ part of Laplace's equation in spherical coordinates. This is a great motivator for further studies in partial differential equations.

2.6 Parametric Surfaces

A vector function $\mathbf{r}$ from a subset $R \in \mathbf{R}^2$ to $\mathbf{R}^3$ locally represents a surface. The function is called a coordinate patch of a surface, because in general, it will take more one than such patch to cover the full surface. More precisely, we have a function

$$(u,v) \xrightarrow{\mathbf{r}} \langle x, y, x\rangle,$$
$$(u,v) \mapsto \langle x(u,v), y(u,v), z(u,v)\rangle,$$

that maps a region in the uv-plane to a two-dimensional surface in $\mathbf{R}^3$, as shown in figure 2.7. The general parametric equation of a curve in $\mathbf{R}^3$ is

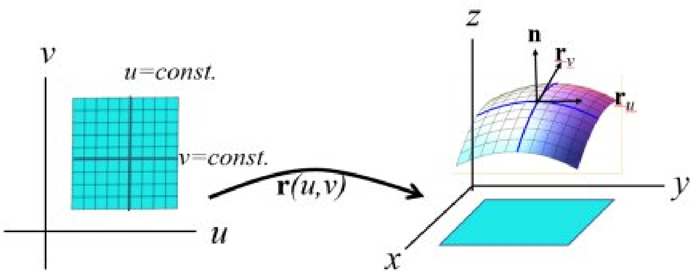


Fig. 2.7: Parametric Surface

$$\mathbf{r}(u,v) = x(u,v)\,\mathbf{i} + y(u,v)\,\mathbf{j} + z(u,v)\,\mathbf{k}, \tag{2.45}$$

where $\mathbf{r}(u,v)$ represents the position vector of a point on the surface. If in the region $R \in \mathbf{R^2}$ one takes a grid of defined by the lines $v =$ constant, then for each constant, $\mathbf{r}$ effectively depends only on the parameter u and the resultant is a set of parametric curves on the surface, as shown in figure 2.7. Therefore, the partial derivative vectors $\mathbf{r}_u$ are tangential to these curves. Similarly, the partial derivative vectors $\mathbf{r}_v$ are tangential to curves $u =$ constant. Thus, the vector

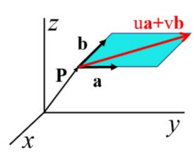
$$\mathbf{N} = \mathbf{r}_u \times \mathbf{r}_v, \tag{2.46}$$

is normal to the surface at each point, We define the unit normal $\mathbf{n}$ to the surface by

$$\mathbf{n} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{\|\mathbf{r}_u \times \mathbf{r}_v\|}. \tag{2.47}$$

2.6.1 Example Planes. The equation of a plane that passes through a point $P(x_0, y_0, y_0)$ and contains the vectors $\mathbf{a} = \langle a_1, a_2, a_3\rangle$ and $\mathbf{b} = \langle b_1, b_2, b_3\rangle$ is given by

$$\mathbf{r}(u,v) = \mathbf{P} + u\mathbf{a} + v\mathbf{b}. \tag{2.48}$$



This equation of a plane is the analog of the equation of a line. If $v = 0$ the equation represents the line that contains P and has direction vector $\mathbf{a}$. Similarly, if $u = 0$ one gets the line that contains P and has direction vector $\mathbf{b}$. The full equation describes the plane by starting at the point P

and then moving along linear combinations of **a** and **b** to get to any point in the plane. If we make the restriction $u, v \in [0,1]$, the equation describes the parallelogram spanned by **a** and **b**.
Unpacking the equation of the plane 2.48 in terms of the vector components, we get

$$\mathbf{r}(u,v) = \langle x_0 + ua_1 + vb_1, y_0 + ua_2 + vb_2, x_3 + ua_3 + vb_3 \rangle.$$

It is worth noticing that if one were to follow the conventions of tensor calculus, we should vectors as column vectors rather than row vectors. Doing so, makes the equation more transparent,

$$\begin{aligned} \mathbf{r}(u,v) &= \mathbf{P} + u\mathbf{a} + v\mathbf{b}, \\ &= \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + u \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + v \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \end{aligned}$$

Most likely, the main reasons why column vectors are not universally used in the exposition of calculus is that it is much more taxing to the typesetter, and much less efficient in space utilization. The parametric equation of the plane here is very elegant, because it does not require cross products. One could in principle eliminate the parameters u and v and find the more common type of equation of the form $Ax + By + Cz = D$, however it is easier to show the equation is consistent with the procedure described in section 1.5.2 by simply noticing that a normal vector is given by

$$\mathbf{N} = \mathbf{r}_u \times \mathbf{r}_v = \mathbf{a} \times \mathbf{b}$$

2.6.2 Example Saddle. The equation

$$\begin{aligned} \mathbf{r}(u,v) &= u\,\mathbf{i} + v\,\mathbf{j} + (u^2 - v^2)\,\mathbf{k}, \quad \text{or in scalar parametric form,} \\ x &= u, \\ y &= v, \\ z &= u^2 - v^2 \end{aligned}$$

is clearly equivalent to the Cartesian equation $z = x^2 - y^2$, so the equation represents a saddle.

2.6.3 Example Explicit surface.
The equation of the saddle above is an example of an explicit equation $z = f(x,y)$ of a surface. Just as a curve $y = f(x)$ in $\mathbf{R}^2$ can be written in parametric form $\mathbf{r}(t) = \langle t, f(t) \rangle$, explicit equations of surfaces can be instantly converted to parametric from by what I will call the **trivial parametrization**

$$\mathbf{r}(u,v) = u\,\mathbf{i} + v\,\mathbf{j} + f(u,v)\,\mathbf{k}.$$

We could even use the coordinates x and y as parameters and write.

$$\mathbf{r}(x,y) = \langle x, y, f(x,y)\rangle. \tag{2.49}$$

2.6.4 Example Paraboloid and Saddle
The trivial parametric equation of the paraboloid $z = x^2 + y^2$ is

$$\mathbf{r}(x,y) = \langle x, y, x^2 + y^2\rangle.$$

Parametric equations are certainly not unique. We could describe that same paraboloid in cylindrical coordinates $z = r$ which in parametric form reads

$$\mathbf{r}(r,\theta) = \langle r\cos\theta, r\sin\theta, r^2\rangle.$$

In components, $x = r\cos\theta$, $y = r\sin\theta$, $z = r^2$. Therefore $x^2 + y^2 = r^2 = z$.
The equation

$$\mathbf{r}(r,\theta) = \langle r\cosh\theta, r\sinh\theta, r^2\rangle.$$

gives the saddle $x^2 - y^2 = z$

2.6.5 Example cone
A minor modification

$$\mathbf{r}(r,\theta) = \langle r\cos\theta, r\sin\theta, r\rangle.$$

gives the equation of a cone. The see this just notice that if $x = r\cos\theta$ and $y = r\sin\theta$, then $x^2 + y^2 = r^2 = z^2$. The big advantage is that in parametric form, the graph gives the full cone, whereas in the Cartesian equation, one has to solve for z to get a function. The positive square root gives only the top half of the cone and the negative square root gives only the bottom half of the cone.

2.6.6 Example Surface of revolution. Circular paraboloids and circular cones are examples of surfaces of revolution obtained by starting with a curve $z = f(x)$ and rotating about the z-axis. This amount to replacing x by r to give the surface $z = f(r)$. Hence, a parametric representation of surfaces of revolution is given by

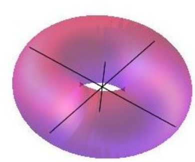
$$\mathbf{r}(r,\theta) = \langle r\cos\theta, r\sin\theta, f(r)\rangle \tag{2.50}$$

More generally, if $x = x(u)$, $z = z(u)$ is the parametric equation of a curve in the xz-plane, then the surface of revolution generated by rotating the curve around the z-axis is given by

$$\mathbf{r}(u,v) = \langle x(u)\cos v, x(u)\sin v, z(u)\rangle. \tag{2.51}$$

2.6.7 Example Torus

A torus is a surface of revolution generated by rotating a circle around an axis that is coplanar with the circle. In the standard torus, the axis does not touch the circle resulting in a shape that looks like doughnut or a bagel. Consider a circle of radius r in the xz-plane, centered at $(R, 0)$ and rotate the circle a full revolution about the z-axis.



We assume that $R > r$ so that the axis of revolution does not touch the circle With this choice, we obtain a standard torus. The parametric equation of the circle in the xz-plane is

$$x - R = r\cos(\theta), \quad z = r\sin\theta.$$

By equation 2.51, the parametric equation of the torus is

$$\mathbf{r}(\theta,\phi) = \langle\ (R + r\cos\theta)\cos\phi, (R + r\cos\theta)\sin\phi, r\sin\theta\ \rangle \tag{2.52}$$

2.6.8 Example Sphere
In spherical coordinates the equation of a sphere of radius a centered at the origin is $r = a$. The spherical coordinates transformation 2.43 suggests an obvious parametrization,

$$\mathbf{r}(\theta,\phi) = \langle a\sin\theta\cos\phi, a\sin\theta\sin\phi, a\cos\theta\rangle. \tag{2.53}$$

This might be the best example of why parametric equations is the ultimate

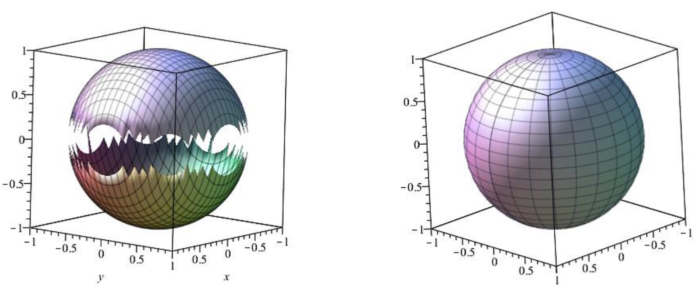


Fig. 2.8: Sphere

way to graph surfaces. Figure 2.8 shows an attempt to plot a sphere in Cartesian coordinates by the two functions $z = \pm\sqrt{a^2 - x^2 - y^2}$. The graph on the right shows a clean graph. A minor observation is that the parametric equation graph is actually missing a point, namely the north pole. It is not possible to get the entire graph of an sphere with a singe coordinate patch because a rectangle and a sphere have different topologies - spheres are not flat. As a bonus, observe that in the parametric equation graph, the grid consists of meridians and parallels.

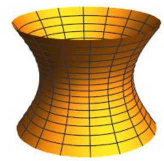
2.6.9 Example More cones. We have also learned that in spherical coordinates, setting $\theta = \theta_o$ yields the equation of a cone with a generator making an angle θ_o. Therefore the transformation equations 2.43 also suggest that setting

θ equal to a constant gives a parametric equation of a cone. For example, let $\theta = \pi/4$. We get

$$\mathbf{r}(r, \phi) = \langle r\tfrac{1}{\sqrt{2}} \cos\phi, r\tfrac{1}{\sqrt{2}} \sin\phi, r\tfrac{1}{\sqrt{2}}\rangle.$$

The common factor of $\frac{1}{\sqrt{2}}$ is just a scaling factor and it can be deleted without affecting the shape of the graph. We now have $x = r\cos\phi$, $y = r\sin\phi$, and $x^2 + y^2 = r^2 = z^2$, which is indeed the equation of a cone at 45 degrees.

2.6.10 Example Hyperboloid
The parametric equation of the sphere 2.53 can be modified to obtain the equation of a circular hyperboloid of one sheet. The trick is convert a sum of squares by a difference of squares. This can be done naturally if one replaces $\cos\phi$

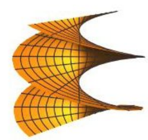


and $\sin\phi$ by $\cosh\phi$ and $\sinh\phi$ respectively. This results in the following parametric equation of a circular hyperboloid of one sheet.

$$\begin{aligned} \mathbf{r}(\theta, \phi) &= \langle a\sin\theta\cosh\phi, a\sin\theta\sinh\phi, a\cos\theta\rangle, \\ &\Rightarrow x^2 - y^2 + z^2 = a^2 \end{aligned}$$

If in addition one replaces $\sin\theta$ and $\cos\theta$ by $\sinh\theta$ and $\cosh\theta$ respectively, one gets a hyperboloid of two sheets. Instead of $\cosh\phi$ and $\sinh\phi$, we could also use $\sec\phi$ and $\tan\phi$, since $\sec^2\phi - \tan^2\phi = 1$

2.6.11 Example Helicoid
A helicoid is a **ruled surface** traced by a line segment rotating about a perpendicular axis, as it rises at a constant rate along the axis. A discrete version of the surface can be visualized by a spiral staircase. Every particular point on



the rotating line segment traces a helix. Suppose the segment is the interval $[-1, 1]$ as it rotates around the z- axis. A point a distance u along the generating line segment traces circles looking from above the xy-plane, but it rises at a constant rate. Thus, the equation of a helicoid in parametric form can be represented by the equation

$$\mathbf{r}(u, v) = \langle\ u\cos v, u\sin v, v\ \rangle. \tag{2.54}$$

In Cartesian coordinates the equation is $y/x = \tan z$.

2.7 Kepler's Laws

In a remarkable feat of empirical observation of planetary motion, Johannes Kepler formulated in the early 1600's, the following laws:

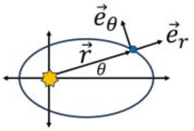
1. Planets move in elliptical orbits around the sun and the sun is located at one of the foci of the ellipses.

2. The line joining each plane with the sun sweeps equal areas in equal time.

3. The square of the periods is proportional to the cube of the length of the major axis.

Newton understood that such magnificence must have a mathematical explanation, and it is mainly for this reason that he invented calculus. It is a shame that this being one of the highest achievements in the history of science, the derivation of this laws is omitted from the standard calculus curriculum. We present here a derivation of the laws of Kepler. It must be noted that the approach we take here is not that which appeared in Newton's 1687 *Principia Mathematica*.

Kepler's Second Law



We prove that Kepler's second law is equivalent to conservation of angular momentum. Suppose the sun sits at the origin of a coordinate system, and the position vector of a planet with respect to the sun is $\mathbf{r}$. In the process we will show that the motion takes place on a plane, so we adapt a polar basis frame $\{\mathbf{e}_r, \mathbf{e}_\theta\}$, as in equation 2.42. Since r and θ depend on time, we have

$$\begin{aligned} \mathbf{e}_r &= \cos\theta\mathbf{i} + \sin\theta\mathbf{j} \\ \dot{\mathbf{e}}_r &= -\sin\theta\,\tfrac{d\theta}{dt}\mathbf{i} + \cos\theta\,\tfrac{d\theta}{dt}\mathbf{j} \\ \dot{\mathbf{e}}_r &= \tfrac{d\theta}{dt}\mathbf{e}_\theta \end{aligned} \qquad \text{and} \qquad \begin{aligned} \mathbf{e}_\theta &= -\sin\theta\mathbf{i} + \cos\theta\mathbf{j} \\ \dot{\mathbf{e}}_\theta &= -\cos\theta\,\tfrac{d\theta}{dt}\mathbf{i} - \sin\theta\,\tfrac{d\theta}{dt}\mathbf{j} \\ \dot{\mathbf{e}}_\theta &= -\tfrac{d\theta}{dt}\mathbf{e}_r \end{aligned}$$

In terms of the polar frame, the position and velocity vectors are

$$\begin{aligned} \mathbf{r} &= r\,\mathbf{e}_r, \\ \mathbf{v} = \dot{\mathbf{r}} &= \tfrac{dr}{dt}\,\mathbf{e}_r + r\dot{\mathbf{e}}_r, \\ &= \tfrac{dr}{dt}\,\mathbf{e}_r + r\tfrac{d\theta}{dt}\mathbf{e}_\theta. \end{aligned}$$

We do not need the acceleration vector in polar coordinates to derive Kepler's first law, but we include it here for completeness.

$$\begin{aligned} \mathbf{a} = \ddot{\mathbf{r}} &= \frac{d}{dt}(\dot{r}\,\mathbf{e}_r + r\dot{\theta}\,\mathbf{e}_\theta), \\ &= \ddot{r}\mathbf{e}_r + \dot{r}\dot{\mathbf{e}}_r + (\dot{r}\dot{\theta} + r\ddot{\theta})\,\mathbf{e}_\theta + r\dot{\theta}\,\dot{\mathbf{e}}_\theta, \\ &= \ddot{r}\mathbf{e}_r + \dot{r}\dot{\theta}\,\mathbf{e}_\theta + (\dot{r}\dot{\theta} + r\ddot{\theta})\,\mathbf{e}_\theta - r\dot{\theta}\dot{\theta}\,\mathbf{e}_r. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{a} = \ddot{\mathbf{r}} &= (\ddot{r} - r\dot{\theta}^2)\mathbf{e}_r + (r\ddot{\theta} + 2\dot{r}\dot{\theta})\mathbf{e}_\theta, \\ &= (\ddot{r} - r\dot{\theta}^2)\mathbf{e}_r + \tfrac{1}{r}\tfrac{d}{dt}(r^2\dot{\theta})\,\mathbf{e}_\theta \end{aligned} \tag{2.55}$$

Now we compute the magnitude L of the angular momentum (see subsection

1.4.4).

$$\begin{aligned}
\mathbf{L} &= m\mathbf{r} \times \mathbf{v}, \\
&= mr\,\mathbf{e}_r \times (\tfrac{dr}{dt}\,\mathbf{e}_r + r\dot{\mathbf{e}}_r), \\
&= mr^2\,(\mathbf{e}_r \times\ \dot{\mathbf{e}}_r), \\
&= mr^2\,\tfrac{d\theta}{dt}(\mathbf{e}_r \times\ \mathbf{e}_\theta), \\
L = \|\mathbf{L}\| &= mr^2\,\tfrac{d\theta}{dt},
\end{aligned} \tag{2.56}$$

since $(\mathbf{e}_r \times\ \mathbf{e}_\theta)$ is a unit vector perpendicular to the plane spanned by the two vectors.
Newton's gravitational law for planetary motion states that the force $\mathbf{F}$ exerted by the sun with mass M on a planet of mass m is given by

$$\mathbf{F} = -\frac{mMG}{r^3}\mathbf{r} = -\frac{mMG}{r^2}\mathbf{e}_r, \tag{2.57}$$

where M is the mass of the sun, m the mass of the planet, G is a universal constant of gravitation, and $\mathbf{r}$ is the position vector of the planet from the sun at the origin of the coordinate system. Since $\mathbf{F} = m\mathbf{a}$, we have

$$\begin{aligned}
\mathbf{a} &= -\frac{MG}{r^3}\mathbf{r}, \\
\mathbf{r} \times \mathbf{a} &= 0.
\end{aligned}$$

Taking the derivative of the angular momentum, we get

$$\begin{aligned}
\frac{d}{dt}\mathbf{L} &= m\frac{d}{dt}(\mathbf{r} \times \mathbf{v}), \\
&= m(\mathbf{v} \times \mathbf{v} + \mathbf{r} \times \mathbf{a}) \\
&= 0.
\end{aligned}$$

Therefore, the angular momentum is constant. It should be noted that the conservation of angular momentum depends only of the fact that $\mathbf{a}$ is a central force, which gives $\mathbf{r} \times \mathbf{a} = 0.$. Since $\mathbf{L}$ is constant, so is its magnitude and direction, hence the motion is on a plane with

$$L = mr^2\frac{d\theta}{dt} = \text{constant.}$$

This is consistent with equation 2.55, from which one infers that the acceleration is purely radial if $(r^2\dot{\theta})$ is constant.
On the other hand, the differential of area in polar coordinates is $dA = \frac{1}{2}r^2\,d\theta$, hence,

$$\frac{dA}{dt} = \tfrac{1}{2}r^2\frac{d\theta}{dt} = \text{constant}$$

Geometrically, this means the radial arm sweeps equal areas in equal times. For a celestial object orbiting the sun with high eccentricity as in the case of a comet, the orbit must whip faster around the sun as the the orbit gets closer to the center of attraction.

Kepler's First Law

One approach to deriving Kepler's first law is to solve directly the differential equation from Newton's laws in polar coordinates for a radial force, The equation is

$$\ddot{r} - r\dot{\theta}^2 = -\frac{MG}{r^2}. \tag{2.58}$$

The process starts by using conservation of angular momentum to write $\dot{\theta}$ as a constant over r^2. Here, we use a well-known vector calculus approach that does not rely much on differential equations. The goal is the same; to integrate the equation of motion to find an expression for r. Since the angular momentum is constant, so is $\mathbf{r} \times \mathbf{v}$. In fact,

$$\|\mathbf{r} \times \mathbf{v}\| = r^2 \tfrac{d\theta}{dt} = \frac{L}{m} = k \quad \text{where } k \text{ is constant.}$$

We begin by computing the cross product of $\mathbf{a}$ with $\mathbf{L}$

$$\begin{aligned}
\mathbf{a} \times \mathbf{L} &= -\frac{MG}{r^2}\,\mathbf{e}_r \times (mr^2\,[\mathbf{e}_r \times \dot{\mathbf{e}}_\theta)], \\
&= -mMG\,[\mathbf{e}_r \times (\mathbf{e}_r \times \dot{\mathbf{e}}_r)], \\
&= -mMG[(\mathbf{e}_r \cdot \dot{\mathbf{e}}_r)\mathbf{e}_r - (\mathbf{e}_r \cdot \mathbf{e}_r)\dot{\mathbf{e}}_r] \quad \text{By the BAC-CAB equation 1.18,} \\
\frac{d\mathbf{v}}{dt} \times \mathbf{L} &= mMG\,\dot{\mathbf{e}}_r.
\end{aligned}$$

Since $\mathbf{L}$ is constant, we can integrate both sides

$$\mathbf{v} \times \mathbf{L} = mMG\,\mathbf{e}_r + \mathbf{c},$$

where $\mathbf{c}$ is a constant vector of integration. We adapt the $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ so that $\mathbf{k}$ points in the direction of $\mathbf{L}$ and $\mathbf{c}$ is aligned with $\mathbf{i}$. Let θ be the angle subtended between $\mathbf{r}$ and $\mathbf{c}$. Now we take the dot product of $\mathbf{r}$ with the last equation, taking notice that on the left hand side with get a triple product we can rearrange. We get

$$\begin{aligned}
\mathbf{r} \cdot (\mathbf{v} \times \mathbf{L}) &= mMG\mathbf{r} \cdot \mathbf{e}_r + \|\mathbf{c}\|\,\|\mathbf{r}\| \cos\theta, \\
\frac{1}{m}(m\mathbf{r} \times \mathbf{v}) \cdot \mathbf{L} &= mMG\,r + rc\cos\theta, \\
\frac{1}{m^2}\|\mathbf{L}\|^2 &= r(MG + c\cos\theta), \\
k^2 &= r(MG + c\cos\theta), \\
r &= \frac{k^2}{MG + c\cos\theta}
\end{aligned}$$

We conclude that

$$r = \frac{ed}{1 + e\cos\theta}, \tag{2.59}$$

where, $ed = k^2/(MG)$ and $e = c/(MG)$. We have found that the motion of the planet is a conic!

Kepler's Third Law

This derivation requires some classical facts about ellipses. Let T be the time it takes for a plane to complete a full orbit. We know that

$$k^2 = r^2 \frac{d\theta}{dt}, \quad \text{and,} \quad \frac{dA}{dt} = \tfrac{1}{2} r^2 \frac{d\theta}{dt},$$

therefore,

$$\frac{dA}{dt} = \tfrac{1}{2} k \qquad \Rightarrow \qquad A = \int_0^T \tfrac{1}{2} k \, dt$$
$$A = \tfrac{1}{2} k T$$

Let a and b be the semi-major and semi-minor axes of the ellipse. The area of the ellipse

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

is $A = \pi ab$. If you have never seen this formula before, peek at equation 4.18 for the derivation of the volume of an ellipsoid, Soon you will be able to prove the area formula yourself by a homologous method. Also, you need to recall that the foci of the ellipse occur at coordinates $F_\pm = (\pm c, 0)$, where $c^2 = a^2 - b^2$. So far we have

$$A = \pi ab = \tfrac{1}{2} kT \qquad \Rightarrow \qquad T = \frac{2\pi ab}{k}$$

Now we compute the **semi-latus rectum** of the ellipse in two different ways. The semi-latus rectum is the positive y-coordinate at the foci. Of course we could define the latus rectum in a coordinate-free manner, more in the spirit of classical geometry. In an ellipse, the latus rectum is the length of the segment perpendicular to the major axis at one of the foci, that is contained inside the ellipse. A similar definition exists for other conics. Back to coordinate geometry, choose the focus $(c, 0)$. We get

$$\frac{c^2}{a^2} + \frac{y^2}{b^2} = 1,$$
$$\frac{y^2}{b^2} = 1 - \frac{c^2}{a^2} = \frac{a^2 - c^2}{a^2} = \frac{b^2}{a^2},$$
$$y = \frac{b^2}{a}$$

Next, in the equation of the celestial conic

$$r = \frac{k^2/MG}{1 + e\cos\theta}$$

set $\theta = \pi/2$. That is also the latus rectum since this is the y-coordinate above the origin where the sun sits at a focus. We get

$$\frac{k^2}{GM} = \frac{b^2}{a} \qquad \Rightarrow \qquad k^2 = \frac{GMb^2}{a}$$

Square the period and insert the value above for k^2.

$$
\begin{aligned}
T^2 &= \frac{4\pi^2 a^2 b^2}{k^2}, \\
&= (4\pi^2 a^2 b^2)\frac{a}{GMb^2}, \\
T^2 &= \frac{4\pi^2}{GM} a^3
\end{aligned}
$$

This is Kepler's third law.

Chapter 3

Partial Derivatives

3.1 Functions of Several Variables

A function of several real variables $f : \mathbf{R}^n \to \mathbf{R}^m$ is a map that takes n real inputs which we can be denoted by a vector $\mathbf{x} = \langle x_0, x_1, \ldots, x_n \rangle$ into a vector that we can denote as $\mathbf{y} = \langle y_0, y_1, \ldots, y_m \rangle$. The standard notation for a function of several variables $f : \mathbf{R}^n \to \mathbf{R}^m$ is $\mathbf{y} = f(\mathbf{x})$. The set of all possible inputs in $\mathbf{R}^n$ is called the domain $D(f)$ of the function and the set of all possible outputs $\mathbf{y} \in \mathbf{R}^m$ is called the range $Rg(f)$. In this section we are primarily interested in real-valued functions of several variables meaning that $m = 1$.
A real-valued function of two variables

$$f : \mathbf{R}^2 \to \mathbf{R},$$
$$(x, y) \in \mathbf{R}^2 \xrightarrow{f} z \in \mathbf{R}.$$

has two inputs and one output. Ordinarily we denote these functions by $z = f(x, y)$. we can graph a real-valued function $z = f(x, y)$ as a surface in $\mathbf{R}^3$ by using the z-axis for the range. Just as in the case of one-variable functions $y = f(x)$, the domain is naturally restricted if:
a) The function has square roots, since the radicand can not be negative,
b) The function has fractions where the denominator has zero's,
c) The function has logarithms, since the argument must be positive.
The domain can also be imposed manually.

Example . Let $z = f(x, y) = \sqrt{4 - x^2 - y^2}$. The domain $D(f)$

$$D(f) = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq 4\},$$

consists of all the points on a disk of radius 4. Squaring both sides, the reader will recognize this as the equation of the top half of a sphere of radius 2. The range is the interval $[0, 2]$. Please note that the domain is a set, so one must use the formal set notation. The range is also a set, but being a subset of the real line, it is simpler to use the equivalent interval notation.

Example Let $z = f(x,y) = \frac{x^2+y^2}{xy}$. The domain consists of all points in $\mathbf{R}^2$ except for those in the x or y-axes. That is,

$$D(f) = \{(x,y)) \in \mathbf{R}^2 : x \neq 0, \text{and } y \neq 0\}.$$

Example Let $w = F(x,y,z) = \ln(36 - 9x^2 - 4y^2 - 36z^2)$. The domain

$$D(F) = \{(x,y,z) \in \mathbf{R}^3 : \frac{x^2}{4} + \frac{y^2}{9} + \frac{x^2}{1} \leq 1\}.$$

consists of all the points on or inside an ellipsoid.
If $z = f(x,y)$, the **level curves** (also called **contours**) of the function are the curves given by setting the z coordinates equal to a constant c. The level curve $z = c$ is the curve of intersection of the surface with the plane $z = c$, that is, the set of all points on the surface at a height $z = c$. If $w = F(x,y,z)$, the contours are **level surfaces** defined implicitly by setting $F(x,y,z) = c$. For example, as shown in figure 1.14 the level surfaces of

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

are elliptic hyperboloids of one sheet for $c > 1$, an elliptic cone for $c = 0$, and elliptic hyperboloids of two sheets for $c < 0$.
By plotting the level curves of $z = f(x,y)$ on a plane, we can visualize the surface as a topographical map

Example

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

. The domain $D(f)$ of this function is all of $\mathbf{R}^2$. The range of the function is $Rg(f) = [0, \infty] \in bfR$, because the sum of two squares is non-negative. The surface is a circular paraboloid, and the level curves is a set of circles as shown in figure 1.14.

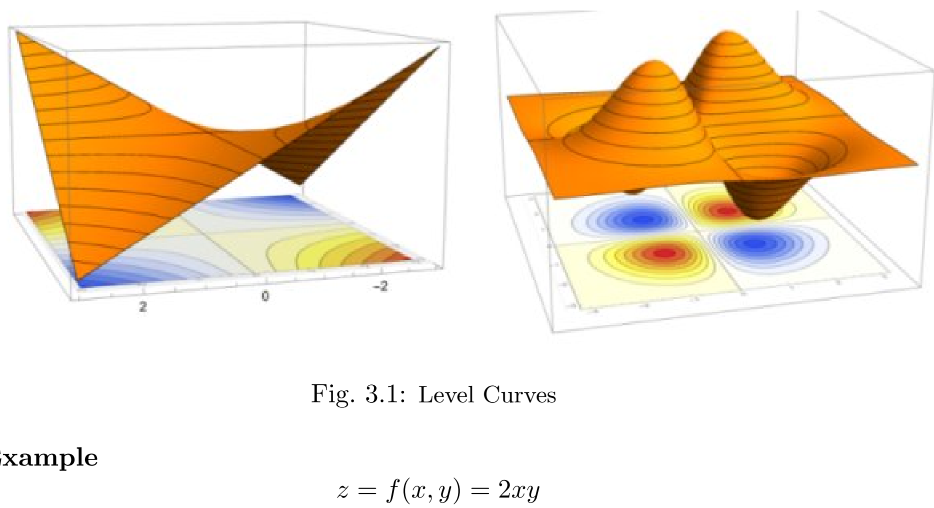


Fig. 3.1: Level Curves

Example

$$z = f(x,y) = 2xy$$

.

The domain $D(f)$ is again all of $\mathbf{R}^2$, because x and y are real numbers that can always be multiplied. The range is $Rg(f) = \mathbf{R}$ because, both x and y can be positive or negative. The set of level curves $z = c$, that is $2xy = c$, is a set of hyperbolas plus the two intersecting lines $x = 0$ and $y = 0$ which correspond to $c = 0$. The level curves are depicted on the left in figure 3.1. The hyperbolas here have exactly the same shape as the hyperbolas $x^2 - y^2 = c$ that appear in 1.14. The surface $z = 2xy$ is a saddle, shown here with some level curves, along with the 2 dimensional topographical map below the surface. Should the surface describe a temperature distribution $T(x, y) = 2xy$, we have picked a color scheme for the level curves on the plane with shades of yellow and red color to indicate hot and shades of blue color to indicate cold. The temperature would be zero on the coordinate axes around which the color is a neutral white. If we think of the surface as the heights of a terrain, the areas shaded shaded in blue would be below sea level.

Example

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

Without the factor xy this would be a surface of revolution generated by a normal (Gaussian) distribution curve. The xy factors cause the surface to dip under the xy-plane when the variables have different signs, as shown in figure 3.1. The level curves are displayed on a plane using the same temperature-color scheme as in the saddle above.

3.2 Limits and Continuity

The concept of a limit of a real-valued function of one variable, as envisioned by the early developers of calculus in the 1600's, is most intuitive. It basically says that the limit of a function $y = f(x)$ as x approaches a point x_0 is L if as x gets arbitrarily close to x_0, the function $f(x)$ gets arbitrarily close to L. The notion of "arbitrarily close" lacks mathematical rigor. The current rigorous definition of limits was introduced by Augustin-Louis Cauchy, a French mathematician who was born in 1789, the same year as the French Revolution. The first rigorous definition of continuity of a function at a point was given by Karl Weierstrass.

The rigorous definition of limits for functions of several variables captures essentially the same spirit as the one variable version. We may as well write the definition for vector-valued function of several variables, although in this section we are primarily concerned with real-valued functions. Denote the distance formula in $\mathbf{R}^n$ by the d_n and the distance formula in $\mathbf{R}^m$ by d_m. In the case of real-valued functions ($m = 1$), the distance formula between two points is the absolute value of their difference. We have

3.2.1 Definition Let $f : \mathbf{R}^n \to \mathbf{R}^m$ be a vector valued function $\mathbf{y} = f(\mathbf{x})$. We say that

$$\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = \mathbf{y}_0. \tag{3.1}$$

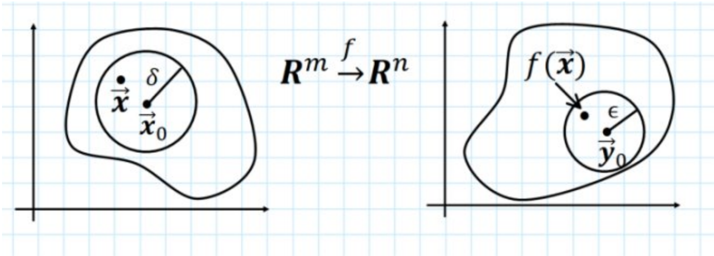


Fig. 3.2: Limits

if given any number $\epsilon > 0$, there exists a number $\delta > 0$, such that $d_m(f(\mathbf{x}), \mathbf{y_0}) < \epsilon$, whenever $d_n(\mathbf{x}, \mathbf{x_0}) < \delta$.

Just as in the case of one-variable, this definition is a mouthful and we need to try to unpack it a bit. The number ϵ is arbitrary, but we usually think of ϵ as a small number. What small number? Any. We can pick ϵ to be 0.1, or 0.001, or 10^{-8} or any other number as small as we desire. We can interpret ϵ as a tolerance of how close we want the value of the function $f(\mathbf{x})$ to be to $\mathbf{y_0}$. We can visualize this as in figure 3.2 in which we show a diagram for a function from $\mathbf{R}^2$ to $\mathbf{R}^2$. In the figure, we want the value of the function to be within a radius ϵ of $\mathbf{y}_0$. The limit exists if this can be always be achieved by choosing a radius δ (which necessarily depends of ϵ) with $\mathbf{x}$ less than δ units of $\mathbf{x}_0$. This captures the idea that $f(\mathbf{x})$ is "ϵ-close" to $\mathbf{y}_0$ whenever $\mathbf{x}$ is "δ-close" to $\mathbf{x_0}$. If the function $\mathbf{y} = f(\mathbf{x})$ is defined at $\mathbf{x}_0$, the limit exists, and $f(\mathbf{x}_0) = \mathbf{y}_0$, we say that the function is **continuous** at that point. In other words, $f(\mathbf{x})$ is continuous at $\mathbf{x}_0$ if

$$\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0). \tag{3.2}$$

Establishing when a limit of function of several variables exists at a point is by far more difficult that for one-variable calculus functions. In the latter, the domain is a set of real numbers, so when we say that x approaches x_0, this can only happen from the left or from the right. In $\mathbf{R}^2$ for example, when we say that $\mathbf{x}$ approaches $\mathbf{x}_0$, this can happen along an infinite set of paths. We also do not have a several variable version of l'Hôpital's rule to compute limits.

The Limit theorem and the main theorems about single-variable continuous functions extend in a natural way to several variables. Thus, for example, given two function which are continuous at a particular point, then so is their sum, difference and product. The quotient is also continuous whenever the denominator is not zero. But of course, the most interesting cases are precisely the latter. The number of examples in this course, for which we can determine that the limit of a quotient of functions of several variables exists at a point where the denominator is zero, is a very small set indeed. Therefore, we will be content with primarily solving problems determining that certain limits do not exist. This is a much, easier task, because if the limit exists, then it has to be unique. Therefore, for the limit not to exist, it suffices to show that the function approaches a different number for just two or more paths.

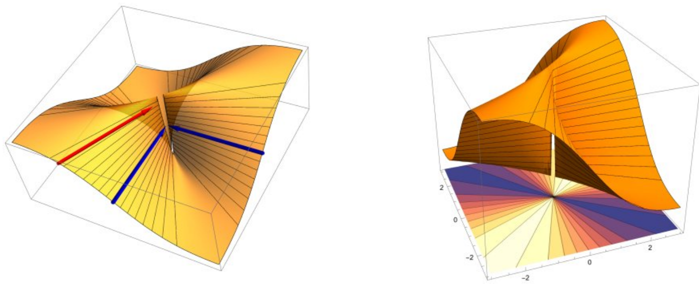


Fig. 3.3: Limit does not exist

Example Show that the limit

$$\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2},$$

does not exist.
Solution

At any point in $\mathbf{R}^2$ other than the origin $(0,0)$, the function is the ratio of two polynomials so the function is continuous. At $(0,0)$ the function is not defined, but the limit might still exist - we show that it doesn't.

The surface with level curves is pictured in figure 3.3. Inspection of the surface shows that there is some "Wonton"-like pinching. This is an indication that something bad is happening at the origin. The figure shows two (blue) paths along which one approaches the same value, but from that, one can conclude nothing. The problem is that as one approaches the origin along other straight lines of different slopes, the function may approach different values, instead of a unique value. We can quantify this as follows. Chose $y = mx$ and investigate what happens as we approach origin along these paths. Along $y = mx$, we have

$$\begin{aligned}
\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2} &= \lim_{(x,mx)\to(0,0)} \frac{x(mx)}{x^2+(mx)^2}, \\
&= \lim_{(x,mx)\to(0,0)} \frac{mx^2}{x^2+m^2x^2}, \\
&= \lim_{x\to 0} \frac{mx^2}{x^2(1+m^2)}, \\
&= \frac{m}{1+m^2}
\end{aligned}$$

This means that for paths of different slopes one gets different numbers. Thus, for example, along $y = 0$ for which $m = 0$ the result is 0. but along $y = x$, for which $m = 1$, the result is $1/2$. That alone does it. The limit can't depend on m, so the limit does not exist (DNE).
Note: Since we are working with limits, it is essential that we adhere strictly to the rules of syntax. A syntax error is a mathematical error. Also note, that this being a proof, it must have a conclusion.

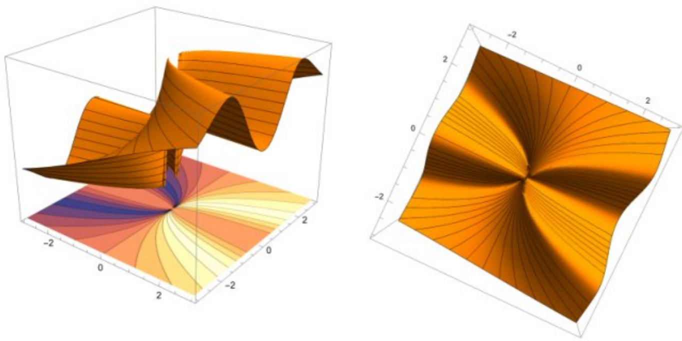


Fig. 3.4: Limit does not exist

Example Show that the limit

$$\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+3y^4},$$

does not exist.
Solution

The pinching at the origin shown in the graph of the function in figure 3.4 suggests that the limit does not exist there. We would like to choose a power function path along which all the variables cancel. We choose the family of parabolas $x = my^2$. Along these paths, we have

$$\begin{aligned} \lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+3y^4} &= \lim_{(my^2,y)\to(0,0)} \frac{(my)^2y^2}{(my^2)^2+3y^4}, \\ &= \lim_{(my^2,y)\to(0,0)} \frac{my^4}{m^2y^4+3y^4}, \\ &= \lim_{x\to 0} \frac{my^4}{y^4(m^2+3)}, \\ &= \frac{m}{m^2+3} \end{aligned}$$

We get different results as we approach the origin along different parabolas. The limit can't depend on m, so the limit does not exist.
At the origin, the two functions above have a sort of higher dimensional version of jump discontinuity. Here is another kind of example with most interesting behavior.

3.2.2 Example Investigate the limit

$$\lim_{(x,y)\to(0,0)} \frac{xy(x^2-y^2)}{x^2+y^2},$$

Solution. the function $f(x,y) = xy(x^2-y^2)/(x^2+y^2)$ is not defined at the origin, so it is not continuous there. The graph of the function along with some

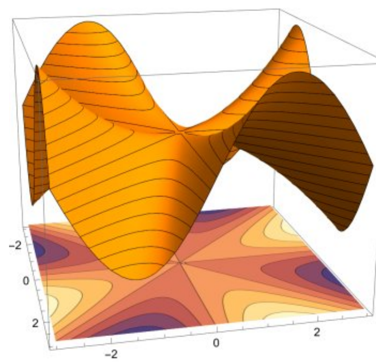


Fig. 3.5: Limit exists

level curves appears in figure 3.5. For lack of a better name we will call this function as a "4-leg-saddle," or a "foul saddle." The reader might be surprised to see that the graph appears to be totally well behaved as we approach the origin; but we know there must be a hole at $(0, 0)$. We should be fairly confident that the limit exists. In general, proving that a limit exists must involve an ϵ-δ argument and this is hard. However, in this special example there is a neat trick - we convert to cylindrical coordinates. Let $x = r\cos\theta$ and $y = r\sin\theta$, so that $x^2 + y^2 = r^2$ We get

$$
\begin{aligned}
\lim_{(x,y)\to(0,0)} \frac{xy(x^2-y^2)}{x^2+y^2} &= \lim_{(r,\theta)\to(0,0)} \frac{r^2\cos\theta\sin\theta(r^2\cos^2\theta - r^2\sin^2\theta)}{r^2}, \\
&= \lim_{(r,\theta)\to(0,0)} \frac{r^4\cos\theta\sin\theta(\cos^2\theta - \sin^2\theta)}{r^2}, \\
&= \lim_{(r,\theta)\to(0,0)} \frac{r^4\sin 2\theta\cos 2\theta}{2r^2}, \\
&= \lim_{(r,\theta)\to(0,0)} \frac{r^4\sin 4\theta}{4r^2}, \\
&= \lim_{(r,\theta)\to(0,0)} \frac{1}{4}r^2\sin 4\theta
\end{aligned}
$$

,

which is 0, independent of θ. By extending the definition of the function so that $f(0, 0) = 0$, we get a continuous function. In the computer generated graph one would not even notice the difference.

If we try to apply the same trick to convert the "wonton" surface to cylindrical coordinates, we get

$$\begin{aligned} z &= \frac{xy}{x^2+y^2}. \\ &= \frac{r\cos\theta\, r\sin\theta}{r^2}, \\ &= \cos\theta\sin\theta, \quad r>0, \\ &= 2\sin 2\theta, \end{aligned}$$

So for any particular value $\theta = \theta_o$, the level curve is the semi-infinite horizontal line of intersection of the surface and the plane $z = 2\sin 2\theta_o$. Thus, the surface is "ruled" by these lines with heights oscillating between -1 and 1 as theta goes around a full period. This explains why the surface looks the way it does, and why the function approaches different values depending on the angle of approach to the origin.

3.3 Partial Derivatives

We recall from calculus of one variable that if one has a function $y = f(x)$, the derivative of the function at a point is defined by limit of the ratio $\Delta y/\Delta x$ as the increment Δx approaches 0. The ratio $\Delta y/\Delta x$ represents the slope of a secant line passing through two nearby points. The limit of this as the increment Δx approaches zero then has the geometric interpretation as the slope of the tangent line to the curve. We would like to capture this concept for a real-valued multivariate function. We introduce partial derivatives for a real-valued function of 2 variables, but the definition can be easily extended to several variables. Let $z = f(x, y)$ be a function, and $P(x_0, y_0)$ be a point in the domain of the that function. We define

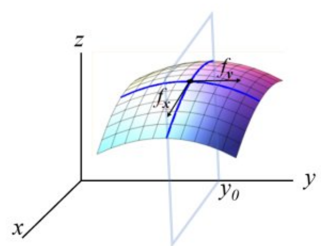


Fig. 3.6: Partial Derivative

$$\begin{aligned} f_x(x, y_0) &= \frac{\partial f}{\partial x}(x, y_0) = \lim_{h\to 0}\frac{f(x+h, y_0) - f(x, y_0)}{h}, \\ f_y(x_0, y) &= \frac{\partial f}{\partial y}(x_0, y) = \lim_{k\to 0}\frac{f(x_0, y+k) - f(x_0, y)}{k}, \end{aligned} \tag{3.3}$$

The limits may, or may not exist. If the limits exist, they are called the **partial derivatives** of the function. The geometry of the definition is illustrated in figure 3.6. When one is extracting the partial derivative with respect to x one holds the value of y constant, so only the variation of the function in the x direction is considered. We may think of $f_x(x, y_0)$ as the slope of the curve of intersection of the surface with the vertical plane $y = y_0$. This is also illustrated in figure 3.6. In a similar manner, when one computes the partial derivative

$f_y(x_0, y)$, one holds the x coordinate constant and only considers the variation of the function in the y direction. Other than that, the definition has exactly the same spirit as the corresponding definition of derivatives in one variable calculus.

It follows that if one can find the derivative of a function of one variable, one can also compute partial derivatives, by the following procedure. When taking partial derivatives with respect to a variable, treat all other variables as if they were constant. Otherwise, all the rules of differentiation and all the formulas for derivatives of functions of one variable, still apply. Several notations for partial derivatives are found in the literature. Here are the most common ones which we will use in this book. Let $z = f(x, y)$

$$
\begin{aligned}
f_x = z_x &= \frac{\partial f}{\partial x} = \frac{\partial z}{\partial x}, \\
f_y = z_y &= \frac{\partial f}{\partial y} = \frac{\partial z}{\partial y}, \\
f_{xx} = z_{xx} &= \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 z}{\partial x^2}, \\
f_{xy} = z_{xy} &= \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x},
\end{aligned}
$$

and so on for other higher order partial derivatives. Note that in the mixed partials, the quantity $\partial^2 f/(\partial y \partial x)$ means that one first differentiates with respect to x and then with respect to y. In the notation in the bottom, the variable y appears first, which is in the opposite lexicographical order of the corresponding quantity f_{xy}. We do not have to worry much about the order of the mixed partial derivatives, because of the following theorem

3.3.1 Theorem (Clairaut) If the functions f_{xy} and f_{yx} are both continuous on an open region containing the point (x_0, y_0), then

$$f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$$

The conditions of this theorem are met by almost all functions $z = f(x, y)$ treated here. The first counter-example to the theorem showing that the continuity condition is necessary was found by Peano circa 1900. The most common function for which the mixed partial are not equal at a particular point is the "4-leg-saddle" at the origin. At the origin the first derivatives $f_x(0, 0)$ and $f_y(0, 0)$ exist and are continuous, but the mixed partials are not and the theorem fails.
Example Let $f(x, y) = x^3 + xy^2 + y^3$. Compute f_x and f_y.
Solution.

$$
\begin{aligned}
f_x &= 3x^2 + y^2, \\
f_y &= 2xy + 3y^2.
\end{aligned}
$$

Example Let $f(x, y) = x \sin(xy)$. Compute f_x and f_y.

Solution

$$\begin{aligned} f_x &= x\frac{\partial}{\partial x}\sin(xy) + \sin(xy)\frac{\partial}{\partial x}x, \\ &= xy\cos(xy) + \sin(xy), \\ f_y &= x^2\cos(xy). \end{aligned}$$

Example Let $f(x,y) = e^{xy^2}\ln(x+y^2)$
Solution

$$\begin{aligned} f_x &= y^2 e^{xy^2} + \frac{1}{x+y^2}, \\ f_y &= 2xye^{xy^2} + \frac{2y}{x+y^2}. \end{aligned}$$

Example Let $f(x,y) = x^2y^3 + e^{x^2y}$. Compute all the first and second partial derivatives.
Solution

$$\begin{aligned} f_x &= 2xy^3 + 2xy\, e^{x^2y}, \\ f_y &= 3x^2y^2 + x^2\, e^{x^2y}, \\ f_{xx} &= 2y^3 + 2xy\, e^{x^2y}(2xy) + 2y\, e^{x^2y}, \\ &= 2y^3 + 2y(2x^2y+1)e^{x^2y}, \\ f_{xy} &= 6xy^2 + 2xy\, e^{x^2y}(x^2) + 2x\, e^{x^2y}, \\ &= 6xy^2 + 2x(x^2y+1)e^{x^2y} = f_{yx}, \\ f_{yy} &= 6x^2y + x^4\, e^{x^2y}. \end{aligned}$$

Example Consider the ideal gas law $PV = nRT$. Find $\partial P/\partial T$ and $\partial P/\partial V$.
Solution

$$\begin{aligned} P &= \frac{nRT}{V}, \\ \frac{\partial P}{\partial T} &= \frac{nR}{V}, \\ \frac{\partial P}{\partial V} &= -\frac{nRT}{V^2}, \end{aligned}$$

3.3.2 Example Mixed Partials not Equal

The function

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

shown in picture 3.5 is exceptional at the origin for it provides a rare example of a function for which $f_{xy} \neq f_{yx}$. In fact, by the quotient rule we get,

$$f_x(x,y) = \frac{y(x^4+4x^2y^2-y^4)}{(x^2+y^2)^2}, \quad \text{if}(x,y) \neq (0,0),$$
$$f_y(x,y) = \frac{x(x^4-4x^2y^2-y^4)}{(x^2+y^2)^2}, \quad \text{if}(x,y) \neq (0,0).$$

For the first partial derivatives at $(0,0)$, we have

$$f_x(0,0) = \lim_{h\to 0} \frac{f(h,0)-f(0,0)}{h} = \lim_{h\to 0} \frac{(0/h^2)-0}{h} = 0,$$
$$f_y(0,0) = \lim_{k\to 0} \frac{f(0,k)-f(0,0)}{k} = \lim_{k\to 0} \frac{(0/k^2)-0}{k} = 0.$$

Therefore, $f_x(0,y) = -y$ for all y, which implies that,

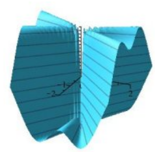
$$f_{xy}(0,y) = -1, \quad f_{xy}(0,0) = -1.$$

Similarly, $f_y(x,0) = x$ for all x, which implies that,

$$f_{yx}(x,0) = 1, \quad f_{yx}(0,0) = 1.$$

In accordance with Clairaut's theorem, it must be the case that f_{xy} or f_{yx} is not continuous at $(0,0)$. For $(x,y) \neq (0,0)$,

$$f_{xy} = f_{yx} = \frac{x^6+9x^4y^2-9x^2y^4-y^6}{(x^2+y^2)^3}.$$



A plot of the function f_{xy} reveals almost the same type of singularity at the origin as the "wonton" surface. The level curves are straight lines but the lines appear to approach the origin at different values depending upon the direction. Indeed, let $y = mx$. Evaluating the function f_{xy} for $(x,y) \neq (0,0)$, along the directions $y = mx$, we find

$$f_{xy}(x,mx) = \frac{1-m^6}{(1+m^2)^3}.$$

Hence, the limit as $(x,mx) \to (0,0)$ gives different values which depend on m. This implies that the limit does not exist and hence the function f_{xy} is not continuous at the origin.

3.3.1 Laplacian

A partial differential equation (PDE) is an equation that contains partial derivatives. If the equation only contains the first partial derivatives, the equation is called a PDE of first order, and if it contains one or more second partial derivatives, it is called PDE of second order. The process of finding solutions of partial differential equations is the subject of a more advanced course. However, the process of establishing whether of not a particular function satisfies a PDE, is elementary, since all one has to do is insert the function into the equation and verify that it holds. For example, we can verify that the function $u(x,t) = \sin(x+3t)$ satisfies the equation $u_x - (1/3)u_t = 0$ because

$$u_x - \tfrac{1}{3}u_t = \cos(x+3t) - \tfrac{1}{3}\cos(x+3t)(3) = 0.$$

Definition The **Laplacian** (in Cartesian coordinates) is the differential operator ∇^2 defined by

$$\begin{aligned}
\nabla^2 u &= \frac{d^2 u}{dx^2}, && \text{if } u = u(x),\\
\nabla^2 u &= \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}, && \text{if } u = u(x,y),\\
\nabla^2 u &= \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}, && \text{if } u = u(x,y,z), \qquad (3.4)
\end{aligned}$$

The quantity ∇^2 is called a differential operator because it is looking for a function input on which to operate. The output is another function. The central application of differential calculus of several variables to STEM fields is mathematical modelling of physical phenomena by partial differential equations. It turns out that the PDE models of the main branches of mathematical physics all involve the Laplacian. The operator enters into the field equations of Newtonian gravitation, classical theory of electricity and magnetism via Maxwell's equations, the wave equation including the models for propagation of light, sound, water and gravitational waves, and Schrödinger's equation in quantum mechanics. We present a very introductory note to the three basic classical models featuring the Laplacian

3.3.2 Laplace's equation

The equation

$$\nabla^2 u = 0$$

is called the **Laplace equation**. It is a special case of the potential theory Poisson equation $\nabla^2 u = \rho$, where ρ is a function typically representing a density of mass, or charge creating the field. The solutions of the Laplace equation are called **harmonic functions**. We show some examples of harmonic functions in two variables

Example Show that the function $u(x,y) = x^3 - 3xy^2$ is harmonic.

Solution.

$$\begin{aligned} u_x &= 3x^2 - 3y^2 \\ u_y &= -6xy \end{aligned} \qquad \text{And} \qquad \begin{aligned} u_{xx} &= 6x \\ u_{yy} &= -6x \end{aligned}$$

So

$$u_{xx} + u_{yy} = 6x - 6x = 0$$

As a teaser to stimulate your thirst for knowledge, we introduce a most remarkable fact one learns in a more advanced course in complex variables. All harmonic functions of two variables arise as the real and imaginary parts of differentiable complex functions $w = f(z)$, where $z = x + iy$ and $i = \sqrt{-1}$. More specifically, if the complex function w is separated into its real and imaginary parts $w = u(x, y) + iv(x, y)$, then, both, u and v are harmonic functions. The pair are called **conjugate harmonic**.

Example Let $w = f(z) = z^3$. Then

$$\begin{aligned} w &= (x+iy)^3, \\ &= x^3 + 3x^2(iy) + 3x(iy)^2 + (iy)^3, \\ &= (x^3 - 3xy^2) + i\,(3x^2y - y^3), \\ u &= (x^3 - 3xy^2), \quad v = 3x^2y - y^3. \end{aligned}$$

The function $u(x, y) = x^3 - 3xy^3$ is the one that appears in the preceding example.

Example Let Let $w = f(z) = z^4$. Then

$$\begin{aligned} w &= (x+iy)^4, \\ &= x^4 + 4x^3(iy) + 6x^2(iy)^2 + 4x(iy)^3 + (iy)^4, \\ &= (x^4 - 6x^2y^2 + y^4) + i\,(4x^3y - 4xy^3), \\ u &= x^4 - 6x^2y^2 + y^4, \quad v = 4x^3y - 4xy^3. \end{aligned}$$

We verify that the function $u(x, y) = x^4 - 6x^2y^2 + y^4$ is harmonic.

$$\begin{aligned} u_x &= 4x^3 - 12xy^2 \\ u_y &= -12x^2y + 4y^3 \end{aligned} \qquad \text{And} \qquad \begin{aligned} u_{xx} &= 12x^2 - 12y^2 \\ u_{yy} &= -12x^2 + 12y^2, \end{aligned}$$

So

$$u_{xx} + u_{yy} = 0$$

.

Example Complex logarithm

Consider complex numbers in polar coordinates

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta. \end{aligned} \qquad \text{and} \qquad \begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \tan^{-1}(y/x). \end{aligned}$$

We assume that $0 \le \theta < 2\pi$ to avoid multiple labels for the points. Then

$$\begin{aligned} z &= x + i\,y, \\ &= r\cos\theta + i\,r\sin\theta, \\ &= r(\cos\theta + i\ \sin\theta), \\ &= r\,e^{i\theta.} \end{aligned}$$

Then, the (principal part of) the complex logarithm function is

$$\begin{aligned} w &= \ln z, \\ &= \ln(r\, e^{i\theta}), \\ &= \ln r + i\,\theta, \\ &= \ln(\sqrt{x^2+y^2}) + i\ \tan^{-1}(y/x). \\ u &= \ln(\sqrt{x^2+y^2}); \quad v = \tan^{-1}(y/x). \end{aligned}$$

. We verify that the function $u = \ln(\sqrt{x^2+y^2}) = (1/2)\ln(x^2+y^2)$ is harmonic.

$$\begin{aligned} u_x &= \frac{x}{x^2+y^2} \\ u_y &= \frac{y}{x^2+y^2} \end{aligned} \qquad \text{And} \qquad \begin{aligned} u_{xx} &= \frac{y^2-x^2}{(x^2+y^2)^2} \\ u_{yy} &= \frac{x^2-y^2}{(x^2+y^2)^2}, \end{aligned}$$

So

$$u_{xx} + u_{yy} = 0.$$

The function $u = \ln(1/r)$ is (up to a constant) the potential due a current on a straight wire. We leave to reader to verify that $v(x,y) = \tan^{-1}(y/x)$ is also harmonic.

Example Complex cosine

Starting with Euler's formula, we have

$$\begin{aligned} e^{ix} &= \cos\theta + i\ \sin\theta, \\ e^{-ix} &= \cos\theta - i\ \sin\theta. \end{aligned}$$

Adding the two equations above and dividing by 2, we get,

$$\cos x = \frac{e^{ix}+e^{-ix}}{2}.$$

Then, the complex cosine function is,

$$\begin{aligned} w = \cos z &= \frac{1}{2}(e^{iz} + e^{-iz}), \\ &= \frac{1}{2}[e^{i(x+iy)} + e^{-i(x+iy)}], \\ &= \frac{1}{2}[e^{ix}e^{-y} + e^{-ix}e^{y}], \\ &= \frac{1}{2}[e^{-y}(\cos x + i\ \sin x) + e^{y}(\cos x - i\ \sin x)], \\ &= \cos x\left(\frac{e^y+e^{-y}}{2}\right) + i\ \sin x\left(\frac{e^{-y}-e^{y}}{2}\right), \\ &= \cos x\cosh y - i\ \sin x \sinh y. \\ u &= \cos x \cosh y, \quad v = -\sin x \sinh y \end{aligned}$$

It follows immediately that

$$u_{xx} + u_{yy} = -\cos x \cosh y + \cos x \cosh y = 0,$$

so u is harmonic; by a similar computation, so is v.
Table 3.1 shows some common conjugate harmonic functions in some domain, with the generating complex function $w = f(z) = u + i\, v$

$w = f(z)$	$u(x, y)$	$v(x, y)$
$w = z^2$	$u = x^2 - y^2$	$v = 2xy$
$w = z^3$	$u = x^3 - 3xy^2$	$v = 3x^2y - y^3$
$w = z^4$	$u = x^4 - 6x^2y^2 + y^4$	$v = 4x^3y - 4xy^3$
$w = e^z$	$u = e^x \cos y$	$v = e^x \sin y$
$w = \ln z$	$u = \ln(\sqrt{x^2 + y^2})$	$v = \tan^{-1}(y/x)$
$w = \cos z$	$u = \cos x \cosh y$	$v = -\sin x \sinh y$
$w = \sin z$	$u = \sin x \cosh y$	$v = \cos x \sinh y$
$w = 1/z$	$u = x/(x^2 + y^2)$	$v = -y/(x^2 + y^2)$

Table 3.1: List of some conjugate harmonics

Harmonic functions in two variables give rise to very neat surfaces in $\mathbf{R}^3$. In figure 3.7 we display some of them in **cylindrical coordinates**. Cylindrical coordinates just means we convert x and y to polar coordinates. The first graph is the saddle $u(x, y) = x^2 - y^2$ that comes from the real part of $f(z) = z^2$. In cylindrical coordinates the equation is

$$z = r^2 \cos^2 \theta - r^2 \sin^2 \theta = r^2 \cos 2\theta.$$

The second graph is called a monkey saddle because it has a place for the tail of the monkey. The monkey saddle comes from the real part of the function $f(z) = z^3$ which in cylindrical coordinates is $z = r^3 \cos 3\theta$. The third graph on the right is the real part of $f(z) = z^4$ with cylindrical equation $z = r^4 \cos 4\theta$. There is no classical name for the latter, this is basically the same we called a "foal-on-horse" saddle because a baby horse riding on papa-horse would have space for the four legs. The function $z = r^8 \cos 8\theta$ would give an "octopus saddle".
We emphasize that the expectation on this topic at this level reduces to verifying whether or not a function of two variables is harmonic.
Example Inverse Square Law
For harmonic functions in three variables we mention only one example, however, this is the most important of all harmonic functions. Let $\mathbf{r} = x\,\mathbf{i} + y\,\mathbf{j} + x\,\mathbf{k}$ be the position vector of a point in $\mathbf{R}^3$ with distance $r = \|\mathbf{r}\|$ to the origin. Let $\phi = -1/r$. We show that

$$\nabla^2 \phi = \nabla^2 \left(-\frac{1}{r}\right) = 0, \quad r \neq 0. \tag{3.5}$$

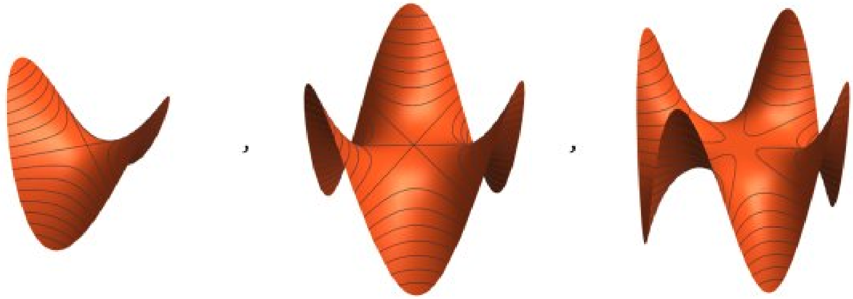


Fig. 3.7: a) Saddle b) Monkey Saddle c) "Foal-on-Horse" Saddle

is a solution of Laplace's equation.

$$\begin{aligned}
\phi(x,y,z) &= -\frac{1}{\sqrt{x^2+y^2+z^2}},\\
\phi_x &= \frac{x}{(x^2+y^2+z^2)^{3/2}},\\
\phi_y &= \frac{y}{(x^2+y^2+z^2)^{3/2}},\\
\phi_z &= \frac{z}{(x^2+y^2+z^2)^{3/2}}.
\end{aligned}$$

In the next section we will learn that the vector whose components are the first partial derivatives, is called the gradient, and is denoted by the symbol $\nabla\phi = \langle \phi_x, \phi_y, \phi_z \rangle$. Let $\mathbf{F}$ be a force vector given by,

$$\mathbf{F} = -\nabla\phi$$

Using this notation, we have

$$\mathbf{F} = \nabla\left(\frac{1}{r}\right) = -\frac{\mathbf{r}}{r^3} \tag{3.6}$$

Up to a constant, this is the **inverse square law** for Newtonian gravitation and electromagnetism. This is because the magnitude $\|\mathbf{F}\|$ of the force vector is $1/r^2$. For gravitation, the force field is

$$\mathbf{F}_g = \nabla\left(\frac{MG}{r}\right) = -\frac{MG\mathbf{r}}{r^3}, \tag{3.7}$$

where M is the mass of the gravitational source and G is Newton's universal gravitation constant.

For the second partial derivatives we use the quotient rule and simplify the result by multiplying top and bottom by the square root term to eliminate the double fraction. We leave the details to the reader and just present the results.

$$\begin{aligned}\phi_{xx} &= \frac{-2x^2+y^2+z^2}{(x^2+y^2+z^2)^{5/2}},\\ \phi_{yy} &= \frac{x^2-2y^2+z^2}{(x^2+y^2+z^2)^{5/2}},\\ \phi_{zz} &= \frac{x^2+y^2-2z^2}{(x^2+y^2+z^2)^{5/2}}.\end{aligned}$$

Adding the results, we get $\phi_{xx}+\phi_{yy}+\phi_{zz}=0$

3.3.3 Wave Equation

Let $u=u(x,y,z,t)$ The wave equation is the PDE

$$\nabla^2 u = \frac{1}{v^2}\,\frac{\partial^2 u}{\partial t^2}, \quad v = \text{constant} \tag{3.8}$$

The constant v represents the speed of the wave. The wave operator

$$\Box = \nabla^2 - \frac{1}{v^2}\,\frac{\partial^2}{\partial^2 t}$$

enters into the physical model of any wave propagation phenomena, whether it refers to light, sound, water, or wave on a string. In this unit we will only be concerned with verifying whether or not a particular function $u(x,t)$ satisfies the wave equation in one dimension

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{v^2}\,\frac{\partial^2 u}{\partial t^2} \tag{3.9}$$

If one were to pluck a string on a guitar, the sound wave generated would be described by such an equation. A more realistic model would include a friction term and boundary conditions.

3.3.3 Example Show that function $u(x,t)=\sin(x-vt)$ is a solution of the one-dimensional wave equation.
Solution. This is a very easy problem. We compute the partial derivatives and insert into the equation to see if it holds. It amounts to a simple application of the chain rule.

$$\begin{aligned} u_x &= \cos(x-vt) \\ u_t &= -v\cos(x-vt)\end{aligned} \quad \text{and} \quad \begin{aligned} u_{xx} &= \sin(x-vt) \\ u_{tt} &= (-v)^2\sin(x-vt)\end{aligned}$$

So,

$$u_{xx} - (1/v^2)\,u_{tt} = 0.$$

The solution has a fairly obvious interpretation. If in a function $y=f(x)$ one replaces x by $(x-h)$, the graph of the function is shifted to the right by an amount h. Here, time t is a continuous variable, so the graph of the function

$u = \sin(x - vt)$ represents a sine wave moving to the right with speed v. The sine function can be replaced by any twice differentiable function

$$u(x,t) = f(x - vt) = f(\xi), \quad \text{where } \xi = x - vt$$

Then, by the chain rule again

$$u_{xx} = f''(\xi), \quad u_{tt} = (-v)^2 f''(\xi)$$

so this is also a solution of the wave equation. The general solution of the one-dimensional wave equation is of the form

$$u(x,t) = f(x - vt) + g(x + vt), \tag{3.10}$$

where, f and g are twice differentiable functions. The solution represents a wave form moving to right plus a wave form moving to the left, both with speed v. One can corroborate this visually by creating a "bump" disturbance on a stretched slinky.

3.3.4 Heat Equation

Let $u = u(x, y, z, t)$. The heat equation, also called the diffusion equation is given by

$$\frac{\partial u}{\partial t} = \alpha^2 \nabla^2 u, \quad \alpha = \text{constant}. \tag{3.11}$$

The generic solution of the heat equation in one dimension

$$\frac{\partial u}{\partial t} = \alpha^2 \frac{\partial^2 u}{\partial x^2}$$

is of the form

$$u(x,t) = e^{-\alpha^2 k^2 t^2}(A \cos kx + B \sin kx),$$

where A, B and k are constants. It is very easy to verify that this is a solution. The solution is sinosoidal with an amplitude that attenuates exponentially. We leave it at that.

3.4 The Differential

3.4.1 The Differential in one Variable

We begin this section with a short review of the fundamentals of one variable calculus in terms of differentials. Anticipating that we will extend the notion to two variables we denote a differentiable real-valued function by $z = f(x)$. The function is graphically represented by a curve, as shown in figure 3.8. Let $z_0 = f(x_0)$ be a point on the curve. We are interested in finding the slope of the tangent line to the curve at the point $P(x_0, z_0)$. We pick a point $x = x_0 + \Delta x$ nearby and denote the change of the function to get to the point $z = f(x)$ by Δz. We define the derivative of the function at the point (x_0, z_0) as the limit of the slope $\Delta z/\Delta x$, as Δx approaches 0. At the risk of a temporary abuse of notation, we interchangeably denote the derivative $dz/dx = f'(x)$ by the symbol z_x or f_x. We have

$$\frac{dz}{dx} = \lim_{\Delta x \to 0} \frac{\Delta z}{\Delta x} = z_x = f_x.$$

The limit is a rigorous concept that can be made very precise with the ϵ-δ language invented by Cauchy in the 1800's. The limit can be computed by step by step process, leading eventually to the list of derivatives rules and the formulas for the derivatives of the elementary functions. We are taught again and again that dy/dx is not a fraction, but rather, a process represented by the limit definition. But now, something rather peculiar is done in standard calculus books. We define the **differential** by the equation

$$dz = z_x\, dx. \tag{3.12}$$

The peculiar object here is dx. Counter to the intuition of Newton and Leibniz who called dx an infinitesimal, we are told that dx is really an independent variable that can be given any value. This is absolute non-sense. If dx were really a variable in the usual sense as we understand it, then we should be able to take the square root, or exponential of dx, which of course would be grievous error! Intuitively, dx should be something "smaller" than Δx, but that makes no sense either, since Δx is approaching 0 and there is no smallest number that is not 0. So, for now, we prefer to stick with Newton and Leibnitz, and just refer to dx as an infinitesimal, which conjures the idea of something very small. We will have to wait until a more advanced course to make rigorous sense of dx as explained by Eli Cartan in the early 1900's. Having accepted dx exists, then the differential dz is fine. The differential leads to three fundamental concepts that we wish to extend to multivariate functions.

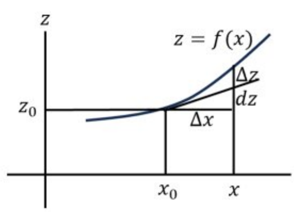


Fig. 3.8: Differentials

1. <u>Linearization</u>

Linearization of a function of one variable means to approximate the value of the function at a point $z = f(x)$ by the equation the tangent line at a nearby point. Let $z_0 = f(x_0)$. If Δx is small, we have the approximation

$$\begin{aligned}
\frac{\Delta z}{\Delta x} &\doteq z_x,\\
\frac{z - z_0}{x - x_0} &\doteq f_x(x_0),\\
z &\doteq z_0 + (x - x_0) f_x(x_0), \quad \text{that is,}\\
f(x) &\doteq f_(x_0) + (x - x_0) f_x(x_0).
\end{aligned}$$

The reader will recognize the quantity on the right hand side as the approximation to the function by the linear term of the Taylor series. So, for example, if we wanted to approximate $\sqrt{101}$, we would take $z = f(x) = \sqrt{x}$, $x_0 = 100$, $x = 101$ and $\Delta x = 1$. Then,

$$\begin{aligned} f(x) = \sqrt{101} &\doteq \sqrt{100} + (101 - 100)\frac{1}{2\sqrt{100}}, \\ &\doteq 10 + \frac{1}{20}, \\ &\doteq 10.05 \end{aligned}$$

Of course, with one push of a button in a calculator which has the Taylor series coded in a chip, we can get by far a much better approximation. In this sense, this particular application to evaluate functions is obsolete. However, finding the equation of the tangent line, is fundamental.

2. Implicit Differentiation

Every expression for the derivative of a function, can be rewritten in terms of differentials. For example,

$$\begin{aligned} \frac{d}{dx}(x^2) &= 2x, \quad \text{then} \quad d(x^2) = 2x\,dx, \\ \frac{d}{dy}(y^2) &= 2y, \quad \text{then} \quad d(y^2) = 2y\,dy, \\ \frac{d}{du}(u^2) &= 2u, \quad \text{then} \quad d(u^2) = 2u\,du. \end{aligned}$$

The differential does not discriminate because of the name of the variable, they are all treated equally. With this simple observation, implicit differentiation of an equation $f(x, y) = c$ is reduces to two instructions

a) Take d

b) Solve for dy/dx

3.4.1 Example . Find dy/dx, given the circle $x^2 + y^2 = 1$.
Solution,

$$\begin{aligned} d(x^2 + y^2) &= 0, \\ 2x\,dx + 2y\,dy &= 0, \\ \frac{dy}{dx} &= -\frac{x}{y}, \quad y \neq 0. \end{aligned}$$

The solution is not valid when $y = 0$, because at those points, the tangent line is vertical. A circle is not a function of x. The derivative of y with respect to x could have been obtained without implicit differentiation by first solving for y to get two explicit functions $y = \sqrt{1 - x^2}$ and $y = -\sqrt{1 - x^2}$. Not only is this more work, but for more complicated expressions it might be difficult or impossible to solve for y explicitly.

3.4.2 Example Find dy/dx, given $y^3 + 4xy^2 - 5x = 1$

In this case it is still possible to solve for y using the cubic formula discovered by Scipione del Ferro and first published by Gerolamo Cardano in 1545. The solutions with Maple cover an entire page - clearly not the thing to do. Instead, we follow the two steps mentioned above, keeping in mind that for the middle term we need the product rule.

$$\begin{aligned} d(y^3 + 4xy^2 - 5x) &= 0, \\ 3y^2\,dy + 4y^2\,dx + 8xy\,dy - 5\,dx &= 0, \\ (3y^2 + 8xy)\,dy &= -(4y^2 - 5)\,dx, \\ \frac{dy}{dx} &= -\frac{(4y^2-5)}{(3y^2+8xy)}, \quad 3y^2 + 8xy \neq 0 \end{aligned}$$

3.4.3 Example Find dy/dx, given $y^2 + x\sin y + x^3 = 1$
Solution. This is a transcendental equation and it is not possible to solve for y in terms of elementary functions. However, as before, the differential always converts a nonlinear equation on the variables to a linear equation of the differential of the variables. We get

$$\begin{aligned} d(y^2 + x\sin y + x^3) &= 0, \\ 2y\,dy + \sin y\,dx + x\cos y\,dy + 3x^2\,dx &= 0, \\ (2y + x\cos y)\,dy &= -(\sin y + 3x^2)\,dx, \\ \frac{dy}{dx} &= -\frac{\sin y + 3x^2}{(2y + x\cos y)}; \quad 2y + x\cos y \neq 0. \end{aligned}$$

We reiterate that the derivative is valid only in the cases where the denominator is not zero.

3. Chain Rule
Suppose that $z = f(x)$ and $x = g(t)$, Then, composition of functions gives z as a function of t and we can take the derivative dz/dt. From the definition of the differential, we have

$$dz = z_x\,dx.$$

Recalling the $z_x = dz/dx$ We get immediately

$$\frac{dz}{dt} = \frac{dz}{dx}\frac{dx}{dt}.$$

This is the chain rule in the notation of Leibnitz. Huge disclaimer! this is not a proof of the chain rule. It completely ignores the subtlety that as $\Delta t \to 0$, we also have $\Delta x \to 0$. However, the standard rigorous proof of the chain rule that appears in all calculus textbooks shows that the formula is correct.

3.4.2 Multivariate Differential

We extend the definition of differentials to functions $z = f(x, y)$ of two variables in a natural way. Assume the partial derivatives of the function exist and are continuous. The function represents a surface in $\mathbf{R}^3$ as shown in figure 3.9. Let $P(x_0, y_0)$ be a point on the domain and z_0 be the value of the function at this point. We are interested in approximating the value of the function at a point (x, y) nearby. As shown in the picture, if we start at P and vary the function holding y_0 constant, we get a curve on the surface. Thus, the definition of differentials for one variable applies. Let $x = x_0 + \Delta x$, where Δx is a small number. The change Δz_1 on the surface resulting from this increment in the x coordinate is given approximately by

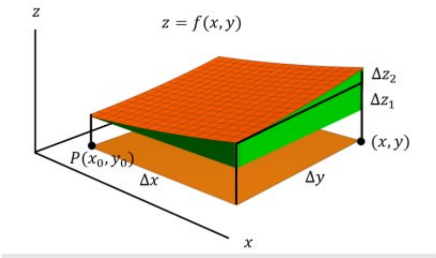


Fig. 3.9: Differentials

$$\Delta z_1 \doteq z_x \, \Delta x.$$

Similarly, if we vary y while holding x constant, we get a curve on the surface that depends only on y. Let $y = y_0 + \Delta y$, where Δy is also a small number. The change Δz_2 on the surface resulting from this increment in the y coordinate is given by

$$\Delta z_2 = z_y \, \Delta y.$$

The composite change Δz resulting from the increments in x and y is the sum

$$\Delta z = z_x \, \Delta x + z_y \, \Delta y. \tag{3.13}$$

The infinitesimal version of the formula is the differential

$$\begin{aligned} dz &= z_x \, dx + z_y \, dy, \quad \text{or,} \\ df &= f_x \, dx + f_y \, dy. \end{aligned} \tag{3.14}$$

Embedded in the differential are three very important concepts:

1. Linearization

Linearization of a function of two variables means to approximate the value of the function $z = f(x, y)$ by the equation of the tangent plane at a nearby point. Let $z_0 = f(x_0, y_0)$, $x = x_0 + \Delta x$ and $y = y_0 + \Delta y$, where the increments are small, Then, equation 3.13 reads

$$(z - z_0) \doteq f_x(x_0, y_0)\,(x - x_0) + f_y(x_0, y_0)\,(y - y_0).$$

which we can rewrite as

$$f(x, y) \doteq f(x_0, y_0) + f_x(x_0, y_0)\,(x - x_0) + f_y(x_0, y_0)\,(y - y_0). \tag{3.15}$$

There is a lot going on here. The equation

$$f_x(x_0, y_0)\,(x - x_0) + f_y(x_0, y_0)\,(y - y_0) - (z - z_0) = 0 \tag{3.16}$$

is the equation of the tangent plane to the surface at the point (x_0, y_0, z_0). Clearly, the normal $\mathbf{N}$ to the plane is given by the vector

$$\mathbf{N} = \langle f_x, f_y, -1\rangle \tag{3.17}$$

You may find it easier to remember the expression for the normal vector and use it when needed, to find the equation of the tangent plane. The approximation 3.15 is just the linear part of the Taylor expansion of the function $f(x, y)$ near the point (x_0, y_0). It is neat that one can use 3.15 to approximate numerically the value of the function near the given point (x_0, y_0), but again, this is fairly pointless nowadays, considering the proliferation of calculators and computers.

The definition of the differential extrapolates in a natural way to functions of more variables. For example, if we have a hypersurface $w = F(x, y, z)$, we write

$$\begin{aligned} dw &= w_x\,dx + w_y\,dy + w_z\,dz. \quad \text{or,}\\ dF &= F_x\,dx + F_y\,dy + F_z\,dz. \end{aligned}$$

3.4.3 Implicit Differentiation and the Gradient

We revisit implicit differentiation in two variables in a more formal manner that sheds light onto the geometry of the precess. Consider a function defined implicitly by the equation $f(x, y) = c$, where c is a constant. These represent the level curves of the function $z = f(x, y)$. Extracting the differential, we find,

$$\begin{aligned} df &= f_x\,dx + f_y\,dy = 0,\\ f_y\,dy &= -f_x\,dx,\\ \frac{dy}{dx} &= -\frac{f_x}{f_y}, \quad \text{where } \ f_y \neq 0. \end{aligned} \tag{3.18}$$

We have reduced implicit differentiation in one variable to the very simple formula 3.18. It is easy because it should be easy. We are talking about differentiation in one variable, but using the more advanced tool of the multivariate differential. The reader may wish to corroborate that the new simple formula is compatible with the examples on implicit differentiation presented in subsection 3.4.1. The expression for df is most interesting. We can rewrite it as a dot product

$$\begin{aligned} df &= f_x\,dx + f_y\,dy,\\ &= \langle f_x, f_y\rangle \cdot \langle dx, dy\rangle = 0, \end{aligned}$$

3.4.4 Definition The differential operator

$$\nabla = \frac{\partial}{\partial x}\,\mathbf{i} + \frac{\partial}{\partial y}\,\mathbf{j}, \tag{3.19}$$

is called the two-dimensional **gradient**. The quantity is called an operator, because it is ready to receive an input on which to operate. It the input is a function, the output is a vector. Rather, we should say the output is a **vector field** because the values of the vector at a point depend of the coordinates at that point. That is, the vector field assigns a vector to each point in the domain. The gradient vector field of $f(x, y)$ is

$$\nabla f = \langle f_x, f_y \rangle$$

We recall that $d\mathbf{r} = \langle dx, dy \rangle$ is just the differential of arc length vector 2.18, Thus, we can write the equation $df = 0$ in the form

$$\nabla f \cdot d\mathbf{r} = 0 \tag{3.20}$$

Since $d\mathbf{r}$ is tangential to the level curves, deduce that the gradient vector field is normal to the level curves!

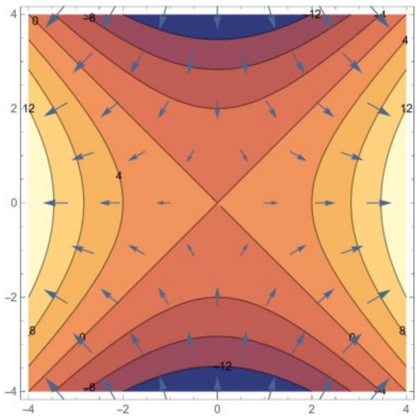


Fig. 3.10: Gradient

3.4.5 Example Let $f(x, y) = x^2 - y^2$. The surface is a saddle and the level curves $x^2 - y^2 = c$ are hyperbolas, except for the case $c = 0$ that consists of the two lines $y = \pm x$. The gradient of the function is the vector field

$$\nabla f = \langle 2x, -2y \rangle$$

Plotting a vector field is not task to be done by hand by human beings. In figure 3.10 we have used Mathematica to render on a grid, the gradient vector field superimposed on a set of level curves. The figure makes obvious the general fact that gradient vectors point in the direction of fastest increase of the surface. This is so, because the fastest way to move from a level curve to one with higher altitude is in a direction orthogonal to the level curve.

We now consider implicit differentiation in three variables. An implicit function of three variables is an equation of the form

$$F(x, y, z) = c$$

For each constant c we may consider $F(x, y, z) = c$ as a level surface of a hypersurface $w = F(x, y, x)$. For example, the implicit function

$$F(x, y, z) = x^2 + y^2 - z^2 = c$$

represents a family of nested circular hyperboloids, except for the case $c = 0$ that represents a circular cone. The surfaces are not regular functions of x and y, but in principle, one one could solve for z and obtain locally explicit functions. In the given case it is easy to solve for z in term of two square root functions, but in general, this may be hard or impossible. Still, it is straightforward to find the partial derivatives z_x and z_y by a three-step process,

1. Take the differential d.
2. Solve for dz. One gets $dz = z_x\,dx + z_y\,dy$
3. Read the partial derivatives from the "coefficients" of dx and dy.

3.4.6 Example Given the hyperboloid $x^2 - y^2 + z^2 + 6z = 8$, find the first partial derivatives z_x and z_y.
Solution,

$$\begin{aligned}
&d(x^2 - y^2 + z^2 + 6z) = 0,\\
&2x\,dx - 2y\,dy + 2z\,dz + 6\,dz = 0,\\
&(2z + 6)\,dz = -2x\,dx + 2y\,dy,\\
&dz = -\frac{2x}{2z+6}dx + \frac{2y}{2z+6}dy,\\
&z_x = -\frac{2x}{2z+6}, \quad z_y = \frac{2y}{2z+6}, \quad z \neq -3
\end{aligned}$$

3.4.7 Example Given $yz^3 + 4xe^y + 5z = 9$, find z_x and z_y.
Solution,

$$\begin{aligned}
&d(yz^3 + 4xe^y + 5z) = 0,\\
&z^3\,dy + 3yz^2\,dz + 4xe^y\,dy + 4e^y\,dx + 5\,dz = 0,\\
&(3yz^2 + 5)\,dz = -4e^y\,dx - (z^3 + 4xe^y)\,dy,\\
&dz = \frac{-4e^y\,dx - (z^3 + 4xe^y)\,dy,}{3yz^2 + 5},\\
&z_x = -\frac{4e^y}{3yz^2+5}, \quad z_y = -\frac{z^3 + 4xe^y}{3yz^2+5}, \quad 3yz^2 + 5 \neq 0.
\end{aligned}$$

These two problems are special cases of the general case of finding z_x and z^y for the equation

$$F(x, y, z) = c$$

. Taking the differential, we get,

$$\begin{aligned}
&dF = F_x\,dx + F_y\,dy + F_z\,dz = 0,\\
&F_z\,dz = -F_x\,dx - F_y\,dy,\\
&dz = -\frac{F_x}{F_z}\,dx - \frac{F_y}{F_z}\,dy,
\end{aligned}$$

Hence, once again we have reduced the implicit differentiation problem to the following compelling formula

$$z_x = -\frac{F_x}{F_z}, \quad z_y = \frac{F_y}{F_z}, \quad F_z \neq 0. \tag{3.21}$$

Also, as before, the intermediate step $dF = 0$ is very interesting. We can rewrite the equation, as

$$\nabla F \cdot d\mathbf{r} = 0,$$

where ∇ is now the three dimensional gradient operator

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}. \tag{3.22}$$

The interpretation is completely analogous. The infinitesimal arc length vector $d\mathbf{r}$ is tangent to any curve $\mathbf{r}(t)$ on the surface, so it is tangent to the surface. Hence, The vector

$$\mathbf{N} = \nabla F = \langle F_x, F_y, F_z \rangle \tag{3.23}$$

is normal to the level surfaces and hence normal to the tangent plane to the surface at each point. This yields an easy to remember procedure to find the equation of the tangent plane to a surface at a given point P, namely, compute the gradient, evaluate it at P, and use this vector as the normal to the plane.

3.4.8 Example Find the equation of the tangent a plane to the ellipsoid $F(x, y, z) = \frac{x^2}{1} + \frac{y^2}{4} + \frac{z^2}{9} = 3$, at the point $P(1, 2, 3)$.
Solution. To find the equation of plane, we need a point and a normal. The point is given. For the normal, we have

$$\begin{aligned} \nabla F &= \langle 2x, y/2, 2z/9 \rangle, \\ \nabla F|_P &= \langle 2, 1, 2/3 \rangle, \\ \mathbf{N} = &= \langle 6, 3, 2 \rangle, \end{aligned}$$

so the equation is

$$\begin{aligned} 6(x-1) + 3(y-2) + 2(z-3) = 0, \\ 6x + 3y + 2z = 18. \end{aligned}$$

We need to reconcile the normal vector to a surface defined implicitly with the normal vector to an explicit surface $z = f(x, y)$. This is easy. Let

$$F(x, y, z) = f(x, y) - z = 0$$

In other words, we can always view an explicit equation $z = f(x, y)$ as the particular level surface of the function $F(x, y, z)$ above. Then

$$\nabla F = \langle f_x, f_y, -1 \rangle$$

which is the result 3.17 we had before.

3.5 Chain Rule

3.5.1 First Order Chain Rule

We use the intuitive approach for the chain rule in one variable using the differential to extrapolate to the chain rule in several variables. Recall that if $z = f(x)$ and $x = g(t)$, then the differential $dz = z_x\,dx$, where $z_x = dz/dx$, leads to the derivative of the composite function

$$\frac{dz}{dt} = \frac{dz}{dx}\frac{dx}{dt}, \quad \text{or in subscript notation,}$$
$$z_t = z_x x_t.$$

Now, suppose that

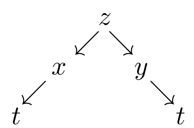
$$z = f(x, y), \quad x = x(t),\ y = y(t).$$

Then $z = f(x(t), y(t))$ is ultimately a function of t, that is, a curve. That is, the composition of the maps is a real-valued function as shown in the diagram here,

$$t \to (x, y) \xrightarrow{f} z,$$
$$\mathbf{R} \to \mathbf{R}^2 \xrightarrow{f} \mathbf{R}.$$

Therefore, it makes sense to take the derivative dz/dt. The derivative is a regular derivative dz/dt and not a partial derivative, since we are just finding the slope of a curve.

As with one variable, we convert the differential into a derivative formula,



$$dz = z_x\,dx + z_y\,dy,$$
$$z_t = z_x x_t + z_y y_t.$$

This is the chain rule in this case. A good mnemonic to remember this formula is to arrange the variable dependence in the form of a tree as shown in the figure. In the tree, z depends on x and y, which themselves depend on t. Now think of a "Hungry Caterpillar" that likes to munch on t-leaves. The caterpillar crawls down one branch to x and then from x to t. Each node of the tree represents a derivative. The caterpillar then goes back to the tree trunk and crawls back down on the y branch looking of the other t-leaf. Writing the equation in the notation of Leibnitz

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}, \tag{3.24}$$

gives a better sense of why this is called the chain rule. We humorously add that the allegorical caterpillar is a chain eater in the sense that it has to chomp on every t-leaf.

We provide an example that although extremely simple, it has all features that need to be illustrated with this case of the bivariate chain rule. The usual suspect is a quadric surface.

3.5.1 Example Let $z = f(x, y) = x^2 - y^2$, $x = \cos t$, $y = \sin t$. Use the chain rule to find dz/dt and then evaluate this at $t = \pi/3$.
Solution

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt},\\ &= 2x(-\sin t) - 2y\cos t,\\ &= -2(x\sin t + y\cos t).\end{aligned}$$

That is all there is to computing dz/dx. Now we evaluate at the given point,

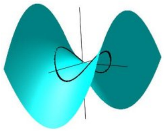
$$\begin{aligned}\left.\frac{dz}{dt}\right|_{t=\pi/3} &= -2(x\sin t + y\cos t)|_{t=\pi/3},\\ &= -2(\cos(\pi/3)\sin(\pi/3) + \sin(\pi/3)\cos(\pi/3),\\ &= -\sin(2\pi/3)), \quad \text{by double angle formula},\\ &= -\sqrt{3}.\end{aligned}$$

The goal here is to understand the new chain rule, and if asked to use it, the problem must be solved as above. In the next lines we "cheat" just for the purpose of verifying the process works. We solve the problem without any use of the multivariate chain rule.

$$\begin{aligned}z &= x^2 - y^2,\\ &= \cos^2 t - \sin^2 t,\\ &= \cos 2t,\\ \frac{dz}{dt} &= -2\sin 2t.\end{aligned}$$

As expected, the function $z = \cos 2t$ is ultimately an elementary single-variable function. We should point out that given $z = f(x, y)$ and $x = x(t), y = y(t)$, we could define a space curve

$$\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + [\,f(x(t), y(t))\,]\mathbf{k}.$$



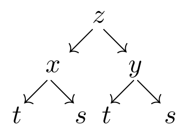
The position vector of a point on the space curve is constrained to satisfy the equation $z = f(x, y)$, therefore, the space curve must lie on the surface. In the example above, the space curve $\mathbf{r}(t) = \langle \cos t, \sin t, \cos 2t\rangle$ indeed lies on the saddle. Since $x^2 + y^2 = 1$, the curve is geometrically described by the intersection of the saddle with a cylinder of radius 1. The projection of the curve onto the xy-plane is a circle, whereas the z-coordinate oscillates with period $T = \pi$ according to the function $z = \cos 2t$.
Now we consider another case in which we have

$$z = f(x, y), \quad x = x(t, s), \quad y = y(t, s),$$

where all the given functions are differentiable. The composite function diagram in this case now needs to be modified a follows,

$$(t,s) \xrightarrow{\mathbf{T}} (x,y) \xrightarrow{f} z,$$
$$\mathbf{R}^2 \xrightarrow{\mathbf{T}} \mathbf{R}^2 \xrightarrow{f} \mathbf{R},$$

where we have denoted by $\mathbf{T}$ the vector function that takes vector $\langle x, y\rangle$ to vector $\langle t, s\rangle$. Such types of vector functions from $\mathbf{R}^n$ to $\mathbf{R}^n$ are called coordinate transformations or change of variables. We have already given a couple of examples of the this type when we changed to polar coordinates to study limits of some functions of two variables



The tree diagram is depicted in the adjacent figure. Since z ultimately depends on t and s, it makes sense to take the partial derivatives z_t and z_s. Recalling that when one computes the partial derivative with respect to one variable , the other variables are treated as if they were constants, we see that the formula for z_t looks just as before, the only difference being that now z_t represents a partial derivative instead of a regular derivative. The same can be said about z_s. Hence, the equations for the two partial derivatives are given by,

$$z_t = z_x x_t + z_y y_t,$$
$$z_s = z_x x_s + z_y y_s,$$

The formulas above are short and elegant, but to emphasize that all the subscripts denote partial derivatives, we rewrite the equations as,

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t},$$
$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}, \tag{3.25}$$

3.5.2 Example Let $z = x^3 - 3xy^2$, $x = r\cos\theta$, $y = r\sin\theta$. Use the chain rule to find z_r and x_θ.
Solution

$$z_r = z_x x_r + z_y y_r,$$
$$z_\theta = z_x x_\theta + z_y y_\theta,$$
$$z_r = (3x^2 - 3y^2)\ \cos\theta - 6xy\,\sin\theta,$$
$$z_\theta = -(3x^2 - 3y^2) r\sin\theta - 6xy\, r\cos\theta,$$

For the purposes of fulfilling the requirements of the curriculum, we are done with this problem. However, you may find the following enrichment perspective interesting. The function here is a monkey saddle which is a harmonic function, so there are complex numbers lurking in the neighborhood. Let's expand z_r by a full substitution. We get,

$$\begin{aligned}
z_r = & \quad (3x^2 - 3y^2)\cos\theta - 6xy\sin\theta, \\
, &= (3r^2\cos^2\theta - 3r^2\sin^2\theta)\cos\theta - (6r\cos\theta\, r\sin\theta)\sin\theta, \\
&= 3r^2[\cos^2\theta - \sin^2\theta]\cos\theta - 3r^2[2\sin\theta\cos\theta]\sin\theta, \\
&= 3r^2[\cos 2\theta\cos\theta - \sin 2\theta\sin\theta], \\
&= 3r^3\cos(2\theta+\theta), \quad \text{using the sum and double angle formulas} \\
&= 3r^2\cos 3\theta
\end{aligned}$$

This is the way it should be because the real part of $w = z^3 = r^3 e^{3i\theta}$ is

$$u = r^3\cos 3\theta \;\Rightarrow\; u_r = 3r^2\cos 3\theta.$$

A similar computation can be done for u_θ.

3.5.3 Example Wave Equation. Let $u = f(x - vt) + g(x + vt)$, where f, g are arbitrary twice differentiable functions and v is a constant. Show that u satisfies the wave equation $u_x x = (1/v^2)u_t t$
Solution. Let $\xi = (x - vt)$ and $\eta = (x + vt)$ so that

$$u = u(\xi, \eta) = f(\xi) + g(\eta).$$

We have,

$$\begin{aligned}
u_x &= u_\xi \xi_x + u_\eta \eta_x = f_\xi + f_\eta, \\
u_t &= u_\xi \xi_t + u_\eta \eta_t = vf_\xi + (-v)f_\eta.
\end{aligned}$$

If u is a function of ξ and η, the so are u_x and u_t, so the same chain rule applies. The mixed partial derivatives of f and g are zero, since these are functions of one variable. Thus

$$\begin{aligned}
u_{xx} &= (u_x)_\xi \xi_x + (u_x)_\eta \eta_x = f_{\xi\xi} + f_{\eta\eta}, \\
u_{tt} &= (u_t)_\xi \xi_t + (u_t)_\eta \eta_t = v^2 f_{\xi\xi} + (-v)^2 f_{\eta\eta}.
\end{aligned}$$

So clearly $v^2 u_{xx} = u_{tt}$

3.5.2 Second Order Chain Rule

The general second order chain rule is more challenging and is not covered satisfactorily in most fat calculus textbooks. But if we don't do it, then it falls through the cracks only to come back and bite you in the back at a later time. So let's do it now and let's do it right. We begin by extending the differential to a full multivariate calculus function from $\mathbf{R}^n$ to $\mathbf{R}$. So that we do not run out of letters we use an index notation for the variables. We write

$$u = f(x^1, x^2, \ldots, x^n) = f(x^k), \quad , k = 1, 2, \ldots n.$$

Then,

$$df = \frac{\partial f}{\partial x^1}dx^1 + \frac{\partial f}{\partial x^2}dx^2 + \cdots + \frac{\partial f}{\partial x^n}dx^n,$$
$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}dx^i. \tag{3.26}$$

Now suppose the we also have $x^i = x^i(t^1, t^2, \ldots, t^m)$. What we have is a composition of maps

$$(t^1, t^2, \ldots, t^m) \xrightarrow{T} (x^1, x^2, \ldots, x^n) \xrightarrow{f} u,$$
$$\mathbf{R}^m \xrightarrow{T} \mathbf{R}^n \xrightarrow{f} \mathbf{R}$$

Thus, it makes sense to take the partial derivatives with respect to each of the t variables. The resulting chain rule is

$$\frac{\partial f}{\partial t^k} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\frac{\partial x^i}{\partial t^k}. \tag{3.27}$$

This more general formula is very slick. It contains the pair of equations 3.25 as a special case in which $(t^1, t^2) = (t, s)$, $(x^1, x^2) = (x, y)$ and the summation is from $i = 1$ to 2. Now comes the "hard" part. We take the second derivatives of equation 3.27, keeping in mind that $\partial f/\partial x^i$ is also ultimately a function of the t's, so for these functions, we have to apply the chain rule recursively. Combining with the product rule, we get the second order chain rule formula

$$\frac{\partial^2 f}{\partial t^l \partial t^k} = \sum_{j=1}^{n}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial x^j \partial x^i}\frac{\partial x^j}{\partial t^l}\frac{\partial x^i}{\partial t^k} + \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}\frac{\partial^2 x^i}{\partial t^l \partial t^k}. \tag{3.28}$$

Equation 3.28 is rather formidable for a first pass in multivariate calculus, so we unpack the equation for f_{tt} and f_{ss} for the two variable case $(t^1, t^2) = (t, s)$, and $(x^1, x^2) = (x, y)$. We could also expand the formula for the mixed partials f_{ts}, but we are after the Laplacian, so we will not need them for now. The explicit formulas are,

$$f_{tt} = f_{xx}(x_t)^2 + 2f_{xt}x_t y_t + f_{yy}(y_t)^2 + f_x x_{tt} + f_y y_{tt},$$
$$f_{ss} = f_{xx}(x_s)^2 + 2f_{xt}x_s y_s + f_{yy}(y_s)^2 + f_x x_{ss} + f_y y_{ss}. \tag{3.29}$$

One of the most difficult problems in this book is to convert the 2-dimensional Laplacian to polar coordinates. The Laplacian in polar coordinates appears in classical models that have cylindrical symmetry such as the vibrations of a membrane, diffraction by a circular aperture, electric potential of charge distribution on a long cylinder, heat conduction on a circular plate, and many more. Most authors leave the conversion of the Laplacian as an advanced exercise for the students, but doing so alienates all but the brightest. All STEM students should be exposed to this problem, so we believe that it is best to work out the problem for all.

3.5.4 Example Show that if $u = f(x, y)$ is a twice differentiable functions and $x = r\cos\theta$, $y = r\sin\theta$, then

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{1}{r}\frac{\partial f}{\partial r} \tag{3.30}$$

Solution
It is better to work with the chain rule on the right hand side. First, we need

$$\begin{aligned} x_r &= \cos\theta, \\ y_r &= \sin\theta, \\ x_{rr} &= 0, \\ y_{rr} &= 0. \end{aligned} \quad \text{and} \quad \begin{aligned} x_\theta &= -r\sin\theta, \\ y_\theta &= \ \ r\cos\theta, \\ x_{\theta\theta} &= -r\cos\theta, \\ y_{\theta\theta} &= -r\sin\theta. \end{aligned}$$

Then, the first partial derivatives are

$$\begin{aligned} f_\theta &= -f_x\, r\sin\theta + f_y\, r\cos\theta, \\ f_r &= \ \ f_x\ \cos\theta + \ \ f_y \sin\theta, \\ \frac{1}{r}f_r &= \frac{1}{r}f_x\cos\theta + \frac{1}{r}f_y\sin\theta. \end{aligned}$$

Computing the second partial derivatives we are lead to the set of equations

$$\begin{aligned} f_{\theta\theta} &= r^2 f_{xx}\sin^2\theta - 2r^2 f_{xy}\cos\theta\sin\theta + r^2 f_{yy}\cos^2\theta - r f_x\cos\theta - r f_y\sin\theta, \\ \frac{1}{r^2}f_{\theta\theta} &= f_{xx}\sin^2\theta - 2f_{xy}\cos\theta\sin\theta + f_{yy}\cos^2\theta - \frac{1}{r}f_x\cos\theta - \frac{1}{r}f_y\sin\theta, \\ f_{rr} &= f_{xx}\cos^2\theta + 2f_{xy}\cos\theta\sin\theta + f_{yy}\sin^2\theta, \\ f_{rr} + \frac{1}{r^2}f_{\theta\theta} &= f_{xx} + f_{yy} - \frac{1}{r}f_x\cos\theta - \frac{1}{r}f_y\sin\theta, \\ &= f_{xx} + f_{yy} - \frac{1}{r}f_r. \end{aligned}$$

Rearranging the last equation, we get the desired result

$$f_{rr} + \frac{1}{r^2}f_{\theta\theta} + \frac{1}{r}f_r = f_{xx} + f_{yy}.$$

3.6 Directional Derivative

3.6.1 Definition and Computation

Given a a surface $z = f(x, y)$, and a point $P(x_0, y_0)$, we have learned that the partial derivative f_x and f_y evaluated at P represent the slopes of the surfaces at P in the x and y directions respectively. Now we are interested in finding the rate of change of the surface in any direction. Let $\mathbf{x} = \langle x, y\rangle$ represent the position vector of an arbitrary point in $\mathbf{R}^2$. It is convenient to use the notation $f(\mathbf{x}) = f(x, y)$. That is, when we write the value of the function f at a vector, we mean the value of the function at the position coordinates of that vector.

Thus, for example, $f(\mathbf{P}) = f(x_0, y_0)$. Let $\mathbf{u} = \langle a, b\rangle$ be a unit vector. In $\mathbf{R}^2$, the equation

$$\mathbf{r}(t) = \mathbf{P} + t\mathbf{u}$$

represents a straight line that contains the point P and and has direction vector $\mathbf{u}$. But in $\mathbf{R}^3$, the graph of the equation is a plane obtained by extruding the line in $\mathbf{R}^2$ into a vertical plane. The plane intersects the surface $z = f(x, y)$ restricting the surface to a curve on the surface, as shown in figure 3.11. We define the directional derivative $D_{\mathbf{u}}f(P)$ of the function f in the direction $\mathbf{u}$ at the point as

$$D_{\mathbf{u}}f(P) = \lim_{t\to 0} \frac{f(\mathbf{P} + t\mathbf{u}) - f(\mathbf{P})}{t} \tag{3.31}$$

The definition here generalizes to functions of more variables by just taking the point P and the vector $\mathbf{u}$ in the appropriate dimension. We have chosen to illustrate the definition in two variables because in this case it is easy to visualize the geometry. The definition of directional derivative parallels the definition of derivatives in one-variable calculus. We evaluate the function $f(x, y)$ at a point near P along the line $\mathbf{r}(t)$. We compute the average slope, then we take the limit as the nearby point approaches P. The result is a number that measures of the rate of change of the surface at the given point in the given direction. The real valued function whose rare of change we are computing can be written

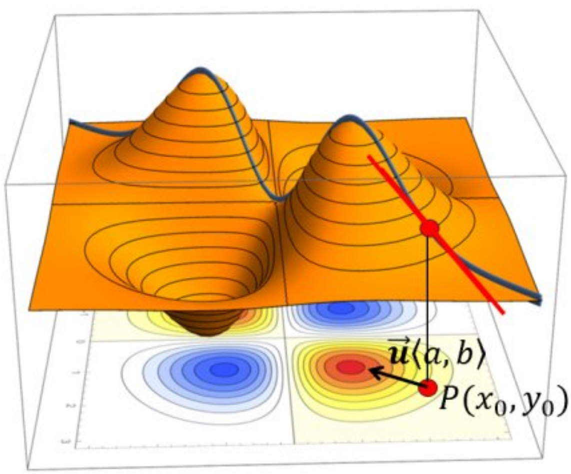


Fig. 3.11: Directional Derivative

as the composition

$$g(t) = f(\mathbf{P} + t\mathbf{u}) = f(x_0 + at, y_0 + bt)$$

of the function $f(x, y)$ with

$$x = x_0 + at,$$
$$y = y_0 + bt.$$

The derivative $g'(0)$ can be computed as a very simple example of the chain rule

$$\begin{aligned} g'(t) &= \frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + + \frac{\partial f}{\partial y}\frac{dy}{dt}, \\ &= a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial x}, \\ &= \nabla f \cdot \mathbf{u}. \end{aligned}$$

So we have just proved that

$$D_{\mathbf{u}} f(P) = \nabla f|_P \cdot \mathbf{u}. \tag{3.32}$$

This very neat formula reduces computation of directional derivatives to a very simple kitchen recipe.

1. Compute the gradient and evaluate at the point
2. Make sure the direction is given by a unit vector, if not, make it so.
3. Compute the dot product of the two vectors above.

The directional derivative formula 3.32 also is also very intuitive from a geometric point of view.

- If the unit vector is $\mathbf{i} = \langle 1, 0\rangle$, then $a = 1$ and $b = 0$. Thus, in this case, the directional derivative is just the partial derivative f_x, or the rate of change in the x-direction.
- If the unit vector is $\mathbf{j} = \langle 0, 1\rangle$, then $a = 0$ and $b = 1$. Thus, in this case, the directional derivative is just the partial derivative f_y, or the rate of change in the y-direction.
- If the unit vector is $\mathbf{u} = \langle a, b\rangle$, the directional derivative is a linear combination of the partial derivatives f_x and f_y

In two variables, there is at least four different ways in which the unit vector is specified.

1. The unit vector is given. Check to make sure $\|\mathbf{u}\|$=1.
2. A direction vector $\mathbf{v}$ is given. Divide it by its length.
3. A direction angle θ is given. Take $\mathbf{u} = \langle\cos\theta, \sin\theta\rangle$.
4. A direction to a point Q is given. Take $\mathbf{u} = \overrightarrow{PQ}/\|\overrightarrow{PQ}\|$.

3.6.1 Example Find the directional derivative of $f(x, y) = y^3 + 3x^2 - 2x$ at the point $(1, 1)$, where $\mathbf{u}$ is the unit vector in the direction vector given by angle $\theta = \pi/3$.
Solution

$$\begin{aligned} \nabla f &= \langle 6x - 2, 3y^2\rangle \\ \nabla f|_P &= \langle 4, 3\rangle, \\ \mathbf{u} &= \langle\cos\tfrac{\pi}{3}, \sin\tfrac{\pi}{3}\rangle = \langle\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}\rangle, \\ D_{\mathbf{u}} f(P) &= 2 + \tfrac{3\sqrt{3}}{2}. \end{aligned}$$

3.6.2 Maximizing Directional Derivative

The main tool for computing the directional derivative of a function f at a point P in the direction of a unit vector is

$$D_{\mathbf{u}}f(P) = \nabla f|_P \cdot \mathbf{u}$$

This is just a dot product. So, if θ is the angle subtended by the two vectors, we can rewrite the expression as

$$\begin{aligned} D_{\mathbf{u}}f(P) &= \nabla f|_P \cdot \mathbf{u}, \\ &= \|\nabla f\|_P \|\mathbf{u}\| \cos\theta, \\ &= \|\nabla f\|_P \cos\theta, \quad \text{since } \|\mathbf{u}\| = 1. \end{aligned}$$

The maximum value the last expression can attain is

$$max[D_{\mathbf{u}}f(P)] = \|\nabla f\|_P$$

and this occurs when $\theta = 0$, that is, when the direction is the direction of the gradient. We already know that gradient vector is orthogonal to the level curves/surfaces, so this is the direction of maximum rate of change. In figure 3.11 we have intentionally taken $\mathbf{u}$ in the direction of the gradient. If the level curves were a topographical of the "mountain" represented by the surface, a climber seeking for the steepest path on the mountain would pick that direction. Thus, in the example above with $f(x,y) = y^3 + 3x^2 - 2x$ at the point $(1,1)$, the gradient at the point is $\nabla f_P = \langle 4, 3\rangle$, so the maximum rate of change at that point is 5 and this is in the direction $\langle 4, 3\rangle$.
In the book "The Spy who Came in from the Cold ", the great English novelist John Le Carré once wrote,

> "It is said that men condemned to death are subject to sudden moments of elation; as if, like moths in the fire, their destruction were coincidental with attainment."

If one were to think of a camp fire as a localized source of heat described by a temperature function $T(x,y,z)$, then the level surfaces $T(x,y,z) = c$ would represent surfaces of equal temperature. It is as if the moths in Le Carré's quote were naturally programmed to seek the shortest path to their morbid attainment.

3.7 Maxima - Minima

One of the most direct applications of differential calculus is that it provides a straightforward procedure to locate local (relative) extrema of functions. For a function of one the variable, the procedure is called the second derivative test. The second derivative test is introduced early in the first semester of calculus, but to really understand why it works, requires Taylor series which typically appear at the very end of the second semester. Thus, it is common for students to come into the vector calculus course without a solid foundation of testing for critical points of smooth functions. For that reason, we begin with a short review of the topic in the context of functions of one variable

3.7.1 Max-Min in One Variable

Let $z = f(x)$ be a smooth, real-valued function of one variable. Anticipating working with two variables, we abuse notation again writing $f_x = f'(x)$. At a point $x = x_0$, the derivatives are either positive, negative or zero. The first derivative $f_x(x_0)$ measures the slope of the curve at that point, and the second derivative $f_{xx}(x_0)$ measures the concavity. We have

$$f_x(x_0)\begin{cases} > 0 & \text{Increasing} \\ < 0 & \text{Decreasing} \\ = 0 & \text{Horizontal} \end{cases} \qquad f_x(x_0)\begin{cases} > 0 & \text{Concave Up} \\ < 0 & \text{Concave Down} \\ = 0 & \text{No information} \end{cases}$$

Combining these two together, we get the second derivative test. To locate the extrema of the function $z = f(x)$ we follow the two-step process,

1) Set $f_x = 0$ and solve. If solutions x_c exist, they are critical points.

2) Test the second derivative at each critical point.

We get the following classification,

$$f_{xx}(x_c)\begin{cases} > 0 & \text{Local minimum at } \; x = x_c, \\ < 0 & \text{Local maximum at } \; x = x_c, \\ = 0 & \text{No information.} \end{cases}$$

When the second derivative at a critical point is equal to zero, most textbooks say that the test fails and leave it a that. The easiest example to motivate this are the functions $f(x) = x^3$, $f(x) = x^4$ and $f(x) = -x^4$. At $x = 0$, all three functions have $f_{xx}(0) = 0$. However, at $x = 0$, $f(x) = x^3$ one has an inflection point, $f(x) = x^4$ has a local minimum and $f(x) = -x^4$ has a local maximum. The second derivative test is inconclusive for these three power functions. Is is not the case that calculus fails, but rather, the information is incomplete without considering the Taylor series,

$$f(x) = f(x_0) + f_x(x_0)(x - x_0) + \frac{1}{2!} f_{xx}(x_0)(x - x_0)^2 + \ldots. \tag{3.33}$$

At a critical point x_c, the first derivative is 0. The idea of the second derivative test is to approximate the function by a parabola. However, if $f_{xx}(x_c) = 0$, the Taylor series says that there is no good parabolic approximation. In such case, one should check if $f_{xxx}(x_c) \neq 0$. If so, the graph is locally approximated by a cubic curve and we have an inflection point. If the third derivative at the critical point is also zero, one needs to continue the Taylor polynomial approximation until one encounters the first integer $n > 3$ so that $f^{(n)}(x_0) \neq 0$. Then, whether one has a local maximum-minimum or infection point is dictated by whether this n is even or odd.

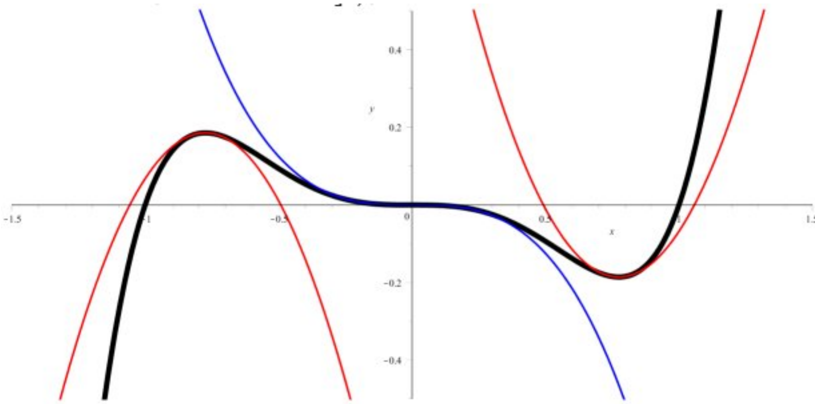


Fig. 3.12: Taylor series for $f(x) = x^3(x^2-1)$

3.7.1 Example Let $f(x) = x^3(x^2-1)$. Then

$$\begin{aligned} f(x) &= x^5 - x^3, \\ f'() &= 5x^4 - 3x^2 = x^2(5x^2-3) = 0, \\ f''(x) &= 20x^3 - 6x = 2x(10x^2-3), \\ x &= -\sqrt{3/5},\ 0,\ \sqrt{3/5}, \quad \text{critical points}, \\ f''(-\sqrt{3/5}) &< 0, \quad \text{Local maximum at } x = -\sqrt{3/5}, \\ f''(-\sqrt{3/5}) &> 0, \quad \text{Local minimum at } x = \sqrt{3/5}, \\ f''(0) &= 0, \quad \text{test fails.} \end{aligned}$$

As shown in figure 3.12 at the points where there is a local maximum or minimum, the second order Taylor series approximates the function by parabolas. But at the critical point $x = 0$, there is no good quadratic approximation, however, the Taylor series of order 3 shows a good cubic approximation. This is the real reason there is an inflection point at $x = 0$.

3.7.2 Max-Min in Two Variables

The second derivative test for maxima and minima of a function of two variables is also based on Taylor series. Let $z = f(x,y)$ be a smooth function in a neighborhood of a point (x_0, y_0), let (x,y) be a point near (x_0,y_0). Using the notation $f(\mathbf{x}) = f(x,y)$, and $f(\mathbf{x_0}) = f(x_0,y_0)$, the Taylor series up to second order about the point (x_0,y_0) is

$$\begin{aligned} f(\mathbf{x}) =& f(\mathbf{x_0}) + [f_x(\mathbf{x_0})(x-x_0) + f_y(\mathbf{x_0})(y-y_0)] + \\ & \frac{1}{2!}[f_{xx}(\mathbf{x_0})(x-x_0)^2 + 2f_{xy}(\mathbf{x_0})(x-x_0)(y-y_0) + f_{yy}(\mathbf{x_0})(y-y_0)^2] + \dots \end{aligned} \tag{3.34}$$

The proof of the Taylor series is an adaptation of the mode of thinking analogous to the proof in one variable. We present the proof in section 3.9. Roughly speaking, if a power series expansion in two variables exists, one proves that the constant coefficients are given as above, by successively computing, term by term, the partial derivatives of the series and then evaluating them at the point (x_0, y_0). We are interested in critical points (x_0, y_0) where the function might have a local maximum or a local minimum. The shape of the graphic is absolutely not affected by translations, so for our purposes is more convenient to consider the Taylor expansion about the origin. At a critical point in one variable, the tangent line is horizontal. In the same manner, at a critical point of $f(x, y)$ the linear approximation is the tangent plane which is also horizontal. Hence, the first partial derivatives are equal to zero. We are left with the problem of analyzing the quadratic approximation,

$$\begin{aligned} f(x,y) &= \frac{1}{2!}[f_{xx}(0,0)x^2 + 2f_{xy}(0,0)xy + f_{yy}(0,0)y^2], \\ &= \frac{1}{2!}[Ax^2 + 2Bxy + Cy^2], \end{aligned}$$

where A, B and C are the constant coefficients. We begin by considering the level curves

$$Ax^2 + 2Bxy + Cy^2 = k \tag{3.35}$$

The equation is quadratic, so the level curves are conics. To classify the conics we begin by writing the equation of the level curve conics by the tantalizing matrix multiplication equation,

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = k, \quad k = \text{constant}$$

We observe that the matrix of coefficients associated with the quadratic form is symmetric. It is almost impossible to overestimate the significance of this fact. To understand the behavior of the quadrics. Define the discriminant $\mathcal{D}$ of the quadratic form by the determinant

$$\mathcal{D} = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = AC - B^2.$$

It turns out that the conics can be classified by whether the discriminant is positive, negative or zero. Consider the simpler situation in which $B = 0$. Then the matrix is diagonal and

$$\mathcal{D} = AC$$

In the case of a diagonal matrix, the entries along the diagonal are called the **eigenvalues**. Here the eigenvalues are A and C and the discriminant is the product of the eigenvalues. There are three possibilities,

$$\mathcal{D} = AC \begin{cases} > 0 & \text{Ellipse} \\ < 0 & \text{Hyperbola} \\ = 0 & \end{cases} \tag{3.36}$$

The explanation is simple.

1. $AC > 0$, then A and C are either both positive or both negative. The quadric $z = Ax^2 + Cy^2$ is a paraboloid, and the level curves are ellipses (or circles if $A = C$) . If A and C are both positive, the paraboloid has upward concavity as in the case $z = x^2+y^2$. If A and C are both negative, the paraboloid has downward concavity as in the case $z = -x^2 - y^2$.

2. $AC < 0$, then A and C have different signs. The quadric is a saddle and the level curves are hyperbolas as in the case $z = x^2 - y^2$.

3. $AC = 0$, then either $A = 0$ or $C = 0$ or both. If one of the two is not zero, the quadric is a parabolic cylinder generated by horizontal lines as in the case $z = x^2$. The level curves are horizontal lines. If both $A = 0$ and $C = 0$, then there is no quadric. We would need to look at the third order Taylor series for the next possible approximation to the surface.

If $B \neq 0$, the situation in general is considerably more complicated because the graph of the level curves are rotated conics for which the axes of symmetry do not align with the coordinate axes. The topic of rotation of conics used to be part of the calculus I curriculum, but it has long been abandoned at this level. There is danger that the important topic might fall through the cracks, unless students have the good fortune of getting the appropriate exposure to rotations of conics in linear algebra. The big **spectral theorem** in linear algebra applied to symmetric 2×2 matrices, says that there are still two eigenvalues and their product is still the given by the discriminant.

$$\mathcal{D} = AC - B^2 \begin{cases} > 0 & \text{Ellipse} \\ < 0 & \text{Hyperbola} \\ = 0 & \end{cases} \tag{3.37}$$

The theory of eigenvalues and eigenvector implies there is a rotation that "diagonalizes" the matrix so that in the rotated axes, there is no cross term xy. After the rotation, the conic axes of symmetry align with the rotated coordinate axes. The discriminant is not affected by a rotation of axes, and neither is the classification scheme of the level curves of equation 3.35. Whereas the proof of this assertion is beyond the scope of this course, a compelling argument can be made by investigating the behavior of a particular family of conics such as

$$2x^2 + 2Bxy + 8y^2 = 1$$

In this case, the discriminant is $\mathcal{D} = 16 - B^2$. We then produce a frame by frame animation of the conics as the value of changes from -9 to 9. We keep an eye on the special frame in which $B = \pm 4$ for which $\mathcal{D} = 0$. Figure 3.13 shows the frames as B changes by a step size equal to -1. It is clear from the video that the shape of the conics is consistent with the classification given by equation 3.37. When the discriminant is negative which happens when $|B| > 4$ the graphs are hyperbolas. When the discriminant is positive with $|B| < 4$, the graphs are ellipses. In both cases the axes of symmetry of the conics are aligned with the rotated coordinate axes. Finally, when $|B| = 4$, the quadrics degenerate into a pair of parallel lines.

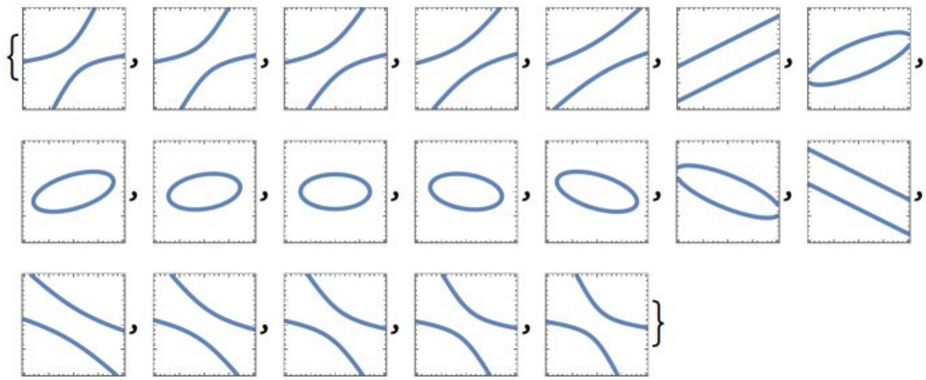


Fig. 3.13: Family of conics $2x^2 + 2Bxy + 8y^2 = 1$, $B = -9 - 8, \dots 9$.

To show the correlation of the conics with of the corresponding quadrics, we render the level curves of the family of surfaces

$$f(x, y) = 2x^2 + 2Bxy + 8y^2.$$

as shown in figure 3.14

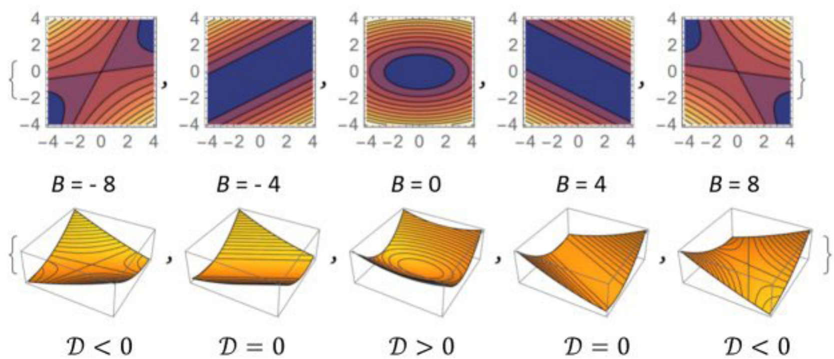


Fig. 3.14: $f(x, y) = Ax^2 + 2Bxy + Cy^2$, $A = 2, C = 8$, Discriminant $D = 16 - B^2$.

1. When the level curves are ellipses, the quadric is a paraboloid and we say that the critical point $(0, 0)$ is **elliptic**. Whether the paraboloid has local minimum or maximum depends on whether $A > 0$ or $A < 0$.

2. When the level curves are hyperbolas, the quadric is a saddle and we say that the critical point $(0, 0)$ is **hyperbolic**.

3. When the discriminant is 0, we say the critical point $(0, 0)$ is **parabolic**. In the example above, the level curves are pairs of parallel lines and the

graph is a parabolic cylinder. It could have happened that the discriminant is zero because all the coefficients are 0, in which case there is no quadric.

Now we have enough intuition to motivate the process for classifying the nature of the critical points of a function $f(x, y)$ of two variables. By classification of a critical point we mean, determining whether at that point the function has relative extremum, a saddle point, or something else. If a relative extremum, we want to know if it is a relative maximum or a relative minimum.
Define the **Hessian** of the function at a point (x, y) by the determinant

$$\mathcal{H}(x,y) = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} = f_{xy}^2. \tag{3.38}$$

Here, we are assuming that the function is smooth within its domain, so that the mixed partial derivatives are equal to each other.

<u>The second derivative test is as follows</u>

1. Set

$$\begin{aligned} f_x &= 0, \\ f_y &= 0. \end{aligned}$$

   We get a system of two equations and two unknowns. Solve the system and find the critical points (x_c, y_c), if they exist. There might be more than one critical point. If so, we test each with the Hessian.

2. Evaluate the Hessian at each critical point, this gives a number. This number can be positive, negative or zero. We classify each critical point by the scheme

$$\mathcal{H}(x_c, y_c) \begin{cases} > 0 & f_{xx}(x_c, y_c) \begin{cases} > 0 & \text{Local minimum at}(x_c, y_c), \\ < 0 & \text{Local maximum at}(x_c, y_c), \end{cases} \\ < 0 & \text{saddle at}(x_c, y_c), \\ = 0 & \text{No information.} \end{cases} \tag{3.39}$$

When the Hessian is zero at a critical point, the test is inconclusive. It means that either, the quadric approximation is cylindrical, or there is no good quadric approximation to the function at that point, and one would need to include the next order term in the Taylor series. The reader might wonder why we do not have to test concavity in the y-direction and why is it not enough to check f_{xx} and f_{yy}. First, the "Foal-on-Horse" surface in figure 3.7 shows a function that is concave up along the x and y-axes, but the function does not have an extremum at the origin. One needs to involve the mixed partial to make sure the other directional derivatives are included. Secondly. if $\mathcal{H} > 0$, then f_{xx} and f_{yy} must have the same sign, else, their produce would be negative and so would be the Hessian.

3.7.2 Example Let $f(x, y) = 2xy$. Classify the critical points.

Solution. Set

$$f_x = 2y = 0,$$
$$f_y = 2x = 0.$$

We only have one critical point at $(0,0)$. We test the critical point.

$$f_{xx} = 0, \quad f_{xy} = 2,$$
$$f_{yx} = 2, \quad f_{yy} = 0.$$

Hence, $\mathcal{H}(0,0) = -4 < 0$ and we have a saddle at $(0,0)$. The problem is rather trivial, because the function is already a quadric and the Hessian is constant. In fact, the saddle has exactly the same shape as the conjugate harmonic saddle $z = x^2 - y^2$, but rotated by 45 degrees. For any quadratic function in two variables, finding the critical point reduces to solving two linear equations and two unknowns and the Hessian will be a constant. What we have is the pleonasm that a quadratic function of two variables is its own Taylor series of order 2.

3.7.3 Example Let $f(x,y) = x^3y + 12x^2 - 8y$. Classify the critical points
Solution. Take the first partial derivatives and solve to find the critical points

$$\begin{aligned} f_x &= 3x^2y + 24x = 0, \\ f_y &= x^3 - 8 = 0, \;\Rightarrow\; x = 2, \\ 12y + 24(2) &= 0, \;\Rightarrow y = -4. \end{aligned}$$

Hence, we have a critical point at $(2,-4)$, We test the critical point.

$$\begin{aligned} f_{xx} &= 6xy + 24, \quad & f_{xy} &= 3x^2, \\ f_{yx} &= 3x^2, & f_{yy} &= 0, \\ \mathcal{H}(x,y) &= -9x^4. \end{aligned}$$

So $\mathcal{H}(2,-4) < 0$ and we have a saddle at $(2,-4)$.

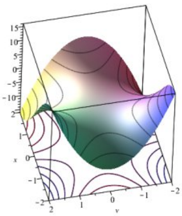
3.7.4 Example Let $f(x,y) = x^3 - 3xy^2$. Find and classify the critical points
Solution. This is an interesting cubic surface.

$$\begin{aligned} f_x &= 3x^2 - 3y^2 = 0, \quad && \Rightarrow\; y = \pm x, \\ f_y &= -6xy = 0, && \Rightarrow x = 0 \;\text{ or }\; y = 0. \end{aligned}$$

Either way, the only critical point is $(0,0)$. We compute the Hessian.

$$f_{xx} = 6x, \quad f_{xy} = -6y,$$
$$f_{yx} = -6y, \quad f_{yy} = 0$$

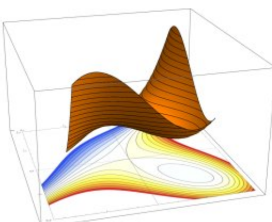
We find that the Hesssian $\mathcal{H}$ is identically 0 and therefore the second derivative test is inconclusive.



The surface in this example is the monkey saddle defined by the harmonic function corresponding to the real part of the complex function $w = z^3$. The surface has no good quadratic approximation at the origin, so one needs to consider the next order of the Taylor series. In this case the function is already third order and it is tautologically its own Taylor series.
The monkey saddle is one of the various possible shapes in the classification of critical points of cubic surfaces.

3.7.5 Example Let $f(x, y) = x^3 - 3xy + y^3$. Find and classify all the critical points
Solution

$$\begin{aligned} f_x &= 3x^2 - 3y = 0, \quad \Rightarrow \quad y = x^2, \\ f_y &= -3x + 3y^2 = 0, \quad \Rightarrow \quad x = y^2. \end{aligned}$$



Therefore, we have

$$x = x^4, \Rightarrow x(x^3 - 1) = 0.$$

The solutions are $x = 0$ and $x = 1$. Since $y = x^2$, We have two critical points $(0, 0)$ and $(1, 1)$. We now test the critical points using the Hessian.

$$\begin{aligned} f_{xx} &= 6x, \quad f_{xy} = -3, & \mathcal{H}(0,0) &< 0, \text{ so } (0,0) \text{ is a saddle point}, \\ f_{yx} &= -3 \quad f_{yy} = 6y & \mathcal{H}(1,1) &= 37 > 0, \\ \mathcal{H} &= 36xy - 9 & f_{xx}(1,1) &= 6 > 0. \quad \text{Local minimum at } (1,1). \end{aligned}$$

The level curve $f(x, y) = x^3 - 3xy + y^3 = 0$ is of historical significance in the development of calculus. The curve is called the **folium of Descartes**. It is said that Descartes challenged Fermat to find the equations of the tangent lines. Fermat solved the problem easily. Students it this class can also solve the problem easily by implicit differentiation.

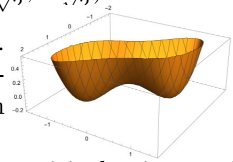
3.7.6 Example Let $f(x, y) = x^2 + y^4 + 2xy$. Find the local maximum and minimum values and saddle points.
Solution

$$\begin{aligned} f_x &= 2x + 2y = 0. \Rightarrow y = -x, \\ f_y &= 4y^3 + 2x = 0 \Rightarrow -4x^3 + 2x = 0, \\ 2x(1 - 2x^2) &= 0 \Rightarrow x = 0, \tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}. \end{aligned}$$

So the critical points are $(0, 0), (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ and $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

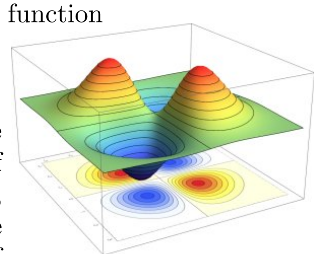
We compute the Hessian

$$\begin{array}{ll} f_{xx} = 2, \quad f_{xy} = 2, & \mathcal{H}(0,0) < 0, \text{ so } (0,0)\text{is a saddle point.} \\ f_{yx} = 2 \quad f_{yy} = 12y^2 & \mathcal{H}(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}) = \mathcal{H}(-\frac{1}{\sqrt{2}}, +\frac{1}{\sqrt{2}}) > 0. \\ \mathcal{H} = 24y^2 - 4 & f_{xx}(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}) = f_{xx}(-\frac{1}{\sqrt{2}}, +\frac{1}{\sqrt{2}}) > 0. \end{array}$$



so there are local minima at the other two critical points. We obtain the minimum values by evaluating the function at these points. In both cases we get a minimum value of $(-1/4)$.
The plot here is rendered in Mathematica. To enhance the critical points we fiddled with the RegionFunction command to control the height of the z-coordinate so as not to overwhelm the size of the bumps.

3.7.7 Example Let's finish this section with a slightly more challenging problem. Find and classify the critical points of the function



$$f(x,y) = xye^{-\frac{1}{2}(x^2+y^2)}$$

Before we proceed with the computations, we observe that the function is symmetric under an exchange of the variables x and y. Hence, if compute f_x and f_{xx}, we immediately get f_y and f_{yy} by such an exchange of variables. The computation is a good exercise of the use of the product rule.

$$\begin{aligned} f_x &= xy\, e^{-\frac{1}{2}(x^2+y^2)}(-x) + y\, e^{-\frac{1}{2}(x^2+y^2)}, \\ f_x &= y(1-x^2)\, e^{-\frac{1}{2}(x^2+y^2)}, \\ f_y &= x(1-y^2)\, e^{-\frac{1}{2}(x^2+y^2)}. \end{aligned}$$

Since the exponential is never 0, setting $f_x = 0$ and $f_y = 0$ gives five critical points, namely $(0,0)$, $(1,\pm 1)$ and $(-1,\pm 1)$. We proceed to compute the second partial derivatives using again the symmetry of the variables to save time in finding f_{yy}. We leave the computation of f_{xy} to the reader

$$\begin{aligned} f_{xx} &= y[(1-x^2)\, e^{-\frac{1}{2}(x^2+y^2)}(-x) - 2x\, e^{-\frac{1}{2}(x^2+y^2)}], \\ &= xy\, e^{-\frac{1}{2}(x^2+y^2)}[-(1-x^2) - 2], \\ f_{xx} &= xy(x^2-3)\, e^{-\frac{1}{2}(x^2+y^2)}, \\ f_{yy} &= xy(y^2-3)\, e^{-\frac{1}{2}(x^2+y^2)}, \\ f_{xy} &= (1-x^2)(1-y^2)\, e^{-\frac{1}{2}(x^2+y^2)} \end{aligned}$$

The Hessian is

$$\mathcal{H}(x,y) = \begin{vmatrix} xy(x^2-3) & (1-x^2)(1-y^2) \\ (1-x^2)(1-y^2) & xy(y^2-3) \end{vmatrix} e^{-(x^2+y^2)}$$

This time we will not write the full Hessian; instead, we evaluate it at each point from the second partial derivatives. We have

- $\mathcal{H}(0,0) = \begin{vmatrix} 0 & (1)(1) \\ (1)(1) & 0 \end{vmatrix} e^0 = -1$. Hence $(0,0)$ is a saddle point.
- $\mathcal{H}(1,1) = \mathcal{H}(-1,-1) = \begin{vmatrix} -2 & 0 \\ 0 & -2 \end{vmatrix} e^{-2} = 4\,e^{-2} > 0.$

  $f_{xx}(1,1) = f_{xx}(-1,-1) = -2\,e^{-1} < 0.$

  Hence we have local maxima at $(1,1)$ and $(-1,-1)$
- $\mathcal{H}(1,-1)\mathcal{H}(-1,1) = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} e^{-2} = 4\,e^{-2} > 0.$

  $f_{xx}(1,-1) = f_{xx}(-1,1) = 2\,e^{-1} < 0.$

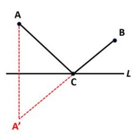
  Hence we have local minima at $(1,1)$ and $(-1,-1)$

3.8 Lagrange Multipliers

Lagrange multipliers is the eponymous name for a method developed in the 1700's by Joseph-Louis Lagrange, to find the extrema of functions subject to constraints. Lagrange is one of the giants in the history of mathematics and physics. He is one of the developer of calculus of variations. The Euler-Lagrange variational principle transformed Newtonian mechanics into a branch of analysis that can be used to obtain the field equations of a system by the extrema of some functional that basically represents the energy of the system. The Euler-Lagrange method can be used to formulate the field equations of mechanics, electricity and magnetism, semi-classical quantum mechanics, quantum field theory and general relativity. We present here a "baby" application to vector calculus.

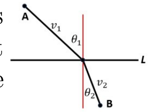
We introduce the subject by posing a question I first encountered in a geometry competition in my fifth grade class. I thought this would be good way to personalize these notes, since winning the competition is what first motivated me to be a mathematician.

Here is the problem. You are given a picture showing two points A and B on one side of a line L. Using only a compass and a straight edge, find the point C on the line such that $\overline{AC} + \overline{CB}$ is as small as possible.



A solution to the problem can be constructed as follows. Find the reflection A' of the point A across the line L and draw a straight line from A to B. The point of intersection C is the desired point. Of course, one needs to prove by Euclidean geometry that the answer is correct. I encountered the problem again in my first year calculus course. It turns out that this problem with such a remarkably simple solution by geometry, can be framed in term of the Fermat principle that states that, a path taken by a ray of light between two points is the one that takes minimum time. If one thinks of the line as a mirror, the solution shows that the angle of incidence is equal to angle of reflection. The calculus solution involves adapting a coordinate plane, labelling an arbitrary point C on L, finding the equation of the sums of the distances from A and B to the arbitrary point, and finally, applying the second derivative test. This is by far more difficult.

A variation of this problem that also appears often in calculus books is this. Suppose the two points are on different sides of the line. Perhaps one could pretend that the line is the border of a beach and that a lifeguard at point A sees a person drowning at point B. The life guard can walk faster than she can swim. If the speed on land is v_1 and on water v_2, at what point on the beach should the lifeguard go into the water so that the time is as short as possible. The problem could just as well refer to a ray of light entering a glass medium from air. If the angles of the light ray with the normal at the point of entry are θ_1 and θ_2, the optimum path occurs when



$$\frac{\sin\theta_1}{v_1} = \frac{\sin\theta_2}{v_2}.$$

The ratio v_1/v_2 is called the index of refraction.

The two problems above are examples of optimizing a function subject to some constraint. We now present a fancier problem based on the same principles as the previous two. There is a farmhouse located at point A and a pigsty at

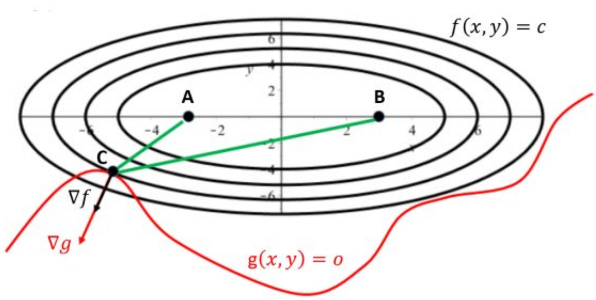


Fig. 3.15: Lagrange Multipliers

point B. Nearby there is a reservoir with the shoreline represented by a curve. Find the point C on the curve such that the distance $\overline{AC} + \overline{CB}$ is shortest. The solution can also be solved by pure geometry by drawing a sequence of confocal conics with foci at A and B and then picking the point C to be the first one on which one of the conics is tangential to the curve, as shown in figure 3.15. Formally with calculus, one adapts a coordinate plane with A and B on the x-axis and the origin in the middle. Let $f(x,y)$ by the function that represents the sum of the distances from an arbitrary point $X(x,y)$ to A and B. The function $z = f(x,y)$ is a surface for which the level curves $f(x,y) = 2a$ are ellipses with semi-major axis a. By definition, the sum of the two distances $\overline{AC} + \overline{CB} = 2a$. Thus, tangency points on the shoreline with conics with larger values of a, are farther away. It is very interesting to note that by the optical properties of ellipses, at the point C, with respect to the tangent line, the angle

of incidence is equal to the angle of reflection; thus the first problem above with the reflection across a flat mirror, is a spacial case.
Since the level curve of $f(x, y)$ is tangential to $g(x, y) = 0$ at C, their normals at that point are multiples of each other.

Perhaps a more intuitive understanding is gained from thinking of $f(x, y)$ as a mountain and the constraint $g(x, y) = 0$ as a trail. In figure 3.16 we have a real picture of El Nevado Del Tolima, one of three amazing volcanos visible from my native city in the central Andes mountains of Colombia. The most impressive of these volcanos is El Nevado Del Ruiz, but in 1987, a major eruption buried the city of Armero causing 21,000 deaths. We chose not to exhibit El Ruiz. Along with the photograph provided by family members, we also show the

Fig. 3.16: Nevado del Tolima, Colombia

volcano as a surface, using geo-elevation data accessed by Mathematica. On the level curves I have drawn a non-existent trail as an illustration. There is a difference between the maximum height of the mountain as opposed to the maximum elevation along the trail. The latter occurs at a point where the trail is tangential to a level curve.
The examples are motivation for the following theorem that we state without proof.

3.8.1 Theorem If the extrema of a function $f(x, y)$ subject to constraint $g(x, y) = 0$ exist, then they occur at the point(s) where ∇f is parallel to ∇g. That is, there exist constant(s) λ such that

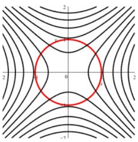
$$\nabla f = -\lambda \nabla g$$

Better yet, let

$$F(x, y, \lambda) = f(x, y) + \lambda(x, y), \quad \text{then } \ \nabla F = 0 \tag{3.40}$$

This is the method of Lagrange multipliers with one constraint. We take all the components of the gradient of F, set them equal to 0 and solve. Then we check the values at the critical points to determine the extrema.

3.8.2 Example Find the extrema of $f(x, y) = x^2 - y^2$ with constraint $x^2 + y^2 = 1$.



Solution
The level curves of the saddle $f(x, y)$ are hyperbolas. The constraint in 3D is a cylinder of radius 1. A saddle has no maximum or minimum, but the constraint chops out a pringle of the surface. The figure shows that there are two points at which we have a maximum, namely $(\pm 1, 0)$ and two at which there is a minimum, namely $(0, \pm 1)$. Now we solve the problem by Lagrange multipliers. Let $F = f + \lambda g$. We have

$$\begin{aligned} F(x, y, \lambda) &= x^2 - y^2 + \lambda(x^2 + y^2 - 1), \\ F_x &= 2x + 2x\lambda = 2x(1 + \lambda) = 0, \\ F_y &= -2y + 2y\lambda = 2y(-1 + \lambda) = 0, \\ F_\lambda &= x^2 + y^2 - 1 = 0 \end{aligned}$$

The equation $F_\lambda = 0$ always gives back the constraint. Solving the equations, we get $x = 0$, or $y = 0$, or $\lambda = \pm 1$. In this simple case the values of λ are not needed. If $x = 0$ the constraint gives $y = \pm 1$, and if $y = 0$, we get $x = \pm 1$.We have 4 critical points. To find the extrema we just evaluate the function at all the critical points. The largest value is the maximum and the smallest is the minimum. We have

$$f(\pm 1, 0) = 1, \quad f(0, \pm 1) = -1$$

The maximum is 1 and it occurs at points $(1, 0)$ and $(-1, 0)$.
The minimum is -1 and it occurs at points $(0, 1)$ and $(0, -1)$.

3.8.3 Example Find the extrema of $f(x, y) = 4x + 6y$ subject to the constraint $x^2 + y^3 = 13$
Solution The function $z = 4x + 6y$ is a plane and the intersection with the cylinder $x^2 + y^3 = 13$ is a slanted ellipse. We should get a maximum and a minimum at the vertices of the ellipse. Let $F = f + \lambda g$

$$\begin{aligned} F &= 4x + 6y + \lambda(x^2 + y^2 - 13), \\ F_x &= 4 + 2x\lambda = 0, && \Rightarrow \quad x = -2/\lambda, \\ F_y &= 6 + 2y\lambda = 0, && \Rightarrow \quad y = -3/\lambda, \\ F_\lambda &= x^2 + y^2 - 13 = 0, && \Rightarrow \quad x^2 + y^2 = 13. \end{aligned}$$

Inserting the values of x and y into the constraint equation, we get

$$\begin{aligned} \frac{4}{\lambda^2} + \frac{9}{\lambda^2} &= 13, \\ \frac{13}{\lambda^2} &= 13, \quad \Rightarrow \lambda = \pm 1. \end{aligned}$$

If $\lambda = 1$, then $(x, y) = (-2, -3)$ and $f(-2, -3) = -26$.
If $\lambda = -1$, then $(x, y) = (2, 3)$ and $f(2, 3) = 26$.
$f_{max} = 26$ and $f_{min} = -26$

3.9 Taylor's Theorem

We present a quick overview of Taylor's theorem for a function of two variables. We will dispense a bit from the rigor of estimating the error term in a Taylor polynomial approximation. Let $z = f(x, y)$ be a real analytic function, so that the function is infinitely differentiable in a neighborhood U of a point $P(x_0, y_0)$; all the partial derivatives exist and are continuous. Following the introduction to directional derivatives, as illustrated in figure 3.11, let $\mathbf{x}(t) = \mathbf{P} + t\mathbf{v}$ be a line containing the point P. we restrict t to range from $t = 0$ to $t = 1$. Since P is the point with $t = 0$, we also use the notation $\mathbf{P} = \mathbf{x}_0$, and $f(\mathbf{x}_0) = f(x_0, y_0))$. We call the components of the direction vector $\mathbf{v} = \langle h, k\rangle$. We choose h and k to be sufficiently small so that the points in the line segment are inside U.
In parametric form, we have

$$x = x_0 + ht, \quad y = y_0 + kt.$$

Define $g(t) = f(\mathbf{x}(t)) = f(x_0 + ht, y_0 + kt)$. By the chain rule,

$$\begin{aligned} df &= f_x\, dx + f_y\, dy, \\ g'(t) &= f_x \frac{dx}{dt} + f_y \frac{dy}{dt}, \\ g'(t) &= hf_x + kh_y. \end{aligned}$$

We introduce an operator D_t defined by

$$\begin{aligned} D_t &= h\tfrac{\partial}{\partial x} + k\tfrac{\partial}{\partial y}, \\ g'(t) = D_t g &= (h\tfrac{\partial}{\partial x} + k\tfrac{\partial}{\partial y})(f) = hf_x + kf_y, \\ g''(t) = D_t^2 g &= (h\tfrac{\partial}{\partial x} + k\tfrac{\partial}{\partial y})^2(f) = h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy}, \\ g'''(t) = D_t^3 g &= (h\tfrac{\partial}{\partial x} + k\tfrac{\partial}{\partial y})^3(f) = h^3 f_{xxx} + +3h^2 k f_{xxy} + 3hk^2 f_{xyy} + k^3 f_{yyy}, \\ g^{(n)}(t) = D_t^n g &= (h\tfrac{\partial}{\partial x} + k\tfrac{\partial}{\partial y})^n(f). \end{aligned}$$

Now we expand $g(t)$ into a Taylor series, with the usual caution that infinitely differentiable does not always imply that a Taylor series exists.

$$g(t) = g(0) + g'(0)t + \tfrac{1}{2!}g''(0)t^2 + \ldots + \tfrac{1}{n!}g^{(n)}(0)t^n + \tfrac{1}{(n+1)!}g^{(n+1)}(c)t^n,$$

where, by the extended mean value theorem, there exists at least one number $c \in [0, 1]$, such that the remainder is given by

$$\tfrac{1}{(n+1)!}g^{(n+1)}(c)t^n.$$

Now, we set $t = 1$

$$g(1) = g(0) + g'(0) + \tfrac{1}{2!}g''(0) + \ldots + \tfrac{1}{n!}g^{(n)}(0) + \tfrac{1}{n!}g^{(n)}(c),$$

At $t = 1$ we have $\mathbf{x} = \langle x_0 + h, y_0 + k\rangle$ and $\mathbf{x}_0 = \langle x_0, y_0\rangle$, so the equation becomes

$$\begin{aligned} f(x_0+h, y_0+k) &= f(x_0, y_0) + [hf_x + kf_y)]_{\mathbf{x}_0} + \tfrac{1}{2!}[h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy}]_{\mathbf{x}_0} \\ &\tfrac{1}{3!}[h^3 f_{xxx} + +3h^2 kf_{xxy} + 3hk^2 f_{xyy} + k^3 f_{yyy}]_{\mathbf{x}_0} + \dots \\ &+ \tfrac{1}{n!}[(h\tfrac{\partial}{\partial x} + k\tfrac{\partial}{\partial y})^n (f)]_{\mathbf{x}_0} + \tfrac{1}{(n+1)!}[(h\tfrac{\partial}{\partial x} + k\tfrac{\partial}{\partial y})^{n+1}(f)]_{(\mathbf{x}_0+\mathbf{c})} \end{aligned} \tag{3.41}$$

where $\mathbf{c} = \langle c, c\rangle,\ c \in [0,1]$. For the analog of Maclaurin series, we take $\mathbf{x}_0 = \vec{0}$. Since the series works for any (h,k), we replace these by variables (x,y). We get

$$\begin{aligned} f(x,y) &= f(0,0) + [xf_x + yf_y)]_{\vec{0}} + \tfrac{1}{2!}[x^2 f_{xx} + 2hxy_{xy} + y^2 f_{yy}]_{\vec{0}}, \\ &\tfrac{1}{3!}[x^3 f_{xxx} + +3x^2 yf_{xxy} + 3xy^2 f_{xyy} + y^3 f_{yyy}]_{\vec{0}} + \dots \\ &+ \tfrac{1}{n!}[(x\tfrac{\partial}{\partial x} + y\tfrac{\partial}{\partial y})^n (f)]_{\vec{0}} + \tfrac{1}{(n+1)!}[(x\tfrac{\partial}{\partial x} + y\tfrac{\partial}{\partial y})^{n+1}(f)]_{\mathbf{c}}. \end{aligned} \tag{3.42}$$

We need to be careful with the expression above. The evaluation of the terms in brackets applies only to the derivatives, and the operator D_t only applies to f and not to the the variables x and y. The first two terms of 3.41 and 3.42 represent the linear approximation. We take a closer look at 3.42 in the context of Max-Min. For an extremum to occur at $(0,0)$ it is a necessary condition that $f_x = f_y = 0$. Therefore the quadratic approximation at the origin is

$$f(x,y) - f(0,0) = \frac{1}{2}(f_{xx}(0,0)x^2 + 2f_{xy}(0,0))xy + f_{yy}(0,0)y^2$$

The behavior of the function at the critical point is dependent on the quadric

$$Q(x,y) = Ax^2 + 2Bxy + Cy^2,\ \ A = f_{xx}, B = f_{xy},\ C = f_{xx}\ \text{ at } (0,0),$$

that measures the value of the difference between $f(x,y)$ and $f(0,0)$. The proper way to analyze such quadratic forms is the theory of eigenvalues and eigenvector in linear algebra, but we can get a reasonable handle on the situation by the elementary process of completing the square. We have

$$\begin{aligned} Q(x,y) &= A\left[\left(x^2 + \frac{2By}{x} + \frac{B^2}{A^2}y^2\right) + \frac{C}{A}y^2 - \frac{B^2}{A^2}y^2\right], \\ &= A\left[\left(x + \frac{B}{A}y\right)^2 + \frac{AC - B^2}{A^2}y^2\right] \end{aligned}$$

The first quantity is a perfect square, so the value of function is determined by the discriminant

$$D = AC - B^2 = f_{xx}f_{yy} - f_{xy}^2\,|_{(0,0)}$$

There are three cases

1. $D > 0$. Then the quantity in brackets is a sum of squares.

   a) If $A > 0$, then $f(x,y) > f(0,0)$ at all points nearby and there is a local minimum at $(0,0)$.

b) If $A < 0$, then $f(x, y) < f(0, 0)$ at all points nearby and there is a local maximum at $(0, 0)$.

2. $D < 0$. Then $Q(x, y)$ is a difference of squares and we have a saddle

3. $D = 0$. Then $Q(x, y) = 0$ along $x + (B/A)y = 0$ and can't tell what happens.

We include a reminder, that as shown in figure 3.14, the presence of a cross term means that the approximating quadric is rotated about the origin.

Chapter 4

Multiple Integrals

4.1 Riemann Sums

4.1.1 Review of Riemann Sums in one Variable

We begin with a brief review of Riemann sums for functions of one variable. Let $z = f(x)$ be a continuous function on the interval $[a, b]$. Assume $f(x) > 0$ on this interval.The assumption is not necessary and it makes absolutely no difference in the definition of integrals by Riemann sums, but if we make the assumption, the integral we define represents the area under the graph between a and b. A partition of the interval $[a, b]$ into n subintervals, is a choice of a set

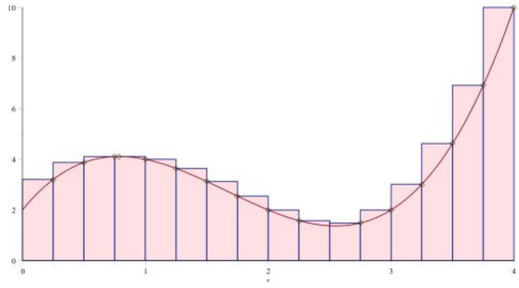


Fig. 4.1: Riemann Sum

of numbers $\{x_0, x_1 \dots x_n\}$ such that

$$a = x_0 < x_1 \dots < x_k < \dots < x_n = b$$

For computational reasons, we choose the partition to be equally spaced into subintervals of width Δx, so that we can write equations for the width of the interval and for the partition coordinates. We have

$$\Delta x = \frac{b-a}{n}, \quad \text{and} \quad x_k = a + k\,\Delta x$$

The second equation says that to obtain the x_k coordinate, one starts at a and takes k steps of size Δx. We approximate the area under the curve by the sum of areas of rectangles with base Δx and height $f(x_k)$, as shown in figure 4.1. We define

$$\int_a^b f(x)\,dx \equiv \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k)\,\Delta x. \tag{4.1}$$

We could make the definition a bit more general by taking the height on each subinterval to be $f(x^*)$ where $x_k^* \in [x_{k-1}, x_k]$ is any coordinate, but then the definition would not be useful for computation. The typical choices of x_k^* for each interval are the left, the middle or the right coordinate. Here we have chosen the right coordinates. The approximation by finite number of rectangles is called the right sum. The theorem of Riemann says that whether or not the partition is equally spaced and how the points are chosen within each subinterval to evaluate the height, the limit exists and gives the same answer. Convergence of the Riemann sums is vastly improved by taking the midpoints or taking the average of the left and right sums. The left sum just requires that we change the index of summation to $k = 0..n-1$. The average of the right and left sums is equivalent to the trapezoidal rule.

Figure 4.1 was created in Maple using the Student[CalculusI] package command Riemann Sum, with the option "method=upper" which in this case is the right sum. If no method is specified, the default is the midpoint method. The package is easy to use, but as it is the case with all packages, it is more instructive to write the code from scratch.

If we had to compute the limit of a Riemann sum every time we needed for all functions, integral calculus would not have survived for long. Fortunately, Newton proved the fundamental theorem of calculus that says that if $f(x)$ is integrable, then

$$\int_a^b f(x)\,dx = F(b) - F(a), \text{where } \; F'(x) = f(x).$$

4.1.2 Double Riemann Sums

We follow a parallel procedure to introduce double integrals. Let $z = f(x, y)$ be a continuous function over a rectangle $R = [a, b] \times [c, d]$. We assume that the function is positive over the rectangle. We wish to compute the volume under the surface above the yx-plane over the rectangle. Let $\{a = x_0, x_1, \ldots, x_n = b\}$ be a partition of $[a, b]$, and $\{c = y_0, y_1, \ldots, y_m = d\}$ a partition of $[c, d]$. For computational simplicity, we choose the partitions to be equally spaced so that

$$\Delta x = \frac{b-a}{n}, \quad x_k = a + k\,\Delta x,$$
$$\Delta y = \frac{d-c}{m}, \quad y_l = c + l\,\Delta y,$$

as shown in figure 4.2. The increment of area is given by $\Delta A = \Delta y \Delta x$. For each sub-rectangle in the grid, multiply ΔA by the value of the function at the

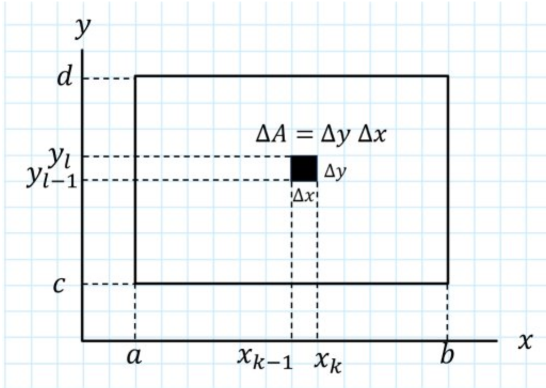


Fig. 4.2: Double Riemann Sum

point at the upper right (UR) corner (x_k, y_l). The product gives the volume of a cuboid with base ΔA and height $f(x_k, y_l)$. We define

$$\int_a^b \int_c^d f(x,y)\,dy\,dx = \lim_{n\to\infty} \lim_{m\to\infty} \sum_{k=1}^{n} \sum_{l=1}^{m} f(x_k, y_l)\Delta y \Delta x, \tag{4.2}$$

provided the limit exists. One could also evaluate the Riemann sums by choosing the height at the lower left (LL) corner of the rectangles just by changing the indices of summation to $k = 0] \ldots, n-1$ and $l = 0, \ldots, m-1$. A better number would be obtained from the average of these two, or by evaluating the function at the midpoints. The integral on the left-hand-side of equation 4.2 is called an **iterated integral**. Technically one should write

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

but the bracket is usually suppressed.]
The summation process can be easily implemented in Maple as illustrated by the following example.

4.1.1 Example Consider the paraboloid $f(x,y) = 9 - x^2 - y^2$, over the rectangle $R = [0,2] \times [0,2]$. For the first finite approximation we take $n = m = 2$ which gives 4 boxes over a 2×2 grid. This makes it very easy to do a pencil and paper computation of the volume of the four boxes. Here $\Delta x = \Delta y = 1$. The approximation to the volume by upper right (UR) corner sums is

$$V \approx [f(1,1) + f(1,2) + f(2,1) + f(2,2)](1)(1) = 7 + 4 + 4 + 1 = 16$$

The approximation by the lower left (LL)corner sums is

$$V \approx [f(0,0) + f(0,1) + f(1,0) + f(1,1)](1)(1) = 9 + 8 + 8 + 7 = 32$$

Of course these will both be a very poor approximations. The average 24 should be a bit better. We would like to increase grid to have much larger number of rectangles such as 12×12. Evaluation of the Riemann sum with 144 terms becomes no task for modern humanoids, but that is the reason we have coded the process into a computer as shown in figure 4.3. The output of the maple

Double integrals by Riemann Sums Visualization

```
> restart : with(plottools) : with(plots) : with(student) :
> a := 0 : b := 2 : c := 0 : d := 2 :
> n := 2 : m := 2 :
> f := (x, y) → 9 - x^2 - y^2 :
> Δx := (b-a)/n : Δy := (d-c)/m :
> plotf := plot3d(f(x, y), x = a..b, y = c..d,
      color = red, style = surface, transparency = 0.3) :
> Approx := { } :
    for i from 1 to n do;
      for j from 1 to m do;
       Approx := Approx union
         { cuboid([a + (i - 1) Δx, c + (j - 1) Δy, 0],
                  [a + i Δx, c + j Δy, f(a + i Δx, c + j Δy)],
      color = green)}
    od: od:
> display(plotf, Approx)
```

$$> \sum_{k=1}^{n}\left(\sum_{l=1}^{m}(f(a+k\,\Delta x, c+l\,\Delta y)\ \ \Delta y\ \ \Delta x)\right)$$

$$> \int_a^b\int_c^d f(x,y)\,\mathrm{d}y\ \mathrm{d}x$$

Fig. 4.3: Maple Code for Riemann Sums

code shown was removed so as to fit the entire code in a smaller window. The graphics output for $n = m = 2$ and for $n = m = 12$ grid are shown in figure 4.4. There is no need to write separate Maple codes for the (UR), (LL) or midpoint sums. The "tutor" on multivariate calculus under the "tools" heading provides a radio button driven menu for the various Riemann sum methods. The applet allows for input of function $f(x, y)$, the rectangle of integration and the grid size. The display shows the boxes, the value of the Riemann sums, and for comparison, the "exact value".
Here are the results for the various Riemann sums for the two grid sizes we have chosen

Grid	UR	LL	(UR + LL)/2	Midpoint
2×2	16	32	24	26
12×12	23.963	26.630	25.297	25.352

Because the definition of the integral is essentially an extension of the definition in one variable, we can use the fundamental theorem of calculus. The key for this to work is that when we integrate with respect to y we are taking the limit

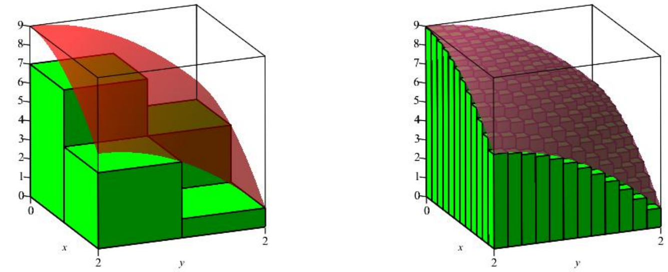


Fig. 4.4: Riemann sum for $f(x,y) = 9 - x^2 - y^2$ with 2×2 and 12×12 grid.

of Riemann sums in the y direction while we keep the x-coordinates fixed. This is exactly the reverse process of taking partial derivatives. When we perform a multiple integral with respect to a variable, one treats the others as if they were constant. Thus, the process is iterative. We first evaluate the integral with respect to y using its limits of integration, then we get a single integral with respect to x. For the example in question, we have

$$\begin{aligned}
\int_0^2 \int_0^2 (9 - x^2 - y^2)\, dy\, dx &= \int_0^2 (9y - x^2 y - \tfrac{1}{3}y^3)\Big|_{y=0}^{y=2}\, dx, \\
&= \int_0^2 (18 - 2x^2 - \tfrac{8}{3})\, dx, \\
&= 18x - \tfrac{2}{3}x^3 - \tfrac{8}{3}x\Big|_0^2, \\
&= 36 - \tfrac{16}{3} - \tfrac{16}{3}, \\
&= \frac{76}{3} \approx 25.333
\end{aligned}$$

As in he case of a single integral, if the function $f(x,y)$ is positive on the rectangle, the double integral gives the volume under the surface. If the $f(x,y)$ is not positive on the all or part of the region, the definition and computation of the integral is not affected in any way, but the result is the volume above the xy- plane minus the volume under the xy-plane. We have the following

4.1.2 Theorem (Fubini) If $f(x,y)$ is continuous over a rectangle $R = [a,b] \times [c,d]$, then

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

The proof of the theorem is beyond the scope of this course and it belongs more in a more theoretical course in analysis. However, the intuition behind the theorem is easy to visualize. Think of an infinitesimal differential of area rectangle $dA = dy\, dx$. Assuming $f(x,y) > 0$, the expression $f(x,y)\, dA$ represents the volume of ΔV a "French fry" with an infinitesimal rectangle as base

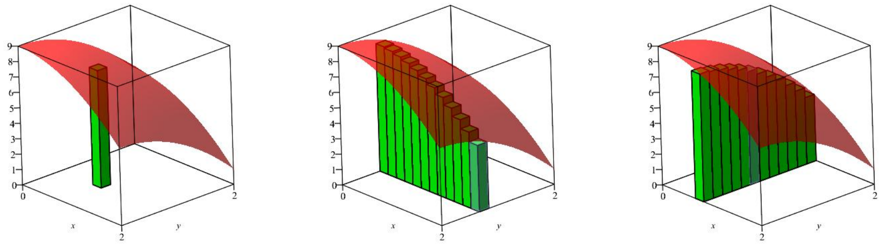


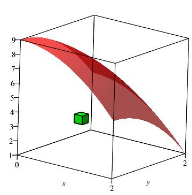
Fig. 4.5: Fubini's theorem - Intuition

and with height equal to the height of the surface above dA, as shown in figure 4.5. The illustrations were obtained by the same Maple code included, but restricting the values of one or both of the do-loops to one number.

- When $dA = dy\,dx$, that is $dV = f(x,y)\,dy\,dx$, the first iterated integral with respect to y adds the volumes of the pillars along the y direction yielding the volume of a "bread slice" of infinitesimal thickness, as also shown in figure 4.5. The second integral with respect to x adds the volume of the bread slices along the x direction to give the entire volume. The formula $V = \int\int f(x,y)\,dA$ just says that the whole is the sum of its pieces.

- When $dA = dx\,dy$, that is $dV = f(x,y)\,dx\,dy$, the first iterated integral with respect to x adds the volumes of the pillars along the x direction yielding the volume of a "bread slice" of infinitesimal thickness. The second integral with respect to y adds the volume of the bread slices along the y direction to give the same volume.

There is not real need to wait several sections to introduce volumes by triple integrals. The idea is much the same.
Define an infinitesimal differential of volume $dV = dz\,dy\,dx$. This is represented by an mini sugar cube. Then the volume is just given by $V = \int\int\int 1\,dV$. The integral with respect to z goes from the bottom surface, ($z = 0$ in this case) to the top surface $z = f(x,y)$. The integral over z just gives $f(x,y)$ and we are immediately reduced to the double integral formula for the volume. The advantage of the triple integral at this point is minimal, except that is more natural to compute areas by double integrals and volumes by triple integrals.



4.1.3 Example Evaluate $I = \int\int_R (6xy^2 + 4y)\,dA$, where R is the square region $R = [0,1] \times [0,1]$
Solution. We evaluate the integral in both orders to verify the theorem of Fubini in this case.

$$
\begin{aligned}
I &= \int_0^1 \int_0^1 (6xy^2 + 4y)\, dy\, dx, \\
&= \int_0^1 \left[2xy^3 + 2y^2\right]_{y=0}^{y=1} dx, \\
&= \int_0^1 (2x + 2)\, dx, \\
&= 3
\end{aligned}
\qquad \text{or} \qquad
\begin{aligned}
I &= \int_0^1 \int_0^1 (6xy^2 + 4y)\, dx\, dy, \\
&= \int_0^1 \left[3x^2y^2 + 4xy\right]_{x=0}^{x=1} dy, \\
&= \int_0^1 (3y^2 + 4y)\, dy, \\
&= 3
\end{aligned}
$$

4.1.4 Example Find the volume under the surface $f(x, y) = y\, e^{xy}$ above the region $R = [0, 1] \times [0, 1]$
Solution. According to Fubini's theorem we get the same answer regardless of the order of integration, however the integral with respect to y requires integration by parts. Thus, it is easier to integrate with respect to x first.

$$
\begin{aligned}
V &= \int_0^1 \int_0^1 \int_0^{y\, e^{xy}} 1\, dz\, dx\, dy, \\
&= \int_0^1 \int_0^1 y\, e^{xy}\, dx\, dy, \\
&= \int_0^1 \left[e^{xy}\right]_{x=0}^{x=1} dy, \\
&= \int_0^1 (e^y - 1)\, dy, \\
&= e^y - y\big]_{y=0}^{y=1}, \\
&= e - 1 - (1) = e - 2.
\end{aligned}
$$

4.2 Volume Integrals

4.2.1 Double Integrals over General Regions

To calculate a double integral over a region that is not rectangular, one has to be more attentive to the order of integration. In figure 4.6 we show three regions with different features In the first two figures, the differential of area dA is shown as a small black rectangle. We will call (a) a type I region, (b) a type II region, and c) is of both types.

1. For regions of type I the natural order of integration is $dA = dy\, dx$ because if the first integral is with respect to y, we may scan the entire region with infinitesimal vertical slices, without running into walls or kinks. If we denote the region of integration by

$$R = \{(x, y) \in \mathbf{R}^2 : a \leq x \leq b; \;\; f(x) \leq y \leq g(x)\}.$$

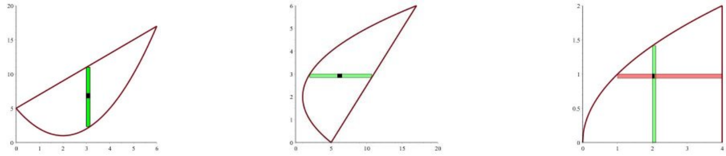


Fig. 4.6: Regions of a) Type I, b) Type II, c) Both

then the integral of a surface $F(x, y)$ over this region is given by

$$I = \int_a^b \int_{f(x)}^{g(x)} F(x, y)\, dy\, dx.$$

2. For regions of type II the natural order of integration is $dA = dx\, dy$ because if the first integral is with respect to x, we may scan the entire region with infinitesimal horizontal slices, without running into walls or kinks. If we denote the region of integration by

$$R = \{(x, y) \in \mathbf{R}^2 : h(y) \leq x \leq k(y); \quad c \leq x \leq d, \}.$$

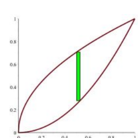
then the integral of a surface $F(x, y)$ over this region is given by

$$I = \int_c^d \int_{h(y)}^{k(y)} F(x, y)\, dx\, dy.$$

3. If region is both of type I and type II, the order of integration is dictated by judgment about which is easier. It might be the case that the degree of difficulty is about the same in either order of integration, or it could happen the in one particular order one encounters an integral that does not exist in terms of elementary functions.

4.2.1 Example Evaluate $\int\int_R (x + y), dA$, where R is the region bounded by $y = x^2$ and $y = \sqrt{x}$.

Solution



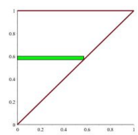
The first step is to graph the region of integration. The region of integration is bounded by a branch of a parabola that opens up and a branch of another parabola that opens to the right. To find the intersection points, we set $x^2 = \sqrt{x}$. In this case we see by inspection that the solutions are $x = 0$ and x=1.

The region is of both types, so we may choose to integrate in the y-direction first. We observe that the parabola $y = x^2$ is at the bottom, so this must reflected in the limits of integration. We have

$$\begin{aligned} I &= \int_0^1 \int_{x^2}^{\sqrt{x}} (x+y)\, dy\, dx, \\ &= \int_0^1 \left[xy + \tfrac{1}{2}y^2\right]_{y=x^2}^{y=\sqrt{x}}\, dx, \\ &= \int_0^1 [x\sqrt{x} + \tfrac{1}{2}x - (x^3 + \tfrac{1}{2}x^4)]\, dx, \\ &= \int_0^1 [x^{3/2} + \tfrac{1}{2}x - x^3 - \tfrac{1}{2}x^4]\, dx, \\ &= \tfrac{2}{5} + \tfrac{1}{4} - \tfrac{1}{4} - \tfrac{1}{10} = \tfrac{3}{10}. \end{aligned}$$

4.2.2 Example Evaluate $\int\int_R e^{y^2}, dA$, where $R = \{(x,y) : 0 \le y \le 1;\ 0 \le x \le y\}$

Solution



The integral with respect to y of e^{y^2} is not doable in terms of elementary functions, so it is clear we should attempt to integrate with respect to x first. The values of x from $x = 0$ to the line $x = y$, so the region of integration is a triangle above the the line $x = y$.

We proceed to compute the integral

$$\begin{aligned} I &= \int_0^1 \int_0^y e^{y^2}\, dx\, dy, \\ &= \int_0^1 y\, e^{y^2}\, dy, \\ &= \tfrac{1}{2}e^{y^2}\Big|_0^1, \\ &= \tfrac{1}{2}(e-1). \end{aligned}$$

4.2.2 Triple Integrals

When setting up an integral to compute a volume over a region R is more natural to define a differential of volume dV and perform a triple iterated integrals. Suppose the volume is over the region R bounded by

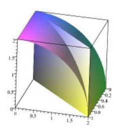
$$\begin{aligned} G(x,y) \le &z \le F(x,y), \\ g(x) \le &y \le f(x), \\ a \le &x \le b. \end{aligned}$$

Then the triple integral for the volume is

$$
\begin{aligned}
dV &= dz\,dy\,dx,\\
V &= \int\int\int_R dV = \int\int\int_R 1\,dz\,dy\,dx,\\
V &= \int_a^b \int_{g(x)}^{f(x)} \int_{G(x,y)}^{F(x,y)} 1\,dz\,dy\,dx.
\end{aligned}
$$

The rules are the same as for every iterated integral. When one integrates over a variable, one treats the other variables as if they were constant.

4.2.3 Example Find the volume in the first octant bounded by $y^2 + z^2 = 4$, $x = 2y$, $x = 0$, $z = 0$
Solution
The first surface $y^2 + z^2 = 4$ is a cylinder extruded in the direction of the missing coordinate x. The second surface $x = 2y$ is a vertical plane. The two surfaces intersect on a quarter of an ellipse in the first octant. The shape looks like a wedge with a triangular base and cylindrical roof.



We set up volumes as triple integrals $V = \int\int\int 1\,dV$. The first integral is always trivial and reduces the problem to a double integral. We have

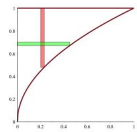
$$
\begin{aligned}
I &= \int_0^2 \int_0^{2y} \int_0^{\sqrt{4-y^2}} 1\,dz\,dx\,dy,\\
&= \int_0^2 \int_0^{2y} \sqrt{4-y^2}, dx\,dy,\\
&= \int_0^2 2y\sqrt{4-y^2}\,dy,\\
&= -2\tfrac{1}{2}\tfrac{2}{3}(4-y^2)^{3/2}\Big|\,0^2,\\
&= -\tfrac{2}{3}(0-8) = \frac{16}{3}
\end{aligned}
$$

4.2.3 Reversing the Order of Integration

Given a function $f(x, y)$ over a general region D in the xy-plane, one can enclose the region by a rectangle R. One then extends the function to the entire rectangle by a new function $g(x, y)$ which is equal to $f(x, y)$ in D and 0 outside of D. Fubini's theorem applies to $\int\int_R g(x, y)\,dA$. One can easily argue the the theorem extends to $\int\int_D f(x, y)\,dA$ as long as reversing the order of integration scans the entire region. Reversing the order of integration may result in an easier integral or in one that would not be doable in terms of elementary functions in the original order.

4.2.4 Example Let $\displaystyle\int_0^1 \int_{\sqrt{x}}^1 \sqrt{y^3+1}\,dy\,dx$. Reverse the order of integration to compute the integral

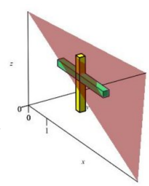
Solution



The integrals of square roots of cubic polynomials result in incomplete elliptic integrals of the first kind that are way above the level of this course. Our only hope is to reverse the order of integration. Since y ranges from $y = \sqrt{x}$ to 1, the region of integration is the triangular-shaped region above the parabola. If we change to horizontal slices, x now ranges from 0 to $x = y^2$. We proceed to reverse the order on integration

$$\begin{aligned} I &= \int_0^1 \int_{\sqrt{x}}^1 \sqrt{y^3+1}\,dy\,dx, \\ &= \int_0^1 \int_0^{y^2} \sqrt{y^3+1}\,dx\,dy, \\ &= \int_0^1 y^2\sqrt{y^3+1}\,dy \\ &= \tfrac{2}{3}\tfrac{1}{3}(y^3+1)^{3/2}\Big|_0^1, \\ &= \tfrac{2}{9}(2\sqrt{2}-1) \end{aligned}$$

4.2.5 Example Volume of Tetrahedron. Consider the tetrahedron in the first octant bounded by $2x+3y+4z = 12$. We set up the triple integral for the the volume in two different orders of integration.



The points of intersection with the coordinate axes are given by $x = 6$, $y = 4$, and $z = 3$. Let $dV = dz\,dy\,dx$ meaning that we wish to integrate first with respect to z resulting on a "French fry" with infinitesimal footprint on the xy-plane of area $dA = dy\,dx$. We then integrate with respect to y to get a bread slice of the volume. The limits of the integral with respect to y range from 0 to the value of y on the line of intersection of the given plane with the plane $z = 0$. The equation of the line is therefore $2x + 3y = 12$. The triple integral for the volume is

$$V = \int_0^6 \int_0^{\frac{1}{3}(12-2x)} \int_0^{\frac{1}{4}(12-2x-3y)} 1\,dz\,dy\,dx.$$

On the other hand, if we set $dV = dy\,dz\,dx$, the "roof" is now given by solving the equation of the original plane for y and the line of intersection with the xz-plane is given by $2x + 4z = 12$. The volume integral is

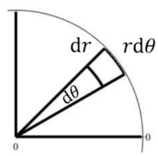
$$V = \int_0^6 \int_0^{\frac{1}{4}(12-2x)} \int_0^{\frac{1}{3}(12-2x-4z)} 1\,dy\,dz\,dx.$$

In total, there are six possible orders of integration corresponding to the six permutations of the variables. The problem is completely academic because no one in their right mind would compute the volume of a tetrahedron by triple integrals when by vector methods all we need is $V = \frac{1}{6}|(\mathbf{abc})|$, where the three vectors represent edges of the tetrahedron. In the case in question, the absolute value of the triple product is,

$$V = \frac{1}{6}\begin{vmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{vmatrix} = 12.$$

4.3 Polar and Cylindrical Coordinates

For integration problems that exhibit symmetry with respect to an axis, it is most convenient to use cylindrical coordinates as in 2.5.1.
The main item with need for integration is the differential of area dA. We divide the region into wedges of infinitesimal central angle $d\theta$ and the wedges into rings of radii r and $r+dr$. The elements of area have dimension dr and $r\,d\theta$. Then the differential of area is $dA = r\,dr\,d\theta$.

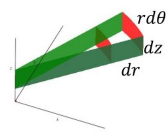


Given a polar curve $r = r(\theta)$, then the area A is given by the double integral

$$A = \int\int 1\,dA = \int\int r\,dr\,d\theta = \frac{1}{2}\int r^2\,d\theta,$$

and we immediately recover the formula one learns in calculus II. The assumption is that students have done a substantial number of polar area problems.

We are interested primarily on triple integrals in coordinates (z, r, θ). This time we first divide the region into pizza wedges of infinitesimal thickness dz and infinitesimal angles $d\theta$ and the wedges into rings of radii r and $r + dr$. The element of volume is a cube so its volume is the base area of the base $dA = r\,dr\,d\theta$ times the thickness dz. Thus, the differential of volume dV is given by



$$dV = dz(r\,dr\,d\theta) = r\,dz\,dr\,d\theta.$$

The extra factor r always appears in the multiple integrals when we transform to polar coordinates on two of the Cartesian coordinates. The factor is called the **Jacobian** of the transformation. While we are at it, the arc length element in cylindrical coordinates is the square of the diagonal of the cube

$$ds^2 = dr^2 + r^2\,d\theta^2 + dz^2 \tag{4.3}$$

As we introduced quadric surfaces in section 1.6, we pointed out that any curve on the xy-plane of the form $z = f(x)$ can be transformed into a **surface of revolution**

$$z = f(r),$$

by just replacing x by r. Here are some basic examples.

1. Spheres. Start with $x^2 + z^2 = a^2$. Replace x^2 by $(r^2 = x^2 + y^2)$ and we get

$$r^2 + z^2 = a^2,$$
$$z = \pm\sqrt{a^2 - r^2}$$

The equation with the positive square root is the upper hemisphere and the one with the negative sign is the lower hemisphere.

2. Circular Paraboloids. Start with $z = ax^2$. Replace x^2 by $(r^2 = x^2 + y^2)$ and we get

$$z = ar^2$$

If $a > 0$, the paraboloid points up and if $a < 0$, the paraboloid points down. Thus, for example, the paraboloid $z = 9 = x^2 - y^2$ in cylindrical coordinates is given by $z = 9 - r^2$

3. Cones. Start with $z = mx$, then $z^2 = m^2x^2$. If we replace x^2 by $(r^2 = x^2 + y^2)$ we get

$$z = \pm mr.$$

If $m > 0$ we get the upper branch of the cone that opens up, and if $m < 0$, we get the lower branch of the cone that opens down. Equations of cones with axial symmetry about a coordinate axis are particularly nice in cylindrical coordinates because the functions do not involve square roots. All we need to know, is the slope of the generating line. Thus, for example

- $z = -r$ is the cone $z = -\sqrt{x^2 + y^2}$ that opens down at 45^o with axis of symmetry along the z-axis.
- $z = \sqrt{3}r$ is the cone $z = \sqrt{3}\sqrt{x^2 + y^2}$ that opens up at 60^o with axis of symmetry along the z-axis.
- $y = \sqrt{x^2 + z^2}$. There is nothing special about the z axis. We can do polar coordinates in any pair of cartesian coordinates. Here we can let $x = r\cos\theta$ and $z = r\sin\theta$. In cylindrical coordinates, the equation simplifies to $y = r$

4.3.1 Example Volume of a Sphere

Find the volume of the sphere $x^2 + y^2 + z^2 = a^2$.

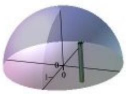
Solution

The equation of a sphere is not a function. Solving for z we actually get two functions, namely $z = \sqrt{a^2 - x^2 - y^2}$ and $z = -\sqrt{a^2 - x^2 - y^2}$. The full integral in Cartesian coordinates is

$$V = \int_{-a}^{a} \int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} \int_{-\sqrt{a^2-x^2-y^2}}^{\sqrt{a^2-x^2-y^2}} 1\, dz\, dy\, dx,$$

We could simplify the integral somewhat by integrating over the first octant and using symmetry, but no matter, the integral is still rather formidable. The square root integrands have no chain rule factors outside, so we are bound to encounter a number of trigonometric substitutions.

The sphere is symmetric with respect to the z-axis, so it is better to switch to cylindrical coordinates. We save half of the work by setting of the integral over the top half and then multiplying by 2.



In cylindrical coordinates the equation of the sphere is $z = \sqrt{a^2 - r^2}$. To find the volume we integrate $V = \int\int\int 1\, dV$. The integral with respect to z gives the volume of a "French fry" with an infinitesimal footprint on a circle of radius a. We scan over the circle using polar coordinates. We get the integral

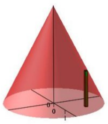
$$\begin{aligned}
V &= 2\int_0^{2\pi}\int_0^a\int_0^{\sqrt{a^2-r^2}} r\, dz\, dr\, d\theta,\\
&= 2\int_0^{2\pi}\int_0^a r\sqrt{a^2-r^2}\, dr\, d\theta \quad \text{(we get the chain rule factor } r\text{)},\\
&= 2(2\pi)\left[-\tfrac{1}{2}\tfrac{2}{3}(a^2-r^2)^{3/2}\right]_0^a,\\
&= -\tfrac{4\pi}{3}((0-(a^2)^{3/2}),\\
&= \tfrac{4}{3}\pi a^3
\end{aligned}$$

Thanks to the Jacobian r, we got just the chain rule factor we needed for the integral of the square root and thus, we avoided the dreaded trigonometric substitutions. It is remarkable that Archimedes was able to obtain the formula for the volume of a sphere 2200 years ago by the method of exhaustion.

Lesson. If you have symmetry, use it!

4.3.2 Example Volume of a Cone

We wish to find the volume of circular cone with base radius R and height h. We adapt the coordinate axes so that the circular base sits on the xy-plane, centered at the origin. Then the generator of the cone on the xz- plane is the line segment



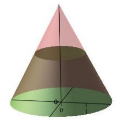
with slope $-h/R$ and z-intercept h. The equation of the cone is obtained by replacing x by r in the equation of the line. Thus, the the cylindrical equation of the cone is,

$$z = -\frac{h}{R}r + h.$$

The volume integral is

$$\begin{aligned} v &= \int_0^{2\pi}\int_0^R\int_0^{-\frac{h}{R}r+h} r\,dz\,dr\,d\theta, \\ &= \int_0^{2\pi}\int_0^R (-\tfrac{h}{R}r^2 + hr)\,dr\,d\theta, \\ &= 2\pi(-\tfrac{h}{R}\tfrac{R^3}{3} + \tfrac{hR^2}{2}), \\ &= 2\pi R^2 h(\tfrac{1}{2} - \tfrac{1}{3}) \\ V &= \tfrac{1}{3}\pi R^2 h. \end{aligned}$$

4.3.3 Example Volume of Frustum of a Cone



We wish to find the volume of a truncated cone with base radius b, top radius a and height h. Again We adapt the coordinate axes so that the circular base sits on the xy-plane, centered at the origin. Then the generator of the cone on the xz- plane is the line segment passing through the points $(b, 0)$ and (a, h). It is tempting to write the equation of line segment in the xy-plane, replace x by r to get the cylindrical equation of the cone, and integrate to find the volume. But why do that when we can solve the problem by geometry. Extend the generator to its point of intersection $(0, H)$ with the z-axis. The volume of the truncated cone is the difference between the full cone with base b and hight H and the top cone with base a and height $H - h$. By ratio and proportions.

$$\begin{aligned} \frac{b}{a} &= \frac{H}{H-h}, \\ H &= \frac{bh}{b-a} \qquad \text{Solving for } H, \\ H - h &= \frac{bh}{b-a} - h = \frac{ah}{b-a}. \end{aligned}$$

The volume is

$$\begin{aligned} V &= \tfrac{1}{3}\pi b^2 H - \tfrac{1}{3}\pi a^2 (H-h), \\ &= \tfrac{1}{3}\pi b^2 \frac{bh}{b-a} - \tfrac{1}{3}\pi a^2 \frac{ah}{b-a}, \\ &= \tfrac{1}{3}\pi h \frac{b^3 - a^3}{b-a}, \\ &= \tfrac{1}{3}\pi h(b^2 + ab + a^2). \end{aligned}$$

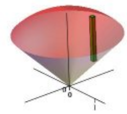
It is interesting to note how very few people including those in academia, are acquainted with this formula which was also known to Archimedes.

4.3.4 Example Volume of a SnoKone

Find the volume of the SnoKone bounded above by the sphere $x^2 + y^2 + z^2 = a^2$ and below by the cone $z = \sqrt{x^2 + y^2}$

Solution

As in the previous example, the integral would be a nightmare in Cartesian coordinates because now the z integral would range from $\sqrt{x^2+y^2}$
In cylindrical coordinates the only changes we need to the integral limits is than now z ranges from r to $\sqrt{a^2-r^2}$ and the r limits from 0 to the radius of the circle of intersection of the sphere and the cone. To find this radius we set $r^2 = a^2 - r^2$, so $r = a/\sqrt{2}$. .



We get the integral

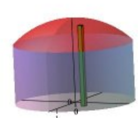
$$\begin{aligned}
V &= \int_0^{2\pi}\int_0^{a/\sqrt{2}}\int_r^{\sqrt{a^2-r^2}} r\,dz\,dr\,d\theta,\\
&= \int_0^{2\pi}\int_0^{a/\sqrt{2}} r[\sqrt{a^2-r^2}-r]\,dr\,d\theta,\\
&= \int_0^{2\pi}\int_0^{a/\sqrt{2}} [r\sqrt{a^2-r^2}-r^2]\,dr\,d\theta,\\
&= -\frac{2\pi}{3}\left[(a^2-r^2)^{3/2}+\frac{r^3}{3}\right]_0^{a/\sqrt{2}},\\
&= -\frac{2\pi}{3}[(a^2-\frac{a^2}{2})^{3/2}+\frac{a^3}{2\sqrt{2}}-a^3],\\
&= -\frac{2\pi}{3}[\frac{a^3}{2\sqrt{2}}+\frac{a^3}{2\sqrt{2}}-a^3],\\
&= \frac{2\pi}{3}[a^3-\frac{a^3}{\sqrt{2}}],\\
&= \frac{\pi}{3}a^3(2-\sqrt{2})
\end{aligned}$$

We will find later that this volume integral is much easier in spherical coordinates.

4.3.5 Example Find the volume of the "pill capsule" bounded inside the sphere $x^2+y^2+z^2=4$ and the cylinder $x^2+y^2=2$

Solution

Using symmetry we evaluate the integral over the plane $z=0$ and multiply by 2. The picture resembling a grain elevator barn represents only half of the volume In cylindrical coordinates integral over z ranges from $z=0$ to $z=\sqrt{4-r^2}$. The integral over z gives the volume of a "French fry" with base on s circle of radius $r=\sqrt{2}$. Thus, to obtain the entire volume of the top half, we integrate r from 0 to $\sqrt{2}$ and θ from 0 to 2π. .



We are ready to evaluate the volume integral,

$$
\begin{aligned}
V &= 2\int_0^{2\pi}\int_0^{\sqrt{2}}\int_0^{\sqrt{4-r^2}} r\,dz\,dr\,d\theta, \\
&= 2\int_0^{2\pi}\int_0^{\sqrt{2}} r[\sqrt{4-r^2}-r]\,dr\,d\theta, \\
&= 2(2\pi)\left[-\tfrac{1}{2}\tfrac{2}{3}(4-r^2)^{3/2}\right]_0^{\sqrt{2}}, \\
&= -\frac{4\pi}{3}(2\sqrt{2}-8) = \frac{8\pi}{3}(4-\sqrt{2})
\end{aligned}
$$

4.4 Applications

4.4.1 Center of Mass

Given some mass distribution in space, the center of mass is that point at which one would have to concentrate the total mass so that the system is in static equilibrium. To introduce the concept we begin with a discrete distribution of masses along the axes. Suppose masses $\{m_1, m_2, m_3, \ldots, m_k, \ldots, m_n\}$ are placed on a weightless horizontal rod at positions $\{x_1, x_2, x_3, \ldots, x_k, \ldots, x_n\}$ respectively, as shown in figure 4.7 We seek a coordinate $\overline{x}$ where the total mass

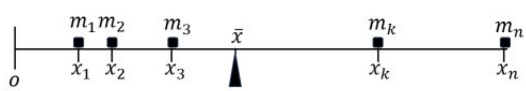


Fig. 4.7: Center of Mass - Discrete Distributin

$M = m_1 + m_2 + \cdots + m_n$ would have to placed so that its torque about the origin would be equal to the sum of the torques of the individual masses. The point $\overline{x}$ is the point at which a negative torque of magnitude $Mg\overline{x}$ counterbalances the total torque of the other masses, so that the rod just balances. The condition we need is,

$$
Mg\overline{x} = m_1 g x_1 + m_2 g x_2 + \ldots + m_k g x_k + \cdots + m_n g x_n.
$$

The gravity constant g cancels out. Solving for $\overline{x}$, we get

$$
\overline{x} = \frac{m_1 x_1 + m_2 x_2 + \ldots + m_k x_k + \cdots + m_n x_n}{m_1 + m_2 + \cdots + m_k + \cdots + m_n}
$$

The point $\overline{x} = x_{c.m}$ is called the center of mass. The equation is the same as used to compute an average such as the GPA (grade point average.) It is best to write the equation using the summation symbol, specially if the number of point masses is large. The more concise and more elegant equation is

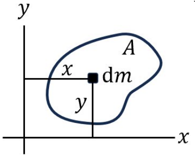
$$
\overline{x} = \frac{\sum_{k=1}^{n} x_k m_k}{\sum_{k=1}^{n} m_k} = \frac{1}{M}\sum_{k=1}^{n} x_k m_k. \tag{4.4}
$$

We are more interested in a continuous distribution of mass. Define the density $\rho(x) = dm/dx$. By the usual limiting process of Riemann sums to go from the discrete to the continuum, the equation for the center of mass becomes,

$$\overline{x} = \frac{\int x\, dm}{\int dm} \tag{4.5}$$

The numerator of this fraction is usually called the first moment. Now suppose that instead, we have 2-dimensional plate (a lamina) of area A of density $\rho(x, y)$.

We subdivide the region into infinitesimal rectangles of mass $dm = \rho\, dA$ at coordinates (x, y). The moments. The center of mass is the point $(\overline{x}, \overline{y})$ where the plate would balance horizontally on a pin. From the point of view physics, this is because according to Newton, the force of gravity on a body acts on the center of mass.

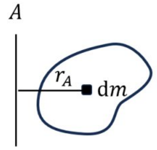


The formula for the center of mass in one dimension naturally extends to two dimensions, because at the end, we are still looking for averages. The equations are

$$\overline{x} = \frac{\int x\, dm}{\int dm}, \qquad \overline{y} = \frac{\int y\, dm}{\int dm}, \tag{4.6}$$

the only difference being that the integrals are now double integrals over the area. Unless there is some symmetry, one needs to compute three double integrals to find the coordinates of the center of mass in two dimensions

We can actually summarize symbolically the formula for the center of mass of any coordinate, whether the mass distribution is one, two or three dimensional. Denote by dm the differential of mass, let A be some axis. Axes are usually taken to be the coordinate axes



parallel to the coordinate axes. Let ρ be the density, dm the differential of mass and r_A the distance from a coordinate of the dm to the axis A. Then, we define

$$\begin{aligned} M_A &= \int r_A\, dm \quad \text{Moment of Mass,} \\ \overline{r}_A &= \frac{\int r_A\, dm}{\int dm} \quad \text{Center of Mass,} \\ I_A &= \int r_A^2\, dm \quad \text{Moment of Inertia,} \end{aligned} \tag{4.7}$$

where the integrals are single, double or triple integrals depending on whether dm

$$dm = \begin{cases} \rho\, ds & \text{For a mass distribution along a curve,} \\ \rho\, dA & \text{For a mass distribution on an area,} \\ \rho\, dV & \text{For a mass distribution on a volume} \end{cases} \tag{4.8}$$

is one, two, or three dimensional.

4.4.1 Example Find the coordinates of the center of mass of a semicircular lamina $y = \sqrt{a^2 - x^2}$, $y \geq 0$ of homogeneous density $\rho = k$.
Solution. By symmetry, we expect the center of mass to be along the y-axis. We could set up the integrals for $\overline{x}$ but the answer would be 0. It suffices to compute $\overline{y}$. The limits of integration will be exactly the same as if we were computing an area, since we need to scan the same region. We also save some time noticing that the total mass $M = \int dm$ is k times the area of the semicircle $\frac{1}{2}\pi a^2$. Thus, we have,

$$
\begin{aligned}
M\overline{y} &= \int\int y\,dm = \int_{-a}^{a}\int_{0}^{y} y(k\,dy\,dx), \\
&= \int_{0}^{\pi}\int_{0}^{a} kr\sin\theta(r\,dr\,d\theta). \\
&= -\tfrac{1}{3}ka^3\,\cos\theta|_0^{\pi}\,, \\
&= \tfrac{2}{3}ka^3, \\
\overline{y} &= \frac{\frac{2}{3}ka^3}{\frac{1}{2}\pi a^2}, \\
&= \tfrac{4}{3\pi}a.
\end{aligned}
$$

This is about right. There is more area below the center, so the center of mass should be located below $a/2$.

4.4.2 Moment of Inertia

We start by providing a most elementary introduction to the motivation for the definition of moments of inertia given in equation 4.7. Consider a point mass m constrained to move in a circle of radius r. We apply a force of magnitude F perpendicular to r to cause the mass to accelerate in circular motion. We have

$$F = ma, \quad \text{and} \quad s = r\theta.$$

We differentiate the arc length s twice to get the speed and acceleration,

$$v = r\omega, \quad \text{and} \quad a = r\alpha,$$

where ω and α are the angular speed and the angular acceleration respectively. To transform $F = ma$ to a rotational second law of motion, we multiply F by the lever arm r to get the magnitude of the torque,

$$
\begin{aligned}
F &= ma, \\
rF &= mra = mr(r\alpha) = mr^2\alpha, \\
\tau &= I\alpha, \quad \text{where} \quad I = mr^2
\end{aligned}
$$

The quantity I is called the moment of inertia of the point mass. This is a good name because in rotational motion, I plays the role that m plays in linear motion. The moment of inertia I is that which opposes angular acceleration when a torque is applied. To find the moment of inertia of a body about an axis A we do what we usually do in calculus. We go to the infinitesimal level and integrate. Thus, given a blob of mass M and an axis A, we subdivide the mass into infinitesimal pieces dm with distance r_A to the axis. Then we integrate the moment of inertia of dm over the whole region occupied by the blob. We can do this because we can effectively consider dm a point mass. Of course this can be made rigorous with the formalism of Riemann sums, but the simplicity of the intuition is sufficient at this point. If dm is located a coordinates (x, y) for a lamina on the plane, or at coordinates (x, y, z) for mass distribution in $\mathbf{R}^3$, the distance to the z axis is $(x^2 + y^2)$, so the moment of inertial with respect to the z-axis is

$$I_z = \int (x^2 + y^2)\, dm$$

This is called the **polar moment of inertia**. It is useful for objects that have axial symmetry. So, lets get on with the work and compute the moment of inertial of some common geometric shapes that appear in physics and engineering. This is by far more productive than the typical academic exercises involving odd shapes that typically appear in standard calculus textbooks.

4.4.2 Example Moment of inertia of solid disk

Consider a solid disk or radius R, thickness h and homogeneous density $\rho = k$. We adapt a coordinate frame so that the equation of the disk is

$$x^2 + y^2 = R^2, \quad 0 \leq z \leq h$$

We set up the polar moment of inertia in cylindrical coordinates.

$$\begin{aligned} I_z &= \iiint (x^2 + y^2)(k\, dz\, dy\, dx), \\ &= \int_0^{2\pi} \int_0^R \int_0^h r^2(k\, r\, dz\, dr\, d\theta, \\ &= 2\pi k(\tfrac{1}{4}R^4)h = \tfrac{1}{2}k\pi R^4 h. \end{aligned}$$

We would like to have the formula in terms of the total mass $M = k\pi R^2 h$, so we perform a common trick of "multiplying by 1".

$$\begin{aligned} I_z &= \tfrac{1}{2}k\pi R^4 h \left[\frac{M}{k\pi R^2 h}\right], \\ I_z &= \tfrac{1}{2}MR^2, \end{aligned}$$

4.4.3 Example Moment of inertia of solid sphere.

Let $x^2 + y^2 + z^2 = R^2$ be the equation of a sphere of constant mass density $\rho = k$. We are interested in the polar moment of inertia I_z. Because of axial symmetry, it is convenient to set up the integral in cylindrical coordinates. Since a sphere is not a function, we compute the integral for the top half and multiply by 2.

$$
\begin{aligned}
I_z &= \iiint (x^2 + y^2)\, dm, \\
&= 2k \int_0^{2\pi} \int_0^{\sqrt{R^2-r^2}} r^2 k(r\, dz\, dr\, d\theta), \\
&= 2k \int_0^{2\pi} \int_0^{R} r^3 \sqrt{R^2 - r^2}\, dr\, d\theta.
\end{aligned}
$$

The integral with respect to r could be done by the standard methods of calculus II using trigonometric substitution or integration by parts. However, one of the tricks one learns through experience is that since the power of r outside the radical is odd, a regular calculus I substitution will do. Call the integral J.

$$
\begin{aligned}
J &= \int_0^R r^3 \sqrt{R^2 - r^2}\, dr, \\
&= \int_0^R r^2 \sqrt{R^2 - r^2} (r\, dr).
\end{aligned}
$$

The obvious substitution is

$$
\begin{aligned}
u &= \sqrt{R^2 - r^2}, \quad \text{that is,} \quad u^2 = R^2 - r^2, \\
2u\, du &= -2r\, dr, \qquad\qquad\qquad r^2 = R^2 - u^2, \\
J &= -\int_R^0 (R^2 - u^2) u (u\, du), \\
&= \int_0^R (R^2 u^2 - u^4)\, du, \\
&= R^5 (\tfrac{1}{3} - \tfrac{1}{5}) = \tfrac{2}{15} R^5
\end{aligned}
$$

The moment of inertia becomes

$$
\begin{aligned}
I_z &= 4\pi k \tfrac{2}{15} R^5, \\
&= \tfrac{8}{15} \pi k R^5 \left[\frac{M}{\tfrac{4}{3}\pi R^3 k} \right], \\
&= \tfrac{2}{5} M R^2
\end{aligned}
$$

4.4.4 Example Moment of inertia of solid cone.
Suppose the cone has base radius R, height h and homogenous density $\rho = k$. Position the cone so that it rests on the xy-plane with axis of symmetry aligned with the z axis. Then, the equation of the cone in cylindrical coordinates is

$$
z = -\frac{h}{R} r + h.
$$

The generator of the cone is a straight line on the xz-plane with the same slope and z-intercept. The polar moment of inertia $I_z = \iiint r^2\,dm$ is

$$\begin{aligned}
I_z &= \int_0^{2\pi}\int_0^R\int_0^{h-\frac{h}{R}r} r^2 k(r\,dz\,dr\,d\theta),\\
&= k\int_0^{2\pi}\int_0^R [hr^3 - \tfrac{h}{R}r^4]\,dr\,d\theta,\\
&= 2\pi k R^4(\tfrac{1}{4}-\tfrac{1}{5}),\\
&= \tfrac{1}{10}h\pi R^4,\\
&= \tfrac{1}{10}h\pi R^4\left[\frac{M}{\frac{1}{3}k\pi R^2 h}\right],\\
&= \tfrac{3}{10}MR^2
\end{aligned}$$

4.4.3 Normal Distribution

We present a brief introduction to the normal distribution. A full treatment of concepts associated with the normal distribution would be the principal topic of an entire first course on probability and there would be no way to do justice to the subject in a few pages. Hence, we restrict ourselves to the modest task of deriving the normalization constants to insure we have indeed a probability density function.

4.4.5 Example Bell-shaped curve. Evaluate

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

Solution
This is a most important integral because it is the foundation of normal distributions in probability. As it stands the indefinite integral is not doable in terms of elementary functions. However, there is most famous trick attributed to the French mathematician and physicist Siméon Denis Poisson, that allows us the compute the integral exactly. We can write

$$I = \int_{-\infty}^{\infty} e^{-y^2}\,dy$$

because the definite integral does not depend on the name of the dummy variable of integration. Then we have

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2}\,dx \int_{-\infty}^{\infty} e^{-y^2}\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dy\,dx,$$

thus transforming into a double improper integral over the entire plane.

The integrand is now an axially symmetric function in the shape of a 2 dimensional, bell-shaped surface, and hence it makes sense to change to polar coordinates. The integral is over the entire plane, so in polar coordinates θ ranges from 0 to 2π and r ranges from 0 to infinity

The double integral then transforms to an improper integral,

$$\begin{aligned} I^2 &= \int_0^{2\pi}\int_0^{\infty} r\, e^{-r^2} dr\, d\theta, \\ &= 2\pi \left[\lim_{R\to\infty} -\tfrac{1}{2} e^{-r^2} \right]_0^R \\ &= -\pi \lim_{R\to\infty} e^{-R^2} - 1, \\ &= \pi, \quad \text{therefore,} \\ I &= \sqrt{\pi} \end{aligned}$$

The appearance of π here is a remarkable result. It is only doable because the Jacobian r once again provides up to a constant, the chain rule factor needed to compute the integral of the exponential function. it means that the function

$$f(x) = \frac{1}{\sqrt{\pi}} e^{-x^2} \tag{4.9}$$

integrates to a value of 1 over the entire real line and hence it constitutes a probability density function. The mean value $\overline{x}$ for the distribution is given by the integral

$$\overline{x} = \int_{\infty}^{\infty} x\, f(x)\, dx = \frac{1}{\sqrt{\pi}} \int_{\infty}^{\infty} x e^{-x^2}$$

The mean is zero because we have an integral of an odd function over a symmetric interval. Some modifications are usually made. Consider the integral

$$J = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}}\, dt.$$

Making the substitution $x^2 = t^2/(2\sigma^2)$, we have $x = t/(\sigma\sqrt{2})$ and $dt = (\sigma\sqrt{2}\, dx)$. After the substitution the integral becomes

$$J = \sigma\sqrt{2}\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2}\, dx = \sigma\sqrt{2}$$

Hence,

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}}\, dt = 1$$

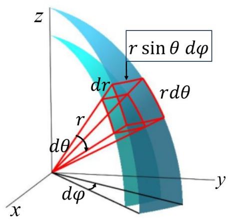
We then perform a translation by replacing t by $(x - x_0)$. The value of the integral is not affected. The function

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_O)^2}{2\sigma^2}} \tag{4.10}$$

is called the **normal distribution**, also called the **Gaussian distribution**. The mean of the distribution is translated to x_0. The quantity σ is called the **standard deviation**. If σ is small, the distribution has a sharp peak at the mean value, and if σ is large, the distribution is spread out. As stated before, the normal distribution constitutes the central topic of discussion on a first course on probability.

4.5 Spherical Coordinates

To integrate in spherical coordinates, we need a formula for the differential of volume. We proceed as usual. We subdivide the volume into infinitesimal pieces and use the principle that the whole is the sum of its pieces. Think of a spherical orange. We subdivide sphere into wedges, the wedges into sectors and the sectors, into vesicles.



The key to integrals in spherical coordinates is the concept depicted in subsection 2.5.2 of a triply orthogonal system. That means that, as in the case of Cartesian coordinates, the "vesicle" differential of volume is a cube. As such all we need to know are the side dimensions of the cube. As shown in the picture. Let dr, $d\theta$ and $d\phi$ be the differentials of the coordinates. One edge of the cube is clearly dr. The edge subtended by the infinitesimal angle $d\theta$ is an arc of a great circle of radius r, so it has dimension $r\,d\theta$. The third edge is an arc of a parallel circle on the sphere and not an arc of a great circle. To find its dimension we must first project the arc into the xy-plane. The projected arc is subtended by the infinitesimal angle $d\phi$ has dimension $r\sin\theta\,d\phi$. Multiplying the lengths of the edges of the cube, we get

$$dV = r^2 \sin\theta\, dr\, d\theta\, d\phi. \tag{4.11}$$

We may as well note that the square of the element of arc length in spherical coordinates the square of the diagonal of the cube

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta\, d\phi^2 \tag{4.12}$$

4.5.1 Example Volume of a Sphere. As the first and most basic example in spherical coordinates We compute the volume of a sphere $x^2 + y^2 + z^2 = R^2$; that is $r = R$. We have

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

4.5.2 Example Volume of SnoKone. We compute the volume of the SnoKone is bounded above by the sphere $x^2 + y^2 + z^2 = a^2$ and below by the cone $z = \sqrt{x^2 + y^2}$. The volume integral is

$$\begin{aligned} V &= \int_0^{2\pi} \int_0^{\pi/4} \int_0^{a} r^2 \sin\theta \, dr \, d\theta \, d\phi, \\ &= -2\pi \tfrac{a^3}{3} \cos\theta \,\big|_0^{\pi/4}, \\ &= -\tfrac{2}{3}\pi a^3 (\tfrac{\sqrt{2}}{2} - 1), \\ &= \tfrac{\pi}{3} a^3 (2 - \sqrt{2}) \end{aligned}$$

Not surprisingly, the computation is much easier than the integral in cylindrical coordinates found in example 4.3.4

4.5.3 Example Volume of Ice Cream Cone. We compute the volume of the ice cream cone bounded above by the shifted sphere $r = a\cos\theta$ and below by the cone $\theta = \pi/6$. The volume integral is

$$\begin{aligned} V &= \int_0^{2\pi} \int_0^{\pi/6} \int_0^{a\cos\theta} r^2 \sin\theta \, dr \, d\theta \, d\phi, \\ &= \int_0^{2\pi} \int_0^{\pi/6} \tfrac{1}{3} a^3 \cos^3\theta \sin\theta \, dr \, d\theta \, d\phi, \\ &= -2\pi \tfrac{1}{12} a^3 \cos^4\theta \,\big|_0^{\pi/6} \\ &= -\tfrac{1}{6}\pi a^3 [(\tfrac{\sqrt{3}}{2})^4 - 1^4)], \\ &= -\tfrac{1}{6}\pi a^3 (\tfrac{9}{16} - 1) = \tfrac{7}{96}\pi a^3 \end{aligned}$$

4.6 Change of Variables

As mentioned in the introduction to chapter 2, a multivariate function

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

is called a change of coordinates. It is common in this context to call the function T and to use the term transformation of coordinates. So far we only have three examples, polar, cylindrical and spherical coordinates, Two examples really, since cylindrical coordinates are not that different from polar coordinates. When doing an integration in calculus I and calculus II we call change of variables a substitution, or in more vernacular terms, a u-substitution. When performing a definite integral by substitution, it is essential to keep in mind that the differential of the variable on integration acquires a chain rule factor, and the limits of integration need to be adjusted to the new variable.

In several variables the situation is more complicated. When changing variables in double of triple integrals what we have is really more like a simultaneous double or triple substitution, depending on whether we are in $\mathbf{R}^2$ or $\mathbf{R}^3$. The multivariate “chain rule” factor is what we have been calling the Jacobian. We have used elegant geometric arguments to obtain the Jacobians,

$$\begin{aligned} dA &= r\,dr\,d\theta, && \text{for polar coordinates,}\\ dV &= r\,dz\,dr\,d\theta && \text{for cylindrical coordinates,}\\ dV &= r^2\sin\theta\;\,dr\,d\theta\,d\phi && \text{for spherical coordinates,} \end{aligned}$$

We were able to make these geometric arguments and change the limits of integration because we could easily visualize by geometry, the effect of the regions of integration due to the coordinate transformations. For example, in polar coordinates, the transformation $T:\mathbf{R}^2\to\mathbf{R}^2$

$$(r,\theta)\xrightarrow{T}(x,y)$$

maps the rectangle $R=[0,2\pi]\times[0,a]$ in the $r\theta$-plane into a circle of radius a in the xy-plane. This change of variables falls in the special category of **orthogonal coordinates** in the sense that the constant coordinate functions $r=$ constant and $\theta=$ constant form an orthogonal grid of concentric circles and radial lines.

The elements of area and volume are stated in terms of differentials, so there must be a way to obtain the Jacobian but a computation using differentials. Lets begin with the differential of area in the Cartesian plane $dA=dx\,dy$. Let's give some thought to this expression. What kind of multiplication is this? If were to treat the differentials dx and dy as variables that can take any value, as defined in calculus textbooks, then why don't we see multiple integrals with squares or cubes of differentials. Certainly something subtle is going on here. There is no elementary answer to these fundamental questions, so we will take a leap of faith and jump into some 20th century mathematics. We define a new type of product of differentials by listing a “multiplication table”. The new type of product is called the **wedge product**, first introduced in 1844 by Hermann Grassmann. Here are the properties

$$\begin{aligned} dx\wedge dx &= 0, & & & dy\wedge dx &= -dx\wedge dy,\\ dy\wedge dy &= 0, & &\text{and} & dy\wedge dz &= -dz\wedge dy,\\ dz\wedge dz &= 0, & & & dz\wedge dx &= -dx\wedge dz. \end{aligned}$$

The wedge product only affects the differentials. Everything else is treated by the usual rules of arithmetic. Thus, for example $f\,dx\wedge g\,dy = fg\,dx\wedge dy$. The name of the variables or the number of variables do not affect the anti-commutativity of the wedge product. The idea of the wedge product is to capture for differentials, the properties of the cross product. After all, what we formally know about computing areas does indeed go back to cross products. If one changes the order of the cross product, the answer changes by a minus sign. The cross product of a vector with itself is zero. For the xy-plane we envision an infinitesimal rectangle with dimensions $dx\,\mathbf{i}$ and $dy\,\mathbf{j}$. Then, the differential of area $\overrightarrow{dA}_{xy}$ in the xy-plane would be the vector given by the cross product

$$\begin{aligned}\overrightarrow{dA}_{xy} &= dx\,\mathbf{i} \times dy\,\mathbf{j},\\ &= dx\,dy\,\mathbf{i} \times \mathbf{j},\\ &= dx\,dy\,\mathbf{k}\end{aligned}$$

What we get is an oriented differential of area vector, perpendicular to the xy-plane and with magnitude equal the area of the infinitesimal rectangle. The is exactly what one expects of the cross product of two vectors. In terms of wedge products, the differential of area is

$$dA_{xy} = dx \wedge dy. \tag{4.13}$$

The one big difference between the cross product and the wedge product is that that the former is only defined in $\mathbf{R}^3$. The differential of volume in terms of wedges is simply

$$dV = dx \wedge dy \wedge dz. \tag{4.14}$$

The wedge product is associative and distributive, so there is no need to include parenthesis. As long as we recall the definition of differentials in several variables, this is all we need to compute Jacobians. Let's compute the Jacobian in polar coordinates. We have

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta, \end{aligned} \qquad \text{and} \qquad \begin{aligned} dx &= \cos\theta\,dr - r\sin\theta\,d\theta, \\ dy &= \sin\theta\,dr + r\cos\theta\,d\theta, \end{aligned}$$

Using the rules of wedge multiplication and the law of distributivity, we compute $dA = dx \wedge dy$,

$$\begin{aligned} dA &= (\cos\theta\sin\theta)(dr \wedge dr) + r\cos^2\theta\,dr \wedge d\theta - r\sin^2\theta\,d\theta \wedge dr - r^2\sin\theta\cos\theta\,d\theta \wedge d\theta,\\ &= r\cos^2\theta\,dr \wedge d\theta - r\sin^2\theta\,d\theta \wedge dr,\\ &= r(\cos^2\theta + \sin^2\theta)dr \wedge d\theta,\\ &= r\,dr \wedge d\theta \end{aligned}$$

It does not take much to realize that if we arrange the coefficients of dx and dy into a matrix, the Jacobian is just the determinant of the coefficients. That is,

$$\begin{aligned} dA = dx \wedge dy &= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} dr \wedge d\theta,\\ &= r\,dr \wedge d\theta. \end{aligned}$$

The change of variable theorem in this case say that

$$\iint_D f(x,y)\,dx\,dy = \iint_R f(r\cos\theta, r\sin\theta)\,r\,dr\,d\theta,$$

where D is the region in the xy-plane which is the image under T of a region in the $r\theta$-plane.

4.6.1 Remark The discussion above appears to contradict Fubini's theorem. The resolution to this is also slightly subtle. When one sets up a double iterated integral $I = \iint_D f(x,y)\,dx\,dy$, what is meant is $I = \int(\int_D f(x,y)\,dx)\,dy$, so technically there is no multiplication of the differentials. The parenthesis is ignored to avoid the clutter. We will worry about orientated areas in chapter 5 when we treat surface integrals.

It is interesting to note that the transformation T maps a rectangle R in the $r\theta$-plane to a circular sector region D in xy-plane. On the other the, the integration over $dA = dx\,dy$ is transformed backward into an integral over $dA = r\,dr\,d\theta$. For this reason, the differential of area $dA = r\,dr\,d\theta$ called the **pull-back** in the literature.

The general equation of the Jacobian for a coordinate transformation in the plane is easily obtained. Let $x = x(u,v)$ and $y = y(u,v)$. Computing the differentials,

$$\begin{aligned} dx &= x_u\,du + x_v\,dv, \\ dy &= y_u\,du + y_v\,dv, \\ dA = dx \wedge dy &= \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} du \wedge dv. \end{aligned}$$

The Jacobian $J = \partial(x,y)/\partial(u,v)$ of the transformation $T : (u,v) \to (x,y)$ is then given by,

$$J \equiv \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}. \tag{4.15}$$

Thus, if $T : \mathbf{R}^2 \to \mathbf{R}^2$ maps a region R in the uv-plane to a region D in the xy-plane, the formula for change of variables in a double integral is given by the **change of variables theorem**

$$\int\int_D f(x,y)\,dx\,dy = \int\int_R f(x(u,v), y(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| du\,dv. \tag{4.16}$$

The Jacobian factor is always taken in absolute value.

4.6.2 Example Use the transformation $T :\ x = u^2 - v^2, y = 2uv$ to find $\int\int_D y\,dx\,dy$, where D is the image under T of the rectangle $R = [0,1] \times [0,1]$.
Solution
the level curves are mutually orthogonal and thus they make a good coordinate grid. The curves that bound the region D are the images of the edges of a unit square, $\{u = 0,\ u = 1,\ v = 0,\ v = 1\}$. The equations of the curves are,

$$\begin{array}{llll} C_1 :\ \mathbf{r}(v) = \langle -v^2, 0 \rangle, & \text{Line}, & y = 0, & x \in [-1,0], \\ C_2 :\ \mathbf{r}(v) = \langle 1 - v^2, 2v \rangle, & \text{Parabola}, & x = 1 - \frac{1}{4}y^2, & x \in [0,1], \\ C_3 :\ \mathbf{r}(u) = \langle u^2, 0 \rangle, & \text{Line}, & y = 0, & x \in [0,1], \\ C_4 :\ \mathbf{r}(v) = \langle u^2 - 1, 2u \rangle, & \text{Parabola}, & x = \frac{1}{4}y^2 - 1, & x \in [-1,0]. \end{array}$$

For the Jacobian, we compute the differentials,

$$\begin{aligned} dx &= 2u\,du - 2v\,dv, \\ dy &= 2v\,du + 2u\,dv, \\ dA = dx \wedge dy &= \begin{vmatrix} 2u & -2v \\ 2v & 2u \end{vmatrix} du \wedge dv, \\ &= 4(u^2+v^2) du \wedge dv. \end{aligned}$$

Finally, the pull-back integral is

$$\begin{aligned} I &= \int_0^1 \int_0^1 2uv[4(u^2+v^2)]\,dv\,du, \\ &= \int_0^1 \int_0^1 8(u^3 v + uv^3)\,dv\,du, \\ &= 8\int_0^1 \tfrac{1}{2}u^3 + \tfrac{1}{4}u\,du, \\ &= 8(\tfrac{1}{8} + \tfrac{1}{8}) = 2 \end{aligned}$$

We should note that the problem is kind of a cheat because we specified the pre-image of D to be a rectangle and we gave away the appropriate transformation. In practice one knows neither. Furthermore, in this case we could have easily done the integral directly if D had been specified. Still, there is much to be gained by the details of what it takes to map a region under a coordinate transformation.

Obtaining the general formula for the change of coordinates of the differential of volume for a transformation $T : \mathbf{R}^3 \to \mathbf{R}^3$ is just as easy. Suppose the map takes coordinates (u, v, w) to (x, y, z). We have

$$\begin{aligned} dx &= x_u\,du + x_v\,dv + x_w\,dw \\ dy &= y_u\,du + y_v\,dv + y_w\,dw \\ dz &= z_u\,du + z_v\,dv + z_w\,dw \\ dV &= dx \wedge dy \wedge dz \\ dV &= \begin{vmatrix} x_u & x_v & x_w \\ y_u & y_v & y_w \\ z_u & z_v & z_w \end{vmatrix} du \wedge dv \wedge dw. \end{aligned} \tag{4.17}$$

4.6.3 Example Volume of Ellipsoid. We wish to compute the volume of the ellipsoid

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

Solution. We perform the following transformation

$$\begin{aligned} x &= aX, \quad & dx &= a\,dX, \\ y &= b\,Y, \quad & dy &= b\,dY, \\ z &= c\,Z, \quad & dz &= c\,dZ. \end{aligned}$$

The transformation maps the unit sphere $S: \; X^2+Y^2+Z^2=1$ into the ellipsoid. The pullback of the differential of volume is

$$dV = dx \wedge dy \wedge dz = abc\, dX \wedge dY \wedge dZ.$$

Thus, the change of variable theorem gives,

$$V = \iiint_S abc\, dX\, dY\, dZ = abc\tfrac{4}{3}\pi(1)^3 = \tfrac{4}{3}abc. \tag{4.18}$$

4.6.4 Example Spherical Coordinates
We compute the differentials and wedge product.

$$\begin{aligned} x &= r\sin\theta\cos\phi & & dx = \sin\theta\cos\phi\, dr + r\cos\theta\cos\phi\, d\theta - r\sin\theta\sin\phi\, d\phi. \\ y &= r\sin\theta\sin\phi & \text{and} \quad & dy = \sin\theta\sin\phi\, dr + r\cos\theta\sin\phi\, d\theta + r\sin\theta\cos\phi\, d\phi. \\ z &= r\cos\theta & & dz = \cos\theta\, dr - r\sin\theta\, d\theta \end{aligned}$$

$$\begin{aligned} dV &= dx \wedge dy \wedge dz, \\ &= \begin{vmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{vmatrix} dr \wedge d\theta \wedge d\phi, \\ J &= \cos\theta \begin{vmatrix} r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ r\cos\theta\sin\phi & r\sin\theta\cos\phi \end{vmatrix} + r\sin\theta \begin{vmatrix} \sin\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\sin\theta\cos\phi \end{vmatrix}, \\ &= \cos\theta(r^2)(\cos\theta\sin\theta) + r^2\sin\theta\sin^2\theta, \\ |J| &= r^2\sin\theta. \end{aligned}$$

So the Jacobian of the transformation is consistent with the differential of volume we obtained with far less effort by geometry,

$$dV = r^2\sin\theta\, dr\, d\theta\, d\phi. \tag{4.19}$$

Perhaps it is appropriate to end this subsection by the tantalizing quote by the famous professor at Oxford, Sir Michael Atiyah

> Algebra is the offer made by the devil to the mathematician. The devil says: "I will give you this powerful machine, and it will answer any question you like. All you need to do is give me your soul: give up geometry and you will have this marvellous machine." ... the danger to our soul is there, because when you pass over into algebraic calculation, essentially you stop thinking: you stop thinking geometrically, you stop thinking about the meaning.
>
> Sir Michael Atiyah

4.7 Surface Area

The differential of surface area on $\mathbf{R}^3$ is a straightforward generalization of the differential of area in the xy-plane given by 4.13. As shown in figure 4.8, let us consider infinitesimal arc length vectors $\mathbf{i}\,dx$, $\mathbf{j}\,dy$ and $\mathbf{k}\,dz$ pointing along the coordinate axes. Recall from the definition, that the cross product of two vectors is a new vector whose magnitude is the area of the parallelogram subtended by the two vectors and which points in the direction of a unit vector perpendicular to the plane containing the two vectors, oriented according to the right hand rule. As discussed previously, since $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are mutually orthogonal vectors,

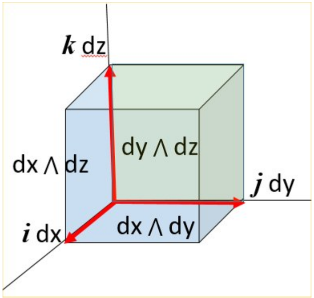


Fig. 4.8: Area Forms

the cross product of any pair is again a unit vector pointed in the direction of the third or the negative thereof. In the xy-plane the differential of area is really an oriented quantity that can computed by the cross product $d\mathbf{S}_{xy} = \mathbf{i}\,dx \times \mathbf{j}\,dy = dx\,dy\,\mathbf{k}$). A similar computation yields the differential of areas in the other two coordinate planes, except that in the xz-plane, the cross product needs to be taken in the reverse order, so that ($d\mathbf{S}_{xz} = -dx\,dz\,\mathbf{j}$). In terms of wedge products, the differential of area in the xy-plane is ($dS_{xy} = dx \wedge dy$), so that the oriented nature of the surface element is built-in. Technically, when reversing the order of variables in a double integral one should introduce a minus sign. This is typically ignored in basic calculus computations of double and triple integrals, but it cannot be ignored in vector calculus in the context of flux of a vector field through a surface. We define the differential of surface area in $\mathbf{R}^3$ as

$$d\mathbf{S} = dy\,dz\,\mathbf{i} - dx\,dz\,\mathbf{j} + dx\,dy\,\mathbf{k}, \tag{4.20}$$

keeping in mind that this is an oriented element of area and the products of differentials are really wedge products. In other words,

$$d\mathbf{S} = dy \wedge dz - dx \wedge dz + dx \wedge dy, \tag{4.21}$$

An equivalent equation for the differential of area for surface in parametric form as in 2.45 can be obtained as follows. If in the equation for the unit normal 2.46, instead of the tangent vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ we take infinitesimal vectors arc length vectors $\mathbf{r}_u\,du$ and $\mathbf{r}_v\,dv$, then the cross product gives

$$d\mathbf{S} = (\mathbf{r}_u \times \mathbf{r}_v)\,du\,dv. \tag{4.22}$$

Of course, we need to verify the assertion that 4.20 and 4.22 are equivalent. Recalling the formula for the Jacobian 4.15 we have

$$\begin{aligned} d\mathbf{S} &= dy\,dz\,\mathbf{i} - dx\,dz\,\mathbf{j} + dx\,dy\,\mathbf{k}, \\ &= \left\{ \begin{vmatrix} y_u & y_v \\ z_u & z_v \end{vmatrix} \mathbf{i} - \begin{vmatrix} x_u & x_v \\ z_u & z_v \end{vmatrix} \mathbf{j} + \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \mathbf{k} \right\} du\,dv \end{aligned}$$

On the other hand

$$\begin{aligned} d\mathbf{S} &= (\mathbf{r}_u \times \mathbf{r}_v)\,du\,dv, \\ &= \left\{ \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} \right\} du\,dv, \\ &= \left\{ \begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix} \mathbf{i} - \begin{vmatrix} x_u & z_u \\ x_v & z_v \end{vmatrix} \mathbf{j} + \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix} \mathbf{k} \right\} du\,dv \end{aligned}$$

Emulating the one-dimensional case for the arc length element $ds^2 = \|d\mathbf{r}\|^2$, we define the scalar differential of surface area by

$$dS = \|d\mathbf{S}\| = \|\mathbf{r}_u \times \mathbf{r}_v\|\,du\,dv. \tag{4.23}$$

The **area of a surface** in $\mathbf{R}^3$ is then given by

$$S = \iint_D dS = \iint_D \|\mathbf{r}_u \times \mathbf{r}_v\|\,du\,dv, \tag{4.24}$$

where D is an appropriate region of integration. I find it more natural to use 4.20 in computation, but some might prefer 4.22. A neat formula for surface area can be obtained using Lagrange's identity 1.22. We have

$$\begin{aligned} \|\mathbf{r}_u \times \mathbf{r}_v\|^2 &= \begin{vmatrix} \mathbf{r}_u \cdot \mathbf{r}_u & \mathbf{r}_u \cdot \mathbf{r}_v \\ \mathbf{r}_v \cdot \mathbf{r}_u & \mathbf{r}_v \cdot \mathbf{r}_v \end{vmatrix}, \\ &= \begin{vmatrix} E & F \\ F & G \end{vmatrix} \\ &= EF - G^2, \end{aligned}$$

Where $E = \mathbf{r}_u \cdot \mathbf{r}_u$, $F = \mathbf{r}_u \cdot \mathbf{r}_v$,and $G = \mathbf{r}_v \cdot \mathbf{r}_v$. Therefore, we have

$$S = \iint_D \sqrt{EG - F^2}\,du\,dv, \tag{4.25}$$

It is much easier to compute dot products than length of cross products. In particular, if the surface grid curves form an orthogonal set, then $F = 0$. The geometric reason is that in this case $\mathbf{r}_u\,du$ and $\mathbf{r}_v\,dv$ are infinitesimal vectors along the two edges of an infinitesimal rectangle, and the area of the rectangle is the product of the base times the height.

While we are at it, we can also get another neat formula for the element of arc length on a surface. We have

$$\begin{aligned} d\mathbf{r} &= \mathbf{r}_u\,du + \mathbf{r}_v\,dv, \\ ds^2 &= d\mathbf{r}\cdot d\mathbf{r}, \\ &= (\mathbf{r}_u\,du + \mathbf{r}_v\,dv)\cdot(\mathbf{r}_u\,du + \mathbf{r}_v\,dv). \end{aligned}$$

After distributing the multiplication, we get,

$$ds^2 = E\,du^2 + 2F\,du\,dv + G\,dv^2. \tag{4.26}$$

This is called the **metric** for parametric surface since the integral of ds along a curve on the surface would yield the length of the curve. If the coordinate grid curves are an orthogonal set, then $F = 0$, and the element of arc length is the square of the hypothenuse of an infinitesimal right triangle,

$$ds^2 = E\,du^2 + G\,dv^2.$$

4.7.1 Example Surface Area for Explicit Function $z = f(x, y)$
We demonstrate that we get the same answer by computing dS for a general explicit function $z = f(x, y)$ using either, equation 4.20 or equation 4.22

$$\begin{aligned} dz &= f_x\,dx + f_y\,dy, \\ dy\wedge dz &= -f_x\,dx\wedge dy, \quad \text{after reversing the wedge order,} \\ -dx\wedge dz &= -f_y\,dx\wedge dy, \\ dx\wedge dy &= \quad\; dx\wedge dy. \end{aligned}$$

Hence, for the surface $z = f(x, y)$

$$\begin{aligned} d\mathbf{S} &= -(f_x\,\mathbf{i} + f_y\,\mathbf{j} - 1\,\mathbf{k})\,dx\,dy, \\ dS &= \sqrt{f_x^2 + f_y^2 + 1}\,dx\,dy, \\ S &= \iint_D \sqrt{f_x^2 + f_y^2 + 1}\,dx\,dy, \end{aligned} \tag{4.27}$$

And just like that, we have a "plug-and-chug" formula to compute surface area for general explicit surfaces. To use the second formula we must first parametrize the surface. We use the trivial parametrization. We have

$$\begin{aligned} \mathbf{r}(x,y) &= x\,\mathbf{i} + y\,\mathbf{j} + f(x,y)\,\mathbf{k}, \\ (\mathbf{r}_x\times\mathbf{r}_y)\,dx\,dy, &= \left\{\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix}\right\}\,dx\,dy, \\ &= -(f_x\,\mathbf{i} + f_y\,\mathbf{j} - 1\,\mathbf{k})\,dx\,dy, \\ dS &= \sqrt{f_x^2 + f_y^2 + 1}\,dx\,dy. \end{aligned}$$

Before doing some examples, we can't resist pointing out that the vector component of $d\mathbf{S}$ is the 3-dimensional gradient of $z = f(x, y)$. We should not be surprised, since $d\mathbf{S}$ must be orthogonal to the surface.

4.7.2 Example Surface Area of Saddle
Find the surface area of the "pringle" $z = x^2 - y^2$, in the region D bounded by $x^2 + y^2 = 4$.
Solution. Applying formula 4.27, we get

$$\begin{aligned} S &= \iint_D \sqrt{(2x)^2 + (-2y)^2 + 1}\, dx\, dy, \\ &= \iint_D \sqrt{4(x^2 + y^2) + 1}\, dx\, dy, \\ &= \int_0^{2\pi} \int_0^2 r\sqrt{4r^2 + 1}\, dr\, d\theta, \\ &= (2\pi)(\tfrac{2}{3})(\tfrac{1}{8})\ (4r^2 + 1)^{3/2}\Big|_0^2, \\ &= \frac{\pi}{6}(17\sqrt{17} - 1). \end{aligned}$$

4.7.3 Example Surface Area of Paraboloid
Find the area of the part of the paraboloid $z = x^2 + y^2$, in the region D bounded by $x^2 + y^2 = 4$.
Solution. I find it better to use the original definition(s) for all problems. To illustrate, we start by parametrizing the surface. We get

$$\begin{aligned} \mathbf{r}(x, y) &= x\,\mathbf{i} + y\,\mathbf{j} + (x^2 + y^2)\,\mathbf{k}, \\ (\mathbf{r}_x \times \mathbf{r}_y)\, dx\, dy, &= \left\{ \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 2x \\ 0 & 1 & 2y \end{vmatrix} \right\} dx\, dy, \\ &= -(2x\,\mathbf{i} + 2y\,\mathbf{j} - 1\,\mathbf{k})\, dx\, dy, \\ dS &= \sqrt{4x^2 + 4y^2 + 1}\, dx\, dy, \\ &= \int_0^{2\pi} \int_0^2 r\sqrt{4r^2 + 1}\, dr\, d\theta, \\ &= (2\pi)(\tfrac{2}{3})(\tfrac{1}{8})\ (4r^2 + 1)^{3/2}\Big|_0^2, \\ &= \frac{\pi}{6}(17\sqrt{17} - 1). \end{aligned}$$

Surprise. We get the same answer. My silly explanation; one gets a pringle by starting with a thin disk of dough and bending two opposite points up and the other two down. One gets a paraboloid by bending all in the same direction. Same amount of dough in both cases.

4.7.4 Example Surface Area of Implicit Function

Let $F(x, y.z) = c$ be a level curve defining an implicit function of (x, y)

$$\begin{aligned}
dF &= F_x\,dx + F_y\,dy + F_z\,dz = 0,\\
dz &= -\frac{F_x}{F_z}\,dx - \frac{F_y}{F_z}\,dy, \quad F_z \neq 0,\\
dy \wedge dz &= \frac{F_x}{F_z}\,dx \wedge dy,\\
-dx \wedge dz &= \frac{F_y}{F_z}\,dx \wedge dy,\\
dx \wedge dy = dx \wedge dy &= \frac{F_z}{F_z}\,dx \wedge dy,\\
d\mathbf{S} &= \frac{F_x\,\mathbf{i} + F_y\,\mathbf{i} + F_z\,\mathbf{k}}{F_z}\,dx\,dy, \qquad F_z \neq 0.
\end{aligned}$$

Hence.

$$\begin{aligned}
d\mathbf{S} &= \frac{\nabla F}{|\nabla F \cdot \mathbf{k}|}\,dx\,dy,\\
dS &= \frac{\|\nabla F\|}{|\nabla F \cdot \mathbf{k}|}\,dx\,dy, \qquad \nabla F \cdot \mathbf{k} \neq 0. \qquad (4.28)
\end{aligned}$$

The formula looks rather mysterious, but it is rather simple to explain by geometry. Let R be the area of the projection onto the xy-plane of the area S of a tilted parallelogram with unit normal $\mathbf{n}$ in $\mathbf{R}^3$. Then $S = R/\|\mathbf{n} \cdot \mathbf{k}\|$, since $\|\mathbf{n} \cdot \mathbf{k}\|$ is just the cosine of the angle between the two planes. The same must be true at the infinitesimal level. If $F(x, y, z) = c$, the unit normal is $\mathbf{n} = \nabla F/\|\nabla F\|$. Hence

$$dS = \frac{dA}{\left|\frac{\nabla F}{\|\nabla F\|} \cdot \mathbf{k}\right|} = \frac{\|\nabla F\|}{|\nabla F \cdot \mathbf{k}|}\,dA, \quad \text{where } dA = dx\,dy$$

The full area S is then the given by summing the infinitesimal pieces $S = \iint dS$. The reader should take a minute to get convinced that if one writes an explicit surface $z = f(x, y)$ as the level surface $F(x, y, z) = f(x, y) - z = 0$, the two formulas for surface area are compatible.

4.7.5 Example Area of Surface of Revolution

Let S be a surface of revolution as in equation 2.50 with generator profile $z = f(r)$. We have

$$\begin{aligned}
\mathbf{r}(r, \theta) &= \langle r\cos\theta, r\sin\theta, f(r)\rangle,\\
\mathbf{r}_r &= \langle \cos\theta, \sin\theta, f'(r)\rangle,\\
\mathbf{r}_\theta &= \langle -r\sin\theta, r\cos\theta, 0\rangle,\\
E &= 1 + [f'(r)]^2,\\
F &= 0,\\
G &= r^2.
\end{aligned}$$

Since $F = 0$ it must be the case that the mesh curves are orthogonal. Indeed, the constant coordinate curves are the meridians and the parallels of the surface of revolution. The formula for the surface area is

$$S = \iint_S r\sqrt{1 + [f'(r)]^2}\, dr\, d\theta. \tag{4.29}$$

If one prefers the equivalent general formula 4.24 for surface area, the details are as follows,

$$\begin{aligned}
\mathbf{r}_r \times \mathbf{r}_\theta &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \cos\theta & \sin\theta & f'(r) \\ -r\sin\theta & r\cos\theta & 0 \end{vmatrix} \\
&= -rf'(r)\cos\theta\,\mathbf{i} - rf'(r)\cos\theta\,\mathbf{j} + r\,\mathbf{k}, \\
\|\mathbf{r}_r \times \mathbf{r}_\theta\| &= \sqrt{r^2[f'(r)]^2 + r^2}, \\
&= r\sqrt{1 + [f'(r)]^2}, \\
dS &= r\sqrt{1 + [f'(r)]^2}\, dr\, d\theta.
\end{aligned}$$

It is worthwhile pointing out that if surface is generated by revolving the curve $z = f(x)$ about the z-axis, the formula reads

$$\begin{aligned}
S &= \int_0^{2\pi} \int r\sqrt{1 + [f'(r)]^2}\, dr\, d\theta, \\
&= 2\pi \int r\sqrt{1 + [f'(r)]^2}\, dr, \\
&= 2\pi \int x\sqrt{1 + [f'(x)]^2}\, dx, \quad \text{relabelling the integration variable} \\
&= 2\pi \int x\, ds, \qquad (4.30)
\end{aligned}$$

where ds is the differential of arc length. The is the formula used in single variable calculus to compute the area of a surface of revolution.

4.7.6 Example <u>Surface Area of Cone</u>

Determine the surface the lateral surface area of a circular cone with base of radius R and height h.

Adapt the coordinate axes such that axis of symmetry is the z-axis and the and the vertex is at the origin. The generator of the cone on the xz-plane is a straight line with equation $z = -\frac{h}{R}x$. The cone of revolution is then given by

$$\mathbf{r}(r, \theta) = \langle r\cos\theta, r\sin\theta, -\tfrac{h}{R}r\rangle$$

By equation 4.29, the surface area integral is

$$
\begin{aligned}
S &= \int_0^{2\pi}\int_0^R r\sqrt{1+\frac{h^2}{R^2}}\,dr\,d\theta,\\
&= 2\pi\tfrac{1}{2}R^2\sqrt{1+\frac{h^2}{R^2}},\\
&= \pi R\sqrt{R^2+h^2},\\
&= \pi RL,
\end{aligned}
$$

where $L = \sqrt{R^2 + h^2}$ is the length of the generator of the cone.

4.7.7 Example Surface area of Torus

Consider the torus of revolution given by equation 2.52

$$\mathbf{r}(\theta,\phi) = \langle\ (R + r\cos\theta)\cos\phi, (R + r\cos\theta)\sin\phi, r\sin\theta\ \rangle$$

We have,

$$
\begin{aligned}
\mathbf{r}_\theta &= \langle -r\sin\theta\cos\phi, -r\sin\theta\sin\phi, r\cos\theta\rangle,\\
\mathbf{r}_\phi &= \langle -(R + r\cos\theta)\sin\phi, (R + r\cos\theta)\cos\phi, 0\rangle,\\
E &= r^2,\\
G &= (R + r\cos\theta)^2, \quad F = 0,\\
S &= \int_0^{2\pi}\int_0^{2\pi} r(R + r\cos\theta)\,d\theta\,d\phi,\\
&= (2\pi)(2\pi)rR, \quad \text{(the integral of the cosine term is zero)},\\
S &= 4\pi^2 rR
\end{aligned}
$$

4.7.8 First Theorem of Pappus

The surprisingly simple answers above could have been obtained by the first **theorem of Pappus**; that states that the surface area of a surface of revolution generated by a curve rotated about an axis is, the product of the length of the circle traced by the center of mass, times the length of the curve. In the case of the torus,

$$S = (2\pi R)(2\pi r).$$

For the cone of base R and generator length L, the center of mass is in the middle of the segment L. The distance of the center of mass to the z-axis is $R/2$m so the surface area of the cone is

$$S = 2\pi(R/2)L = \pi RL.$$

4.7.9 Example Surface Area of Sphere I

In this computation we make repeated mental use of the trigonometric version of Pythagoras theorem encapsulated in the equation $\cos^2 x + \sin^2 x = 1$.

$$\begin{aligned}
\mathbf{r}(\theta,\phi) &= \langle R\sin\theta\cos\phi, R\sin\theta\sin\phi, R\cos\theta\rangle,\\
\mathbf{r}_\theta &= \langle R\cos\theta\cos\phi, R\cos\theta\sin\phi, -R\sin\theta\rangle,\\
\mathbf{r}_\phi &= \langle -R\sin\theta\sin\phi, R\sin\theta\cos\phi, 0\rangle,
\end{aligned}$$

$$\begin{aligned}
E &= \mathbf{r}_\theta\cdot\mathbf{r}_\theta = R^2,\\
G &= \mathbf{r}_\phi\cdot\mathbf{r}_\phi = R^2\sin^2\theta,\\
F &= \mathbf{r}_\theta\cdot\mathbf{r}_\phi = 0,\\
\sqrt{EG-F^2} &= \sqrt{R^2(R^2\sin^2\theta)} = R^2\sin\theta
\end{aligned}$$

Therefore,

$$\begin{aligned}
S &= \int_0^{2\pi}\int_0^{\pi} R^2\sin\theta\,d\theta\,d\phi,\\
&= -2\pi R^2\left|\cos\phi\right|_0^{\pi},\\
&= 4\pi R^2.
\end{aligned}$$

It is remarkable that this formula was also obtained by Archimedes. For a unit sphere the surface area is 4π. Comparing, a circle of radius one has arc length $s = 2\pi$, so we have a natural measure for arcs and angles, namely, there are 2π radians in a circle, or 2π radians in 360^o. In the same manner we have a natural measure for solid angles. There are 4π "somethings" on a sphere. The "something" is called a **steradian**. One steradian corresponds to a central cone subtending an spherical cap of area 1 on a sphere of radius 1. Here is a cute question, how big does the moon appear to the naked eye. Is it the "size" of a half- dollar, a quarter, a dime, a round pin head? If you have not done this before, try putting a circular obstacle a meter away just big enough to cover the moon. You might be surprised and perhaps better appreciate the job required to scan the celestial sphere with powerful, finely-focused optical telescopes.

4.7.10 Example Surface Area of Sphere II

Consider a sphere $x^2 + y^2 + z^2 = R^2$ of radius R. In spherical coordinates, the equation is $r = R$. Following Atiyah's philosophy, we use simple geometry instead of convoluted algebra to find dS. All we have to do is look at the differential of volume 4.11 and set $r = R$ and integrate. We get

$$\begin{aligned}
dS &= R^2\sin\theta\,d\theta\,d\phi,\\
S &= \int_0^{2\pi}\int_0^{\pi} R^2\sin\theta\,d\theta\,d\phi,\\
&= 4\pi R^2.
\end{aligned}$$

Now, this is neat.

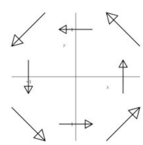
Chapter 5

Integral Vector Calculus

5.1 Vector Fields

A vector field is a smooth assignment of a vector to each point in space. We have a very good intuition on how vector fields are represented graphically because we are bombarded daily with meteorological diagrams of wind patterns, ocean currents, and propagation of weather fronts. Plain vectors in $\mathbf{R}^n$ are represented by n-tuples of numbers; vector fields are represented by n-tuples of smooth functions. A good way to interpret vector fields is that a each point, the corresponding vector represents the velocity vector of a fluid, the force vector of some gravitational or electromagnetic field or some other physical phenomenon that exhibits vector-like qualities.

We begin with a very easy example. Consider the vector field

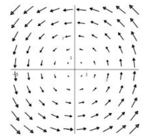


$$\mathbf{F} = \langle -y, x \rangle$$

If we wanted to render this vector field by hand we would prescribe a grid on the xy-plane, make a table of the coordinates of each point in the grid, and for each point, compute the the components of the vector to be drawn at that point. Here we choose the grid with integer coordinates on $[-1, 1] \times [-1, 1]$. We compute

x	y	$\langle -y, x \rangle$
1	0	$\langle 0, 1 \rangle$
1	1	$\langle -1, 1 \rangle$
0	1	$\langle -1, 0 \rangle$
-1	1	$\langle -1, -1 \rangle$
-1	0	$\langle 0, -1 \rangle$

I already got tired of doing this and I have only computed the vectors in the top half plane. What we need is to render the vector field on a sizeable grid like 10×10 or larger if needed. This is yet another task nowadays better done by computers.



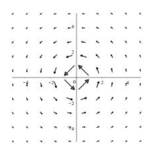
Here is the vector field rendered in Maple by the "fieldplot" command on a $[-5,5] \times [-5,5]$ grid. Even for this small size window, one would need draw 100 vectors by hand. This would be what we call a **rotational vector field**. In this case the rotation is counterclockwise as it is apparent by visual inspection. Maple automatically scales the relative length of the arrow to fit the longest ones on the grid. This would not represent a typical vector field in physics as the vectors get larger as one recedes from the origin. True vector fields in physics we observe locally, usually tend to 0 in magnitude far away from the origin. This is the case for gravitational and electromagnetic fields. There are always exceptions. In the modern view of the universe, galaxies tend to recede faster, the farther away from us, so the velocity vectors would indeed get larger.

The notion of vector fields is not new in this course. Given a function $f(x,y)$, we have already encountered the gradient vector field ∇f. If one can render the level curves, as in the case of the function $z = x^2 - y^2$, we can easily visualize the vector field since at each point, the gradient vector is orthogonal to the level curve. The gradient vector field in the case in question is plotted in figure 3.10. Here is another neat vector field. Consider the harmonic function

$$f(x,y) = \tan^{-1}(y/x)$$

which corresponds to the imaginary part of the of the main branch of the complex function $w = \ln z$, as discussed in subsection 3.3.2. The gradient of this function is

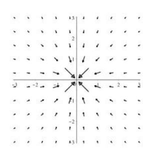
$$\nabla f = \langle \frac{-y}{x^2+y^2}, \frac{x}{x^2+y^2} \rangle = \frac{\langle -y, x \rangle}{(x^2+y^2)}.$$



The gradient field of course has the same rotation feature as the vector $\langle -y, x \rangle$, the difference being that this one is normalized by the square of the length. The gradient vectors are orthogonal to the gradient vectors of the conjugate harmonic function $g(x,y) = \ln\sqrt{x^2+y^2}$.
In physics, the vector field $\mathbf{F}$ associated with a potential function is given by the negative of the gradient of that function. A short computation gives

$$\phi = \ln(1/r) = -\ln\sqrt{x^2+y^2},$$

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



The vector field points in the radial direction $\mathbf{r} = \langle x, y \rangle$ from the origin. As $r \to 0$ the magnitude of the vector field goes to infinity, so the field is singular at the origin. The function $\phi = \ln(1/r)$, up to a constant, represents the potential of a (long) uniformly charged wire. The vector field is an example of one with either a **source** or a **sink**. In section 5.4 we write simple conditions to determine whether or not a vector field is rotational, and whether or not a vector field has a source or a sink.

Talking about vector fields with a source or a sink, the most important of all in $\mathbf{R}^3$ is $\mathbf{F} = \nabla(1/r)$. The potential $\phi = 1/r$, up to a constant, yields the inverse square law, as shown in equation 3.6.

5.2 Line Integrals

As a snake biting its tail, we finally come back to the topic of work and line integrals that was introduced in equation 1.13. Let C be a curve whose position vector is given as in equation 2.1

$$\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}.$$

There are two versions of the differential of arc length; a vector form and a scalar form, given by

$$d\mathbf{r} = dx\,\mathbf{i} + dy\,\mathbf{j} + dz\,\mathbf{k}, \qquad \text{Vector form,} \tag{5.1}$$

$$ds^2 = \|d\mathbf{r}\|^2 = dx^2 + dy^2 + dz^2, \quad \text{Scalar form.} \tag{5.2}$$

The vector form $d\mathbf{r}$ is an infinitesimal vector tangential to the curve. The unit tangent vector $\mathbf{T}$ given by equation 2.23

$$\mathbf{T} = \frac{d\mathbf{r}}{ds} \quad \text{where} \quad \frac{ds}{dt} = v.$$

Let $C:\ \mathbf{r}(t),\ a \leq t \leq b$ be a segment of a curve, and

$$\begin{aligned} w &= f(x,y,z) && \text{be a scalar field, and} \\ \mathbf{F} &= P\,\mathbf{i} + QP\,\mathbf{j} + R\,\mathbf{k} && \text{be a vector field,} \end{aligned}$$

where P, Q and R are functions of (x,y,z). We have two types of line integrals, scalar and vector types,

$$\begin{aligned} I &= \int_C f(x,y,z)\,ds, && \text{Scalar type ,} \\ W &= \int_C \mathbf{F}\cdot d\mathbf{r}, && \text{Vector type.} \end{aligned}$$

In both cases, the integrals can be computed directly by parametrizing the curve, computing the appropriate differentials, and then performing the one-variable integral with respect to t. we should note that in most physics textbooks the vector differential of arc length is denoted by $d\,\mathbf{l}$ as in the case of the formula for the voltage drop $\triangle V = -\int_C \mathbf{E}\cdot d\,\mathbf{l}$ over a path in an electric field.

In some sense, we have already encountered line integrals of type I when we computed the length of a curve, which is the special case in which $f(x,y,z) = 1$, or in computing surface area by the single variable formula 4.30. Among the few other physical applications of scalar-type line integrals, we have moment integrals for density distributions along a curve. That is, we may think of $f(x,y,z)$ as some sort of density function. We provide but a single example in $\mathbf{R}^2$ to illustrate that is just a "plug-and-chug" procedure. This is a good time to review how to parametrize basic curves as we learned in chapter 2.

5.2.1 Example Moment of inertia of ring

Let C be a mass of constant density $\rho = k$ in the shape of a circle of radius R centered at the origin. Compute the polar moment of inertia I_z.

Solution. We have,

$$\begin{aligned}\mathbf{r}(t) &= \langle R\cos t, R\sin t\rangle, \quad 0 \le t \le 2\pi, \\ \mathbf{v}(t) &= \langle -R\sin t, R\cos t\rangle, \\ v &= \|\mathbf{v}\| = R, \\ ds &= v\,dt = R\,dt\end{aligned}$$

Next, we recall that $I_z = \int (x^2+y^2)\,dm$, with $dm = \rho\,ds$

$$\begin{aligned}I_z &= \int_C R^2 k\,ds, \\ &= \int_0^{2\pi} R^2(kR\,dt), \\ &= 2\pi k R^3 = \pi k R^3\left[\frac{M}{k\pi R^2}\right], \\ &= M\,R^2.\end{aligned}$$

Possible difficulty in doing scalar-type line integrals most likely can be traced back to insufficient practice parametrizing curves, computing the differential of arc length, or calculating a one variable integral. Again, good time to review that material. If a curve C is piecewise smooth, that is, a union of smooth curves $\{C_1, C_2, \ldots\}$ which connect one to the next at a point, then, symbolically

$$\int_C = \int_{C_1} + \int_{C_2} + \ldots.$$

Basically, if the curve C has n different smooth pieces, then, effectively, one has n line integrals to perform.

Line Integrals of type $W = \int_C \mathbf{F}\cdot d\mathbf{r}$ represent work done by a force field along the curve. At the infinitesimal level $dW = \mathbf{F}\cdot d\mathbf{r}$ is the differential of work done by the force in moving a particle an infinitesimal distance along the curve. The formal definition of the computation is

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

The definition of the integral might look intimidating at first, but all it says is "plug-in" the curve. If the vector field has components $\mathbf{F} = \langle P, Q, R\rangle$, we can expand the dot product

$$W = \int_C \mathbf{F}\cdot d\mathbf{r} = \int_C P\,dx + Q\,dy + R\,dz.$$

The quantity α inside the integrand

$$\alpha = \mathbf{F}\cdot d\mathbf{r} = P\,dx + Q\,dy + R\,dz, \tag{5.3}$$

is called a **differential one-form**.

If the curve is parametrized by arc length $\mathbf{r}(s)$, one can write the integral as

$$W = \int_C \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(s)) \cdot \frac{d\mathbf{r}}{ds}\, ds = \int_C \mathbf{F} \cdot \mathbf{T}\, ds.$$

This alternative formulation offers absolutely no computational advantage, but it brings out two important points. First, it shows that there is really only kind of line integral type, since $\mathbf{F} \cdot \mathbf{T}$ is just a scalar field (a function). Secondly, the properties of a dot product with a unit vector, show that it is only the projection component of the force along the tangent vector that has any contribution to pushing the particle along the curve. It is important to note that the curves must be treated as oriented curves, with natural (ccw) orientation along increasing parameter t.

5.2.2 Example Compute $W = \int_C \mathbf{F} \cdot d\mathbf{r}$, where $\mathbf{F} = x\,\mathbf{i} + y\,\mathbf{j} + xy\,\mathbf{k}$ and C is the curve $\mathbf{r}(t) = \cos t\,\mathbf{i} + \sin t\,\mathbf{j} + t\,\mathbf{k}$, with $t \in [0, \pi]$.
Solution. Not that it matters in the computation, but we observe that the curve is a helix. We have

$$\begin{aligned} x &= \cos t & & & dx &= -\sin t\, dt \\ y &= \sin t & \quad \text{and} \quad & & dy &= \cos t\, dt, \\ z &= t & & & dz &= dt, \end{aligned}$$

Substitution into the line integral gives

$$\begin{aligned} W &= \int_C \mathbf{F} \cdot d\mathbf{r} = \int_C x\, dx + y\, dy + xy\, dz, \\ &= \int_0^\pi [\cos t(-\sin t) + \sin t(\cos t) + \cos t \sin t]\, dt, \\ &= \int_0^\pi \cos t \sin t\, dt, \\ &= \tfrac{1}{2} \sin^2 t \Big|_0^\pi, \\ &= 0 \end{aligned}$$

5.2.3 Example Compute $W = \int_C xy\, dx + (x - y)\, dy$ where C consists of the line segment joining $(0, 0)$ to $(2, 0)$ followed by the line segment from $(2, 0)$ to $(3, 2)$
Solution. we call the line segments C_1 and C_2. We start by writing parametric equations for the line segments.

$$\begin{aligned} C_1 :\ \mathbf{r}(t) &= \langle 0, 0\rangle + t\langle 2, 0\rangle, \quad 0 \le t \le 1, \\ C_2 :\ \mathbf{r}(t) &= \langle 2, 0\rangle + t\langle 1, 2\rangle, \quad 0 \le t \le 1, \end{aligned}$$

In coordinates, we have

$$\begin{array}{llll} C_1 : & & C_2 : & \\ x = 2t \text{ and} & dx = 2\, dt & x = 2 + t, \text{ and} & dx = dt, \\ y = 0 & dy = 0, & y = 2t, & dy = 2\, dt. \end{array}$$

In the integral over C_1 the first term dies because $y = 0$ and the second because $dy = 0$. The line integral is,

$$\begin{aligned} W &= \int_{C_2} xy\,dx + (x-y)\,dy \\ &= \int_0^1 [(2+t)(2t) + ((2+t)-2t)(2)]\,dt, \\ &= \int_0^1 [4t + 2t^2 + 4 - 2t]\,dt = \int_0^1 [4 + 2t + 2t^2]\,dt, \\ &= 4 + 1 + \tfrac{2}{3} = \tfrac{17}{3} \end{aligned}$$

5.3 Conservative Vector Fields

5.3.1 Definition A vector field $\mathbf{F}$ is called **conservative** if there exists a function f such that $\mathbf{F} = \nabla f$. We will call f a potential. This differs slightly from the use of the term in physics where, if ϕ is a scalar potential, then the field is given by $\mathbf{F} = -\nabla\phi$. I know it is repetitious, but the most well-know example is the scalar potential $\phi = 1/r$ that gives rise to the $1/r^2$ law.

Line integrals of conservative vector fields have very special properties. Let $\mathbf{r}(t)$ be smooth or piecewise smooth curve with end points $A = \mathbf{r}(a)$ and $B = \mathbf{r}(b)$. Suppose that

$$\mathbf{F} = \nabla f = f_x\,\mathbf{i} + f_y\,\mathbf{j} + f_z\,\mathbf{k}.$$

Recall that

$$d\mathbf{r} = dx\,\mathbf{i} + dy\,\mathbf{j} + dz\,\mathbf{k}.$$

Then

$$\begin{aligned} W &= \int_C \mathbf{F}\cdot d\mathbf{r}, \\ &= \int_C f_x\,dx + f_y\,dy + f_z\,dz, \\ &= \int_C df, \quad \text{by the definition of the differential,} \\ &= f|_A^B = f(B) - f(A). \end{aligned}$$

This application of the fundamental theorem of calculus is a very significant result. It says that line integral of a conservative vector field is **path-independent**! Another way of saying this is that the integral depends only on the **boundary** δC, which in the case of a segment of a curve, consists of the two endpoints. For a conservative vector field, the integral over any two paths that have the same end points yields the same amount of work. The prototype of conservative vector fields is the gravitational field. If in the laboratory, where we can consider the acceleration due to gravity to have constant magnitude g, we lift a mass from point A at height 0 to a point B at height h, the magnitude of the work is the same as the potential energy $PE = mgh$, no matter what path we take.

For a line integral of a conservative vector field, being able to replace a complicated the curve by the simplest curve with same boundary, namely a line segment, already results in a major simplification. But we can do better. First, we need a tool that will guarantee that the vector field is the gradient of a potential function. Then we need a process to find the potential. We "lay on the table" direct computation of the line integral while we introduce more vector tools.

5.4 Curl and Divergence

5.4.1 The Del Operator

Let $\mathbf{F}$ be a vector field with smooth components.

$$\mathbf{F} = P(x, y, z)\mathbf{i} + Q(x, y, z)\mathbf{j} + R(x, y, z)\mathbf{k}.$$

Recall the definition of the Del operator

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}$$

We already know that the Del operator acting on function f produces a vector field ∇f called the gradient. But, we can also apply the Del operator to vector field in two different ways. First, we define

$$\begin{aligned} \nabla \times \mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R, \end{vmatrix} \\ &= (R_y - Q_z)\,\mathbf{i} - (R_x - P_z)\,\mathbf{j} + (Q_x - P_y)\,\mathbf{k}. \end{aligned} \tag{5.4}$$

This operation yields a new vector $(\nabla \times \mathbf{F})$ called the **Curl** of the vector field. A second way we can apply the Del operator is the by definition

$$\nabla \cdot \mathbf{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \tag{5.5}$$

In this case, the result of the operation is a function $\nabla \cdot \mathbf{F}$ that we call the **Divergence** of the vector field. We have the following theorem

5.4.1 Theorem Let f be a smooth function and $\mathbf{F}$ a smooth vector field. Then

$$\begin{aligned} &\text{a)}\ \ \nabla \times \nabla f = \overrightarrow{0}, \\ &\text{b)}\ \ \nabla \cdot (\nabla \times \mathbf{F}) = 0 \end{aligned} \tag{5.6}$$

The proofs are by simple direct computation. First, if

$$\mathbf{F} = \nabla f = f_x\mathbf{i} + f_y\mathbf{j} + f_z\mathbf{k}$$

equation 5.4 yields

$$\begin{aligned}\nabla \times \mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f_x & f_y & f_z \end{vmatrix}, \\ &= (f_{zy} - f_{yz})\,\mathbf{i} - (f_{zx} - f_{xz})\,\mathbf{j} + (f_{yx} - f_{xy})\,\mathbf{k}, \\ &= \overrightarrow{0}\end{aligned}$$

Secondly,

$$\begin{aligned}\nabla \cdot (\nabla \times \mathbf{F}) &= \nabla \cdot [(R_y - Q_z)\,\mathbf{i} - (R_x - P_z)\,\mathbf{j} + (Q_x - P_y)\,\mathbf{k}], \\ &= \frac{\partial}{\partial x}(R_y - Q_z) - \frac{\partial}{\partial y}(R_x - P_z) + \frac{\partial}{\partial z}(Q_x - P_y), \\ &= (R_{yx} - Q_{zx}) - (R_{xy} - P_{zy}) + (Q_{xz} - P_{yz}), \\ &= 0\end{aligned}$$

The most important information to extract from this section are the following facts.

1. (a) $\nabla \times \mathbf{F} = 0 \Leftrightarrow \mathbf{F}$ is **irrotational** (Conservative).
   (b) $\nabla \cdot \mathbf{F} = 0 \Leftrightarrow \mathbf{F}$ has **no sources or sinks** (Incompressible).

2. For the situations that arise in this book, the reverse of equations 5.4.1 also hold.
   (a) $\nabla \times \mathbf{F} = 0 \Leftrightarrow \mathbf{F} = \nabla f$ for some function.
   (b) $\nabla \cdot \mathbf{G} = 0 \Leftrightarrow \mathbf{G} = \nabla \times \mathbf{F}$ for some vector field $\mathbf{F}$.

Theorem 5.4.1 implies that
a) If $\mathbf{F}$ is conservative, that is, if $\mathbf{F} = \nabla f$, then $\mathbf{F}$ is irrotational.
b) If $\mathbf{G} = \nabla \times \mathbf{F}$, then $\nabla \times \mathbf{F}$ has no sinks or holes.

The reader might find it helpful to consider the following diagram. Let $\mathcal{F}$ be the set of functions in $\mathbf{R}^3$ and $\mathcal{VF}$ the set of vector fields. Let $f \in \mathcal{F}$ and $\mathbf{F} \in \mathcal{VF}$. The Del operator acts in three different ways according to the scheme

$$\mathcal{F} \xrightarrow[\text{Grad}]{\nabla f} \mathcal{VF} \xrightarrow[\text{Curl}]{\nabla\times\mathbf{F}} \mathcal{VF} \xrightarrow[\text{Div}]{\nabla\cdot\mathbf{F}} \mathcal{F}. \tag{5.7}$$

In terms of the diagram, theorem 5.4.1 says that if we apply the Del operator twice in a row along the diagram, the result is zero. On the other hand, if one skips the middle arrow and applies the divergence the gradient, the result in general is not zero. In fact one gets the Laplacian.

$$\nabla \cdot \nabla f = \nabla^2 f. \tag{5.8}$$

We will show the relation between curl of a vector field and rotation later in a more advanced exercise. The jewel relating vector field to sources or sinks is captured in the first two of Maxwell's equations

$$\nabla \cdot \mathbf{E} = 4\pi\rho \qquad \nabla \cdot \mathbf{B} = 0.$$

The former states that the vector field has a source/sink, namely a charge density ρ that generates the electric field $\mathbf{E}$. The latter is indicative that the magnetic field $\mathbf{B}$ is the curl $\mathbf{B} = \nabla \times \mathbf{A}$ of some vector $\mathbf{A}$ called the vector potential. The magnetic field has no sources or sinks. Field lines of a magnet do not begin at the north pole and end at the south pole. The field lines are actually loops that continue inside the magnet. There is no magnetic monopole!

The proof of the converse equations 5.4.1 is beyond the scope of this course. The argument requires some subtle topological considerations without which the result may be false. Fortunately other than the vector field in $\mathbf{R^2}$ given by

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

there are no problems with converse of the theorem in this book. If the curl of a vector field is zero, then the vector field is the gradient of some scalar potential.

5.4.2 Path Independent Integrals

We now have the tool we need to treat properly line integrals $W = \int_C \mathbf{F}\cdot d\mathbf{r}$ of conservative vector fields.

A 1-form $\alpha = \mathbf{F}\cdot d\mathbf{r}$ is called an **exact differential 1-form**, if there exists some function f called a potential such that

$$\alpha = df \tag{5.9}$$

In terms of differential forms, a vector field $\mathbf{F} = \langle P, Q, R\rangle$ is conservative if the 1-form

$$\alpha = \mathbf{F}\cdot d\mathbf{r} = P\,dx + Q\,dy + R\,dz,$$

is exact. To find the work line integral is a three step process.

1. Compute $\nabla \times \mathbf{F}$. If the answer is the zero vector, then $\mathbf{F} = \nabla f$ for a function f, unique up to the addition of a constant. The line integral is path independent.

2. Integrate to find the potential f.

3. Evaluate the line integral by the fundamental theorem of calculus.

Finding a potential for a conservative vector field in $\mathbf{R}^2$ is equivalent to solving an exact, first order ordinary differential equation

$$P(x,y)\,dx + Q(x,y)\,dy = 0$$

This sums up just about all the applications of calculus III that appear in a first course in ordinary differential equations.

5.4.2 Example Find $W = \int_C \mathbf{F}\cdot d\mathbf{r}$, where $\mathbf{F} = \langle y^2z^3, 2xyz^3, 3xy^2z^2\rangle$, and C is the curve $\mathbf{r}(t) = \langle \cos t, \sin t, \sin 2t\rangle,\ t \in [0,\pi]$.

Solution. All line integrals can be done by the direct method, but in this case, the integral is nasty. It is wise to start all line integrals by first establishing whether or not the vector field is conservative.

$$\begin{aligned}
\nabla \times \mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y^2z^3 & 2xyz^3 & 3xy^2z^2 \end{vmatrix}, \\
&= (6xyz^2 - 6xyz^2)\,\mathbf{i} - (3y^2z^2 - 3y^2z^2)\,\mathbf{j} + (2yz^3 - 2yz^3)\,\mathbf{k}, \\
&= \vec{0}.
\end{aligned}$$

We conclude that $\mathbf{F}$ is conservative and $\mathbf{F} = \nabla f$ for some function. That is

$$\mathbf{F} = \langle f_x, f_y, f_z \rangle.$$

We start by picking the easiest of the partial derivatives to integrate, and keep only the terms that are not redundant. In this case, all integrals are easy

$$\begin{aligned}
f_x &= y^2z^3 & f &= xy^2z^3 & &\text{(integrating with respect to } x\text{)}, \\
f_y &= 2xyz^3 & f &= xy^2z^3 & &\text{(integrating with respect to } y\text{)}, \\
f_z &= 3xy^2z^2 & f &= xy^2z^3 & &\text{(integrating with respect to } z\text{)}.
\end{aligned}$$

The potential is $f = xy^2z^3 + C$, where is C is constant. the constant of integration plays no role when we evaluate the line integral. Formally, when one performs an iterated integral with respect to a variable, say x, one should add an arbitrary function of the other two variables. However, in most cases, it is much more practical to integrate with respect to each variable and just pick the terms that make f consistent with $\nabla f = \mathbf{F}$, as we have done here. The endpoints are $A = \mathbf{r}(0)$ and $B = \mathbf{r}(\pi)$, so the value of the line integral is,

$$\begin{aligned}
W &= \int_C \mathbf{F} \cdot d\mathbf{r}, \\
&= xy^2z^3\big|_A^B = xy^2z^3\big|_{(1,0,0)}^{(-1,0,0)} = 0.
\end{aligned}$$

5.4.3 Example Find the work done by $\mathbf{F} = (2xz + \sin y)\,\mathbf{i} + (x\cos y)\,\mathbf{j} + x^2\,\mathbf{k}$, over the curve C given by one cycle of the helix $\mathbf{r}(t) = \cos t\,\mathbf{i} + \sin t\,\mathbf{j} + t\,\mathbf{k}$, where $t \in [0, 2\pi]$.
Solution.

$$\begin{aligned}
\nabla \times \mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ 2xz + \sin y & x\cos y & x^2 \end{vmatrix}, \\
&= (0)\,\mathbf{i} - (2x - 2x)\,\mathbf{j} + (\cos y - \cos y)\,\mathbf{k}, \\
&= \vec{0}.
\end{aligned}$$

So $\mathbf{F}$ is conservative and $\mathbf{F} = \nabla f = \langle f_x, f_y, f_z \rangle$ for some potential function. We proceed to find f,

$$\begin{aligned} f_x &= 2xz + \sin y, & f &= x^2 z + x \sin y \quad \text{(integrating with respect to } x\text{)},\\ f_y &= x \cos y & f &= x \sin y \quad \text{(we already got this term)},\\ f_z &= x^2 & f &= x^2 z \quad \text{(we already got this one as well)}. \end{aligned}$$

Hence $f = x^2 z + x \sin y + C$. The endpoints are $A = \mathbf{r}(0)$ and $B = \mathbf{r}(2\pi)$, so the value of the line integral is,

$$\begin{aligned} W &= \int_C \mathbf{F} \cdot d\mathbf{r},\\ &= x^2 z + x \sin y \Big|_A^B,\\ &= x^2 z + x \sin y \Big|_{(1,0,0)}^{(1,0,2\pi)} = 2\pi. \end{aligned}$$

Occasionally, the vector field is generated from a potential for which the gradient involves the product rule with respect to one of the variables, say for example $f = xy\, e^{y^2}$. In such cases the practical way to find the potential is to integrate f_x with respect to x, which does not require integrating by parts. Then one computes f_y to check for consistency.

5.4.3 Applications to Physics

Gravitational Field

Newton's Law of gravitation is based on the $1/r$ potential. Specifically, the gravitational potential of a mass m due to a celestial body of mass M, is given by

$$\varphi = -\frac{mGM}{r} \tag{5.10}$$

where $r = \sqrt{x^2 + y^2 + z^2}$ is the distance between the center of masses, and G is a universal constant. It is convenient to set the origin at the center of mass of the larger mass M. The gravitational force that the mass M exerts on the mass m is

$$\mathbf{F} = -\nabla \varphi \tag{5.11}$$

implying that the gravitational vector field is a conservative vector field. We compute the force field,

$$\begin{aligned}
\mathbf{F} &= \nabla\left(\frac{mMG}{r}\right), \\
&= mMG\nabla\left(\frac{1}{\sqrt{x^2+y^2+z^2}}\right), \\
&= mMG\left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right)\left(\frac{1}{\sqrt{x^2+y^2+z^2}}\right), \\
&= -mMG\frac{x\mathbf{i}+y\mathbf{j}+z\mathbf{k}}{(x^2+y^2+z^2)^{3/2}}.
\end{aligned}$$

Hence, the gravitational force is

$$\mathbf{F} = -mMG\frac{\mathbf{r}}{r^3} \tag{5.12}$$

The magnitude F of the force is $F = mMG/r^2$, so this is a $1/r^2$ law.

Conservation of Energy

The most fundamental law of physics is conservation of energy. No physical system has ever been observed in which energy is not conserved. Some times energy might hide, most commonly in the form of heat, but when all the energy is counted, the final energy is the same as the initial energy. A version of conservation of kinetic energy can be easily derived from line integrals. Let $\mathbf{F}$ be a force acting on a particle of mass m along a curve $\mathbf{r}(t)$, starting at an initial point $A = \mathbf{r}(t_i)$, and ending at a final point $B = \mathbf{r}(t_f)$. The work done by the force is

$$W = \int_C \mathbf{F}\cdot d\mathbf{r}.$$

By Newton's law,

$$\mathbf{F} = \mathbf{a} = m\ddot{\mathbf{r}}$$

Inserting into the work line integral, we get,

$$\begin{aligned}
W &= \int_{t_i}^{t_f} \mathbf{F}\cdot\frac{d\mathbf{r}}{dt}\,dt, \\
&= \int_{t_i}^{t_f} m\,\ddot{\mathbf{r}}\cdot\dot{\mathbf{r}}\,dt, \\
&= m\int_{t_i}^{t_f} \tfrac{1}{2}\frac{d}{dt}(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}})\,dt \\
&= \tfrac{1}{2}m\int_{t_i}^{t_f} \frac{d}{dt}v^2\,dt \qquad \text{where } v \text{ is the speed.}
\end{aligned}$$

Hence the work done,

$$W = \tfrac{1}{2}m(v_f^2 - v_i^2) = \Delta(KE) \tag{5.13}$$

is the difference between the final and the initial kinetic energy, that is, the change $\Delta(KE)$ in kinetic energy. For example, if a mass is sliding on a table with initial speed v and comes to a complete stop, then it must be that the work W done by friction has magnitude $\frac{1}{2}mv^2$. The kinetic energy is converted into heat.

To connect the conservation equation 5.13 with an equation more familiar in first year physics, consider the case of uniform motion in which the acceleration a is constant, and the work is done along a straight line segment of magnitude x. The work is given by $W = Fx$, where F is the magnitude of the constant force vector. We get

$$\begin{aligned} max &= \tfrac{1}{2}m(v_f^2 - v_i^2), \\ 2ax &= (v_f^2 - v_i^2) \end{aligned}$$

This is one of the most basic equations in elementary physics. For example, in a conservative vector field like gravity in which the work is path independent, if a frustrated physics student climbs to the second floor by any path and drops the book from a height h, the equation reads $2gh = v_f^2$, so the book hits the ground with speed $\sqrt{2gh}$. Fortunately, the concept is so easy, that this silly situation is most unlikely to occur. More generally, if the conservative vector field $\mathbf{F}$ is the gradient of a potential

$$\mathbf{F} = -\nabla\varphi,$$

resulting on a path independent work integral, then equation 5.13 reads,

$$\begin{aligned} -\varphi(B) + \varphi(A) &= KE(B) - KE(A) \quad \text{or,} \\ \varphi(A) + KE(A) &= \varphi(B) + KE(B) \end{aligned} \tag{5.14}$$

This is the **law of conservation of energy**. Under the action of a conservative vector field, the sum of the potential and kinetic energy is constant.

Maxwell's Equations with no Sources or Currents

Here is a quote from the first edition of the Encyclopedia Britannica first published in 1768.

> Of Light
>
> Light consists of an inconceivably great number of particles flowing from a luminous body in all manner of directions; and these particles are so small, as to surpass all human comprehension.
>
> Encyclopedia Britannica - Optics (1768)

Before treating Maxwell's equations in the special case noted in the title, we first need to mention some important vector identities. The Del operator on functions is a linear derivation, meaning that it satisfies the basic linear properties of derivatives and Leibnitz rule for derivatives

$$\begin{aligned}\nabla(kf) &= k\nabla f, \quad \text{where } k \text{ is constant,}\\ \nabla f \pm g &= \nabla f \pm \nabla g,\\ \nabla(fg) &= f\nabla g + g\nabla f.\end{aligned}$$

The curl and divergence are linear operators, but not surprisingly the derivatives of "products" are slightly more complicated. We leave as a non-inspiring computation to verify the following identities,

$$\begin{aligned}\nabla \times (f\mathbf{F}) &= f(\nabla \times \mathbf{F}) + \mathbf{F} \times \nabla f,\\ \nabla \cdot (f\mathbf{F}) &= f(\nabla \cdot \mathbf{F}) + \mathbf{F} \cdot \nabla f,\\ \nabla \cdot (\mathbf{F} \times \mathbf{G}) &= \mathbf{G} \cdot (\nabla \times \mathbf{F}) - \mathbf{F} \cdot (\nabla \times \mathbf{G}),\\ \nabla \times (\nabla \times \mathbf{F}) &= \nabla(\nabla \cdot \mathbf{F}) - \nabla^2\mathbf{F}.\end{aligned} \tag{5.15}$$

In the last equation in 5.15, by the Laplacian of a vector we just mean we apply the Laplacian to each component. The proofs of these identities are much more elegant in the tensor formalism which is beyond the scope of this course. On the other hand, there are just four computations that can be done directly by the definitions. There are no clever mysteries to be discovered here.

Now, consider Maxwell's equations for electromagnetic fields in a region with no charges or currents, as it is the case in vacuum. The equations in Gaussian units are

$$\begin{aligned}\nabla \cdot \mathbf{E} &= 0, & & & \nabla \cdot \mathbf{B} &= 0,\\ \nabla \times \mathbf{E} &= -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, & &\text{and} & \nabla \times \mathbf{B} &= \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}.\end{aligned}$$

These equations must have looked odd when they were first considered by Maxwell in the 1860's, since there is no charge density to act as a source for the electric field in Gauss's law. and no current to generate a magnetic field in Ampére's law. To complicate matters, cross products had only recently been discovered by Hamilton in the context of quaternions and the formalism was still not trusted by the general physics community. Hence, Maxwell worked with the explicit set of 20 partial differential equations. Let's compute the curl of the curl of the electric field.

$$\begin{aligned}
\nabla \times (\nabla \times \mathbf{E}) &= \nabla \times \left(-\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}\right), \\
\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} &= -\frac{1}{c}\nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t}\right), \quad \text{by equation 5.15,} \\
-\nabla^2 \mathbf{E} &= -\frac{1}{c}\frac{\partial}{\partial t}(\nabla \times \mathbf{B}), \quad \text{since} \quad \nabla \cdot \mathbf{E} = 0. \\
&= -\frac{1}{c}\frac{\partial}{\partial t}\left(\frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}\right), \\
&= -\frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2}.
\end{aligned}$$

Hence

$$\nabla^2 \mathbf{E} = \frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{5.16}$$

By a homologous computation we can also deduce that

$$\nabla^2 \mathbf{B} = \frac{1}{c^2}\frac{\partial^2 \mathbf{B}}{\partial t^2}. \tag{5.17}$$

Therefore, the electric and magnetic fields satisfy the wave equation with a speed of light. Actually, in MKS units the coefficient that appears in the wave equations is $(\epsilon_o \mu_0)$, where ϵ_o is the permitivity of free space and the μ_o is the permeability. The revolutionary conclusion of Maxwell is that light must be an electromagnetic phenomenon propagating with speed $c = 1/\sqrt{\epsilon_o \mu_0}$. Let there be light! The discovery is probably the most significant event in science and mathematics of the 19th century.

5.5 Stokes' Theorem

Traditional calculus textbooks typically take a repetitious approach to Stokes' theorem. They first introduce the theorem in two dimensions, then they reintroduce the theorem in two dimensions using curls of vectors, and finally they do the theorem in three dimensions. We see no reason for that. Our approach is to present the theorem once, but do it right. Then it is easy to specialize to dimension two. The ingredients that we need are a vector field $\mathbf{F}$ a surface S whose boundary is a curve $C = \delta S$. Think for example of the surface $S : z = 4 - x^2 - y^2, \;\; x \geq 0$. This surface is a concave down circular paraboloid with boundary $C : \;\; x^2 + y^2 = 4$. We will also need the differential of surface area in $\mathbf{R}^3$.

5.5.1 Stoke's Theorem Let S be an oriented piecewise smooth surface whose boundary $C = \delta S$ is a simple, piecewise smooth closed curve. Let $\mathbf{F}$ be a vector field whose components have continuous partial derivatives. Then

$$W = \oint_C \mathbf{F} \cdot d\mathbf{r} = \iint_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S}. \tag{5.18}$$

The loop on the line integral indicates that the curve is closed. We do not often see Stokes' theorem expanded in terms of the components, but for the sake of completeness, here it is,

$$\oint_C Pdx+Qdy+Rdz = \iint_S (R_y-Q_z)dydz-(P_z-R_x)dxdz+(Q_x-P_y)dxdy. \quad (5.19)$$

Consider the the special case where the vector field and the curve are in the xy- plane. We have

$$\mathbf{F} = P(x,y)\,\mathbf{i} + Q(x,y)\,\mathbf{j}$$

Since the curl is only defined in three dimensions, we have to embed the vector field in $\mathbf{R}^3$ by taking the z-component equal to zero. We get,

$$\begin{aligned}\nabla\times\mathbf{F} &= \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z}\\ P(x,y) & Q(x,y) & 0,\end{vmatrix}\\ &= (Q_x-P_y)\,\mathbf{k}.\end{aligned}$$

The differential of surface area in the xy- plane is

$$d\mathbf{S}_{xy} = dx\,dy\,\mathbf{k}.$$

Therefore Stokes' theorem on the xy-plane reduces to

5.5.2 Green's theorem

$$\oint_{\delta D} P\,dx + Q\,dy = \iint_D \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right] dA. \quad (5.20)$$

This version of Stokes' theorem in dimension two in the xy-plane is called **Green's Theorem**. We present a proof of Green's Theorem. We first prove that for a type I region such as the one bounded between a and b shown in 5.1 We say that a region D in the plane is of type I if it is enclosed between the

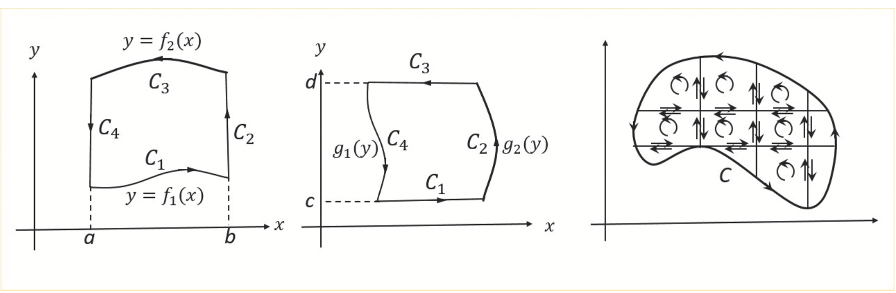


Fig. 5.1: Simple closed curve.

graphs of two continuous functions of x. The region inside the simple closed curve in figure 5.1 bounded by $f_1(x)$ and $f_2(x)$, between a and b, is a region of type I. A region in the plane is of type II if it lies between two continuous functions of y. The region in 5.1 bounded between $c \leq y \leq d$, would be a region of type II. For a region of type I, we claim that

$$\oint_C P\,dx = -\iint_D \frac{\partial P}{\partial y}\,dA \tag{5.21}$$

Where C comprises the curves C_1, C_2, C_3 and C_4. By the fundamental theorem of calculus, we have on the right,

$$\begin{aligned}\iint_D \frac{\partial P}{\partial y}\,dA &= \int_a^b \int_{f_1(x)}^{f_2(x)} \frac{\partial P}{\partial y}\,dydx,\\ &= \int_a^b [P(x, f_2(x)) - P(x, f_1(x))]\,dx.\end{aligned}$$

On the left, the integrals along C_2 and C_4 vanish, since there is no variation on x. The integral along C_3 is traversed in opposite direction of C_1, so we have,

$$\begin{aligned}\oint_C P(x,y)\,dx &= \int_{C_1} + \int_{C_2} + \int_{C_3} + \int_{C_4} P(x,y)\,dx,\\ &= \int_{C_1} P(x,y)\,dx - \int_{C_3} P(x,y)\,dx,\\ &= \int_a^b P(x, f_1(x))\,dx - \int_a^b P(x, f_2(x))\,dx\end{aligned}$$

This establishes the veracity of equation 5.21 for type I regions. By a completely analogous process on type II regions, we find that

$$\oint_C Q\,dy = \iint_D \frac{\partial Q}{\partial x}\,dA. \tag{5.22}$$

The theorem follows by subdividing R into a grid of regions of both types, all oriented in the same direction as shown on the right in figure 5.1. Then one applies equations 5.21 or 5.22, as appropriate, for each of the subdomains. All contributions from internal boundaries cancel since each is traversed twice, each in opposite directions. All that remains of the line integrals is the contribution along the boundary δD.
It is possible to extend Green's Theorem to more complicated regions that are not simple connected, such as the region between two concentric circles.

5.5.3 Example Find $\int_C y^3\,dx - x^3\,dy$, where C is the circle $x^2 + y^2 = 4$. Solution. The vector field here is $\mathbf{F} = y^3\,\mathbf{i} - x^3\,\mathbf{j}$. First we compute the curl of the vector field to see if by chance $\mathbf{F}$ is conservative.

$$\begin{aligned}\nabla \times \mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y^3 & -x^3 & 0, \end{vmatrix}\\ &= (-3x^2 - 3y^2)\,\mathbf{k}.\end{aligned}$$

Since curve C is closed and it is the boundary $C = \delta D$ of a disk in the xy-plane, the differential of surface is $d\mathbf{S} = dx\,dy\,\mathbf{k}$. By Stokes' theorem we have,

$$\begin{aligned} W = \oint_C \mathbf{F}\cdot d\mathbf{r} &= \iint_D -3(x^2+y^2)\,dx\,dy, \\ &= -3\int_0^{2\pi}\int_0^2 r^2(r\,dr\,d\theta), \\ &= -3(2\pi)(\tfrac{2^4}{4}), \\ &= -24\pi. \end{aligned}$$

5.5.4 Example Evaluate $\oint_C (2y+\cos\sqrt{x})\,dx + (4x+e^{y^2})\,dy$, where C is the boundary of the region enclosed by $y = x$ and $y = x^2$.

Solution. The vector field is $\mathbf{F} = (2y+\cos\sqrt{x})\,\mathbf{i} + (4x+e^{y^2})\,\mathbf{j}$. We have.

$$\begin{aligned} \nabla\times\mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ 2y+\cos\sqrt{x} & 4x+e^{y^2} & 0, \end{vmatrix} \\ &= (4-2)\,\mathbf{k} = 2\mathbf{k} \\ d\mathbf{S} &= dx\,dy\,\mathbf{k} \end{aligned}$$

Stokes' theorem gives

$$\begin{aligned} W = \oint \mathbf{F}\cdot d\mathbf{r} &= \int_0^1\int_{x^2}^x 2\,dy\,dx, \\ &= 2\int_0^1 (x-x^2)\,dx, \\ &= 2(\tfrac{1}{2}-\tfrac{1}{3}) = \tfrac{1}{3}. \end{aligned}$$

5.5.5 Example Find $W = \oint_C (x^3-y^3)\,dx + (x^3+y^3)\,dy$, where C is the boundary of the annulus D between $x^2+y^2 = 4$ and $x^2+y^2 = 1$.

Solution. $\mathbf{F} = (x^3-y^3)\,\mathbf{i} + (x^3+y^3)\,\mathbf{j}$.

$$\begin{aligned} \nabla\times\mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x^3-y^3 & x^3+y^3 & 0, \end{vmatrix} \\ &= (3x^2+3y^2)\,\mathbf{k}, \\ d\mathbf{S} &= dx\,dy\,\mathbf{k} \end{aligned}$$

Stokes' theorem gives

$$\begin{aligned} W = \oint \mathbf{F}\cdot d\mathbf{r} &= \iint_D (3x^2+3y^2)\,dy\,dx, \\ &= 3\int_0^{2\pi}\int_1^2 r^2(r\,dr\,d\theta) \\ &= 3(2\pi)(\tfrac{2^4}{4}-\tfrac{1}{4}), \\ &= 6\pi(\tfrac{15}{4}) = \tfrac{45}{2}\pi. \end{aligned}$$

Much of the power of Stokes' theorem 5.5.1 comes from the following observation. Let S and S' be two surfaces that satisfy the hypothesis of the theorem, with common boundary $C = \delta S = \delta S'$. Then

$$\iint_S (\nabla\times\mathbf{F})\cdot d\mathbf{S} = W = \oint_C \mathbf{F}\cdot d\mathbf{r} = \iint_{S'} (\nabla\times\mathbf{F})\cdot d\mathbf{S}'. \tag{5.23}$$

This means that the surface S can be replaced by another surface S' that could make the integral easier to compute. The physics behind this process is very neat. Interpret a vector field G as some sort of flow. A surface integral of the form

$$\Phi = \iint_S \mathbf{G}\cdot dS$$

then represents what call the **flux** of the vector field through the surface. In crude terms, the surface integral gives the amount of $\mathbf{G}$ goo passing through the surface. But if $\mathbf{G} = \nabla\times\mathbf{F}$ is the curl of some vector field, by theorem 5.4.1, the divergence of $\nabla\cdot\mathbf{G}$ is zero and the vector field $(\nabla\times\mathbf{F})$ has no sources or sinks inside the surface. Thus, as with a fishing net, the flux through the net does not depend of the shape of the net, as long the boundary does not change. The main payoff is that if the boundary lies on a flat plane that is parallel to one of the coordinate planes, the simplest surface with the same boundary is a plane, and by equation 4.20, the differential of surface requires absolutely no computation. The surface integral collapses to a two-dimensional case as in Green's theorem.

5.5.6 Example Let $\mathbf{F} = y\,\mathbf{i} - x\,\mathbf{j}$. Compute $W = \iint_S \nabla\times\mathbf{F}\cdot d\mathbf{S}$, where S is the hemisphere $x^2+y^2+z^2 = 4,\ z \geq 0$.
Solution. It is always true that $\nabla\cdot(\nabla\times\mathbf{F}) = 0$. Therefore the vector field $\nabla\times\mathbf{F}$ has no sources or sinks and we can replace S by the simplest surface S' with the same boundary. The boundary of S is the circle $C : x^2+y^2 = 4$ on the xy-plane. The simplest surface S' with the same boundary $\delta S' = \delta S = C$ is $z = 0$.

$$\begin{aligned} \nabla\times\mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ y & -x & 0, \end{vmatrix} \\ &= -2\,\mathbf{k}, \\ d\mathbf{S}' &= dx\,dy\,\mathbf{k} \end{aligned}$$

So, Stokes' theorem gives $W = \iint_{S'} -2\,dy\,dx, = -2(\pi 2^2) = -8\pi$. The surface integral is (-2) times the area of a circle of radius 2.

5.5.7 Example Let $\mathbf{F} = (z-y)\,\mathbf{i} + (z+x)\,\mathbf{j} - (x+y)\,\mathbf{k})$. Compute $W = \iint_S \nabla \times \mathbf{F} \cdot d\mathbf{S}$, where S is the portion of the paraboloid $z = 4 - x^2 - y^2,\ z \geq 0$.

Solution. $\nabla \cdot (\nabla \times \mathbf{F}) = 0$. Therefore the vector field $\nabla \times \mathbf{F}$ has no sources or sinks and we can replace S by the surface $S' : z = 0$ with the same boundary. The common boundary is the circle $C : x^2 + y^2 = 4$ on the xy-plane.

$$\begin{aligned}
\nabla \times \mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x-y & z+x & -x-y, \end{vmatrix} \\
&= -2\,\mathbf{i} + 2\,\mathbf{j} + 2\,\mathbf{k}, \\
d\mathbf{S}' &= dx\,dy\,\mathbf{k}
\end{aligned}$$

So, Stokes' theorem gives $W = \iint_{S'} 2\,dy\,dx, = 2(\pi 2^2) = 8\pi$. The surface integral is (2) times the area of a circle of radius 2.

5.5.8 Example Let $\mathbf{F} = xz\,\mathbf{i} + yz\,\mathbf{j} + xy\,\mathbf{k}$. Compute $W = \iint_S \nabla \times \mathbf{F} \cdot d\mathbf{S}$, where S is the spherical cap of the sphere $x^2 + y^2 + z^2 = 4,\ z > 0$, inside the cylinder $x^2 + y^2 = 1$.

Solution. It is always true that $\nabla \cdot (\nabla \times \mathbf{F}) = 0$. Therefore the vector field $\nabla \times \mathbf{F}$ has no sources or sinks and we can replace S by the simplest surface S' with the same boundary. The boundary of S is a circle C at height given by solving $1 + z^2 = 4$, that is $z = \sqrt{3}$. This is a plane S' parallel to the xy-plane with boundary $\delta S' = \delta S = C$. We have

$$\begin{aligned}
\nabla \times \mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ xz & yz & xy, \end{vmatrix} \\
&= (x-y)\,\mathbf{i} - (y-x)\mathbf{j} + 0\,\mathbf{k}, \\
d\mathbf{S}' &= dx\,dy\,\mathbf{k}
\end{aligned}$$

So, Stokes' theorem gives $W = \oint \mathbf{F} \cdot d\mathbf{r} = \iint_{S'} 0\,dy\,dx, = 0$. It is hard to overstress how powerful this theorem is. Without stokes' theorem we would work with the complicated differential of the sphere in spherical coordinates and a resulting proliferation of integrals of powers of sines and cosines.

5.5.9 Example Let $\mathbf{F} = \tan^{-1}(x^2yz^2)\,\mathbf{i} + x^2y\,\mathbf{j} + x^2z^2\,\mathbf{k}$. Compute $W = \iint_S \nabla \times \mathbf{F} \cdot d\mathbf{S}$, where S is the cone $x = \sqrt{y^2 + z^2},\ 0 \leq x \leq 2$.

Solution. Since $\nabla \cdot (\nabla \times \mathbf{F}) = 0$, the integral depends only on the boundary. We can replace S by the plane surface $S' : x = 2$ provided $\delta S' = \delta S = C$. The curve C is a circle of radius 2.

$$d\mathbf{S}' = dy\,dz\,\mathbf{i},$$

$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \tan^{-1}(x^2yz^2) & x^2y & x^2z^2, \end{vmatrix}$$

$$= 0\,\mathbf{i} + \ldots \quad \text{All we need is the } \mathbf{i} \text{ component}$$

$$W = \iint_{S'} 0\;dy\,dz = 0$$

5.5.10 Example Compute $W = \oint \mathbf{F} \cdot d\mathbf{r}$, where $\mathbf{F} = x^2z\,\mathbf{i} + xy^2\,\mathbf{j} + z^2\,\mathbf{k}$, and C is the curve of intersection of the plane $x + y + z = 1$ and the cylinder $x^2 + y^2 = 9$.

Solution. The curve of intersection is neither a circle nor does it lie on a plane parallel to one of the coordinate planes. What do we do now? Answer, we use the same process. The ellipse C is the boundary of the surface $S: \ z = 1 - x - y$ with circular projection $x^2 + y^2 = 9$ on the xy-plane. From the equation of the plane we get $dz = -dx - dy$. From formula 4.20, we get $dy \wedge dz = dx \wedge dy$ and $-dx \wedge dz = dx \wedge dy$. Hence

$$d\mathbf{S} = (1\,\mathbf{i} + 1\,\mathbf{j} + 1\,\mathbf{j})\,dx\,dy$$

Notice that the vector components of $d\mathbf{S}$ are the same as the gradient of $f = x + y + z - 1$, so that indeed, the $d\mathbf{S}$ is normal to the surface, as it should be. We also have

$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ x^2z & xy^2 & z^2, \end{vmatrix}$$

$$= 0\,\mathbf{i} + x^2\,\mathbf{i} + y^2\,\mathbf{i}$$

$$W = \iint_S \nabla \times \mathbf{F} \cdot d\mathbf{S},$$

$$= \int\int_S (x^2 + y^2)\,dx\,dy,$$

$$= \int_0^{2\pi}\int_0^3 r^2(r\,dr\,d\theta) = 2\pi(\tfrac{3^4}{4}) = \tfrac{81\pi}{2}$$

5.6 Surface Integrals

A surface integral is tautologically, an integral over a surface. As in the case of line integrals, there are two flavors of surface integrals

a) Scalar: $I = \iint_S g(x, y, z)\,dS$, where $dS = \|d\mathbf{S}\|$,

b) Vector: $\Phi = \iint \mathbf{F} \cdot d\mathbf{S}$

Again, completely analogous to line integrals, we can compute all surface integrals by parametrizing the surface, and substituting dS or $d\mathbf{S}$ from the definitions in section 4.7. Computations of scalar surface integrals are set up exactly as one sets up surface area problems, except that instead of integrating $\iint_D dS$ one inserts into the integrand the given function $g(x, y, z)$ evaluated over the surface. For the differential of surface, we may use either of the two equivalent formulas 4.20, or 4.22.

$$I = \iint_D g(x(u,v), y(u,v), z(u,v)) \|\mathbf{r}_u \times \mathbf{r}_v\| \, du \, dv. \tag{5.24}$$

If the surface S is given by an explicit function $z = g(x, y)$ then the surface integral is obtained by substitution into

$$I = \iint_D g(x, y, f(x, y)) \sqrt{f_x^2 + f_y^2 + 1} \, dx \, dy. \tag{5.25}$$

Of course, the most interesting problems are when $g(x, y, z) = 1$ which give the surface area. If $g(x, y, z) \neq 1$, the most applicable problems are those involving moments of mass or moments of inertia.

We have also encountered vector type surface integrals in the context of Stokes' theorem. In such cases the surface S has a boundary $C = \delta S$ and the surface integral is the flux of the curl of a vector field. We already know that a vector field of the form $\nabla \times \mathbf{F}$ has zero divergence, and we can replace the surface by the simplest surface with the same boundary. The meaning of a flux integral of a vector field $\mathbf{F}$ over a surface S parametrized by $\mathbf{r}(u, v)$ is more clear if one writes,

$$\begin{aligned}
\Phi &= \iint_S \mathbf{F} \cdot d\mathbf{S}, \\
&= \iint_S \mathbf{F} \cdot (\mathbf{r}_u \times \mathbf{r}_v) \, du \, dv, \\
&= \iint_S \mathbf{F} \cdot \frac{(\mathbf{r}_u \times \mathbf{r}_v)}{\|\mathbf{r}_u \times \mathbf{r}_v\|} \|\mathbf{r}_u \times \mathbf{r}_v\| \, du \, dv, \\
&= \iint_S \mathbf{F} \cdot \mathbf{n} \, \|\mathbf{r}_u \times \mathbf{r}_v\| \, du \, dv, \\
\Phi &= \iint_S \mathbf{F} \cdot \mathbf{n} \, dS
\end{aligned}$$

Just as in the case of work line integrals, the formula offers absolutely no computational advantage, but it elucidates two things:

1. There is really only one kind of surface integral since $\mathbf{F} \cdot \mathbf{n}$ is just a scalar function.

2. Since $\mathbf{n}$ is a unit vector normal to the surface, the dot product shows that only the component of $\mathbf{F}$ onto the normal, contributes to the flux. The tangential components does not result in any flow through the surface.

5.6.1 Scalar Surface Integrals

From the point of view of physics and engineering, the most interesting problems of scalar surface integrals are those related to center of mass and moments of inertia. For that reason, we limit the examples to these type of applications.

5.6.1 Example Center of Mass of a Thin Spherical Shell.

We wish to compute the center of mass of the hollow hemisphere $\Sigma: \ x^2+y^2+z^2=R^2, \ \ z \geq 0$, with constant density $\rho = k$. The total mass M is the density times the surface area $M = 2\pi R^2 k$. By symmetry, it suffices to compute $z_{c.m.}$. We have,

$$\begin{aligned}
M z_{c.m.} &= \iint_\Sigma z\, dm = \iint_\Sigma z\, k\, dS, \\
&= \int_0^{2\pi} \int_0^{\pi/2} R\cos\theta (kR^2 \sin\theta)\, d\theta\, d\phi, \\
&= -2\pi R^3 k \tfrac{1}{2} \cos^2\theta \,\Big|_0^{\pi/2}, \\
&= \pi R^3 k, \\
z_{c.m.} &= \frac{\pi R^3 k}{2\pi R^2 k} = \frac{1}{2} R.
\end{aligned}$$

5.6.2 Example Moment of Inertia of a Thin Spherical Shell.

We compute the moment of inertia of a thin hollow sphere $\Sigma: \ x^2+y^2+z^2-R^2$ of constant density $\rho = k$ about an axis through the origin.

$$\begin{aligned}
I_z &= \iint_\Sigma (x^2+y^2)\, k\, dS, \\
&= \int_0^{2\pi} \int_0^{\pi} (R^2 \sin^2\theta)(kR^2 \sin\theta)\, d\theta\, d\phi, \\
&= \int_0^{2\pi} \int_0^{\pi} (R^2 \sin^2\theta)(kR^2 \sin\theta)\, d\theta\, d\phi, \\
&= \int_0^{2\pi} \int_0^{\pi} R^2 k(1-\cos^2\theta)(kR^2 \sin\theta)\, d\theta\, d\phi, \\
&= -2\pi R^4 k \left[\cos\theta - \tfrac{1}{3}\cos^3\theta\right]_0^{\pi}, \\
&= -2\pi R^4 [-1 + \tfrac{1}{3} - (1 - \tfrac{1}{3})], \\
&= \tfrac{8}{3}\pi R^4 k, \\
&= \tfrac{8}{3}\pi R^4 k \frac{M}{4\pi R^2 k}, \\
I_z &= \tfrac{2}{3} M R^2
\end{aligned}$$

5.6.3 Example Moment of Inertia of weighted Spherical Shell

We compute the moment of inertia of a thin hollow hemisphere $\Sigma: \ x^2+y^2+$

$z^2 - R^2, \;\; z \geq 0$, with density $\rho = z$ about an axis through the origin.

$$\begin{aligned}
I_z &= \iint_{\Sigma} (x^2+y^2)\, z\, dS, \\
&= \int_0^{2\pi}\int_0^{\pi} (R^2\sin^2\theta) R^2\cos\theta(R^2\sin\theta\, d\theta\, d\phi,) \\
&= \int_0^{2\pi}\int_0^{\pi} (R^2\sin^2\theta)(kR^2\sin\theta)\, d\theta\, d\phi, \\
&= \int_0^{2\pi}\int_0^{\pi} R^5\sin^3\theta)\cos\theta\, d\theta\, d\phi, \\
&= 2\pi R^5 k \tfrac{1}{4}\, \sin^4\theta \Big]_0^{\pi/2}, \\
I_z &= \tfrac{1}{2}\pi R^5
\end{aligned}$$

The total mass M is

$$\begin{aligned}
M &= \int_0^{2\pi}\int_0^{\pi/2} (R\cos\theta)(R^2\sin\theta\, d\theta\, d\phi), \\
&= 2\pi\tfrac{1}{2}\, \sin^2\theta\,\Big]_0^{\pi/2}, \\
&= \pi R^3.
\end{aligned}$$

Hence

$$I_z = \tfrac{1}{2}\pi R^5 \frac{M}{\pi R^3} = \tfrac{1}{2} M R^2.$$

5.6.4 Example Moment of Inertia of a Thin Paraboloid Shell

We find the moment of inertia I_x of the paraboloid shell $x = 4 - y^2 - z^2, \;\; x \geq 0$ with density $\rho = 1$. First, we need to compute the differential of surface area. As usual, we prefer to start from the general definition. We have,

$$\begin{aligned}
dx &= -2y\, dy - 2z\, dz, \\
dy \wedge dz &= 1 dy \wedge dz, \\
-dx \wedge dz &= 2y\, dy \wedge dz, \\
dx \wedge dy &= 2z\, dy \wedge dz, \\
d\mathbf{S} &= \langle 1, 2y, 2z\rangle\, dy\, dz, \\
dS &= \sqrt{1+4y^2+4z^2}\, dy\, dz.
\end{aligned}$$

The surface integral is

$$\begin{aligned}
I_x &= \int\int (y^2+z^2)\,dm,\\
&= \int\int (y^2+z^2)\sqrt{1+4y2+4z^2}\;dy\,dz,\\
&= \int_0^{2\pi}\int_0^2 r^2\sqrt{1+4r^2}\,r\,dr\,d\theta,\\
&= \int_0^2\int_1^{\sqrt{17}} \tfrac{1}{4}(u^2-1)\,u\,(\tfrac{1}{4}u\,du)\,d\theta \quad \text{(with the substitution, } u^2=1+4r^2\text{)},\\
&= 2\pi(\tfrac{1}{16})\int_0^{\sqrt{17}}(u^4-u^2)\,du,\\
&= \tfrac{1}{8}\pi\left[\tfrac{1}{5}u^5-\tfrac{1}{3}u^3\right]_1^{\sqrt{17}},\\
&= \frac{1}{60}\pi[23(17)\sqrt{17}+1] = .\frac{1}{60}\pi(391\sqrt{17}+1)
\end{aligned}$$

We spared the reader some arithmetic on the last line.

5.6.2 Vector Surface Integrals

All vector surface integrals can be computed directly from the definition, by “parametrizing and integrating”. It might not always be the best way but it can always be done. Let

$$\mathbf{F} = P(x,y,z)\mathbf{i} + Q(x,y,z)\mathbf{j} + R(x,y,z)\mathbf{k}$$

be a vector field in $\mathbf{R}^3$ and S be a surface

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

Recalling the definition(s) of the vector differential of surface 4.20 we can write the flux of the vector field across the surface as

$$\Phi = \iint_S \mathbf{F}\cdot d\mathbf{S} = \iint_S P\,dy\,dz - Q\,dy\,dz + R\,dx\,dy.$$

The expression should really be written more precisely as,

$$\Phi = \iint_S P\,dy\wedge dz - Q\,dy\wedge dz + R\,dx\wedge dy. \tag{5.26}$$

The quantity α inside the integrand

$$\alpha = \mathbf{F}\cdot d\mathbf{S} = P\,dy\wedge dz - Q\,dy\wedge dz + R\,dx\wedge dy, \tag{5.27}$$

is called a **differential 2-form**. The process of computation is simply to substitute the surface components into 5.26 and integrate. Alternatively, we can use 4.22 and just compute

$$\Phi = \iint_S \mathbf{F}\cdot d\mathbf{S} = \iint_S \mathbf{F}\cdot(\mathbf{r}_u\times\mathbf{r}_v)\,du\,dv. \tag{5.28}$$

5.6.5 Example Flux of Constant Vector Field across a Plane

The simplest example of a line integral is the case in which the vector field $\mathbf{F}$ is constant, and the path is a straight line $\mathbf{r}(t) = \mathbf{P} + t\mathbf{v}$. The work done by the force to move a particle from $\mathbf{P}$ to $(\mathbf{P} + \mathbf{v})$ is just the dot product $\mathbf{F} \cdot \mathbf{v}$. The analogous example for flux is the case where again, the vector field $\mathbf{F}$ is constant and the surface is the parallelogram R spanned by vectors $\mathbf{a}$ and $\mathbf{b}$ on the plane

$$\mathbf{r}(u, v) = \mathbf{P} + u\mathbf{a} + v\mathbf{b}$$

Then, the differential of surface is also constant,

$$\begin{aligned} d\mathbf{S} &= (\mathbf{a} \times \mathbf{b})\, du\, dv, \\ dS &= \|\mathbf{a} \times \mathbf{b}\|\, du\, dv, \quad u \in [0, 1],\ v \in [0, 1]. \end{aligned}$$

The flux integral gives

$$\begin{aligned} \Phi &= \iint_R \mathbf{F} \cdot d\mathbf{S} = \iint_R \mathbf{F} \cdot \mathbf{n}\, dS, \\ &= \iint_R \mathbf{F} \cdot \frac{(\mathbf{a} \times \mathbf{b})}{\|\mathbf{a} \times \mathbf{b}\|} \|\mathbf{a} \times \mathbf{b}\|\, du\, dv, \\ &= \mathbf{F} \cdot (\mathbf{a} \times \mathbf{b}) \int_0^1 \int_0^1 du\, dv, \\ &= \mathbf{F} \cdot (\mathbf{a} \times \mathbf{b}). \end{aligned}$$

The integral is trivial because were able to pull the constant triple product outside the integral. The answer is most intuitive. Suppose that $\mathbf{F}$ represents the velocity of some fluid of density $\rho = 1$. The flux is the triple product $(\mathbf{Fab})$ which is the volume of the parallelepiped where the base is the parallelogram spanned by $\mathbf{a}$ and $\mathbf{b}$. This is the rate at which the fluid flows through the parallelogram. To illustrate the concept, we post the following problem.

5.6.6 Example Calculate the flux of the vector field $\mathbf{F} = \langle 2, -1, 3\rangle$ across the triangle with vertices at $A(0, 1, -2), B(3, 1, 0), C(-2, 2, 1)$.
Solution. Let

$$\mathbf{a} = \overrightarrow{AB} = \langle 3, 0, 2\rangle, \quad \mathbf{b} = \overrightarrow{AC} = \langle -2, 1, 3\rangle,$$

Then

$$\Phi = \frac{1}{2}\begin{vmatrix} 2 & -1 & 3 \\ 3 & 0 & 2 \\ -2 & 2 & 3 \end{vmatrix} = \frac{1}{2}[2(-2) + 1(13) + 3(3)] = 9.$$

5.6.7 Example Flux of Non-constant Vector Field across a Plane We calculate the flux of the vector field $\mathbf{F} = \langle x.xy, xz\rangle$ across the part of the plane $S:\ 3x + 2y + z = 6$ with the natural orientation and which lies in the first octant. To be clear, the natural orientation is the orientation of the gradient.

Although a direct computation of a flux integral is a totally routine matter in the sense that all it requires is to substitute into the formulas and compute, many students struggle because of the various ways to manifest the differential of surface area vector. For the purposes of showing the equivalence of the methods, we compute $d\mathbf{S}$ in three different ways

Method 1. We use the wedge definition given by equation 4.20. This is my favorite method because it is based precisely on the definition, the definition is intuitive, and it always works. We have

$$\begin{aligned} z &= 6 - 3x - 2y, \\ dz &= -3\,dx - 2\,dy, \\ dy \wedge dz &= 3\,dx \wedge dy, \\ -dx \wedge dz &= 2\,dx \wedge dy, \\ dx \wedge dy &= 1\,dx \wedge dy, \\ d\mathbf{S} &= (3\,\mathbf{i} + 2\,\mathbf{j} + 1\,\mathbf{k})\,dx\,dy \end{aligned}$$

Method 2. We use the parametric equation definition 4.22. This method also works every time. We parametrize and compute,

$$\begin{aligned} \mathbf{r}(x,y) &= \langle x, y, 6 - 3x - 2y \rangle, \\ \mathbf{r}_x &= \langle 1, 0, -3 \rangle, \\ \mathbf{r}_y &= \langle 0, 1, -2 \rangle, \\ \mathbf{r}_x \times \mathbf{r}_y &= \langle 3, 2, 1 \rangle, \\ d\mathbf{S} &= (3\,\mathbf{i} + 2\,\mathbf{j} + 1\,\mathbf{k})\,dx\,dy \end{aligned}$$

Method 3. We use equation for implicit functions given by 4.28. Here, the implicit function is $F(x,y,z) = 3x + 2y + z = 6$, so

$$\begin{aligned} \nabla F &= \langle 3, 2, 1 \rangle, \\ \nabla F \cdot \mathbf{k} &= 1, \\ d\mathbf{S} &= (3\,\mathbf{i} + 2\,\mathbf{j} + 1\,\mathbf{k})\,dx\,dy \end{aligned}$$

This method is specialized for implicit functions and it is very efficient. But we must not forget the reason it works fast is that in the derivation of the formula, we have already applied the definition 4.20. In addition, the formula must be revised if the differential of surface is projected instead in another coordinate plane.

We proceed to calculate the flux integral. The region R of integration is the triangle in the first quadrant given $z = 0$, that is $3x + 2y = 6$. We also need to evaluate the vector field on the surface S.

$$\begin{aligned}
\mathbf{F}|_S &= \langle x, xy, x(6-3x-2y)\rangle,\\
d\mathbf{S} &= \langle 3,2,1\rangle\, dy\, dx,\\
\Phi &= \iint_R \mathbf{F}\cdot d\mathbf{S},\\
&= \int_0^2 \int_0^{3-\frac{3}{2}x} [3x+2xy+x(6-3x-2y)]\, dy\, dx,\\
&= \int_0^2 \int_0^{3-\frac{3}{2}x} (9x-3x^2)\, dy\, dx,\\
&= \int_0^2 (3-\tfrac{3}{2}x)(9x-3x^2)\, dx,\\
&= \tfrac{9}{2}\int_0^2 (6-5x^2+x^3)\, dx,\\
&= \tfrac{9}{2}\left[3x^2-\tfrac{5}{2}x^3+\tfrac{1}{4}x^4\right]_0^2 = 12
\end{aligned}$$

5.6.8 Example Flux across a Cone

We compute the flux of the vector field $\mathbf{F} = \langle xy, 0, z\rangle$ across the part S of the cone $z = \sqrt{x^2+y^2}$ $x \leq 1$ with the natural orientation.

Solution. There is nothing special about this surface integral. It just illustrates that direct integration always works, provided one has enough temperance to work out the double integrals. We start with the differential dz and follow up by an easy mental computation of the wedges.

$$\begin{aligned}
dz &= \frac{x}{\sqrt{x^2+y^2}}\, dx + \frac{y}{\sqrt{x^2+y^2}}\, dy,\\
d\mathbf{S} &= \langle \frac{-x}{\sqrt{x^2+y^2}}, \frac{-y}{\sqrt{x^2+y^2}}, 1\rangle\, dx\, dy,\\
\mathbf{F}|_S &= \langle xy, 0, -\sqrt{x^2+y^2}\rangle,\\
\Phi &= \int\int_S \left(\frac{-x^2y}{\sqrt{x^2+y^2}} - \sqrt{x^2+y^2}\right) dx\, dy,\\
&= \int_0^{2\pi}\int_0^1 (-r^2\cos^2\theta\sin\theta - r)(r\, dr\, d\theta),\\
&= -(2\pi)\left[\tfrac{1}{3}r^3\right]_0^1 = -\tfrac{2}{3}\pi.
\end{aligned}$$

The first integral makes no contribution over a full cycle of the $(\cos^3\theta)$ function.

5.6.9 Example Find the flux of a vector field over a closed surface
Solution. First learn Gauss's theorem in section 5.7.

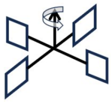
5.6.3 Curl and Circulation

Consider a positively oriented, simple closed curve $\{C : \mathbf{r}(t)\}$. Let $\mathbf{v}$ the velocity vector and $\mathbf{T}$ the unit tangent vector. Think of the vector field $\mathbf{v}$ as representing the velocity field of a fluid. The line integral

$$\oint \mathbf{v} \cdot d\mathbf{r} = \oint \mathbf{v} \cdot \mathbf{T}\, ds$$

represents the rotation tangentially around the curve. The line integral is therefore called the **circulation**.

We use Stoke's theorem to show the relation between the curl of a vector field and the circulation. The figure here illustrates the notion of circulation. If $\mathbf{v}$ represents the velocity field of a fluid, then the field has a circulation if $\oint_C \mathbf{v} \cdot d\mathbf{r}$ is not zero.



Let C be the boundary of region S. The region could be any region in $\mathbf{R}^3$, but we might as well assume the region is the simplest one that has the same boundary. By Stokes' theorem, we have

$$\oint_C \mathbf{v} \cdot d\mathbf{r} = \iint_S (\nabla \times \mathbf{v}) \cdot d\mathbf{S}.$$

To find the circulation at a point P consider the curve C to be a very small circle C_ϵ of radius R_ϵ, centered at P. If ϵ is sufficiently small, the vector field will not vary much within the disk enclosed by C_ϵ so the curl of the vector field is approximately a constant. Hence

$$\begin{aligned}\iint_S (\nabla \times \mathbf{v}) \cdot d\mathbf{S} &= \iint_S (\nabla \times \mathbf{v}) \cdot \mathbf{n}\, dS, \\ &\approx (\nabla \times \mathbf{v}) \cdot \mathbf{n}) \iint_S dS, \\ &\approx [(\nabla \times \mathbf{v}) \cdot \mathbf{n}](\pi R_\epsilon^2).\end{aligned}$$

In the limit as $\epsilon \to 0$, we get

$$[(\nabla \times \mathbf{v}) \cdot \mathbf{n})](P) = \lim_{\epsilon \to 0} \frac{1}{(\pi R_\epsilon^2)} \oint_C \mathbf{v} \cdot d\mathbf{r}. \tag{5.29}$$

This means that the curl of the velocity vector is a measure of the circulation of the fluid at the point P.

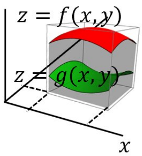
5.7 Gauss' Theorem

5.7.1 Gauss' Theorem Let V is a simple solid region with a piecewise smooth boundary $S = \delta V$ with outward orientation. Let F be a vector field with continuously differentiable components in an open region that contains V. then,

$$\iint_S \mathbf{F}\cdot d\mathbf{S} = \iiint_V \nabla\cdot\mathbf{F}\,dV. \tag{5.30}$$

Proof To prove the theorem we first state it in its full components, namely

$$\iint_S P\,dy\,dz - Q\,dx\,dz + R\,dx\,dy = \iiint_V (P_x + Q_y + R_z)\,dV$$



The proof of Gauss' theorem here follows the same mode of thinking as was done for the proof of Green's theorem. We star with a proof that $\iint R\,dx\,dy = \iiint R_z\,dV$. For this, take the special case where the surface is the boundary of a volume over a rectangle D enclosed between the functions $z = f(x,y)$ and $z = g(x,y)$. The idea is to compute the surface and the volume integral and show they are the same. For the vertical walls, either $dx = 0$ or $dy = 0$, so the only contribution to the flux is the z component of the flux corresponding to $d\mathbf{S} = dx\,dy\,\mathbf{k}$.

$$\Phi_z = \iint_S R\,dx\wedge dy = \iint_D [R(x,y,f(x,y)) - R(x,y,f(x,y)]\,dx\,dy.$$

On the other hand,

$$\begin{aligned}\iiint_V \frac{\partial R}{\partial z}\,dz\,dy\,dx &= \iint_D \left(\int_{g(x,y)}^{f(x,y)} \frac{\partial R}{\partial z}\,dz\right)dy\,dx,\\ &= \iint_D [R(x,y,f(x,y)) - R(x,y,f(x,y)]\,dx\,dy.\end{aligned}$$

By taking special regions over rectangles over the other two coordinate planes, we can establish the corresponding parts of the theorem

$$\begin{aligned}\Phi_x &= \iint P\,dy\,dz = \iiint P_x\,dV,\\ \Phi_y &= -\iint Q\,dx\,dz = \iiint Q_y\,dV.\end{aligned}$$

For general volumes bounded by simple closed surfaces, we dice the solid into small regions of each of the three types and then on argues that all the interior wall surface integrals cancel out, because in an interior cube, there is an equal flux in each direction.

The physics about the theorem is most intuitive once we understand that the divergence of a vector field is associated with the existence of sources or sinks. If we think of a vector field $\mathbf{F}$ as flow, and $\nabla\cdot\mathbf{F} = 0$ inside a closed surface such as a sphere, then there are no sources or sinks inside the region and what ever flows in, must flow out. Thus, the net flux is zero. On the other hand, if $\nabla\cdot\mathbf{F}$ is not zero, then the volume integral generates a net flux $\Phi = \iint_S \mathbf{F}\cdot d\mathbf{S}$. A more quantitative explanation appears in section 5.8.

5.7.2 Example Find $\Phi = \iint_S \mathbf{F}\cdot d\mathbf{S}$, where $\mathbf{F} = y^4\,\mathbf{i} + 2x^3\,\mathbf{j} + 3z\,\mathbf{k}$, and S is the unit sphere $x^2 + y^2 + z^2 = 1$.

Solution. The sphere is a closed surface and it bounds a solid sphere V, so Gauss' theorem applies. Hence

$$\begin{aligned}\iint_S \mathbf{F}\cdot d\mathbf{S} &= \iiint_V \nabla\cdot\mathbf{F}\,dV,\\ &= \iiint_V 3\,dV,\\ &= 3(\tfrac{4}{3}\pi 1^3) = 4\pi.\end{aligned}$$

Yes, this one is as simple as that. Computing the surface integral would have been laborious leading to a proliferation of integrals of power of sines and cosines stemming from the parametrization of the sphere.

5.7.3 Example Find $\Phi = \iint_S \mathbf{F}\cdot d\mathbf{S}$, where $\mathbf{F} = y\,\mathbf{i} + z\,\mathbf{j} + z\,\mathbf{k}$ and S is the boundary of the solid region V bounded above by $z = 1 - x^2 - y^2$ and below by the plane $z = 0$.
Solution. Since $S = \delta V$ is the boundary of a volume, it is a closed surface and Gauss' theorem applies. We get,

$$\begin{aligned}\iint_S \mathbf{F}\cdot d\mathbf{S} &= \iiint_V \nabla\cdot\mathbf{F}\,dV,\\ &= \iiint_V 1\,dV,\\ &= \int_0^{2\pi}\int_0^1\int_0^{1-r^2} r\,dz\,dr\,d\theta,\\ &= \int_0^{2\pi}\int_0^1 (r - r^3)\,dr\,d\theta,\\ &= 2\pi(\tfrac{1}{2} - \tfrac{1}{4}) = \tfrac{\pi}{2}.\end{aligned}$$

5.7.4 Example Find the flux $\Phi = \iint_S \mathbf{F}\cdot dS$, where

$$\mathbf{F} = \frac{\mathbf{r}}{r^3} = \frac{\langle x, y, z\rangle}{(x^2+y^2+z^2)^{3/2}},\quad S : 0 < a^2 \le x^2 + y^2 + z^2 \le b^2.$$

Solution. The surface is the boundary of the solid region V between two hemispheres, so Gauss's theorem applies. Computing the the divergence, we get

$$\begin{aligned}\nabla\cdot\mathbf{F} &= \left(\frac{1}{r^3} - \frac{3x^2}{r^5}\right) + \left(\frac{1}{r^3} - \frac{3y^2}{r^5}\right) + \left(\frac{1}{r^3} - \frac{3z^2}{r^5}\right),\\ &= \frac{3}{r^3} - \frac{3}{r^5}(x^2+y^2+z^2),\\ &= \frac{3}{r^3} - \frac{3r^2}{r^5} = 0,\\ \Phi &= \int\int\int_V \nabla\cdot\mathbf{F}\,dV = 0\end{aligned}$$

The vector field, represents a $1/r^2$ point source, but the source is outside the region bounded inside the two hemispheres, so the total outward flux is zero. Whatever "goo" flows in, flows out.

5.7.5 Example Find the flux $\Phi = \iint_S \mathbf{F}\cdot dS$, where,

$$\mathbf{F} = (5x^3 + 12xy^2)\,\mathbf{i} + (y^3 + e^y \sin z)\,\mathbf{j} + (5z^3 + e^y \cos z)\,\mathbf{j},$$

and S is the boundary of the solid region $1 \leq x^2 + y^2 + z^2 \leq 2$.
Solution

$$\begin{aligned} \nabla\cdot\mathbf{F} &= (15x^2 + 12y^2) + (3y^2 + e^y \sin z) + (15z^2 - e^y \sin z), \\ &= 15x^2 + 15y^2 + 15z^2 = 15r^2 \end{aligned}$$

By Gauss's theorem,

$$\begin{aligned} \Phi &= \iiint_V 15r^2(r^2 \sin\theta\, dr\, d\theta\, d\phi), \\ &= \int_0^{2\pi}\int_0^{\pi}\int_1^{2} r^4 \sin\theta\, dr\, d\theta\, d\phi, \\ &= -2\pi(\tfrac{2^5}{5} - \tfrac{1^5}{5})\,\cos\theta\,|_0^{\pi}\,, \\ &= -2\pi\tfrac{31}{5}(-1-1) = \tfrac{124}{5}\pi \end{aligned}$$

5.7.6 Example <u>Gauss's Law</u>. The first equation of Maxwell for electric fields in Gaussian units reads

$$\nabla\cdot\mathbf{E} = 4\pi\rho,$$

where ρ is the charge density per unit volume. In elementary physics one gets away with not doing a fancy flux integral by assuming the source is a point charge. Then we can enclose the total charge Q centered a at the origin by a sphere $S:\ x^2 + y^2 + z^2 = r^2$. By symmetry, the electric field $\mathbf{E}$ is normal to the sphere and it has constant strength $E = \mathbf{E}\cdot\mathbf{n}$ on the sphere. We can easily compute the flux by Gauss' theorem,

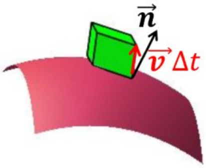
$$\begin{aligned} \iiint_V \nabla\cdot\mathbf{E}\, dV &= \iint\int_V 4\pi\rho\, dV, \\ \iint_S \mathbf{E}\cdot d\mathbf{S} &= 4\pi Q, \\ \iint_S \mathbf{E}\cdot\mathbf{n}\, dS &= 4\pi Q, \\ E\iint_S dS &= 4\pi Q, \\ E(4\pi r^2) &= 4\pi Q, \\ E &= \frac{Q}{r^2} \end{aligned}$$

This is the $1/r^2$ law.

5.8 Vector Theorems in Physics

5.8.1 Continuity Equation

In this section we elaborate more on the assertion that the divergence of a vector field has something to do with sources. We do this in the context of hydrodynamics of smooth flow of vector fields.



Let $\mathbf{v}$ represent the velocity vector field of a fluid of density ρ. Suppose the fluid is flowing through a surface S. As it is usual in calculus we zoom to a very small rectangle centered at a point P on the surface. We consider the flux of the vector field $\mathbf{F} = \rho\mathbf{v}$. Denote the unit normal to the surface by the vector $\mathbf{n}$. Since the region is very small, as in example 5.6.2, the amount of fluid passing through the rectangle of area ΔS in time Δt is given by the volume of the parallelepiped with base ΔS and height equal to the projection of $\mathbf{v}\,\Delta t$ onto the normal

$$\Delta V = \mathbf{v}\,\Delta t \cdot \mathbf{n}\,\Delta S.$$

Recalling that density is mass per unit volume. that is $\Delta m = \rho\,\Delta V$, we get

$$\Delta m = \rho\mathbf{v} \cdot \mathbf{n}\,\Delta t\,\Delta S.$$

Now, with the magic of Riemann sums, we add all the contributions and take the limit as the increments go to zero. The sums become integrals and the result is

$$\begin{aligned} \frac{dm}{dt} &= \iint_S \rho\mathbf{v} \cdot \mathbf{n}\,dS, \\ &= \iint_S \rho\mathbf{v} \cdot d\mathbf{S}, \\ &= \iint_S \mathbf{F} \cdot d\mathbf{S}. \end{aligned}$$

Now, take a small sphere B_ϵ, with radius r_ϵ. Let V_ϵ be the volume of the sphere. The average value of $\nabla \cdot \mathbf{F}$ over the sphere is

$$\frac{1}{V_\epsilon} \iiint_{B_\epsilon} \nabla \cdot \mathbf{F}\,dV.$$

The divergence of the vector field is a continuous function, so it must attain the average value at at last one point $P \in B_\epsilon$,

$$\begin{aligned} (\nabla \cdot \mathbf{F})_P &= \frac{1}{V_\epsilon} \iiint_{B_\epsilon} \nabla \cdot \mathbf{F}\,dV, \\ &= \frac{1}{V_\epsilon} \iint_{S_\epsilon} \mathbf{F} \cdot d\mathbf{S} \end{aligned}$$

The right hand side represents the rate of decrease of mass per unit volume. Now we take the limit as ϵ goes to 0. The left hand side approaches $(\nabla \cdot \mathbf{F})_P$.

We get

$$
\begin{aligned}
(\nabla \cdot \mathbf{F})_P &= -\frac{\partial \rho}{\partial t}, \\
\nabla \cdot (\rho \mathbf{v}) + \frac{\partial \rho}{\partial t} &= 0.
\end{aligned}
\tag{5.31}
$$

This is the hydrodynamics continuity equation. The divergence represents a flow through the surface, but this can only happen at the expense of the rate of change of the matter inside the volume. The conclusion is that Gauss's theorem,

$$
\iint_S \mathbf{F} \cdot d\mathbf{S} = \iiint_V \nabla \cdot \mathbf{F}\, dV,
$$

at least in the context of fluids, is a statement of conservation of mass.

5.8.2 Continuity Equation in E & M

The complete set of Maxwell's equations in Gaussian units is

$$
\begin{aligned}
\nabla \cdot \mathbf{E} &= 4\pi\rho, & & & \nabla \cdot \mathbf{B} &= 0, \\
\nabla \times \mathbf{E} &= -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, & &\text{and} & \nabla \times \mathbf{B} &= \frac{4\pi}{c}\mathbf{J} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t},
\end{aligned}
$$

where ρ is the charge density $\rho = dQ/dV$, and $\mathbf{J}$ is the charge current density per unit area passing through a surface. In other words, if we integrate $\mathbf{J}$ over a surface Σ we get the net electric current on the surface

$$
I_\Sigma = \int\int_\Sigma \mathbf{J} \cdot d\mathbf{S}
$$

Now, let's take the divergence of the curl of $\mathbf{B}$,

$$
\begin{aligned}
\nabla \cdot (\nabla \times \mathbf{B}) &= \frac{4\pi}{c}\nabla \cdot \mathbf{J} + \frac{1}{c}\nabla \cdot \frac{\partial \mathbf{E}}{\partial t}, \\
0 &= \frac{4\pi}{c}\nabla \cdot \mathbf{J} + \frac{1}{c}\frac{\partial}{\partial t}(\nabla \cdot \mathbf{E}), \\
0 &= \frac{4\pi}{c}\nabla \cdot \mathbf{J} + \frac{4\pi}{c}\frac{\partial \rho}{\partial t}.
\end{aligned}
$$

Hence

$$
\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \tag{5.32}
$$

This is the continuity equation for the electric field. If we integrate over a volume V whose boundary is a closed surface $S = \delta V$ and apply Gauss's theorem, we get

$$
\begin{aligned}
\iiint_V \nabla \cdot \mathbf{J}\, dV &= -\iiint_V \frac{\partial \rho}{\partial t}\, dV, \\
\iint_S \mathbf{J} \cdot d\mathbf{S} &= -\frac{d}{dt}\iiint_V \rho\, dV, \\
I_S &= -\frac{d}{dt}Q_S
\end{aligned}
$$

The physics is that the net rate of change of the charge over the volume V is the negative of the net current flowing through the boundary δV.

5.8.3 Green's Identities

Let f and g be smooth functions (continuous second partial derivatives suffices), and S be a simple closed surface that is the boundary of a volume V. The surface integral

$$\iint_S f\nabla g\cdot d\mathbf{S}=\iint_S f\nabla g\cdot\mathbf{n}dS$$

is often written

$$\iint_S f\nabla g\cdot d\mathbf{S}=\iint_S f\frac{\partial g}{\partial\mathbf{n}}dS$$

because $\nabla g\cdot\mathbf{n}$ is the directional derivative of g in the direction of the normal. The notation is meant be read as the normal derivative of g. Applying Gauss's theorem, we get

$$\begin{aligned}\iint_S f\nabla g\cdot d\mathbf{S}&=\iiint_S \nabla\cdot(f\nabla g)\,dV,\\ &=\iiint_S [f\nabla\cdot(\nabla g)+\nabla f\cdot\nabla g]\,dV,\\ \iint_S f\nabla g\cdot d\mathbf{S}&=\iiint_S (f\nabla^2 g+\nabla f\cdot\nabla g)\,dV.\end{aligned}\tag{5.33}$$

Equation 5.33 is called **Green's first identity**
If we interchange f and g in 5.33 and subtract, we get

$$\begin{aligned}\iint_S f\nabla g\cdot d\mathbf{S}&=\iiint_S (f\nabla^2 g+\nabla f\cdot\nabla g)\,dV,\\ \iint_S g\nabla f\cdot d\mathbf{S}&=\iiint_S (g\nabla^2 f+\nabla g\cdot\nabla f)\,dV\\ \iint_S (f\nabla g-g\nabla f)\cdot d\mathbf{S}&=\iiint_S (f\nabla^2 g-g\nabla^2 f)\,dV\end{aligned}\tag{5.34}$$

Equation 5.34 is called **Green's second identity**. I am not sure I should call Green's identities an application in physics, because it is just as well an application to partial differential equations. With the addition Green's third identity which we do not include here, we are lead to a neat process to treat solutions of equations that involve the Laplacian. The general procedure is called the method of Green's functions. More appetizers to pursue further acquisition of knowledge. There are analogous formulas in other dimensions, so the method can be made quite general.

5.9 General Stokes' Theorem in $\mathbf{R}^3$

As an appetizer for continuing the pursuit of knowledge, we present a tantalizing result that consolidate the three major vector integration theorems into a single one. Let f by a smooth function and $\mathbf{F} = \langle P, Q, R\rangle$ be a smooth vector field in $\mathbf{R}^3$. Following equations 5.3 in connection with line integrals and 5.27 in connection we surface integrals, we extend the definition of forms to also include the vector integrands in one and three dimensions. That is,

$$\begin{array}{ll} \text{0-form} & \omega = f\,\mathbf{1}, \\ \text{1-form} & \omega = \mathbf{F}\cdot d\mathbf{r} = P\,dx + Q\,dy + R\,dz, \\ \text{2-form} & \omega = \mathbf{F}\cdot d\mathbf{S} = P\,dy\wedge dz - Q\,dx\wedge dz + R\,dx\wedge dy, \\ \text{3-form} & \omega = f\,dx\wedge dy\wedge dz. \end{array}$$

To keep track of where things live, we introduce the following notation for the spaces of forms in $\mathbf{R}^3$

$$\begin{aligned} \mathbf{\Lambda}^0(\mathbf{R}^3) = \mathbf{\Lambda}^0 = &\quad \text{Space of 0-forms}, \\ \mathbf{\Lambda}^1(\mathbf{R}^3) = \mathbf{\Lambda}^1 = &\quad \text{Space of 1-forms}, \\ \mathbf{\Lambda}^2(\mathbf{R}^3) = \mathbf{\Lambda}^2 = &\quad \text{Space of 2-forms}, \\ \mathbf{\Lambda}^3(\mathbf{R}^3) = \mathbf{\Lambda}^3 = &\quad \text{Space of 2-forms}, \end{aligned}$$

There are no 4-forms in $\mathbf{R}^3$ because that would require 4 wedges and two of the them would necessarily be repeated. The key to consolidation is a simple extension of the notion of differential. Recall the definition of differential of a 0-form, which we may think as a map from 0-forms to 1-forms.

$$df = f_x\,dx + f_y\,dy + f_z\,dz = \nabla f\cdot d\mathbf{r}.$$

We can write the fundamental theorem of line integrals (essentially the fundamental theorem of calculus) on terms of forms. Let C be a curve with end points at A and B. We say that the boundary δC of the curve is the set of points $\{A, B\}$. If $\mathbf{F} = \nabla f$, then $\mathbf{F}\cdot d\mathbf{r} = df$, and we have

$$W = f\,|_{\delta C} = \int_C df. \tag{5.35}$$

Just as reminder, this says that for a conservative vector field, the work line integral is path independent.

We now define define the extended differential

$$\begin{aligned} &d : \mathbf{\Lambda}^k \longrightarrow \mathbf{\Lambda}^{k+1}, \\ &d(f\,\omega) = df\wedge\omega, \quad \text{where,} \quad \omega\in\mathbf{\Lambda}^k \end{aligned} \tag{5.36}$$

We assume that d satisfies the usual linearity properties

$$\begin{aligned} d(k\,\omega) &= k\,d\omega, \quad \text{for } k = \text{constant}, \\ d(\omega_1 \pm \omega_2) &= d\omega_1 \pm d\omega_2 \quad \text{where } \alpha, \beta\in\mathbf{\Lambda}^k. \end{aligned}$$

Example Let $\omega = x^2y^3\,dx + 3x\,dy$. Then

$$\begin{aligned} d\omega &= d(x^2y^3)\ \wedge dx + d(3x) \wedge dy, \\ &= (2xy^3\,dx + 3x^2y^3\,dy) \wedge dx + 3\,dx \wedge dy, \\ &= 2xy^3\,dx \wedge dy + 3\,dx \wedge dy, \\ &= (2xy^3 + 3)\,dx \wedge dy. \end{aligned}$$

Easy. If one can compute the differential of a function and one knows the antisymmetry rule for the multiplication of dx, dy and dz, one can compute the differential of any k-form. Specifically, let us compute the differential,

$$\begin{aligned} \omega &= \mathbf{F} \cdot d\mathbf{r} = P\,dx + Q\,dy + R\,dz, \\ d\omega &= dP \wedge dx + dQ \wedge dy + +dR \wedge dz, \\ &= (P_x\,dx + P_y\,dy + P_z\,dz) \wedge dx + \\ &\quad (Q_x\,dx + Q_y\,dy + Q_z\,dz) \wedge dy + \\ &\quad (R_x\,dx + R_y\,dy + R_z\,dz) \wedge dz. \end{aligned}$$

Now we pick out the coefficients of the differential of surface 2-forms $(dy \wedge dz)$, $(dx \wedge dz)$ and $(dx \wedge dy)$, respectively. In doing so we need to be careful that the order of the basis wedge products are as stated. We get

$$d\omega = (R_y - Q_z)\,dy \wedge dz - (P_z - R_x)\,dx \wedge dz + (Q_x - P_y)\,dx \wedge dy. \qquad (5.37)$$

We recognize the the coefficients are precisely the components of $\nabla \times \mathbf{F}$. In terms of the dot product, $d\omega$ is the integrand of Stokes' theorem 5.19. Hence, we have just shown that if $C = \delta R$ is a simple closed curve that is the boundary of some surface R, Stokes' theorem can be rewritten as

$$\oint_{\delta R} \omega = \iint_R d\omega. \qquad (5.38)$$

Finally, we compute the differential of the flux 2-form 5.26. This one is actually much easier since up to permutations, there is one one 3-form, so we only have to pick out the terms that have distinct variables in the wedges.

$$\begin{aligned} \omega &= \mathbf{F} \cdot d\mathbf{S} = P\,dy \wedge dz - Q\,dx \wedge dz + R\,dx \wedge dz, \\ d\omega &= dP \wedge dy \wedge dz - dQ \wedge dx \wedge dz + dR \wedge dx \wedge dy, \\ &= P_x\,dx \wedge dy \wedge dz - Q_y\,dy \wedge dx \wedge dz + R_z\,dz \wedge dx \wedge dz, \\ &= (P_x + Q_y + R_z, dx \wedge dy \wedge dz, \\ &= \nabla \cdot \mathbf{F}\,dV. \end{aligned}$$

Thus, in terms of forms, if a simple close surface S is the boundary of some region $S = \delta R$, Gauss's divergence theorem reads

$$\iint_{\delta R} \omega = \iiint_R d\omega. \qquad (5.39)$$

We see that the gradient, curl and divergence are three manifestation of the same differential operator d depending on the order of the differential form on which it acts. We can replace the Del operator diagram by 5.7

$$\mathbf{\Lambda}^0(\mathbf{R}^3) \xrightarrow[\text{Grad}]{d} \mathbf{\Lambda}^1(\mathbf{R}^3) \xrightarrow[\text{Curl}]{d} \mathbf{\Lambda}^2(\mathbf{R}^3) \xrightarrow[\text{Div}f]{d} \mathbf{\Lambda}^3(\mathbf{R}^3). \tag{5.40}$$

If one applies d twice to a form, the answer is 0. That is, theorem 5.4.1 can be summarized as follows

$$(d \circ d)\, \omega = 0 \tag{5.41}$$

At the end there is really one theorem - Stokes' theorem, which we symbolically write as

$$\int_{\delta R} \omega = \iint_R d\omega. \tag{5.42}$$

where the integral of $d\omega$ is a single, double or triple integral, depending on the dimension of the space.

5.10 Summary of Vector Integrals

5.10.1 Hints on Line Integrals

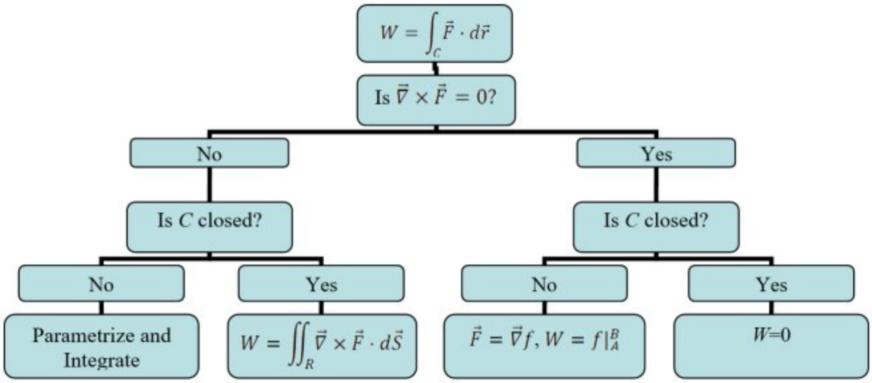


Fig. 5.2: Line Integrals Flow Chart

The computation of a work line integrals is facilitated if the answer is yes to either of two binary questions: a) Is the curl of the vector field zero? b) Is the curve closed? As a result, there are essentially four possible scenarios.

- Case 1. The curl is not zero and the curve is not closed. The integral is done directly from the definition $W = \int_C \mathbf{F}(\mathbf{r}(t))\dot{\mathbf{r}}\,dt$. We call this the "parametrize and integrate method". It is presumed that the student has mastered the equations of the basic parametric curves if not given explicitly in the problem. This is the kind of problem students would call "plug'n-chug". The integrals are reduced to integral of a single variable

- Case 2. The curl is not zero but the curve is closed. Apply Stoke's theorem. Chose simplest surface with same boundary. The differential of surface can be computed by two general but equivalent methods:

  a) $d\mathbf{S} = dy\,dz\,\mathbf{i} - dx\,dz\,\mathbf{j} + dx\,dy\,\mathbf{k}$. Best method if the curve is a plane curve and parallel to a coordinate plane. Remember that the products or differentials are really wedges.

  b) $d\mathbf{S} = (\mathbf{r}_u \times \mathbf{r}_v)\,du\,dv$.

- Case 3. The curl is zero but the curve is not closed. The vector field is conservative, so the integral os path independent. Find a potential $\mathbf{F} = \nabla f$. Then the work integral is $f(B) - f(A)$, where A and B are the beginning and end points of the curve respectively.

- Case 4. The curl is zero and the curve is closed. The work integral is zero.

5.10.2 Hints on Surface Integrals

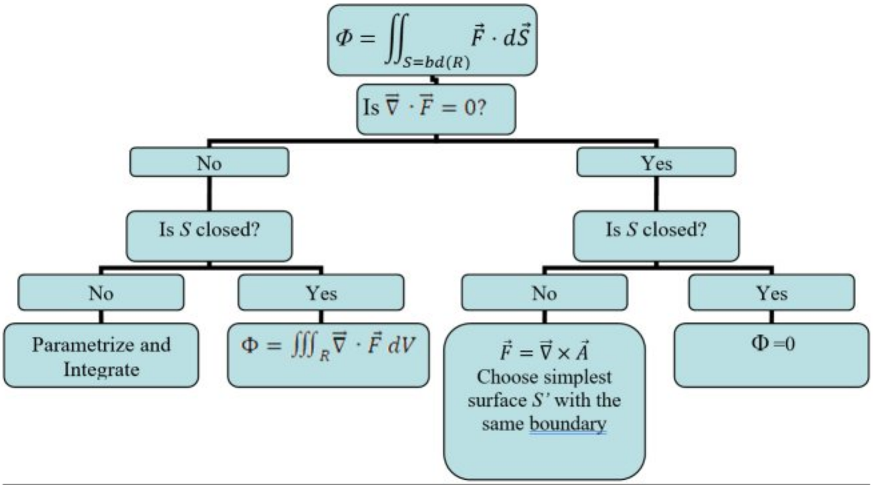


Fig. 5.3: Surface Integrals Flow Chart

The computation of a flux surface integrals is facilitated if the answer is yes to either of two binary questions: a) Is the divergence of the vector field zero? b) Is the surface closed? As a result, there are essentially four possible scenarios.

- Case 1. The divergence is not zero and the surface is not closed. The integral is done directly from the definition $W = \int_C \mathbf{F} \cdot d\mathbf{S}$. We call this the "parametrize and integrate method". It is presumed that the student is well acquainted with parametric equations of surfaces if not given explicitly in the problem. If the surface is of the form $z = f(x, y)$ one can parametrize by $\mathbf{r} = \langle x, y, f(x, y)\rangle$ This is the also kind of problem students would call "plug'n-chug". The differential of surface can be computed by two general but equivalent methods:

  a) $d\mathbf{S} = dy\, dz\, \mathbf{i} - dx\, dz\, \mathbf{j} + dx\, dy\, \mathbf{k}$. Best method if the curve is a plane curve and parallel to a coordinate plane. Remember that the products or differentials are really wedges.

  b) $d\mathbf{S} = (\mathbf{r}_u \times \mathbf{r}_v)\, du\, dv$.

  The integrals are reduced to a double integral

- Case 2. The divergence is not zero but the surface is closed. Apply Gauss's theorem.

- Case 3. The divergence is zero but the surface is not closed. The vector field is incompressible, so the integral depends only on the boundary. The analogous idea of finding vector a potential $\mathbf{F} = \nabla \times \mathbf{A}$ is not practical. If the integrand is already of he form $\nabla \times \mathbf{A}$, then this is a Stokes' theorem problem

- Case 4. The divergence is zero and the surface is closed. The work integral is zero.

Index