Calculus of Higher Dimension

James S. Cook
Liberty University
Department of Mathematics

Fall 2025 and beyond

preface

There are two primary audiences for which these notes are intended:

1. students enrolled in my Math 231 residential course through Liberty University,
2. students interested in self-study and/or preparation for some equivalency test.

Most of the course-specific comments in these notes are clearly intended for (1.). Futhermore, the comments concerning study-habits are also targeted primarily at (1.). Clearly, if you are studying this course on your own then some of my comments intended to push you to review basic content are probably out of place. In general, please keep in mind as you read these notes, the comments about individual study habits and understanding may not be relevant to you. Thus, when faced with such a superfluous comment, simply read on. I merely make comments to reach certain subsets of the spectrum of students who might benefit from various advice.

how to succeed in calculus

From past experience I can tell you that the students who excelled in my course were those students who both studied my notes and completed most of the homework. They also came to every class and paid attention. I recommend the following course of study:

1. submit yourself to learn, keep a positive attitude. This course is a lot of work. Yes, probably more than 3 others for most people. Most people have a lot of work to do in getting up to speed on real mathematical thinking. There is no substitute for time and effort. If you're complaining in your mind about the workload etc... then you're wasting your time.
2. come to class, take notes, **think**.
3. read these notes.
4. attempt the homework, you will likely find forming a study group is essential for success here. I ask some difficult questions. On occasion, an open-ended question appears. In other words, the homework is often more difficult than the test. Note, most of the end-of-chapter Problems have solutions posted at www.supermath.info. Completion of the end-of-chapter problems is a good starting point for self-study in this course.

educational philosophy

My educational philosophy is markedly different than some educators. I believe that the strongest students are as important to challenge as the weakest students. It has become increasingly in fashion to focus on the lowest quintile of students as we design future courses. We (instructors) belabor questions such as:

1. how can we get students to read the text ?
2. how can we get students to study before the test?
3. how can we get students to see their weakness before it's too late?

I don't dismiss such questions. In fact, I seek to answer all of these with homework paired with the accountability of quizzes and tests. However, these questions are of secondary importance to questions such as:

1. how can we get students to think critically ?
2. how can we get students to understand our work in a larger context ?
3. how can we get students to build a foundation of knowledge which extends across the entirety of their core course work?

The difference between these questions and the former set is that the latter focuses on the content and coherence of mathematics assimilated by the student whereas the former is simply focused on getting the student to study as an adult with a modicum of common sense. Well, I refuse to make the focus of my life's work to be about the mollification of sloth. I am interested in seeing my students succeed in the broadest sense. If I satiate bad study habits by making time to do homework in class and broadcasting exactly what is on the test before the test then I may increase the retention rate in this course and I may have more students make good grades. But, at what cost? I want this course to help you towards your independence both as a student and as a worker. Ideally, I encourage you to realize your potential. I assume you are an adult and you ought not expect I teach you as a child. You are not a child and I refuse to treat you as such. That said, please ask for help as you need it. I do not refuse help to those who have already tried to help themselves. I have numerous office hours and good questions are often welcome in lecture.

format of my notes

These notes were prepared with LaTeX. You'll notice a number of standard conventions in my notes:

1. definitions are usually in green.
2. remarks are in red.
3. theorems, propositions, lemmas and corollaries are in blue.
4. proofs start with a **Proof:** and are concluded with a $\square$.
5. often figures in these notes were prepared with **Graph**, a simple and free math graphing program, or **Maple**, or **Mathematica**. Or some online math tool, of which there are dozens.

By now the abbreviations below should be old news, but to be safe I replicate them here once more:

Notation	Meaning of Notation
§	section
$\exists$	there exists
$\nexists$	there does not exist
w.r.t.	with respect to
l.h.s.	left hand side
r.h.s.	right hand side
$x \in B$	the element x is inside the set B
$A \implies B$	A implies B
$A \iff B$	A and B are equivalent statements
$\therefore$	therefore

Notation	Meaning of Notation
$\forall$	for all
$\approx$	approximately
$eq^{\underline{n}}$	equation
$sol^{\underline{n}}$	solution
$\mathbb{N}$	natural numbers; 1,2,3,...
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\mathbb{C}$	complex numbers
$\mathbb{Z}$	integers
$\mathbb{R}^2$	the Cartesian plane

In some sense these lecture notes resemble a text. However, it is best to read these in parallel with a standard text. The standard texts have a a few thousand homework problems which I lack the patience to to include in these notes. In contrast to some "lecture notes" these are not a minimal collection of salient examples. In fact, there are far too many examples for lecture. My goal here is to provide the reader a rather large set of worked examples and a presentation of the theory of multivariable calculus which supercedes typical mainstream texts.

My understanding of multivariate calculus was forged in the fire of my undergraduate physics courses. I do try to include a good part of our University Physics course in our work this semester. It does not take much time since basic physical laws are naturally phrased in the language of this course; vectors and vector calculus.

Please understand that I do assume you have a complete and working knowledge of Calculus I and II.

The old notes and much more can be found at my calculus III webpage. Every so often I mention the existence of an animation on my webpage. You can find a zoo of gif-files where you can go see all sorts of mathematical creatures. If you create interesting creatures I will (with your permission) happily add them to my collection.

sources

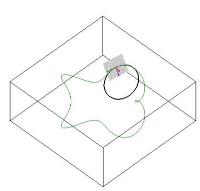
There are many mainstream calculus texts which do a reasonable job of presenting the computational core of multivariate calculus. Certainly Stewart's and Thomas' texts have had considerable influence both in what they cover as well as what they left out. When I initially rewrote these notes we were using Salas and Hille for calculus at Liberty University thus I have made a fair number of references to that text in this edition. The main additions outside the mainstream of calculus are inspired in part from my instruction of Advanced Calculus from C.H. Edwards excellent text. I have made a point to mention texts as I use them in the body of the text so there is currently no global bibliography. Finally, I should thank my students and my teachers as they both have encouraged me to think.

status and plans

Last, I should caution the reader these notes are a work in progress. If you find an error then please send me an email so I can fix it. I do appreciate your input.

The final Chapter on Differential Forms should be added later in the term.
James Cook, July 14, 2025.

In the diagram below (created with Maple) we see a curve as well as the infinitesimal plane of motion with the osculating circle attached. See my multivariable calculus website for an annoying background gif which puts this in motion. Notice I intend to collect material specific to MH 408-61 both in Blackboard as well as at course website for MH 408-61 Calculus of Higher Dimension.



Contents

Chapter 1

analytic geometry

Euclidean space is named for the ancient mathematician Euclid. In Euclid's view geometry was a formal system with axioms and constructions. For example, the fact two parallel lines never interect is called the *parallel postulate*. If you take the course in modern geometry then you'll study more of the history of Euclid. Fortunately for us the axiomatic/constructive approach was replaced by a far easier view about 400 years ago. Rene Descartes made popular the idea of using numbers to label points in space. In this new Cartesian geometry the fact two parallel lines in a plane never intersect could be checked by some simple algebraic calculation. More than this, all sorts of curves, surfaces and even more fantastic objects became constructible in terms of a few simple algebraic equations. The idea of using numbers to label points and the resulting geometry governed by the analysis of those numbers is called **analytic geometry**. We study the basics of analytic geometry in this chapter.

In your previous mathematical studies you have already dealt with analytic geometry in the plane. Trigonometry provided a natural framework to decipher all sorts of interesting facts about triangles. Moreover, the study of trigonometric functions in turn has allowed us solutions to otherwise intractable integrals in calculus II. Trigonometric substitution is an example of where the geometry of triangles has allowed deeper analysis. It goes both ways. Geometry inspires analysis and analysis unravels geometry. These are two sides of something deeper.

In this course we need to tackle three dimensional problems. The proper notation which groups together concepts in the most efficient and clean manner is the vector notation. Historically, it was predated by the quaternionic analysis of Hamilton, but for about 120 year the vector notation has been the dominant framework for two and three dimensional analytic geometry[1]. In particular, the dot and cross products allow us to test for how parallel two lines are, or to project a line onto a plane, or even to calculate the direction which is perpendicular to a pair of given directions. Engineering and basic everyday physics are all written in this vector language.

We also continue our study of functions in this chapter. We have studied functions $f : U \subseteq \mathbb{R} \to \mathbb{R}$ in the first two semesters of calculus. One goal in this course is to extend the analysis to **functions of many variables**. For example, $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x, y) = x^2 + y^2$. What can we say about this function? What calculus is known to analyze the properties of f? Before we begin to answer such questions in future chapters, we need to spend some time on the basic geometry of $\mathbb{R}^n$ and

[1]general geometries are more naturally understood in the language of differential forms and manifolds, but this is where we all must begin

then especially $\mathbb{R}^3$. Naturally, we use $\mathbb{R}^3$ to model the three spatial dimensions which frame our everyday existence[2].

In this chapter we are concerned with understanding how to analytically describe points, curves and surfaces. We will examine what the solution set of $z = f(x, y)$ looks like for various f. Or, what is the solution set of $F(x, y) = k$, or $F(x, y, z) = k$? We learn how think about mappings $t \mapsto \langle x(t), y(t), z(t) \rangle$ or $(u, v) \mapsto \langle x(u, v), y(u, v), z(u, v) \rangle$. What is the geometry of such mappings? These are questions we hope to answer, at least in part, in this chapter.

1.1 vectors in euclidean space

This is a rather lengthy section because I wanted to group all the basic concepts and formulas for vectors in this section. We begin with the geometric meaning of vectors. In the first subsection we define two, three and n-dimensional space. We discuss the interchange of vectors and points as well as the standard formulas for distance between points. Also, the tip-to-tail vector addition is given both geometrically and later as a careful algebraic definition. The distinction between scalar and vector components with respect to Cartesian coordinates is detailed. In the second subsection the dot-product is introduced. This operation plays a central role in many applications throughout the course. The dot-product gives a convenient formula to calculate vector length and hence the distance between two points. We introduce the unit-vector which gives us a vector-formula to back-up the assertion that a non-zero vector is given by a magnitude and a direction. Properties for dot-products as well as norms are given and a number are proved in an efficient index notation. Finally, orthogonality as well as orthonormal-decompositions are introduced. In the third subsection, the dot-product is utilized to provide a simple method to calculate angles between nonzero vectors in 2, 3 or even n-dimensions. We see how the Law of Cosines is implicitly contained within the mathematics of dot-products. Finally, in the fourth subsection we study projections onto vectors and simple planes. This section is a simple application of orthonormality and the geometry of unit-vectors. The application to less trivial planes later in the course is covered in Section 1.3.5.

[2] I would not say we live in $\mathbb{R}^3$, it's just a model, it's not reality. Respected philosophers as recently as 200 years ago labored under the delusion that euclidean space must be reality since that was all they could imagine as reasonable

1.1.1 geometry of vectors and points

We denote the real numbers as $\mathbb{R} = \mathbb{R}^1$. Naturally $\mathbb{R}$ is identified with a line as we are taught in our previous study of the number line. The *Cartesian products* of $\mathbb{R}$ with itself give us natural models for the plane, 3 dimensional space and more abstractly n-dimensional space:

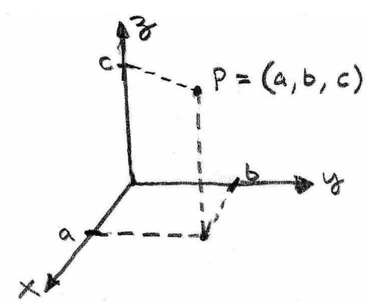
Definition 1.1.1. *.*

1. **two-dimensional space:** is the set of all **ordered pairs** of real numbers:
$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) \mid x, y \in \mathbb{R}\}$$
2. **three-dimensional space:** is the set of all **ordered triples** of real numbers:
$$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$$
3. **n-dimensional space:** is the set of all n**-tuples** of real numbers:
$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \ copies} = \{(x_1, x_2, \dots, x_n) \mid x_j \in \mathbb{R} \ for \ each \ j \in \mathbb{N}_n\}$$

The fact that the n-tuples above are ordered means that two n-tuples are equal iff each and every entry in the n-tuple matches.

Definition 1.1.2. *vector equality, components.*

In particular, $(v_1, v_2, \dots, v_n) = (w_1, w_2, \dots, w_n)$ iff $v_1 = w_1, \ v_2 = w_2, \ \dots, v_n = w_n$. In the context of $\mathbb{R}^2$ we say a is the **x-component** of (a, b) whereas b is the **y-component** of (a, b). In the context of $\mathbb{R}^3$ we say a is the **x-component** of (a, b, c) whereas b is the **y-component** of (a, b, c) and c is the **z-component** of (a, b, c). Generally, we say v_j the j**-th component** of $(v_1, v_2, \dots, v_n)$.

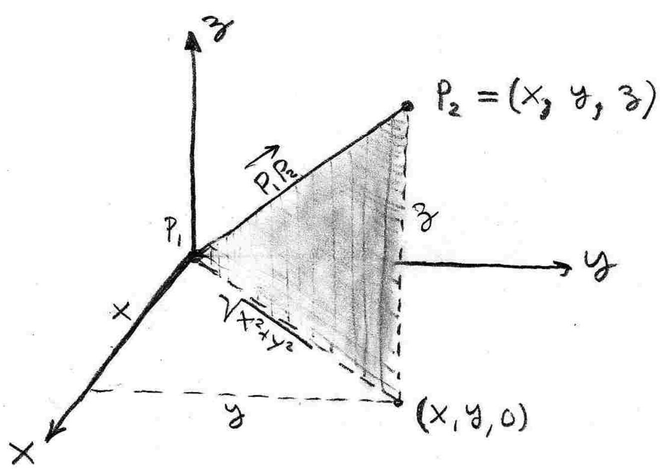


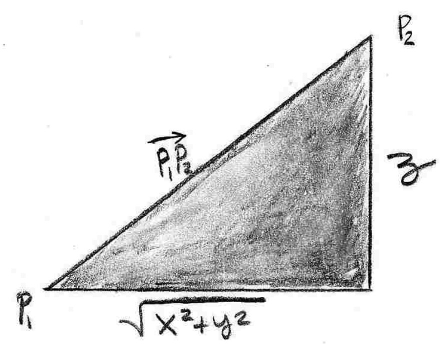
Sometimes the term **euclidean** is added to emphasize that we suppose distance between points is measured in the usual manner. Recall that in the one-dimensional case the distance between $x, y \in \mathbb{R}$ is given by the absolute value function; $d(x, y) = |y - x| = \sqrt{(y - x)^2}$. We define distance in n-dimensions by similar formulas:

Definition 1.1.3. *euclidean distance.*

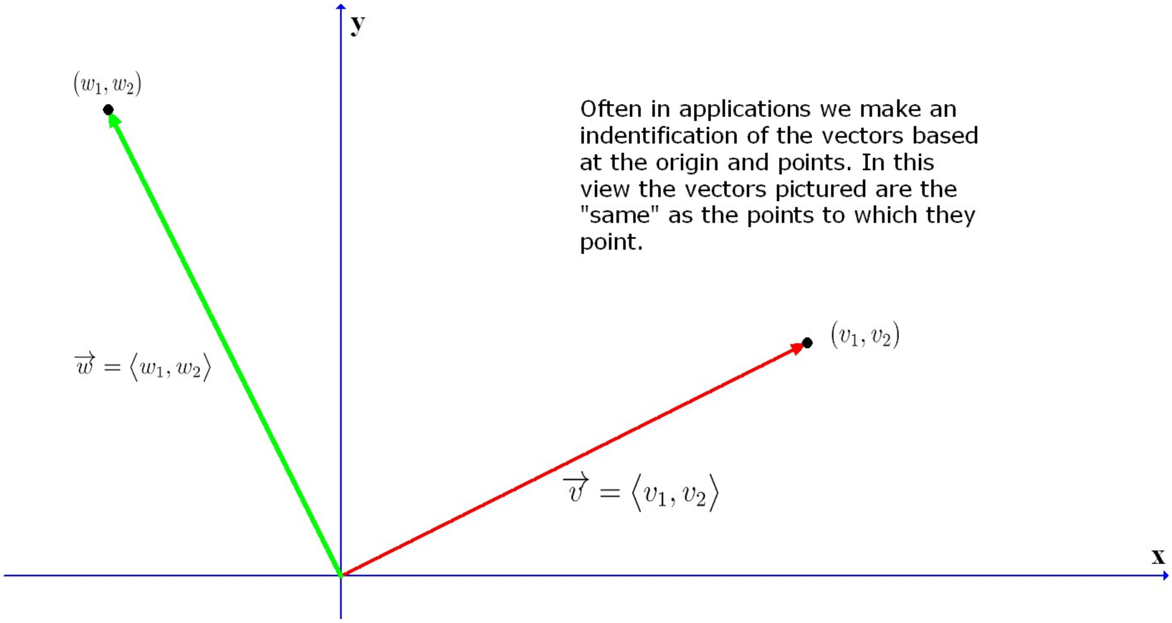
1. **distance in two-dimensional euclidean space:** if $p_1 = (x_1, y_1), p_2 = (x_2, y_2) \in \mathbb{R}^2$ then the distance between points p_1 and p_2 is
$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$
2. **distance in three-dimensional euclidean space:** if $p_1 = (x_1, y_1, z_1), p_2 = (x_2, y_2, z_2) \in \mathbb{R}^3$ then the distance between points p_1 and p_2 is
$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$
3. **distance in n-dimensional euclidean space:** if $a, b \in \mathbb{R}^n$ where $a = (a_1, a_2, \dots, a_n)$ and $b = (b_1, b_2, \dots, b_n)$ then the distance between points a and point b is
$$d(a, b) = \sqrt{\sum_{j=1}^{n} (b_j - a_j)^2} = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}.$$

It is simple to verify that the definition above squares with our traditional ideas about distance from previous math courses. In particular, notice these follow from the Pythagorean theorem applied to appropriate triangles. The picture below shows the three dimensional distance formula is consistent with the two dimensional formula.

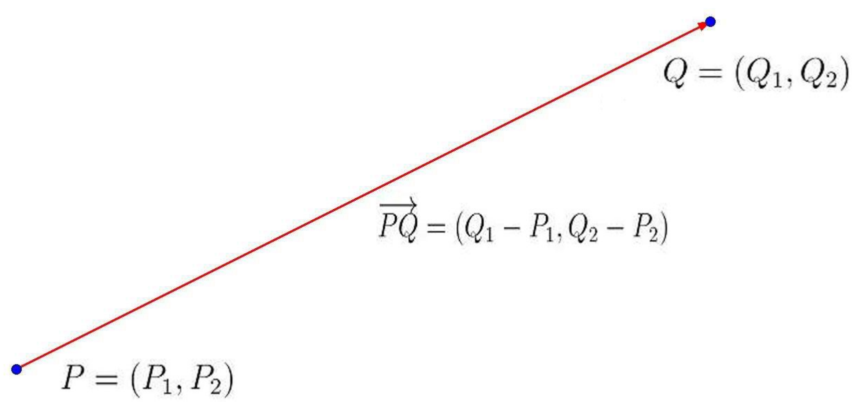




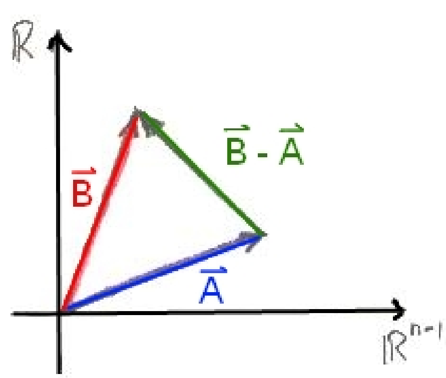
Notice that there is a natural correspondence between points and directed line-segments from the origin. We can view an n-tuple p as either representing the point $(p_1, p_2, \dots, p_n)$ or the directed line-segment from the origin $(0, 0, \dots, 0)$ to the point $(p_1, p_2, \dots, p_n)$.



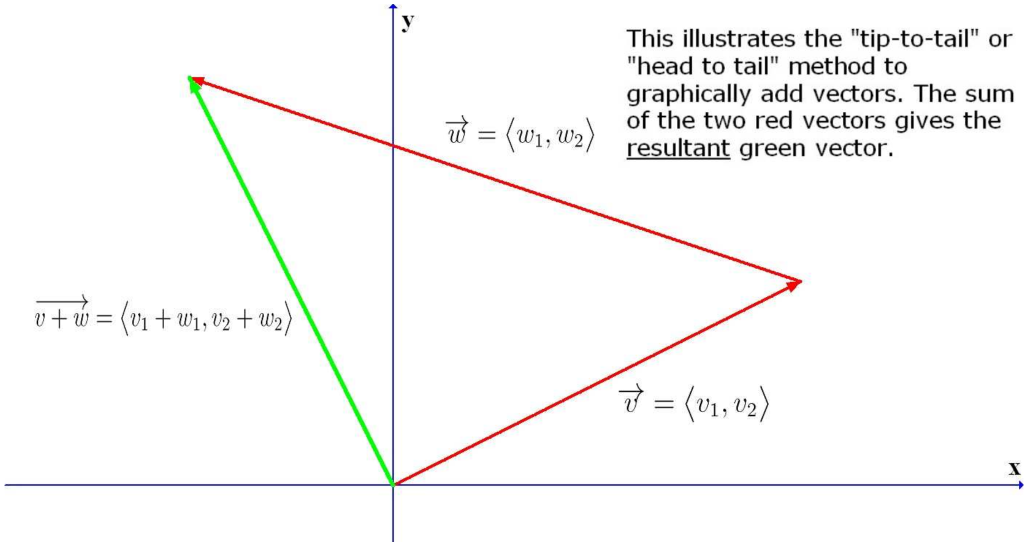
We will use the notation $\vec{p}$ for n-tuples throughout the remainder of these notes to emphasize the fact that $\vec{p}$ is a vector. Some texts use **bold** to denote vectors, but I prefer the over-arrow notation which is easily duplicated in hand-written work. The directed line-segment from point P to point Q is naturally identified with the vector $\vec{Q} - \vec{P}$ as illustrated below:



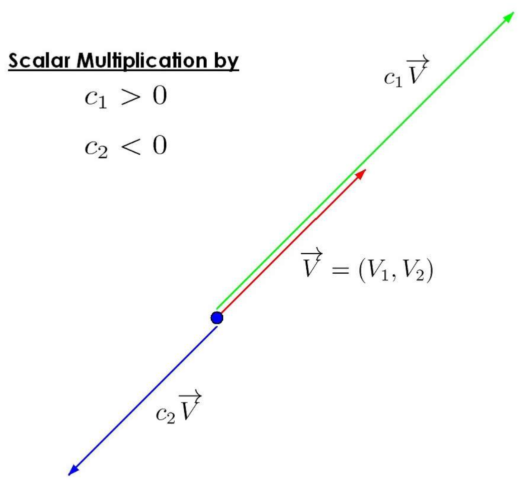
This is consistent with the identification of points and vectors based at the origin. See how the vector $\vec{a}$ and $\vec{b}$ are connected by the vector $\vec{b} - \vec{a}$



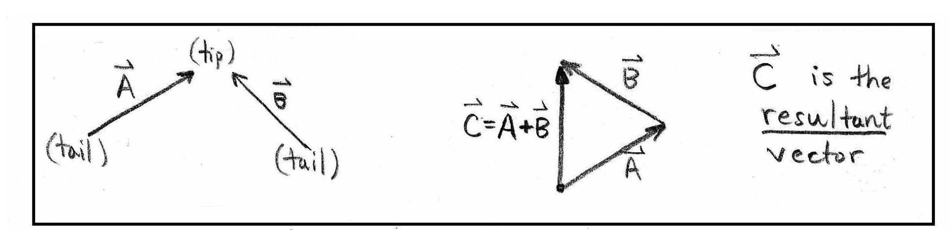
We add vectors geometrically by the tip-to-tail method as illustrated below.

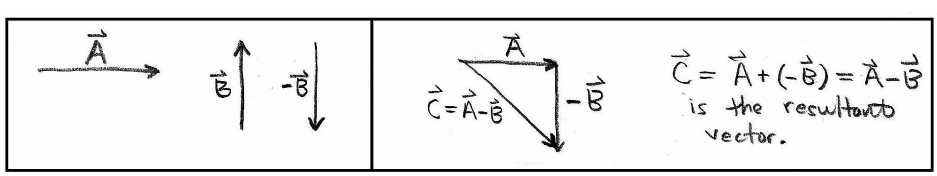


Also, we rescale them by shrinking or stretching their length by a scalar multiple:



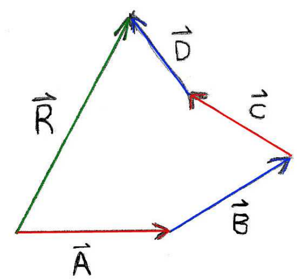
More pictures usually help:





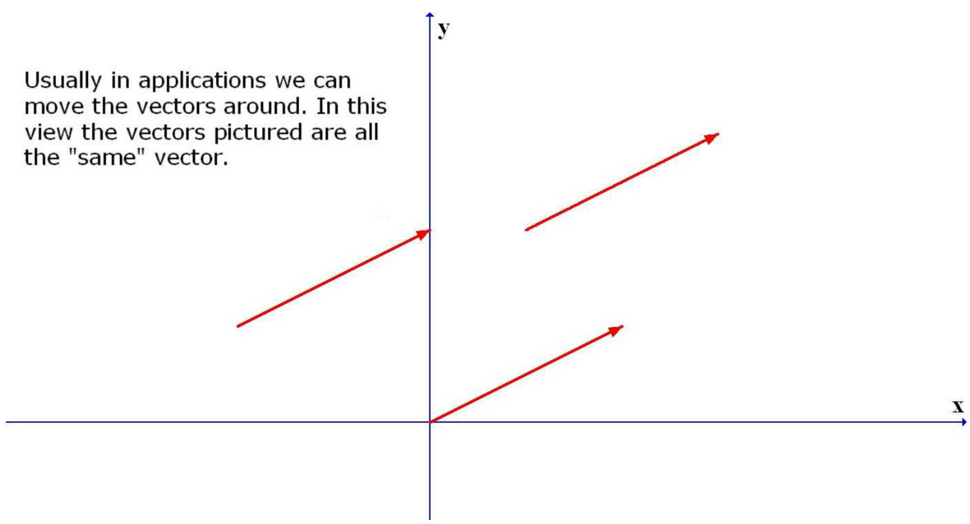
In the diagram below we illustrate the geometry behind the vector equation

$$\vec{R} = \vec{A} + \vec{B} + \vec{C} + \vec{D}.$$



Continuing in this way we can add any finite number of vectors in the same *tip-2-tail* fashion.

One important point to notice here is that we can naturally move vectors from the origin to other points. Moreover, some people define vectors from the outset as directed-line-segments between points. In particular, the vector from A to B in $\mathbb{R}^n$ is denoted $\overrightarrow{AB}$ and is defined by $\overrightarrow{AB} = B - A$. It is natural to suppose the vector $\overrightarrow{AB}$ is based at A, however, we can equally well picture the vector $\overrightarrow{AB}$ based at any point in $\mathbb{R}^n$.



If we wish to keep track of the base point of vectors then additional notation is required. We could say that $(p, \vec{V})$ denotes a vector $\vec{V}$ which has basepoint p. Then addition, scalar multiplication, dot-products and vector lengths are all naturally defined for such objects. You just do those operations to the vector $\vec{V}$ and the point rides along. We will not use this notation in this course. Instead, we will use words or pictures to indicate where a given vector is based. Sometimes vectors are based at the origin, sometimes not. Sorry if this is confusing, but this is the custom of almost all authors and if I invent notation and am more careful on this point then I'm afraid I may put you at a disadvantage in other courses.

Algebraically, vector addition and scalar multiplication are easier to summarize concisely:

Definition 1.1.4. *vector addition and scalar multiplication.*

in $\mathbb{R}^2$,

$$\begin{aligned} \langle x_1,\ y_1\rangle + \langle x_2,\ y_2\rangle &= \langle x_1+x_2,\ y_1+y_2\rangle \\ c\langle x_1,\ y_1\rangle &= \langle cx_1,\ cy_1\rangle \end{aligned}$$

Or for $\mathbb{R}^3$,

$$\begin{aligned} \langle x_1,\ y_1,\ z_1\rangle + \langle x_2,\ y_2,\ z_2\rangle &= \langle x_1+x_2,\ y_1+y_2,\ z_1+z_2\rangle \\ c\langle x_1,\ y_1,\ z_1\rangle &= \langle cx_1,\ cy_1,\ cz_1\rangle. \end{aligned}$$

Generally, for $\vec{x}, \vec{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ we define $\vec{x}+\vec{y}$ and $c\vec{x}$ component-wise as follows:

$$(\ \vec{x}+\vec{y}\ )_j = \vec{x}_j + \vec{y}_j$$

$$(\ c\vec{x}\ )_j = c\vec{x}_j$$

for $j = 1, 2, \dots, n$.

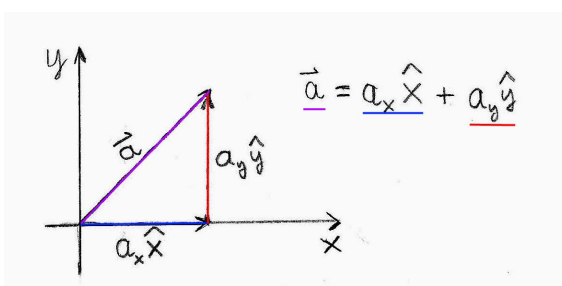
Given these definitions it is often convenient to break a vector down into its **vector components**. In particular, for $\mathbb{R}^2$, **define** $\widehat{x} = \langle 1, 0\rangle$ **and** $\widehat{y} = \langle 0, 1\rangle$ hence:

$$\begin{aligned} \langle a, b\rangle &= \langle a, 0\rangle + \langle 0, b\rangle \\ &= a\langle 1, 0\rangle + \ b\langle 0, 1\rangle \\ &= a\,\widehat{x} + \ b\,\widehat{y} \end{aligned} \tag{1.1}$$

Definition 1.1.5. *vector and scalar components of two-vectors.*

The **vector component** of $\langle a, b\rangle$ in the x-direction[3] is simply $a\,\widehat{x}$ whereas the **vector component** of $\langle a, b\rangle$ in the y-direction is simply $b\,\widehat{y}$. In contrast, a, b are the **scalar components** of $\langle a, b\rangle$ in the x, y-directions respective.

Scalar components are scalars whereas vector components are vectors. These are entirely different objects if $n \neq 1$, please keep clear this distinction in your mind. Notice that the vector components are what we use to reproduce a given vector by the tip-to-tail sum:



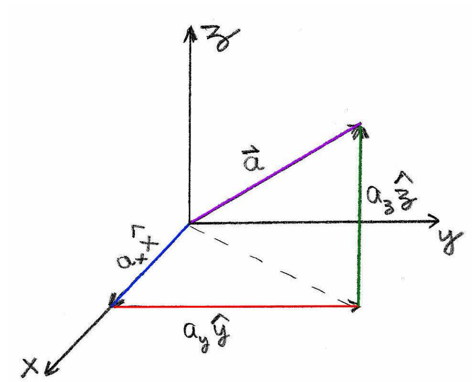
For $\mathbb{R}^3$ we define the following notation: $\widehat{x} = \langle 1, 0, 0\rangle$**,** $\widehat{y} = \langle 0, 1, 0\rangle$**, and** $\widehat{z} = \langle 0, 0, 1\rangle$ hence:

$$\begin{aligned} \langle a, b, c\rangle &= \langle a, 0, 0\rangle + \langle 0, b, 0\rangle + \langle 0, 0, c\rangle \\ &= a\langle 1, 0, 0\rangle + \ b\langle 0, 1, 0\rangle + \ c\langle 0, 0, 1\rangle \\ &= a\,\widehat{x} + \ b\,\widehat{y} + \ c\,\widehat{z} \end{aligned}$$

Definition 1.1.6. *vector and scalar components of three-vectors.*

> The **vector components** of $\langle a, b, c\rangle$ are: $a\,\widehat{x}$ in the x-direction[4], $b\,\widehat{y}$ in the y-direction and $c\,\widehat{z}$ in the z-direction. In contrast, a, b, c are the **scalar components** of $\langle a, b, c\rangle$ in the x, y, z-directions respective.

Again, I emphasize that vector components are vectors whereas components or scalar components are by default scalars.



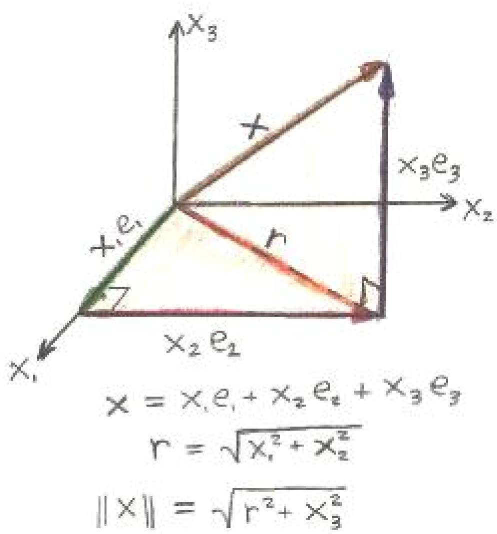
The story for $\mathbb{R}^n$ is not much different. **Define for** $j = 1, 2, \dots, n$ **the vector** $\widehat{x}_j = \langle 0, 0, \dots, 1, \dots, 0\rangle$ **where the** 1 **appears in the** j**-th entry**, hence:

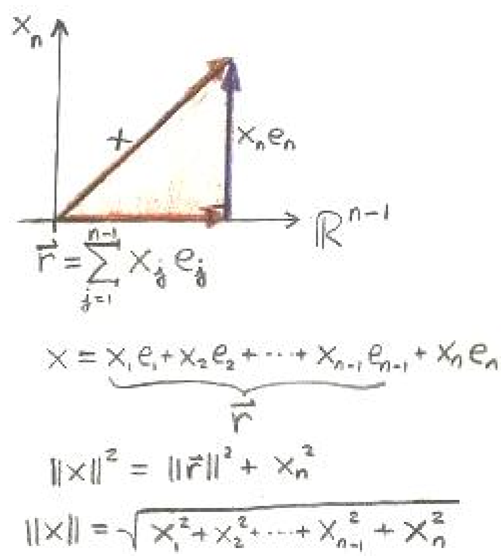
$$\begin{aligned} \langle a_1, a_2, \dots, a_n\rangle &= \langle a_1, 0, \dots, 0\rangle + \langle 0, a_2, \dots, 0\rangle + \cdots + \langle 0, 0, \dots, a_n\rangle \\ &= a_1\langle 1, 0, \dots, 0\rangle + a_2\langle 0, 1, \dots, 0\rangle + \cdots + a_n\langle 0, 0, \dots, 1\rangle \\ &= a_1\,\widehat{x}_1 + \; a_2\,\widehat{x}_2 + \cdots + a_n\,\widehat{x}_n. \end{aligned} \tag{1.2}$$

Definition 1.1.7. *vector and scalar components of n-vectors.*

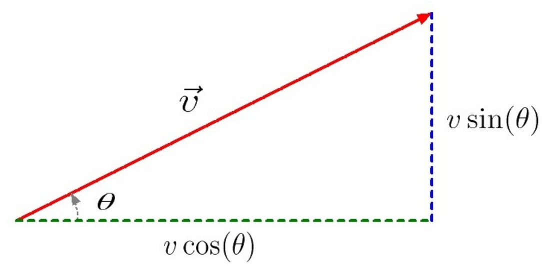
> If $\vec{v} = \langle a_1, a_2, \dots, a_n\rangle$ then $a_j\,\widehat{x}_j$ is the **vector component** in the x_j-direction of $\vec{v}$ whereas a_j is the **scalar component** in the x_j-direction of $\vec{v}$.

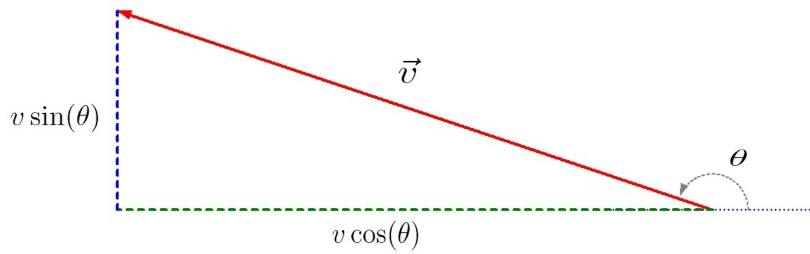
Here's an attempt at the picture for $n > 3$ (I use the linear algebra notation of $e_1 = \widehat{x}$ etc...):





Trigonometry is often useful in applied problems. It is not uncommon to be faced with vectors which are described by a length and a direction in the plane. In such a case we need to rely on trigonometry to *break-down* the vector into it's Cartesian components.





Example 1.1.8. *Suppose a vector $\vec{v}$ has a length $v = 5$ at $\theta = \pi/3$ then $v\cos\theta = 5\cos(\pi/3) = \frac{5}{2}$ and $v\sin\theta = 5\sin(\pi/3) = \frac{5\sqrt{3}}{2}$. Therefore, in view of the diagram above this example, $\vec{v} = \langle \frac{5}{2}, \frac{5\sqrt{3}}{2}\rangle$.*

Example 1.1.9. *Suppose a vector $\vec{v}$ has a length $v = 2$ at $\theta = 4\pi/3$ then $v\cos\theta = 2\cos(4\pi/3) = -1$ and $v\sin\theta = 2\sin(4\pi/3) = \frac{-2\sqrt{3}}{2} = -\sqrt{3}$. Therefore, in view of the diagram above this example, $\vec{v} = \langle -1, -\sqrt{3}\rangle$. Notice, $\theta = 4\pi/3$ is in Quadrant III where both x and y-components are negative.*

1.1.2 dot products

Definition 1.1.10. *dot product.*

The **dot-product** is a useful operation on vectors. In $\mathbb{R}^2$ we define,

$$\langle V_1, V_2\rangle \bullet \langle W_1, W_2\rangle = V_1W_1 + V_2W_2.$$

In $\mathbb{R}^3$ we define,

$$\langle V_1, V_2, V_3\rangle \bullet \langle W_1, W_2, W_3\rangle = V_1W_1 + V_2W_2 + V_3W_3.$$

In $\mathbb{R}^n$ we define, for $\vec{V} = \sum_{j=1}^n V_j\,\widehat{x}_j$ and $\vec{W} = \sum_{j=1}^n W_j\,\widehat{x}_j$

$$\vec{V} \bullet \vec{W} = V_1W_1 + V_2W_2 + \cdots + V_nW_n.$$

It is important to notice that the dot-product takes in two *vectors* and outputs a *scalar*. It has a number of interesting properties which we will often use:

Example 1.1.11. *Let $\vec{A} = \langle 1, 2, 3\rangle$ and $\vec{B} = \langle 1, -1, 5\rangle$. We calculate,*

$$\vec{A} \bullet \vec{B} = \langle 1, 2, 3\rangle \bullet \langle 1, -1, 5\rangle = 1 - 2 + 15 = 14.$$

Proposition 1.1.12. *properties of the dot-product.*

let $\vec{A}, \vec{B}, \vec{C} \in \mathbb{R}^n$ be vectors and $c \in \mathbb{R}$

1. **commutative:** $\vec{A} \bullet \vec{B} = \vec{B} \bullet \vec{A}$,
2. **distributive:** $\vec{A} \bullet (\vec{B} + \vec{C}) = \vec{A} \bullet \vec{B} + \vec{A} \bullet \vec{C}$,
3. **distributive:** $(\vec{A} + \vec{B}) \bullet \vec{C} = \vec{A} \bullet \vec{C} + \vec{B} \bullet \vec{C}$,
4. **scalars factor out:** $\vec{A} \bullet (c\vec{B}) = (c\vec{A}) \bullet \vec{B} = c\vec{A} \bullet \vec{B}$,
5. **non-negative:** $\vec{A} \bullet \vec{A} \geq 0$,
6. **no null-vectors:** $\vec{A} \bullet \vec{A} = 0$ iff $\vec{A} = 0$.

Proof: The proof of these properties is simple if we use the right notation. Observe

$$\vec{A} \bullet \vec{B} = \sum_{j=1}^{n} A_j B_j = \sum_{j=1}^{n} B_j A_j = \vec{B} \bullet \vec{A}.$$

Thus the dot-product is commutative. Next, note that $(\vec{B} + \vec{C})_j = B_j + C_j$ hence,

$$\vec{A} \bullet (\vec{B} + \vec{C}) = \sum_{j=1}^{n} A_j (B_j + C_j) = \sum_{j=1}^{n} A_j B_j + \sum_{j=1}^{n} A_j C_j = \vec{A} \bullet \vec{B} + \vec{A} \bullet \vec{C}.$$

The proof of item (3.) actually follows from the commutative property and the right-distributive property we just proved since

$$(\vec{A} + \vec{B}) \bullet \vec{C} = \vec{C} \bullet (\vec{A} + \vec{B}) = \vec{C} \bullet \vec{A} + \vec{C} \bullet \vec{B} = \vec{A} \bullet \vec{C} + \vec{B} \bullet \vec{C}.$$

The proof of (4.) is left to the reader. Continue to (5.), note that

$$\vec{A} \bullet \vec{A} = \sum_{j=1}^{n} A_j A_j = A_1^2 + A_2^2 + \cdots + A_n^2$$

hence it is clear that $\vec{A} \bullet \vec{A}$ is the sum of squares of real numbers and consequently $\vec{A} \bullet \vec{A} \geq 0$. Moreover, if $\vec{A} \bullet \vec{A} = 0$ and $\vec{A} \neq 0$ then there must exist at least one component, say $A_j \neq 0$ hence $\vec{A} \bullet \vec{A} \geq A_j^2 > 0$ which is a contradiction. Therefore, (6.) follows. $\square$

The length of a vector $\vec{A}$ is simply the distance from the origin to the point which the vector points. In particular, we denote the length of the vector $\vec{A}$ by $||\vec{A}||$ and it's clear from the formula in the proof for $\vec{A} \bullet \vec{A}$ that

$$||\vec{A}|| = \sqrt{\vec{A} \bullet \vec{A}}$$

this formula holds for $\mathbb{R}^n$. Sometimes the length of the vector $\vec{A}$ is also called the *norm* of $\vec{A}$. The norm also has interesting properties which are quite similar to those which are known for the absolute value function on $\mathbb{R}$ (in fact, $||x|| = |x|$ for $x \in \mathbb{R}$).

Proposition 1.1.13. *properties of the norm (also known as length of vector).*

Suppose $\vec{A}, \vec{B} \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

1. **absolute value of scalar factors out:** $||c\vec{A}|| = |c|||\vec{A}||$,
2. **triangle inequality:** $||\vec{A} + \vec{B}|| \leq ||\vec{A}|| + ||\vec{B}||$,
3. **Cauchy-Schwarz inequality:** $|\vec{A} \bullet \vec{B}| \leq ||\vec{A}|| \, ||\vec{B}||$.
4. **non-negative:** $||\vec{A}|| \geq 0$,
5. **only zero vector has zero length:** $||\vec{A}|| = 0$ iff $\vec{A} = 0$.

Proof: The proof of (1.) is simple,

$$||c\vec{A}|| = \sqrt{(c\vec{A}) \bullet (c\vec{A})} = \sqrt{c^2 \vec{A} \bullet \vec{A}} = \sqrt{c^2}\sqrt{\vec{A} \bullet \vec{A}} = |c| \; ||\vec{A}||.$$

To prove the triangle Let $\vec{x}, \vec{y} \in \mathbb{R}^n$,

$$\begin{aligned}
||\vec{x} + \vec{y}||^2 &= |(\vec{x} + \vec{y}) \bullet (\vec{x} + \vec{y})| && \text{defn. of norm} \\
&= |\vec{x} \bullet (\vec{x} + \vec{y}) + \vec{y} \bullet (\vec{x} + \vec{y})| && \text{prop. of dot-product} \\
&= |\vec{x} \bullet \vec{x} + \vec{x} \bullet \vec{y} + \vec{y} \bullet \vec{x} + \vec{y} \bullet \vec{y}| && \text{prop. of dot-product} \\
&= | \; ||\vec{x}||^2 + 2\vec{x} \bullet \vec{y} + ||\vec{y}||^2 \; | && \text{prop. of dot-product} \\
&\leq ||\vec{x}||^2 + 2|\vec{x} \bullet \vec{y}| + ||\vec{y}||^2 && \text{triangle ineq. for } \mathbb{R} \\
&\leq ||\vec{x}||^2 + 2||\vec{x}|| \, ||\vec{y}|| + ||\vec{y}||^2 && \text{Cauchy-Schwarz ineq.} \\
&\leq (||\vec{x}|| + ||\vec{y}||)^2 && \text{algebra}
\end{aligned}$$

Notice that both $||\vec{x}+\vec{y}||$ and $||\vec{x}||+||\vec{y}||$ are nonnegative hence the inequality above yields $||\vec{x}+\vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||$. Continue to item (3.). Let $\vec{x}, \vec{y} \in \mathbb{R}^n$. If either $\vec{x} = 0$ or $\vec{y} = 0$ then the inequality is clearly true. Suppose then that both $\vec{x}$ and $\vec{y}$ are nonzero vectors. It follows that $||\vec{x}||, ||\vec{y}|| \neq 0$ and we can define vectors of unit-length; $\hat{x} = \frac{\vec{x}}{||\vec{x}||}$ and $\hat{y} = \frac{\vec{y}}{||\vec{y}||}$. Notice that $\hat{x} \bullet \hat{x} = \frac{\vec{x}}{||\vec{x}||} \bullet \frac{\vec{x}}{||\vec{x}||} = \frac{1}{||\vec{x}||^2}\vec{x} \bullet \vec{x} = \frac{\vec{x} \bullet \vec{x}}{\vec{x} \bullet \vec{x}} = 1$ and likewise $\hat{y} \bullet \hat{y} = 1$. Consider,

$$\begin{aligned}
0 \leq ||\hat{x} \pm \hat{y}||^2 &= (\hat{x} \pm \hat{y}) \bullet (\hat{x} \pm \hat{y}) \\
&= \hat{x} \bullet \hat{x} \pm 2(\hat{x} \bullet \hat{y}) + \hat{y} \bullet \hat{y} \\
&= 2 \pm 2(\hat{x} \bullet \hat{y}) \\
&\Rightarrow \quad -2 \leq \pm 2(\hat{x} \bullet \hat{y}) \\
&\Rightarrow \quad \pm \hat{x} \bullet \hat{y} \leq 1 \\
&\Rightarrow \quad |\hat{x} \bullet \hat{y}| \leq 1
\end{aligned}$$

Therefore, noting that $\vec{x} = ||\vec{x}||\hat{x}$ and $\vec{y} = ||\vec{y}||\hat{y}$,

$$|\vec{x} \bullet \vec{y}| = | \; ||\vec{x}||\hat{x} \bullet ||\vec{y}||\hat{y} \; | = ||\vec{x}|| \, ||\vec{y}|| \, |\hat{x} \bullet \hat{y}| \leq ||\vec{x}|| \, ||\vec{y}|| \; \square.$$

As I introduced in the proof above[5], if a vector has a length of one then it is called a **unit-vector**.

[5] I happened to find this argument in Insel, Spence and Friedberg's undergraduate linear algebra text.

Definition 1.1.14. *unit vectors.*

If $\vec{A} \neq 0$ then we define the **direction vector** of $\vec{A}$ as follows:

$$\widehat{A} = \frac{1}{||\vec{A}||}\vec{A}.$$

I invite the reader to check that $||\widehat{A}|| = 1$. Moreover, we should observe that any nonzero vector can be written as the product of its unit-vector $\widehat{A}$ and its length $||\vec{A}||$:

$$\vec{A} = ||\vec{A}||\widehat{A}$$

When it is convenient and unambiguous we use the notation $||\vec{A}|| = A$ and it follows $\vec{A} = A\widehat{A}$. This formula expresses the vector as the product of its magnitude A and its direction $\widehat{A}$. In the two-dimensional case it is possible to uniquely describe the direction in terms of the standard angle from polar coordinates.

Example 1.1.15. *Let $\vec{A} = \langle 3, 4\rangle$ then $||A|| = \sqrt{3^2+4^2} = \sqrt{25} = 5$. Therefore, $\widehat{A} = \frac{1}{5}\langle 3, 4\rangle$.*

Example 1.1.16. *Find a vector $\vec{B}$ with length 7 and the same direction as $\vec{A} = \langle 1, 1\rangle$. Observe $\widehat{A} = \frac{1}{\sqrt{2}}\langle 1, 1\rangle$ hence $\vec{B} = B\widehat{B} = \frac{7}{\sqrt{2}}\langle 1, 1\rangle$.*

The solution given in the preceding example is *geometrically* motivated. An alternative *algebraic* approach would be to solve $\vec{B} = k\vec{A}$ and $B = 7$ for k. Both approaches have merit. I used the geometric approach to induce insight for the direction vector concept.

You might have noticed that we already used unit-vectors in the vector component discussion. Note that:

$$\widehat{x} \bullet \widehat{x} = 1, \qquad \widehat{y} \bullet \widehat{y} = 1, \qquad \widehat{z} \bullet \widehat{z} = 1,$$
$$\widehat{x} \bullet \widehat{y} = 0, \qquad \widehat{x} \bullet \widehat{z} = 0, \qquad \widehat{y} \bullet \widehat{z} = 0.$$

We can calculate dot-products by using the properties of the dot-product paired with the results above. For example:

Example 1.1.17. *Let $\vec{A} = \widehat{x} - 2\widehat{y} + 3\widehat{z}$ and $\vec{B} = 5\widehat{x} + 9\widehat{z}$*

$$\begin{aligned}
\vec{A} \bullet \vec{B} &= (\widehat{x} - 2\widehat{y} + 3\widehat{z}) \bullet (5\widehat{x} + 9\widehat{z}) \\
&= 5\widehat{x} \bullet \widehat{x} - 10\widehat{y} \bullet \widehat{x} + 15\widehat{z} \bullet \widehat{x} + 9\widehat{x} \bullet \widehat{z} - 18\widehat{y} \bullet \widehat{z} + 27\widehat{z} \bullet \widehat{z} \\
&= 5(1) - 10(0) + 15(0) + 9(0) - 18(0) + 27(1) \\
&= 32.
\end{aligned}$$

It is easier to use the $\langle a, b, c\rangle$ notation for examples such as the one above, but the notation $\widehat{x}$, $\widehat{y}$, $\widehat{z}$ (or the equivalent $\widehat{i}$, $\widehat{j}$, $\widehat{k}$ used in many other texts) is often used to emphasize that the object in consideration is a **vector**.

In summary, we have that $\widehat{x}_i \bullet \widehat{x}_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$. This is a very interesting formula. It shows that set of vectors $\{\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_n\}$ are all of unit-length and distinct pairs have vanishing dot-products.

Definition 1.1.18. *orthogonal vectors.*

We say $\vec{A}$ is **orthogonal** to $\vec{B}$ iff $\vec{A} \bullet \vec{B} = 0$. A set of vectors which is both orthogonal and all of unit length is said to be an **orthonormal set** of vectors.

Notice $\widehat{x}_i \bullet \widehat{x}_j = \delta_{ij}$ compactly expresses the orthonormality of the standard basis[6] $\{\widehat{x}_1, \widehat{x}_2, \dots, \widehat{x}_n\}$. Orthogonality makes for interesting formulas. Let $\vec{V} = \langle V_1, V_2 \rangle \in \mathbb{R}^2$ and calculate,

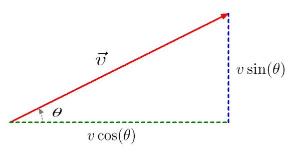
$$\vec{V} \bullet \widehat{x} = (V_1 \widehat{x}_1 + V_2 \widehat{x}_2) \bullet \widehat{x}_1 = V_1 \widehat{x}_1 \bullet \widehat{x}_1 + V_2 \widehat{x}_2 \bullet \widehat{x}_1 = \delta_{11} V_1 + \delta_{12} V_2 = V_1$$

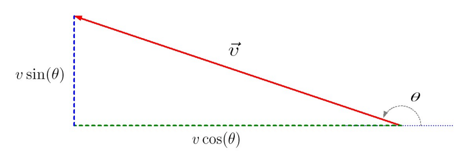
$$\vec{V} \bullet \widehat{x}_2 = (V_1 \widehat{x}_1 + V_2 \widehat{x}_2) \bullet \widehat{x}_2 = V_1 \widehat{x}_1 \bullet \widehat{x}_2 + V_2 \widehat{x}_2 \bullet \widehat{x}_2 = \delta_{12} V_1 + \delta_{22} V_2 = V_2$$

This means we can use the dot-product to select the scalar components of a given vector.

$$\vec{V} = \langle \vec{V} \bullet \widehat{x}_1, \ \vec{V} \bullet \widehat{x}_2 \rangle = (\ \vec{V} \bullet \widehat{x}_1\ )\, \widehat{x}_1 + (\ \vec{V} \bullet \widehat{x}_2\ )\, \widehat{x}_2.$$

Let's pause to make a connection to the standard angle θ and the cartesian components.





Note that $\vec{V} = \cos(\theta)\, \widehat{x} + \sin(\theta)\, \widehat{y}$ and $\vec{V} = (\ \vec{V} \bullet \widehat{x}\ )\, \widehat{x} + (\ \vec{V} \bullet \widehat{y}\ )\, \widehat{y}$. It follows that:

$$\cos(\theta) = \vec{V} \bullet \widehat{x} \qquad \text{and} \qquad \sin(\theta) = \vec{V} \bullet \widehat{y}.$$

You could use these equations to define the standard angle in retrospect. Alternatively, we can use the standard angle for a two-dimensional vector to derive its unit-vector: observe

$$\vec{A} = \langle A\cos\theta, \ A\sin\theta \rangle = A\langle \cos\theta, \ \sin\theta \rangle \ \ \& \ \ \vec{A} = A\widehat{A} \ \ \Rightarrow \ \ \boxed{\widehat{A} = \langle \cos\theta, \ \sin\theta \rangle}.$$

Example 1.1.19. *If $\vec{v}$ has length 10 and $\theta = -\pi/6$ then $\widehat{v} = \langle \cos(\pi/6), -\sin(\pi/6) \rangle = \langle \frac{\sqrt{3}}{2}, \frac{-1}{2} \rangle$.*

Notice, I did not need to use $v = 10$ to find the unit-vector in the $\vec{v}$-direction. The standard angle and the direction-vector are equivalent in the **two**-dimensional context.

Example 1.1.20. *If $\vec{v}$ has length 10 and $\theta = -\pi/6$ then $\widehat{v} = \langle \cos(\pi/6), -\sin(\pi/6) \rangle = \langle \frac{\sqrt{3}}{2}, \frac{-1}{2} \rangle$.*

In $\mathbb{R}^n$ we also have an decomposition of a vector into its vector-cartesian components:

$$\vec{V} = \langle \vec{V} \bullet \widehat{x}_1, \ \vec{V} \bullet \widehat{x}_2, \ \dots \ , \ \vec{V} \bullet \widehat{x}_n \rangle = \sum_{j=1}^{n} (\ \vec{V} \bullet \widehat{x}_j\ )\, \widehat{x}_j.$$

This is called an **orthogonal decomposition** of $\vec{V}$ because it gives $\vec{V}$ as a sum of vectors which are pairwise orthogonal. Intuitively, I think of this as breaking the vector into its basic parts. So far, all of this is with respect to Cartesian coordinates. Perhaps we will also see how similar decompositions are possible for curvelinear coodinate systems or moving coordinate systems.

[6]the theory of orthogonal complements is one of the most interesting chapters in linear algebra, it explains least squares data fitting, general calculation of normal spaces and gives insight into the basic idea of Fourier analysis. See my notes on linear algebra for further comment.

1.1.3 angle measure in higher dimensions

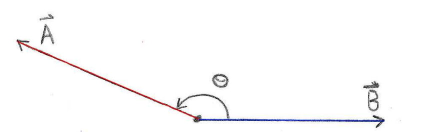
The study of geometry involves lengths and angles of shapes. We have all the tools we need to define the angle θ between nonzero vectors

Definition 1.1.21. *angle between a pair of vectors.*

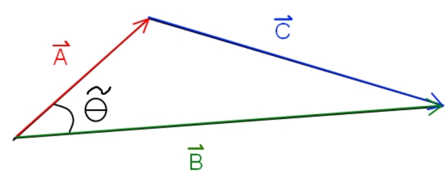
Let $\vec{A}, \vec{B}$ be nonzero vectors in $\mathbb{R}^n$. We define the angle between $\vec{A}$ and $\vec{B}$ by

$$\theta = \cos^{-1}\left[\frac{\vec{A} \bullet \vec{B}}{||\vec{A}||\,||\vec{B}||}\right].$$

Note nonzero vectors $\vec{A}, \vec{B}$ have $||\vec{A}|| \neq 0$ and $||\vec{B}|| \neq 0$ thus the Cauchy-Schwarz inequality $|\vec{A} \bullet \vec{B}| \leq ||\vec{A}||\,||\vec{B}||$ implies $\frac{\vec{A} \bullet \vec{B}}{||\vec{A}||\,||\vec{B}||} \leq 1$. It follows that the argument of the inverse cosine is within its domain. Moreover, since the standard inverse cosine has range $[0, \pi]$ it follows the angle which is given by the formula above is the smallest angle *between* the vectors. Of course, if θ is the angle between $\vec{A}, \vec{B}$ then geometry clearly indicates $2\pi - \theta$ is the angle on the other side of the θ vertex. I think a picture helps:



The careful reader will question how I know the formula really recovers the idea of angle that we have previously used in our studies of trigonometry. All I have really argued thus far is that the formula for θ is reasonable. Examine the triangle formed by $\vec{A}, \vec{B}$ and $\vec{C} = \vec{B} - \vec{A}$. Notice that $\vec{A} + \vec{C} = \vec{B}$. Picture $\vec{A}$ and $\vec{B}$ as adjacent sides to an angle $\tilde{\theta}$ which has opposite side $\vec{C}$. Let the lengths of $\vec{A}, \vec{B}, \vec{C}$ be A, B, C respective.



Applying[7] the **Law of Cosines** to the triangle above yields

$$C^2 = A^2 + B^2 - 2AB\cos(\tilde{\theta}).$$

Solve for $\tilde{\theta}$,

$$\tilde{\theta} = \cos^{-1}\left[\frac{A^2 + B^2 - C^2}{2AB}\right]$$

Is this consistent, does $\theta = \tilde{\theta}$? Choose coordinates[8] which place the vectors $\vec{A}, \vec{B}, \vec{C}$ are in the xy-plane and let $\vec{A} = \langle A_1, A_2\rangle, \vec{B} = \langle B_1, B_2\rangle$ hence $\vec{C} = \langle B_1 - A_1, B_2 - A_2\rangle$ we calculate

$$C^2 = (B_1 - A_1)^2 + (B_2 - A_2)^2 = B_1^2 - 2A_1B_1 + A_1^2 + B_2^2 - 2A_2B_2 + A_2^2$$

[7] if you had Math 131 with me then you proved the Law of Cosines in one of your first Problem Sets.

[8] even in the context of $\mathbb{R}^n$ we can place $\vec{A}, \vec{B}$ and $\vec{B} - \vec{A}$ in a particular plane, this argument actually extends to n-dimensions provided you accept the Law of Cosines is known in any plane

Thus, $C^2 = A^2 + B^2 - 2\vec{A} \bullet \vec{B}$ and we find:

$$\tilde{\theta} = \cos^{-1}\left[\frac{2\vec{A} \bullet \vec{B}}{2AB}\right] = \cos^{-1}\left[\frac{\vec{A} \bullet \vec{B}}{||\vec{A}||\,||\vec{B}||}\right] = \theta.$$

Thus, we find the generalization of angle for $\mathbb{R}^n$ agrees with the two-dimensional concept we've explored in previous courses. Moreover, we discover a geometrically lucid formula for the dot-product:

$$\boxed{\vec{A} \bullet \vec{B} = ||\vec{A}||\,||\vec{B}||\cos(\theta)}$$

or if we denote $\vec{A} = A\widehat{A}$ and $\vec{B} = B\widehat{B}$ then

$$\boxed{\vec{A} \bullet \vec{B} = AB\cos(\theta).}$$

The connection between this formula and the definition is nontrivial and is essentially equivalent to the Law of Cosines. This means that this is a powerful formula which allows deep calculation of geometrically non-obvious angles through the machinery of vectors. Notice:

$$\boxed{\text{If } \vec{A}, \vec{B} \text{ are nonzero orthogonal vectors then the angle between them is } \pi/2.}$$

this observation is an immediate consequence of the the definition of orthogonal vectors and the fact $\cos(\pi/2) = 0$. We find that orthogonal vectors are in fact perpendicular (which is a known term from geometry). In addition,

$$\boxed{\text{If } \vec{A}, \vec{B} \text{ are parallel vectors then } \vec{A} \bullet \vec{B} = AB \text{ and } \theta = 0.}$$

likewise,

$$\boxed{\text{If } \vec{A}, \vec{B} \text{ are antiparallel vectors then } \vec{A} \bullet \vec{B} = -AB \text{ and } \theta = \pi.}$$

The dot-product gives us a concrete method to test for whether two vectors point in the same direction, opposite directions or are purely perpendiular.

Example 1.1.22. *Let $\vec{A} = \langle -5, 3, 7\rangle$ and $\vec{B} = \langle 6, -8, 2\rangle$. Are these vectors parallel, antiparallel or orthogonal? We can calculate the dot-product to answer this question. Observe,*

$$\vec{A} \bullet \vec{B} = \langle -5, 3, 7\rangle \bullet \langle 6, -8, 2\rangle = -30 - 24 + 14 = -40 \neq 0.$$

Thus, we know $\vec{A}$ and $\vec{B}$ are not orthogonal. Furthermore, they cannot be parallel as the dot-product's sign indicates they point in directions more than 90^o oppposed. Are they antiparallel? Consider,

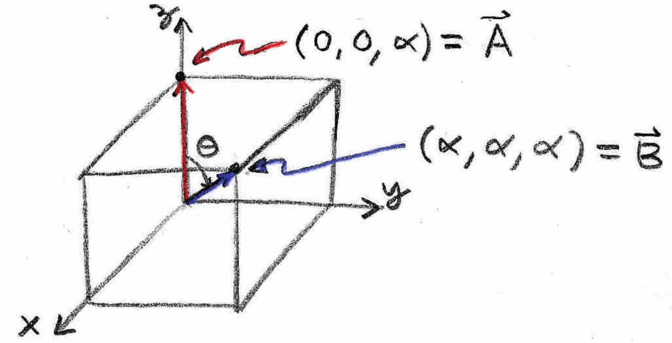
$$AB = \sqrt{25 + 9 + 49}\sqrt{36 + 64 + 4} = \sqrt{8632} \neq -40$$

Therefore, the given pair of vectors is neither parallel, antiparallel nor orthogonal. Of course, we could have ascertained all these comments by simply calculating the angle between the given vectors:

$$\theta = \cos^{-1}\left(\frac{\vec{A} \bullet \vec{B}}{AB}\right) = \cos^{-1}\left(\frac{-40}{\sqrt{8632}}\right) = 115.5^o.$$

I hope the reader can forgive me for abusing notation and sometimes using radian and other times angle measure. When I use degree measure it is primarily to emphasize geometric content.

Example 1.1.23. *Consider a cube of side-length α. What is the angle between the interior diagonal of the cube and the edge of the cube? We place the cube at the origin and envision the diagonal from $(0,0,0)$ to (α,α,α). The edge goes from $(0,0,0)$ to $(0,0,\alpha)$. Let us label the diagonal and edge by $\vec{B}$ and $\vec{A}$ respectively:*



Observe $A = \alpha$ and $B = \alpha\sqrt{3}$ whereas $\vec{A} \bullet \vec{B} = \alpha^2$. We find $\frac{\vec{A}\bullet\vec{B}}{AB} = \frac{\alpha^2}{\alpha^2\sqrt{3}} = \frac{1}{\sqrt{3}}$. Thus $\cos\theta = \frac{1}{\sqrt{3}}$ hence $\theta = \cos^{-1}\left(\frac{1}{\sqrt{3}}\right) \approxeq 54.74^o$.

The reason the angle is not 45^o in the example above is that the vectors $\vec{A}$ and $\vec{B}$ lie on the edge and diagonal of a nonsquare-rectangle. The larger point here: **use vectors** to escape wrong intuition in three-dimensional geometry. The mathematics of vectors allows us to solve problems step-by-step which defy direct geometric methods.

1.1.4 projections and applications

We can do more than just measure angles. We can also use the dot-product to project vectors to lines or even planes. In particular:

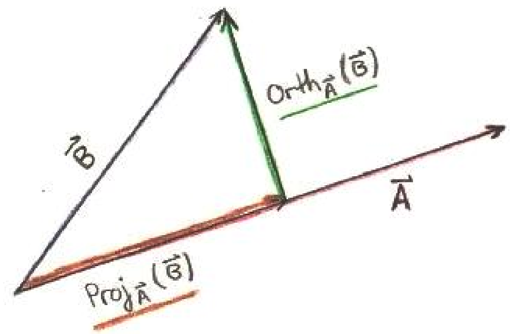
Definition 1.1.24. *projection onto vector.*

$\vec{A} \neq 0$ then $\text{Proj}_{\vec{A}}(\vec{B}) = (\vec{B}\bullet\widehat{A})\widehat{A}$ defines the **vector projection of $\vec{B}$ onto $\vec{A}$**. We also define the orthogonal complement of $\vec{B}$ with respect to $\vec{A}$ by $\text{Orth}_{\vec{A}}(\vec{B}) = \vec{B} - \text{Proj}_{\vec{A}}(\vec{B})$. We also define $\text{Comp}_{\vec{A}}(\vec{B})$ to be the **component of $\vec{B}$ in the $\vec{A}$-direction**.

We've already seen the projection formula implicitly in the formula $\vec{V} = (\vec{V}\bullet\widehat{x})\,\widehat{x} + (\vec{V}\bullet\widehat{y})\,\widehat{y}$ note that $\text{Proj}_{\widehat{x}}(\vec{V}) = (\vec{V}\bullet\widehat{x})\,\widehat{x}$ since the unit vector of the unit vector $\widehat{x}$ is just $\widehat{x}$. Likewise, $\text{Proj}_{\widehat{y}}(\vec{V}) = (\vec{V}\bullet\widehat{y})\,\widehat{y}$. Also, the usage of the term *component* matches our earlier usage of Cartesian components. The x-component of $\vec{B}$ is indeed $\text{Comp}_{\widehat{x}}(\vec{B})$ and so forth. Very well, you might not find this terribly interesting. However, if I were to ask where a line perpendicular to the line through the origin in the direction of $\langle 2,2,1\rangle$ intersects $\langle 9,9,9\rangle$ based at the origin then I doubt you could do it with geometry alone. It's simple with the projection formula:

$$\begin{aligned}
\text{Proj}_{\langle 2,2,1\rangle}\big(\langle 9,9,9\rangle\big) &= \big[\langle 9,9,9\rangle\bullet\langle \tfrac{2}{3},\tfrac{2}{3},\tfrac{1}{3}\rangle\big]\langle \tfrac{2}{3},\tfrac{2}{3},\tfrac{1}{3}\rangle \\
&= 9\cdot\tfrac{1}{3}\cdot\tfrac{1}{3}\big[\langle 1,1,1\rangle\bullet\langle 2,2,1\rangle\big]\langle 2,2,1\rangle \\
&= (2+2+1)\langle 2,2,1\rangle \\
&= \langle 10,10,5\rangle.
\end{aligned}$$

That was actually an accident. I'll try to do an uglier example in lecture. We'll use the projection from time to time, it is a really nice tool to calculate things that are at times hard to picture directly. Generically the projection can be pictured by:



You should think of $\text{Orth}_{\vec{A}}(\vec{B})$ as the way to obtain the piece of $\vec{B}$ which is perpendicular to $\vec{A}$.

Example 1.1.25. *Let $\vec{A} = \langle 3, 4\rangle$ and $\vec{B} = \langle 5, 12\rangle$. Find the angle between the vectors, the direction vectors for each and the vector and scalar projections of $\vec{B}$ onto $\vec{A}$. We begin by calculating the lengths of the vectors*[9]*:*

$$A = \sqrt{3^2 + 4^2} = \sqrt{25} = 5 \qquad \& \qquad B = \sqrt{5^2 + 12^2} = \sqrt{169} = 13.$$

Continuing, the direction vectors are found by dividing by A, B respective:

$$\widehat{A} = \frac{1}{5}\langle 3, 4\rangle \qquad \& \qquad \widehat{B} = \frac{1}{13}\langle 5, 12\rangle.$$

Notice $\widehat{A} \bullet \widehat{B} = \cos\theta$ as $|\widehat{A}| = |\widehat{B}| = 1$. Thus calculate:

$$\theta = \cos^{-1}\left(\frac{15+48}{5(13)}\right) = \cos^{-1}\left(\frac{63}{65}\right) \approxeq 14.25^o$$

The component is in $\vec{A}$ direction is revealed by the dot-product with $\widehat{A}$:

$$Comp_{\vec{A}}(\vec{B}) = \vec{B} \bullet \widehat{A} = \frac{1}{5}\langle 5, 12\rangle \bullet \langle 3, 4\rangle = \frac{63}{5}$$

The projection of $\vec{B}$ into the $\vec{A}$ direction is a vector of length $comp_{\vec{A}}(\vec{B})$ in the $\vec{A}$-direction:

$$Proj_{\vec{A}}(\vec{B}) = (\vec{B} \bullet \widehat{A})\widehat{A} = \frac{63}{25}\langle 3, 4\rangle.$$

Naturally, we could have written $\text{Proj}_{\vec{A}}(\vec{B}) = \langle 189/25, 252/25\rangle$. However, it is wise to present the answer as the explicit scalar multiple of a simple vector. In physics or engineering applications, a decimal approximation $\text{Proj}_{\vec{A}}(\vec{B}) = \langle 7.56, 10.08\rangle$ may be the desired answer. Context is key. Notice that in the preceding example and the example to follow, if I absorbed the scalar multiples into the vectors then we would be faced with a myriad of fractions. Since adding fractions is a major source of errors this is wise to avoid. Use properties of dot-products to simplify your work!

Example 1.1.26. *Let $\vec{A} = \widehat{x} + \widehat{y} + \widehat{z}$ and $\vec{B} = \widehat{x} - \widehat{y} + \widehat{z}$. Find the scalar and vector projection of $\vec{B}$ onto the $\vec{A}$ direction. I begin by noting $\vec{A} = \langle 1, 1, 1\rangle$ and $\vec{B} = \langle 1, -1, 1\rangle$. Furthermore,*

$$\widehat{A} = \frac{1}{\sqrt{3}}\langle 1, 1, 1\rangle.$$

[9] Ah, the joy of Pythagorean triples

Thus, the component of $\vec{B}$ in the $\vec{A}$-direction is given by:

$$Comp_{\vec{A}}(\vec{B}) = \vec{B} \bullet \widehat{A} = \frac{1}{\sqrt{3}}\langle 1,1,1\rangle \bullet \langle 1,-1,1\rangle = \frac{1}{\sqrt{3}}.$$

and the projection of $\vec{B}$ in the $\vec{A}$-direction is found to be

$$Proj_{\vec{A}}(\vec{B}) = (\vec{B} \bullet \widehat{A})\widehat{A} = \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}}\langle 1,1,1\rangle = \frac{1}{3}\langle 1,1,1\rangle.$$

Definition 1.1.27. *projection and orthogonal complement with respect to a plane.*

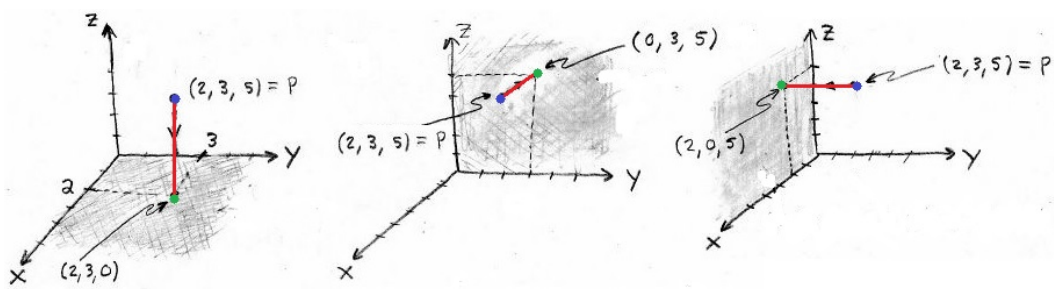
If $\vec{A}, \vec{B}$ are non-parallel vectors in some plane S then $\text{Proj}_S(\vec{R}) = \big(\vec{R} \bullet \widehat{A}\big)\widehat{A} + \big(\vec{R} \bullet \widehat{B}\big)\widehat{B}$ defines the **projection of $\vec{R}$ onto the plane S**. The orthogonal projection of $\vec{R}$ off the plane S is given by $\text{Orth}_S(\vec{R}) = \vec{R} - \text{Proj}_S(\vec{R})$.

The projections onto the coordinate planes are sometimes interesting. Clearly $\widehat{x}$, $\widehat{y}$ fit in the xy-plane hence

$$\text{Proj}_{xy-\text{plane}}(\langle v_1, v_2, v_3\rangle) = (\,\widehat{x} \bullet \langle v_1, v_2, v_3\rangle)\,\widehat{x} + (\,\widehat{y} \bullet \langle v_1, v_2, v_3\rangle)\,\widehat{y} = \langle v_1, v_2, 0\rangle.$$

Likewise, $\text{Proj}_{zx-\text{plane}}(\langle v_1, v_2, v_3\rangle) = \langle v_1, 0, v_3\rangle$ and $\text{Proj}_{yz-\text{plane}}(\langle v_1, v_2, v_3\rangle) = \langle 0, v_2, v_3\rangle$.

Example 1.1.28. *I illustrate the projections of $P = (2,3,5)$ onto the xy, yz and xz-planes from left to right in the diagram below: the blue dot is P the red line is the action of the projection and the green dot is the projected image of P.*



In particular, in the obtuse notation I have introduced in this section,

$$Proj_{xy-plane}(2,3,5) = (2,3,0), \quad Proj_{yz-plane}(2,3,5) = (0,3,5), \quad Proj_{zx-plane}(2,3,5) = (2,0,5).$$

Perhaps this material belongs with the larger discussion of planes. I included it here simply to illustrate the utility of the dot-product.

Example 1.1.29. *Judging the colinearity of two vectors is important to physics. The work done by a force is maximized when the force is applied over a displacement which is precisely parallel to the force. On the other hand, the work done by a perpendicular force is zero. The dot-product captures all these concepts in a nice neat formula: the work W done by a constant force $\vec{F}$ applied to an object undergoing a displacement $\triangle\vec{r}$ is given by $\boxed{W = \vec{F} \bullet \triangle\vec{r}}$. For example, if a force $\vec{F} = \langle 1,1,1\rangle\ N$ is applied to a particle displaced under $\triangle\vec{r} = \langle 1,-2,4\rangle m$ then the work done is:*

$$W = \vec{F} \bullet \triangle\vec{r} = \langle 1,1,1\rangle\ N \bullet \langle 1,-2,4\rangle m = 3\,Nm = 3\,J.$$

Here N is the unit of force called a Newton, m is the unit of distance called a meter and J is the unit of energy called a Joule.

Much later in this course we turn to the question of calculating work done by nonconstant forces over arbitrary curves. Furthermore, often in this course we omit units to reduce clutter. For example:

Example 1.1.30. *Let $\vec{F} = \langle 10, 18, -6 \rangle$ be a constant force field. Find the work done by the given force field on an object which moves from $(2,3,0)$ to $(4,9,15)$. It turns out*[10] *that the same work is done by the given force no matter which path is taken from $(2,3,0)$ to $(4,9,15)$. So, we assume a linear path for our convenience and note $\triangle\vec{r} = (4,9,15) - (2,3,0) = \langle 2,6,15 \rangle$. The dot-product of the force and displacement give the work done by the force:*

$$W = \vec{F} \bullet \triangle\vec{r} = \langle 10, 18, -6 \rangle \bullet \langle 2, 6, 15 \rangle = 20 + 108 - 90 = 38.$$

Naturally, we could assume the points are given in terms of meters and the force in Newtons then our answer above would indicate $38\,J$ of work done. Of course, you could use other units. I leave further discussion of this matter for your physics course(s).

There are many dot-products in basic physics.

Example 1.1.31. *If $\vec{v}$ is the velocity of a mass m then the kinetic energy is given by $K = \frac{1}{2}m\vec{v} \bullet \vec{v}$.*

Example 1.1.32. *Or, if $\vec{v}$ is the velocity of a mass m and $\vec{F}$ is the net-force on m then the power developed by $\vec{F}$ is given by $P = \vec{v} \bullet \vec{F}$.*

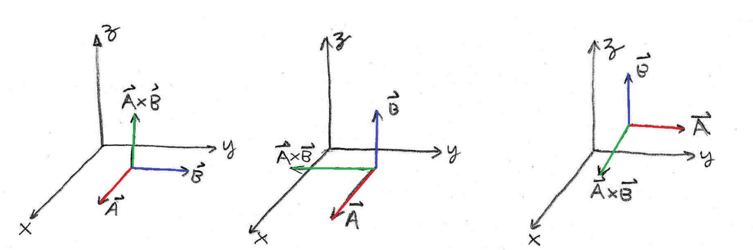
Example 1.1.33. *If $\vec{J}$ is a constant current density then $\vec{J} \bullet (A\widehat{n})$ gives the current flowing through an area A with unit-normal $\widehat{n}$.*

Example 1.1.34. *If $\vec{E}$ is the electric field then $\vec{E} \bullet (A\widehat{n})$ gives the electric flux through an area A with unit-normal $\widehat{n}$.*

Example 1.1.35. *If $\vec{B}$ is the magnetic field then $\vec{B} \bullet (A\widehat{n})$ gives the magnetic flux through an area A with unit-normal $\widehat{n}$.*

1.2 the cross product

We saw that the dot-product gives us a natural way to check if a pair of vectors is orthogonal. You should remember: $\vec{A}, \vec{B}$ are orthogonal iff $\vec{A} \bullet \vec{B} = 0$. We turn to a slightly different goal in this section: given a pair of nonzero, nonparallel vectors $\vec{A}, \vec{B}$ how can we find another vector $\vec{A} \times \vec{B}$ which is perpendicular to both $\vec{A}$ and $\vec{B}$? Geometrically, in $\mathbb{R}^3$ it's not too hard to picture it:



My intent in this section is to motivate the standard formula for this product and to prove some of the standard properties of this cross product. These calculations are special to $\mathbb{R}^3$. The material from here to Definition 1.2.2 is simply to give some insight into where the mysterious formula for the cross product arises. If you insist on remaining unmotivated, feel free to skip to the definition.

[10] for reasons we only complete understand towards the conclusion of this course!

Remark 1.2.1.

Forbidden jutsu ahead, In $\mathbb{R}^n$ the story is a bit more involved, we can calculate the orthogonal complement to $span\{\vec{A},\vec{B}\}$ and this produces an $(n-2)$-dimensional space of orthogonal vectors to $\vec{A},\vec{B}$. If $n=4$ this means there is a whole plane of vectors which we could choose. Only in the case $n=3$ is the orthogonal complement simply a one-dimensional space, a line.

Therefore, suppose $\vec{A},\vec{B}$ are nonzero, nonparallel vectors in $\mathbb{R}^3$. I'll calculate conditions on $\vec{A}\times\vec{B}$ which insure it is perpendicular to both $\vec{A}$ and $\vec{B}$. Let's denote $\vec{A}\times\vec{B}=\vec{C}$. We should expect $\vec{C}$ is some function of the components of $\vec{A}$ and $\vec{B}$. I'll use $\vec{A}=\langle A_1,A_2,A_3\rangle$ and $\vec{B}=\langle B_1,B_2,B_3\rangle$ whereas $\vec{C}=\langle C_1,C_2,C_3\rangle$

$$0=\vec{C}\bullet\vec{A}=C_1A_1+C_2A_2+C_3A_3$$

$$0=\vec{C}\bullet\vec{B}=C_1B_1+C_2B_2+C_3B_3$$

Suppose $A_1\neq 0$, then we may solve $0=\vec{C}\bullet\vec{A}$ as follows,

$$C_1=-\frac{A_2}{A_1}C_2-\frac{A_3}{A_1}C_3$$

Suppose $B_1\neq 0$, then we may solve $0=\vec{C}\bullet\vec{B}$ as follows,

$$C_1=-\frac{B_2}{B_1}C_2-\frac{B_3}{B_1}C_3$$

It follows, given the assumptions $A_1\neq 0$ and $B_1\neq 0$,

$$\frac{A_2}{A_1}C_2+\frac{A_3}{A_1}C_3=\frac{B_2}{B_1}C_2+\frac{B_3}{B_1}C_3$$

Multiply by A_1B_1 to obtain:

$$B_1A_2C_2+B_1A_3C_3=A_1B_2C_2+A_1B_3C_3$$

Thus,

$$(A_1B_2-B_1A_2)C_2+(A_1B_3-B_1A_3)C_3=0$$

One solution is simply $C_2=A_3B_1-A_1B_3$ and $C_3=A_1B_2-B_1A_2$ and it follows that $C_1=A_2B_3-B_2A_3$. Of course, generally we could have vectors which are nonzero and yet have $A_1=0$ or $B_1=0$. The point of the calculation is not to provide a general derivation. Instead, my intent is simply to show you how you might be led to make the following definition:

Definition 1.2.2. *cross product.*

Let $\vec{A},\vec{B}$ be vectors in $\mathbb{R}^3$. The vector $\vec{A}\times\vec{B}$ is called the **cross product** of $\vec{A}$ with $\vec{B}$ and is defined by

$$\vec{A}\times\vec{B}=\ \langle\ A_2B_3-A_3B_2,\ A_3B_1-A_1B_3,\ A_1B_2-A_2B_1\ \rangle.$$

We say $\vec{A}$ cross $\vec{B}$ is $\vec{A}\times\vec{B}$.

It is a simple exercise to verify that

$$\vec{A}\bullet(\vec{A}\times\vec{B})=0 \qquad \text{and} \qquad \vec{B}\bullet(\vec{A}\times\vec{B})=0.$$

Both of these identities should be utilized to check your calculation of a given cross product. Let's think about the formula for the cross product a bit more. We have

$$\vec{A} \times \vec{B} = (A_2B_3 - A_3B_2)\,\widehat{x}_1 + (A_3B_1 - A_1B_3)\,\widehat{x}_2 + (A_1B_2 - A_2B_1)\,\widehat{x}_3$$

distributing,

$$\vec{A} \times \vec{B} = A_2B_3\,\widehat{x}_1 - A_3B_2\,\widehat{x}_1 + A_3B_1\,\widehat{x}_2 - A_1B_3\,\widehat{x}_2 + A_1B_2\,\widehat{x}_3 - A_2B_1\,\widehat{x}_3$$

The pattern is clear. Each term has indices $1, 2, 3$ without repeat and we can generate the signs via the antisymmetric symbol ϵ_{ijk} which is defined be zero if any indices are repeated and

$$\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1 \qquad \text{whereas} \qquad \epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1.$$

With this convenient shorthand we find the nice formula for the cross product that follows:

$$\boxed{\vec{A} \times \vec{B} = \sum_{i,j,k=1}^{3} A_iB_j\epsilon_{ijk}\,\widehat{x}_k}$$

Interestingly the Cartesian unit-vectors $\widehat{x}_1$, $\widehat{x}_2$, $\widehat{x}_3$ satisfy the simple relation:

$$\boxed{\widehat{x}_i \times \widehat{x}_j = \sum_{k=1}^{3} \epsilon_{ijk}\,\widehat{x}_k,}$$

which is just a fancy way of saying that

$$\boxed{\widehat{x} \times \widehat{y} = \widehat{z}, \qquad \widehat{y} \times \widehat{z} = \widehat{x}, \qquad \widehat{z} \times \widehat{x} = \widehat{y}}$$

There are many popular mnemonics to remember these. The basic properties of the cross product together with these formula allow us to quickly calculate some cross products (see Example 1.2.8)

Proposition 1.2.3. *basic properties of the cross product.*

Let $\vec{A}, \vec{B}, \vec{C}$ be vectors in $\mathbb{R}^3$ and $c \in \mathbb{R}$

1. **anticommutative:** $\vec{A} \times \vec{B} = -\vec{B} \times \vec{A}$,
2. **distributive:** $\vec{A} \times (\vec{B} + \vec{C}) = \vec{A} \times \vec{B} + \vec{A} \times \vec{C}$,
3. **distributive:** $(\vec{A} + \vec{B}) \times \vec{C} = \vec{A} \times \vec{C} + \vec{B} \times \vec{C}$,
4. **scalars factor out:** $\vec{A} \times (c\vec{B}) = (c\vec{A}) \times \vec{B} = c\vec{A} \times \vec{B}$,

Proof: once more, the proof is easy with the right notation. Begin with (1.),

$$\vec{A} \times \vec{B} = \sum_{i,j,k=1}^{3} A_iB_j\epsilon_{ijk}\,\widehat{x}_k = -\sum_{i,j,k=1}^{3} A_iB_j\epsilon_{jik}\,\widehat{x}_k = -\sum_{i,j,k=1}^{3} B_jA_i\epsilon_{jik}\,\widehat{x}_k = -\vec{B} \times \vec{A}.$$

The key observation was that $\epsilon_{ijk} = -\epsilon_{jik}$ for all i, j, k. If you don't care for this argument then you could also give the brute-force argument below:

$$\begin{aligned}
\vec{A} \times \vec{B} &= \langle\ A_2B_3 - A_3B_2,\ A_3B_1 - A_1B_3,\ A_1B_2 - A_2B_1\ \rangle \\
&= -\langle\ A_3B_2 - A_2B_3,\ A_1B_3 - A_3B_1,\ A_2B_1 - A_1B_2\ \rangle \\
&= -\langle\ B_2A_3 - B_3A_2,\ B_3A_1 - B_1A_3,\ B_1A_2 - B_2A_1\ \rangle \\
&= -\vec{B} \times \vec{A}.
\end{aligned}$$

Next, to prove (2.) we once more use the compact notation,

$$\begin{aligned}
\vec{A} \times (\vec{B} + \vec{C}) &= \sum_{i,j,k=1}^{3} A_i(B_j + C_j)\epsilon_{ijk}\,\widehat{x}_k \\
&= \sum_{i,j,k=1}^{3} (A_iB_j\epsilon_{ijk}\,\widehat{x}_k + A_iC_j\epsilon_{ijk}\,\widehat{x}_k) \\
&= \sum_{i,j,k=1}^{3} A_iB_j\epsilon_{ijk}\,\widehat{x}_k + \sum_{i,j,k=1}^{3} A_iC_j\epsilon_{ijk}\,\widehat{x}_k \\
&= \vec{A} \times \vec{B} + \vec{A} \times \vec{C}.
\end{aligned}$$

The proof of (3.) follows naturally from (1.) and (2.), note:

$$(\vec{A} + \vec{B}) \times \vec{C} = -\vec{C} \times (\vec{A} + \vec{B}) = -\vec{C} \times \vec{A} - \vec{C} \times \vec{B} = \vec{A} \times \vec{C} + \vec{B} \times \vec{C}.$$

I leave the proof of (4.) to the reader. $\square$

The properties above basically say that the cross product behaves the same as the usual addition and multiplication of numbers with the caveat that the order of factors matters. If we switch the order then we must include a minus due to the anticommutivity of the cross product.

Example 1.2.4. *Consider,* $\vec{A} \times \vec{A} = -\vec{A} \times \vec{A}$ *hence* $2\vec{A} \times \vec{A} = 0$. *Consequently,* $\vec{A} \times \vec{A} = 0$.

We often use the result of the example above in future work. For example:

Example 1.2.5. *Let* $\vec{A}, \vec{B}$ *be two three dimensional vectors. Simplify* $(\vec{A} - \vec{B}) \times (\vec{A} + \vec{B})$.

$$\begin{aligned}
(\vec{A} - \vec{B}) \times (\vec{A} + \vec{B}) &= \vec{A} \times (\vec{A} + \vec{B}) - \vec{B} \times (\vec{A} + \vec{B}) \\
&= \vec{A} \times \vec{A} + \vec{A} \times \vec{B} - \vec{B} \times \vec{A} - \vec{B} \times \vec{B} \\
&= 2\vec{A} \times \vec{B}.
\end{aligned}$$

There are a number of popular tricks to remember the rule for the cross-product. Let's look at a particular example a couple different ways:

Example 1.2.6. *Let* $\vec{A} = \langle 1, 2, 3\rangle$ *and* $\vec{B} = \langle 4, 5, 6\rangle$. *Calculate* $\vec{A} \times \vec{B}$ *directly from the definition:*

$$\begin{aligned}
\vec{A} \times \vec{B} &= \langle 1, 2, 3\rangle \times \langle 4, 5, 6\rangle \\
&= \langle 2(6) - 3(5),\ 3(4) - 1(6),\ 1(5) - 2(4)\rangle \\
&= \langle -3,\ 6,\ -3\rangle.
\end{aligned}$$

There are at least 6 opportunities to make an error in the calculation of a cross product. It is important to check our work before we continue. A simple check is that $\vec{A}$ *and* $\vec{B}$ *must be orthogonal to the cross product. We can easily calculate that* $\vec{A} \bullet (\vec{A} \times \vec{B}) = 0$ *and* $\vec{B} \bullet (\vec{A} \times \vec{B}) = 0$. *This almost guarantees we have correctly calculated the cross product.*

The other popular method to calculate the cross product is based on an abuse of notation with the **determinant**. A determinant can be calculated for any $n \times n$ matrix A. The significance of the determinant is that it gives the signed-volume of the n-piped with edges taken as the rows or columns of A. A simple formula for the determinant in general is given by:

$$\det(A) = \sum_{i_1, i_2, \dots, i_n} \epsilon_{i_1 i_2 \dots i_n} A_{1i_1} A_{2i_2} \cdots A_{ni_n}$$

Ok, I jest. This formula takes a bit of work to really appreciate. So, typically we introduce the determinant in terms of the **expansion by minors** due to Laplace. We begin with a 2×2 matrix:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

Next, a 3×3 can be calculated by an expansion across the top-row,

$$\begin{aligned} \det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} &= a \cdot \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \cdot \det \begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \cdot \det \begin{bmatrix} d & e \\ g & h \end{bmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg). \end{aligned}$$

The minus sign in the middle term is part of the structure of the expansion. It is also one of the most common places where students make an error in their computation of a determinant [11]. We can express the cross product by following the patterns introduced for the 3×3 case. In particular,

$$\begin{aligned} \langle A_1, A_2, A_3 \rangle \times \langle B_1, B_2, B_3 \rangle &= \det \begin{bmatrix} \widehat{x} & \widehat{y} & \widehat{z} \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{bmatrix} \\ &= \widehat{x}(A_2B_3 - A_3B_2) - \widehat{y}(A_1B_3 - A_3B_1) + \widehat{z}(A_1B_2 - A_2B_1) \\ &= (A_2B_3 - A_3B_2)\,\widehat{x} + (A_3B_1 - A_1B_3)\,\widehat{y} + (A_1B_2 - A_2B_1)\,\widehat{z}. \end{aligned}$$

I invite the reader to verify this aligns perfectly with Definition 1.2.2.

Example 1.2.7. *Let $\vec{A} = \langle 1, 2, 3 \rangle$ and $\vec{B} = \langle 4, 5, 6 \rangle$. Calculate $\vec{A} \times \vec{B}$ via the determinant formula:*

$$\begin{aligned} \langle 1, 2, 3 \rangle \times \langle 4, 5, 6 \rangle &= det \begin{bmatrix} \widehat{x} & \widehat{y} & \widehat{z} \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \\ &= \widehat{x}(2(6) - 3(5)) - \widehat{y}(1(6) - 3(4)) + \widehat{z}(1(5) - 2(4)) \\ &= -3\,\widehat{x} + 6\,\widehat{y} - 3\,\widehat{z}. \end{aligned}$$

This result matches $\vec{A} \times \vec{B} = \langle -3, 6, -3 \rangle$ as we found in Example 1.2.6.

Technically, this formula is not really a determinant since genuine determinants are formed from matrices filled with objects of the same type. In the hybrid expression above we actually have one row of vectors and two rows of scalars. That said, I include it here since many people use it and

[11] If we go on, a 4×4 matrix breaks into a signed-weighted-sum of 4 determinants of 3×3 submatrices. More generally, an $n \times n$ martrix has a determinant which requires on the order of $n!$ arithmetic steps. You'll learn more in your linear algebra course, I merely initiate the discussion here. Fortunately, we only need $n = 2$ and $n = 3$ for the majority of the topics in this course.

I also have found it useful in past calculations. If nothing else at least it helps you learn what a determinant is. That is a calculation which is worthwhile since determinants have application far beyond mere cross products.
We can also use the basic relations:

$$\widehat{x} \times \widehat{y} = \widehat{z}, \qquad \widehat{y} \times \widehat{z} = \widehat{x}, \qquad \widehat{z} \times \widehat{x} = \widehat{y}$$

and the properties of cross products to work out cross products algebraically:

Example 1.2.8. *Let $\vec{A} = \widehat{x} + 2\widehat{y} + 3\widehat{z}$ and $\vec{B} = 4\widehat{x} + 5\widehat{y} + 6\widehat{z}$. Calculate $\vec{A} \times \vec{B}$ as follows:*

$$\begin{aligned}
\vec{A} \times \vec{B} &= \widehat{x} \times (4\widehat{x} + 5\widehat{y} + 6\widehat{z}) + 2\widehat{y} \times (4\widehat{x} + 5\widehat{y} + 6\widehat{z}) + 3\widehat{z} \times (4\widehat{x} + 5\widehat{y} + 6\widehat{z}) \\
&= \widehat{x} \times (5\widehat{y} + 6\widehat{z}) + 2\widehat{y} \times (4\widehat{x} + 6\widehat{z}) + 3\widehat{z} \times (4\widehat{x} + 5\widehat{y}) \\
&= 5\widehat{x} \times \widehat{y} + 6\widehat{x} \times \widehat{z} + 8\widehat{y} \times \widehat{x} + 12\widehat{y} \times \widehat{z} + 12\widehat{z} \times \widehat{x} + 15\widehat{z} \times \widehat{y} \\
&= 5\widehat{z} + 6(-\widehat{y}) + 8(-\widehat{z}) + 12\widehat{x} + 12\widehat{y} + 15(-\widehat{x}) \\
&= -3\widehat{x} + 6\widehat{y} - 3\widehat{z}.
\end{aligned}$$

This agrees with the conclusion of the previous pair of examples.

The calculation above is probably not the quickest for the example at hand here, but it is faster for other computations. For example:

Example 1.2.9. *Suppose $\vec{A} = \langle 1, 2, 3 \rangle$ and $\vec{B} = \widehat{x}$ then*

$$\begin{aligned}
\vec{A} \times \vec{B} &= (\widehat{x} + 2\widehat{y} + 3\widehat{z}) \times \widehat{x} \\
&= 2\widehat{y} \times \widehat{x} + 3\widehat{z} \times \widehat{x} \\
&= -2\widehat{z} + 3\widehat{y}.
\end{aligned}$$

Example 1.2.10. *Let $\vec{A} = \langle 3, 2, 4 \rangle$ and $\vec{B} = \langle 1, -2, -3 \rangle$. We calculate,*

$$\begin{aligned}
\vec{A} \times \vec{B} &= det \begin{bmatrix} \widehat{x} & \widehat{y} & \widehat{z} \\ 3 & 2 & 4 \\ 1 & -2 & -3 \end{bmatrix} \\
&= \widehat{x}(-6+8) - \widehat{y}(-9-4) + \widehat{z}(-6-2) \\
&= 2\widehat{x} + 13\widehat{y} - 8\widehat{z}.
\end{aligned}$$

As a check on our computation, note that $\vec{A} \bullet (\vec{A} \times \vec{B}) = 0$ and $\vec{B} \bullet (\vec{A} \times \vec{B}) = 0$.

There are a number of identities which connect the dot and cross products. These formulas require considerable effort if you choose to use brute-force proof methods.

Proposition 1.2.11. *nontrivial properties of the cross product.*

Let $\vec{A}, \vec{B}, \vec{C}$ be vectors in $\mathbb{R}^3$

1. $\vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \bullet \vec{C})\vec{B} - (\vec{A} \bullet \vec{B})\vec{C}$
2. **Jacobi Identity:** $\vec{A} \times (\vec{B} \times \vec{C}) + \vec{B} \times (\vec{C} \times \vec{A}) + \vec{C} \times (\vec{A} \times \vec{B}) = 0,$
3. **cyclicity of triple product:** $\vec{A} \bullet (\vec{B} \times \vec{C}) = \vec{B} \bullet (\vec{C} \times \vec{A}) = \vec{C} \bullet (\vec{A} \times \vec{B})$
4. **Lagrange's identity:** $||\vec{A} \times \vec{B}||^2 = ||\vec{A}||^2 \, ||\vec{B}||^2 - \left[\, \vec{A} \bullet \vec{B} \, \right]^2$

Proof: I leave proof of (1.) and (2.) to the reader. Let's see how (3.) is shown in the compact notation. Note $(\vec{B} \times \vec{C})_k = \sum_{ij} B_i C_j \epsilon_{ijk}$ hence

$$\begin{aligned}
\vec{A} \bullet (\vec{B} \times \vec{C}) = \sum_{k=1}^{3} A_k \sum_{i,j=1}^{3} B_i C_j \epsilon_{ijk} &= \sum_{i,j,k=1}^{3} B_i C_j A_k \epsilon_{ijk} \\
&= \sum_{i,j,k=1}^{3} C_j A_k B_i \epsilon_{jki} \\
&= \sum_{i,j,k=1}^{3} A_k B_i C_j \epsilon_{kij}
\end{aligned}$$

where we have used the cyclicity of the antisymmetric symbol $\epsilon_{ijk} = \epsilon_{jki} = \epsilon_{kij}$. The cyclicity of the triple product follows. Now we turn our attention to Lagrange's identity. I begin by quoting a useful identity connecting the antisymmetric symbol and the Kronecker delta,

$$\sum_{k=1}^{3} \epsilon_{ijk} \epsilon_{lmk} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl} \quad \star$$

Consider that,

$$\begin{aligned}
||\vec{A} \times \vec{B}||^2 = \sum_{k=1}^{3} (\vec{A} \times \vec{B})_k^2 &= \sum_{k=1}^{3} \Big(\sum_{i,j=1}^{3} A_i B_j \epsilon_{ijk} \Big) \Big(\sum_{l,m=1}^{3} A_l B_m \epsilon_{lmk} \Big) \\
&= \sum_{i,j,k,l,m=1}^{3} \sum_{k=1}^{3} A_i A_l B_j B_m \epsilon_{ijk} \epsilon_{lmk} \\
&= \sum_{i,j,l,m=1}^{3} A_i A_l B_j B_m (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}) \\
&= \sum_{i,j,l,m=1}^{3} A_i A_l \delta_{il} B_j B_m \delta_{jm} - \sum_{i,j,l,m=1}^{3} A_i B_m \delta_{im} B_j A_l \delta_{jl} \\
&= \sum_{i,j=1}^{3} A_i^2 B_j^2 - \sum_{i,j=1}^{3} A_i B_i B_j A_j \\
&= \sum_{i=1}^{3} A_i^2 \sum_{i=1}^{3} B_j^2 - \sum_{i=1}^{3} A_i B_i \sum_{j=1}^{3} B_j A_j \\
&= ||\vec{A}||^2 ||\vec{B}||^2 - \big[\ \vec{A} \bullet \vec{B}\ \big]^2.
\end{aligned}$$

I leave derivation of the crucial $\star$ identity to the reader. $\square$.

Use Lagrange's identity together with $\vec{A} \bullet \vec{B} = AB\cos(\theta)$,

$$||\vec{A} \times \vec{B}||^2 = A^2B^2 - [AB\cos(\theta)]^2 = A^2B^2(1 - \cos^2(\theta)) = A^2B^2\sin^2(\theta)$$

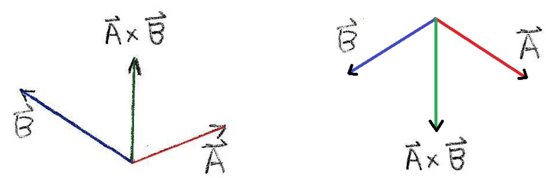
It follows there exists some unit-vector $\hat{\mathbf{n}}$ such that

$$\boxed{\vec{A} \times \vec{B} = AB\sin(\theta)\hat{\mathbf{n}}}$$

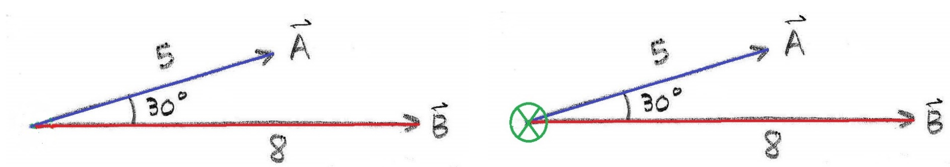
The direction of the unit-vector $\hat{\mathbf{n}}$ is conveniently indicated by the **right-hand-rule**. I typically perform the rule as follows:

1. point fingers of **right hand** in direction $\vec{A}$
2. cross the fingers into the direction of $\vec{B}$
3. the direction your thumb points is the approximate direction of $\hat{\mathbf{n}}$

I say *approximate* because $\vec{A} \times \vec{B}$ is strictly perpendicular to both $\vec{A}$ and $\vec{B}$ whereas your thumb's direction is a little ambiguous. But, it does pick one side of the plane in which the vectors $\vec{A}$ and $\vec{B}$ reside. In the picture below, $\vec{A} \times \vec{B}$ is in the plane of the page whereas I intend for us to visualize $\vec{A}$ and $\vec{B}$ as pointing into or out of the page:



Example 1.2.12. . *Consider $\vec{A}$ and $\vec{B}$ pictured below. Find the magnitude of $\vec{A} \times \vec{B}$ and describe its direction. We produce the right picture by the right hand rule:*



Note $||\vec{A} \times \vec{B}|| = AB \sin\theta = 40 \sin 30^o = 20$. By the right hand rule, we find the direction of $\vec{A} \times \vec{B}$ is into the page. The $\otimes$ symbol intends we visualize the vector as an arrow pointing into the page.

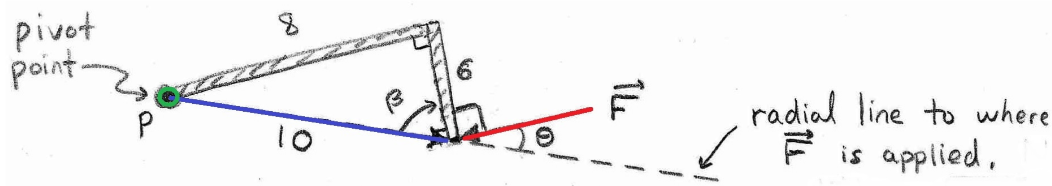
Example 1.2.13. *Let $\vec{u}$ and $\vec{v}$ be as pictured below with $u = 5$ and $v = 4\sqrt{3}$. Find the magnitude and direction vector of $\vec{v} \times \vec{u}$: we use the right hand rule to produce the diagram on the right:*

Note $||\vec{v} \times \vec{u}|| = vu \sin\theta = 20\sqrt{3} \sin 60^o = 30$. By the right hand rule, we find the direction of $\vec{v} \times \vec{u}$ is out of the page. The $\odot$ symbol indicates a vector pointing out of the page.

The cross product is also found in many physical applications. I give two common examples.

Example 1.2.14. *In rotational physics the direction of a rotation is taken to be the axis of the rotation where a counter-clockwise-rotation (CCW) is taken to be positive. To decide which direction is CCW we grip the rotation axis and point our right-hand's thumb in the direction of the positive axis. Once that grip is made the fingers on the right hand encircle the axis in the CCW-rotational sense. A torque on a body allowed to rotate around some axis makes it rotate. In particular, if $\vec{r}$ is the* **moment arm** *and $\vec{F}$ is the force applied then $\vec{\tau} = \vec{r} \times \vec{F}$ is the torque produced by $\vec{F}$ relative to the given axis.*

Problem: *Find the torque due to the force $\vec{F}$ pictured below. Describe the rotation produced as CCW or CW given the axis of rotation points out of the page*



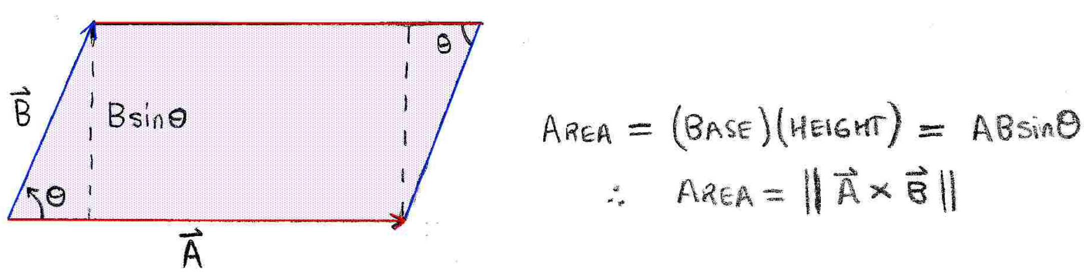
Solution: *Imagine moving $\vec{F}$ to P while maintaining its direction. This is called* **parallel transport**. *We calculate $\vec{r} \times \vec{F}$ as if they are both attached to P. The right hand rule reveals the direction is into the page ($\otimes$) and we can determine θ from trigonometry and the given geometric data. Observe θ is also interior to the triangle at P hence $\sin\theta = \frac{6}{10}$. Also, by pythagorean theorem, $r = \sqrt{8^2 + 6^2} = 10$. Therefore, $\tau = rF\sin\theta = 6F$. The direction of the torque is $\otimes$ which indicates a CW-rotation relative to the outward pointing axis through P.*

Example 1.2.15. *Another important application of the cross product to physics is the Lorentz force law. If a charge q has velocity $\vec{v}$ and travels through a magnetic field $\vec{B}$ then the force due to the electromagnetic interaction between q and the field is $\vec{F} = q\vec{v} \times \vec{B}$.*

Finally, we should investigate how the dot and cross product give nice formulas for the area of a parallelogram or the volume of a parallel piped. Suppose $\vec{A}, \vec{B}$ give the sides of a parallelogram.

$$Area = ||\ \vec{A} \times \vec{B}\ ||$$

The picture below shows why the formula above is true:



On the other hand, if $\vec{A}, \vec{B}, \vec{C}$ give the corner-edges of a parallelogram then[12]

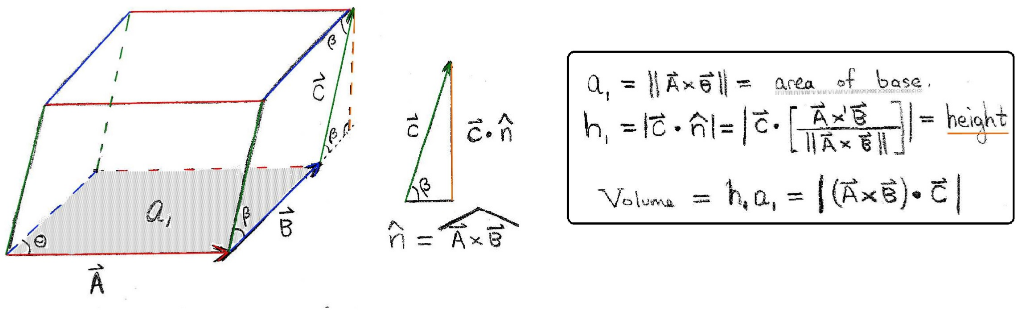
$$Volume = \left|\ \vec{A} \bullet (\vec{B} \times \vec{C})\ \right|$$

These formulas are connected by the following thought: the volume subtended by $\vec{A}, \vec{B}$ and the unit-vector $\hat{\mathbf{n}}$ from $\vec{A} \times \vec{B} = AB\sin(\theta)\hat{\mathbf{n}}$ is equal to the area of the parallelogram with sides $\vec{A}, \vec{B}$. Algebraically:

$$\left|\ \hat{\mathbf{n}} \bullet (\vec{A} \times \vec{B})\ \right| = \left|\ \hat{\mathbf{n}} \bullet (AB\sin(\theta)\hat{\mathbf{n}})\right| = |AB\sin(\theta)| = ||\vec{A} \times \vec{B}||.$$

The picture below shows why the triple product formula is valid.

[12] we could also show that $\det[\vec{A}|\vec{B}|\vec{C}] = \vec{A} \bullet (\vec{B} \times \vec{C})$ thus the determinant of the three edge vectors of a parallel piped yields its signed-volume. We can define the sign of the volume to be positive if the edges are ordered to respect the right hand rule. Respecting the right hand rule means the angle between $\vec{A} \times \vec{B}$ and $\vec{C}$ is less than 90^o.



Example 1.2.16. *Find the volume of a parallel-piped with edge-vectors $\vec{A} = \langle 0,1,1\rangle$ and $\vec{B} = \langle 1,0,0\rangle$ and $\vec{C} = \langle 0,1,0\rangle$. We calculate $\vec{B}\times\vec{C} = \widehat{x}\times\widehat{y} = \widehat{z}$. Therefore, the volume of the solid is $V = \vec{A}\bullet(\vec{B}\times\vec{C}) = \langle 0,1,1\rangle\bullet\widehat{z} = 1$.*

Moreover, given this geometric interpretation we find a new proof (up to a sign) for the cyclic property. By the symmetry of the edges it follows that $|\ \vec{A}\bullet(\vec{B}\times\vec{C})\ | = |\ \vec{B}\bullet(\vec{C}\times\vec{A})\ | = |\ \vec{C}\bullet(\vec{A}\times\vec{B})\ |$. We should find the same volume no matter how we label width, depth and height.

1.3 lines and planes in $\mathbb{R}^3$

There are two main viewpoints to describe lines and planes. The parametric viewpoint introduces parameters which label points on the object of interest. For a line we need one parameter, for a plane we need two parameters. On the other hand, we can view lines and planes just in terms of the solution sets of the cartesian coordinates x, y, z. In contrast, we need one equation to describe a plane whereas we need two equations to fix a line. In between these two viewpoints is the concept of a graph. A graph takes one or more of the Cartesian coordinates as parameter(s) and as such it can easily be thought of as a parametrization. On the other hand, a graph is given by an equation involving only cartesian coordinates so it is easy to think of it as a solution set[13]. Connecting these viewpoints and gaining a geometric appreciation for both is one of the main themes of this course. Finally, I return to the projection to the plane and we examine the connection between the dot-product based projection and the normal to the plane.

1.3.1 parametrized lines and planes

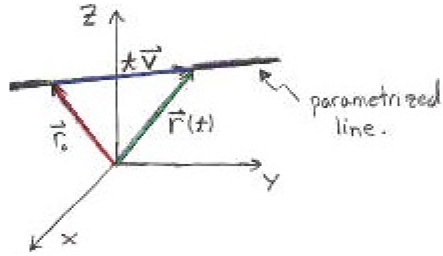
The parametric equations[14] for lines and planes are very natural if you have a proper understanding of vector addition.

Definition 1.3.1. *parametrized line.*

The line L which points in the $\vec{v}$-**direction** and passes through some **base-point** $\vec{r}_o$ has the natural parametrization given by $\vec{r}(t) = \vec{r}_o + t\vec{v}$.

[13] in later sections we will refer to the solution set formulation as a **level set** formulation

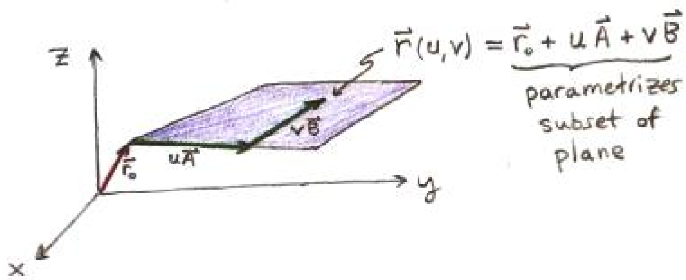
[14] I assume the reader has some familarity with these terms from calculus II, although, I do provide general definitions later in this chapter. These are the two most basic examples



The direction vector $\vec{v}$ points along the line. The direction vector is *tangent* to the line.

Definition 1.3.2. *parametrized plane.*

The plane S which contains the **base-point** $\vec{r}_o$ and the vectors $\vec{A}, \vec{B}$ has a natural parametrization is given by $\vec{r}(u,v) = \vec{r}_o + u\vec{A} + v\vec{B}$.



Notice the vectors $\vec{A}, \vec{B}$ lie along the surface of the plane. We say such vectors are **tangent** to the plane. Vectors which are perpendicular to the plane are called **normal vectors**. For example, $\vec{A} \times \vec{B}$ is normal to the plane pictured above.

If you wish to select a subset of the line or plane above you can appropriately restrict the domain of the parameters. For example, one is often asked to find the parametrization of a line-segment from a point P to a point Q. I recommend the following approach: for $0 \leq t \leq 1$ let

$$\vec{r}(t) = P(1-t) + tQ.$$

It's easy to calculate $\vec{r}(0) = P$ and $\vec{r}(1) = Q$. This formula can also be written as

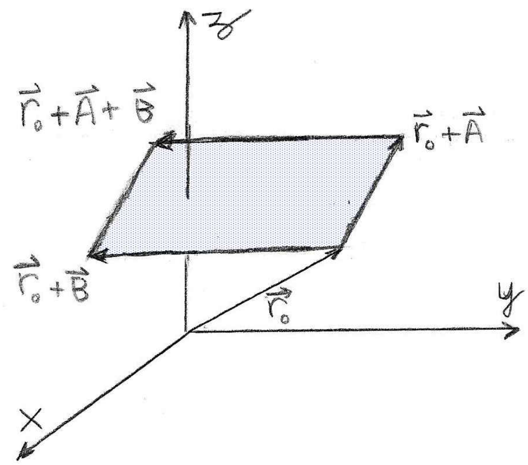
$$\vec{r}(t) = P + t(Q-P) = P + t\ [\overrightarrow{PQ}].$$

If we let t go beyond the unit-interval then we trace out the line which contains the line-segment PQ.

Example 1.3.3. Find the parametrization of a line segment which goes from $(1,3)$ **to** $(5,2)$. *We use the comment preceding this example and construct:*

$$\vec{r}(t) = (1,3) + t[(5,2) - (1,3)] = \langle 1 + 4t, 3 - t \rangle$$

On the other hand, if we wish to parametrize just the parallelogram in the plane with corners $\vec{r}_o$, $\vec{r}_o + \vec{A}$, $\vec{r}_o + \vec{B}$ and $\vec{r}_o + \vec{A} + \vec{B}$ we may limit the values of the parameters u, v to the unit square $[0,1] \times [0,1]$; that is, we demand $0 \leq u \leq 1$ and $0 \leq v \leq 1$.



Example 1.3.4. Find parametrization of plane containing the vectors $\langle 1,0,0\rangle$, $\langle 0,1,0\rangle$ and the point $(1,2,0)$. *We use the natural parametrization:*

$$\vec{r}(u,v) = (1,2,0) + u\langle 1,0,0\rangle + v\langle 0,1,0\rangle = (1+u, 2+v, 0).$$

If we allow (u,v) to trace out all of $\mathbb{R}^2$ then we will find the parametrization above covers the xy-plane. Scalar equations which capture the same are $x = 1+u, y = 2+v, z = 0$. If we restrict the parameters to $0 \leq u,v \leq 1$ then the mapping $\vec{r}$ just covers $[1,2]\times[2,3]\times\{0\}$.

Example 1.3.5. Suppose $\vec{r}(u,v) = (1+u+v, 2-u, 3+v)$ with $(u,v)\in\mathbb{R}^2$ parametrizes a plane. Find the two vectors which lie in the plane and a point on its surface. *The solution is to work backwards in comparison to the last example. We wish to rip apart the formula so that we can identify $\vec{r}_o$ and $\vec{A}, \vec{B}$ for the given $\vec{r}$.*

$$\vec{r}(u,v) = (1+u+v, 2-u, 3+v) = (1,2,3) + u(1,-1,0) + v(1,0,1)$$

Identify the point $(1,2,3)$ is on the plane, in fact, $\vec{r}(0,0) = \vec{r}_o$. Moreover, the vectors $\langle 1,-1,0\rangle$ and $\langle 1,0,1\rangle$ lie on the plane. You can verify that $\langle 1,-1,0\rangle$ connects $\vec{r}(0,0) = (1,2,3)$ and $\vec{r}(1,0) = (2,1,3)$ whereas $\langle 1,1,0\rangle$ connects $\vec{r}(0,0) = (1,2,3)$ and $\vec{r}(0,1) = (2,2,4)$.

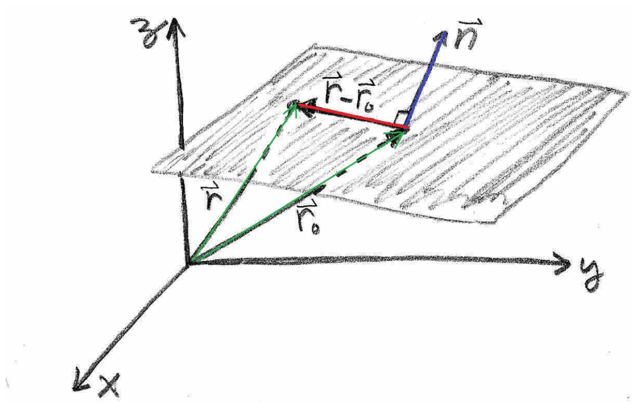
1.3.2 lines and planes as solution sets

Definition 1.3.6. *vector equation of plane.*

We say that $S \subset \mathbb{R}^3$ is a **plane** with **base point** $\vec{r}_o$ and **normal vector** $\vec{n}$ if and only if each point $\vec{r}$ in S satisfies

$$\boxed{(\vec{r}-\vec{r}_o)\bullet\vec{n} = 0.} \qquad \textbf{(vector equation of plane)}$$

The geometric motivation for this definition is simple enough: the normal vector is a vector which is perpendicular to all vectors in the plane. If we take the difference $\vec{r}-\vec{r}_o$ then this will be a vector which lies tangent to the plane and consequently we must insist they are orthogonal. Here's a picture of why the definition is reasonable:



Note that the same set of points S can be given many different base points and many different normals. This reflects the fact that we can choose the base point anywhere on the plane and the normal either above or below the plane and can be given many different lengths.

Let $\vec{r} = \langle x, y, z \rangle$ be an arbitrary point on the plane S with base point $\vec{r}_o = \langle x_o, y_o, z_o \rangle$ and normal $\vec{n} = \langle a, b, c \rangle$ then we can write our plane equation explicitly:

$$(\vec{r} - \vec{r}_o) \bullet \vec{n} = 0 \quad \Leftrightarrow \quad \langle x - x_o, y - y_o, z - z_o \rangle \bullet \langle a, b, c \rangle = 0$$
$$\Leftrightarrow \quad \boxed{a(x - x_o) + b(y - y_o) + c(z - z_o) = 0.} \quad \textbf{(scalar equation of plane)}$$

It is wise to learn both the vector and scalar equations for a plane. These basic examples continue to appear throughout the remainder of this course.

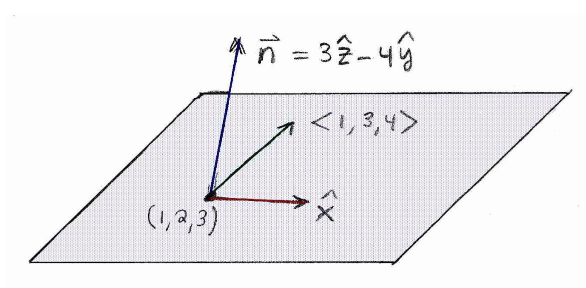
Example 1.3.7. *Suppose the plane S contains the point $(1, 2, 3)$ and the vectors $\vec{A} = \widehat{x}$ and $\vec{B} = \langle 1, 3, 4 \rangle$.* **Find the scalar equation of S.**

Solution: *we are given the base point $(1, 2, 3)$ so we need only find a normal for S. Recall that the cross product of $\vec{A}$ with $\vec{B}$ gives a vector which is perpendicular to both vectors and by the geometry of $\mathbb{R}^3$ it must be colinear with the normal vector we seek. After all, one we use up two of the dimensions then there is only one left to use in $\mathbb{R}^3$. Calculate:*

$$\vec{A} \times \vec{B} = \widehat{x} \times \left[\widehat{x} + 3\,\widehat{y} + 4\,\widehat{z} \right] = \widehat{x} \times \widehat{x} + 3\,\widehat{x} \times \widehat{y} + 4\,\widehat{x} \times \widehat{z} = 3\,\widehat{z} - 4\,\widehat{y}.$$

Thus, we choose $\vec{n} = \langle 0, -4, 3 \rangle$ meaning $a = 0, b = -4, c = 3$ hence

$$\boxed{-4(y - 2) + 3(z - 3) = 0.}$$



1.3.3 equations of intersecting planes

In this section I explore how the equation of the line of intersection for two planes can be characterized as a solution set or parametrically. We also discover another motivation for the formula for the cross product as a side-effect of our inquiry[15].

Given two planes they may intersect in a line. Suppose S_1, S_2 are planes which intersect along some line L then we have that L is the simultaneous solution to the equations of both planes. That is to say $(x,y,z) \in L$ iff $(x,y,z) \in S_1 \cap S_2$. In particular, if $(x_o, y_o, z_o) \in S_1 \cap S_2$ then we can write the equations of S_1, S_2 as:

$$a_1(x-x_o) + b_1(y-y_o) + c_1(z-z_o) = 0 \qquad \text{and} \qquad a_2(x-x_o) + b_2(y-y_o) + c_2(z-z_o) = 0$$

We must solve both at once to find an equation for L. Generally there is no simple formula, however if $a_1, b_1, c_1, a_2, b_2, c_2 \neq 0$ then we are free to divide by those constants. First to get the equations to match divide both by the coefficient of their $(z - z_o)$ factor,

$$\frac{a_1}{c_1}(x-x_o) + \frac{b_1}{c_1}(y-y_o) + z - z_o = 0 \qquad \text{and} \qquad \frac{a_2}{c_2}(x-x_o) + \frac{b_2}{c_2}(y-y_o) + z - z_o = 0$$

Thus, solving both equations for $z - z_o$ we find,

$$\frac{a_1}{c_1}(x-x_o) + \frac{b_1}{c_1}(y-y_o) = \frac{a_2}{c_2}(x-x_o) + \frac{b_2}{c_2}(y-y_o)$$

Multiply by c_1c_2 and rearrange to find

$$\big[a_1c_2 - a_2c_1\big](x-x_o) + \big[b_1c_2 - b_2c_1\big](y-y_o) = 0$$

Consequently,

$$\frac{x-x_o}{b_1c_2 - b_2c_1} = \frac{y-y_o}{a_2c_1 - a_1c_2} \quad \star\,.$$

Following the same algebra we can equally well solve for $x - x_o$,

$$\frac{b_1}{a_1}(y-y_o) + \frac{c_1}{a_1}(z-z_o) = \frac{b_2}{a_2}(y-y_o) + \frac{c_2}{a_2}(z-z_o)$$

and multiply by a_1a_2 to find

$$\big[b_1a_2 - b_2a_1\big](y-y_o) + \big[c_1a_2 - c_2a_1\big](z-z_o) = 0$$

Hence,

$$\frac{y-y_o}{c_1a_2 - c_2a_1} = \frac{z-z_o}{b_2a_1 - b_1a_2} \quad \star\star.$$

We combine $\star$ and $\star\star$ to obtain the **symmetric equations for the line** L

$$\boxed{\frac{x-x_o}{b_1c_2 - b_2c_1} = \frac{y-y_o}{a_2c_1 - a_1c_2} = \frac{z-z_o}{a_1b_2 - a_2b_1}}$$

If we denote $\vec{n}_1 = \langle a_1, b_1, c_1 \rangle$ and $\vec{n}_2 = \langle a_2, b_2, c_2 \rangle$ then recognize that[16]

$$\vec{w} = \vec{n}_1 \times \vec{n}_2 = \langle b_1c_2 - b_2c_1, a_2c_1 - a_1c_2, a_1b_2 - a_2b_1 \rangle$$

[15] the format of this section is discussion/discovery to understand it you must read the entirety of the section.

[16] note, if you didn't already know what the cross product is then this might motivate why we defined it as we did modulo a signed-scale factor

Therefore, with the notation $\vec{w} = \langle a, b, c \rangle$, the symmetric equation is simply:

$$\boxed{\frac{x - x_o}{a} = \frac{y - y_o}{b} = \frac{z - z_o}{c}.}$$

We can use these equations to parametrize the line L. Let $t = \frac{x - x_o}{a}$ hence $x = x_o + at$ is the parametric equation for x, likewise, $y = y_o + bt$ and $z = z_o + ct$. We identify that $\langle a, b, c \rangle$ is precisely the direction-vector for the line L since we can group the scalar parametric equations above to obtatin the vector parametric equation below:

$$\vec{r}(t) = \langle x_o + at, \ y_o + bt, \ z_o + ct \rangle = \langle x_o, y_o, z_o \rangle + t\langle a, b, c \rangle.$$

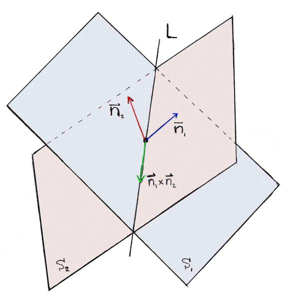
We find the following interesting geometric result:

The line of intersection for two planes has a direction vector which is colinear with the cross product of the normals of the intersected planes.

If we take a step back and analyze this by pure geometric visualization this is rediculously obvious. The line of intersection lies in both planes. Therefore, if $\vec{v}$ is the direction vector of L and $\vec{n}_1$ is the normal of plane S_1 and $\vec{n}_2$ is the normal of plane S_2 then

1. $\vec{v} \bullet \vec{n}_1 = 0$ because $\vec{v}$ lies on S_1
2. $\vec{v} \bullet \vec{n}_2 = 0$ because $\vec{v}$ lies on S_2
3. if $\vec{v}$ is perpendicular to both $\vec{n}_1$ and $\vec{n}_2$ then it must be colinear with $\vec{n}_1 \times \vec{n}_2$.

This is an example of how geometry is sometimes easier than algebra. In fact, that is often the case, however, you must get used to both lines of logic in this course. This is the beauty of analytic geometry.



Let's examine how we can get from the parametric viewpoint to the symmetric equations. Suppose we are given the vector parametric equations for a line with base point $\vec{r}_o = (x_o, y_o, z_o)$ and direction vector $\vec{v} = \langle a, b, c \rangle$ with $a, b, c \neq 0$:

$$\vec{r}(t) = \vec{r}_o + t\vec{v} = (x_o + ta, y_o + tb, z_o + tc).$$

Further suppose we denote $\vec{r} = \langle x, y, z \rangle$ then the scalar parametric equations for the line are:

$$x = x_o + ta, \quad y = y_o + tb, \quad z = z_o + tc.$$

These can be solved for t,

$$\frac{x - x_o}{a} = \frac{y - y_o}{b} = \frac{z - z_o}{c}.$$

and once more we find that the symmetric equations for a line reveal the direction of the line. Well, to be careful, we can multiply this equation by $1/k \neq 0$ and we'd still have the same solution set but $a \to ka$, $b \to kb$ and $c \to kc$ hence the direction naturally identified would be $\langle ka, kb, kc\rangle$. This is an ambiguity we always face with lines. The direction vector is not unique, unless we add further criteria. For example,

$$\frac{x}{2} = \frac{y}{3} = \frac{z}{4}$$

suggest the line has direction $\langle 2, 3, 4\rangle$ whereas

$$\frac{x}{-2} = \frac{y}{-3} = \frac{z}{-4}$$

suggests the line has direction $\langle -2, -3, -4\rangle$.

Finally, recall that I insisted that the intersection of the planes was a line from the outset of this discussion. There is another possibility. It could be that two planes are either parallel and have no point of intersection, or they could simply be the same plane. In both of those cases the cross product of the normals is trivial since the normals of parallel planes are colinear.

1.3.4 lines and planes as graphs

Suppose S is the plane $a(x - x_o) + b(y - y_o) + c(z - z_o) = 0$ then if $c \neq 0$ we can solve for z to find

$$z = z_o - \frac{a}{c}(x - x_o) - \frac{b}{c}(y - y_o).$$

If we define $f(x, y) = z_o - \frac{a}{c}(x - x_o) - \frac{b}{c}(y - y_o)$ then we can write S as the graph $z = f(x, y)$.

Definition 1.3.8. *graphs of $f(x, y), g(x, z)$ or $h(y, z)$.*

$$graph(f) = \{(x, y, f(x, y)) \mid (x, y) \in dom(f)\}$$
$$graph(g) = \{(x, g(x, z), z) \mid (x, z) \in dom(g)\}$$
$$graph(h) = \{(h(y, z), y, z) \mid (y, z) \in dom(h)\}$$

clearly $S = graph(f)$ in this context. You could also write S as the graph $y = g(x, z)$ or $x = h(y, z)$ provided $b \neq 0$ and $a \neq 0$ respective. I hope you can find the formulas for g or h. These graphs provide parametrizations as follows, once more consider the case $c \neq 0$, let

$$x = u, \quad y = v, \quad z = f(u, v)$$

Equivalently,

$$\boxed{\vec{r}(u, v) = \langle u, v, f(u, v)\rangle.}$$

In this way we find a natural parametrization of a graph. Likewise, if $a \neq 0$ or $b \neq 0$ then

$$\vec{r}(u, v) = \langle u, g(u, v), v\rangle \qquad \text{or} \qquad \vec{r}(u, v) = \langle h(u, v), u, v\rangle$$

provide natural parametrizations. Parametrizations created in these ways are said to be **induced** from the graphs g, h respective.

Writing the line as a graph requires us to solve the symmetric equations for two of the cartesian coordinates in terms of the remaining coordinate. For example, solve for y, z in terms of x:

$$\frac{x-x_o}{a} = \frac{y-y_o}{b} = \frac{z-z_o}{c} \quad \Rightarrow \quad a(y-y_o) = b(x-x_o) \;\;\&\;\; a(z-z_o) = c(x-x_o).$$

hence,

$$\underbrace{y = y_o + \frac{b}{a}(x-x_o)}_{\text{let this be } h(x)} \quad \& \quad \underbrace{z = z_o + \frac{c}{a}(x-x_o)}_{\text{let this be } g(x)}$$

Define $f(x) = (h(x), g(x))$ then $\boxed{graph(f) = \{(x, f(x)) \mid x \in \mathbb{R}\}}$ and it is clear that $graph(f) = L$. We say $f : \mathbb{R} \to \mathbb{R}^2$ is a $\mathbb{R}^2$-valued function of a real variable. Some texts call such functions mappings whereas they insist that *functions* are real-valued. I make no such restriction in these notes. In any event, there is a natural parametrization which is induced from the graph for L, use $x = t$ hence[17]

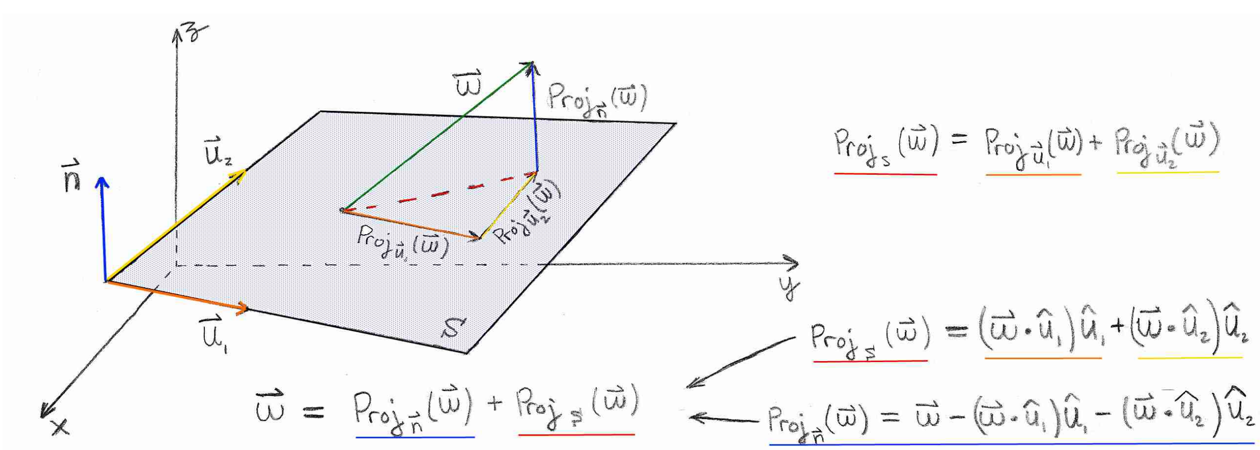
$$\boxed{\vec{r}(t) \;=\; \langle t, f(t)\rangle \;=\; \langle t, g(t), h(t)\rangle}$$

parametrizes L. We could also solve for y or z provided $b \neq 0$ or $c \neq 0$. I leave those to the reader.

1.3.5 on projections onto a plane

There are two ways to look at projections onto a plane. Let S be the plane of interest. If we have a pair of orthogonal unit-vectors $\widehat{u}_1, \widehat{u}_2$ on S then we can project a general vector $\vec{v}$ to the plane by

$$\text{Proj}_S(\vec{v}) = (\vec{v} \bullet \widehat{u}_1)\widehat{u}_1 + (\vec{v} \bullet \widehat{u}_2)\widehat{u}_2.$$

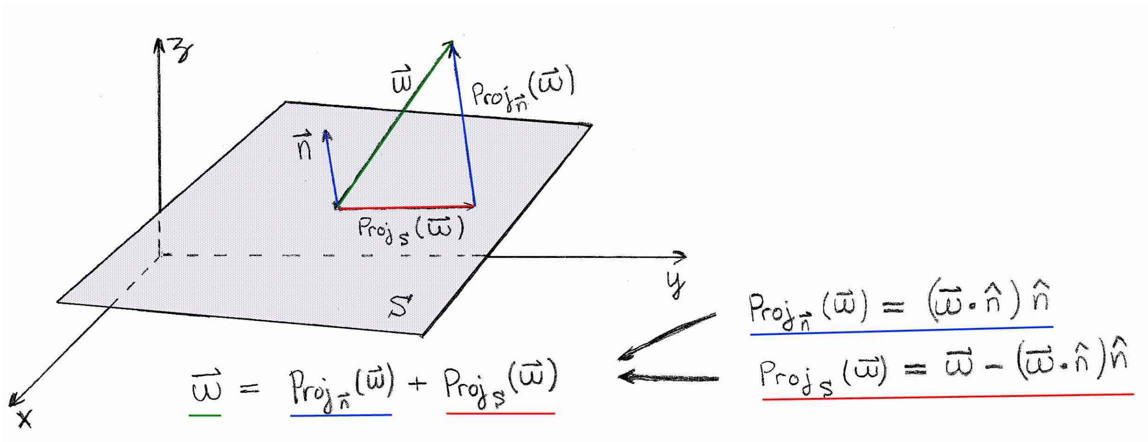


In three dimensions we can calculate this projection via subtracting off the piece of the vector which is in the normal direction. If $\vec{n}$ is the normal to S then I claim

$$\text{Proj}_S(\vec{v}) = \vec{v} - \text{Proj}_{\vec{n}}(\vec{v})$$

Here is a picture for this formula:

[17] we use the notation $((a,b),c) = (a,(b,c)) = (a,b,c)$ which implicits a bijective correspondence between these technically distinct triples, this is a common notation.



Let's see why in $n = 3$ these formulas are equivalent. Construct the normal corresponding to the unit vectors in the usual manner: $\vec{n} = \widehat{u}_1 \times \widehat{u}_2$. Orthogonality of the unit vectors implies $\theta = \frac{\pi}{2}$ hence $||\widehat{u}_1 \times \widehat{u}_2|| = 1$ and it follows $\vec{n} = \widehat{n}$. Let us define

$$\overline{\text{Proj}}_S(\vec{v}) = \vec{v} - (\vec{v} \bullet \widehat{n})\widehat{n} \quad \star$$

Observe this formula produces a vector on the plane,

$$\overline{Proj}_S(\vec{v}) \bullet \widehat{n} = \vec{v} \bullet \widehat{n} - (\vec{v} \bullet \widehat{n})\widehat{n} \bullet \widehat{n} = \vec{v} \bullet \widehat{n} - \vec{v} \bullet \widehat{n} = 0.$$

Notice that $\widehat{n} \bullet \widehat{u}_1 = 0$ and $\widehat{n} \bullet \widehat{u}_2 = 0$. Take the dot-product of $\star$ with $\widehat{u}_1$ and $\widehat{u}_2$ to obtain,

$$\overline{Proj}_S(\vec{v}) \bullet \widehat{u}_1 = \vec{v} \bullet \widehat{u}_1 \qquad \text{and} \qquad \overline{Proj}_S(\vec{v}) \bullet \widehat{u}_2 = \vec{v} \bullet \widehat{u}_2$$

It follows that $\overline{Proj}_S(\vec{v}) = \big(\vec{v} \bullet \widehat{u}_1\big)\widehat{u}_1 + \big(\vec{v} \bullet \widehat{u}_2\big)\widehat{u}_2$. The formulas agree. They both produce the same projection onto the plane. If we attach $\vec{v}$ to the plane at some point $P \in S$ then $Proj_S(\vec{v})$ attached to P will point to the point on the plane S which is closest to the end of the vector $\vec{v}$.

Perhaps the example that follows will help you understand the discussion above.

Example 1.3.9. *Let S be the plane through $(0, 0, 1)$ with normal $\vec{n} = \langle 1, 0, 1\rangle$. Notice that $\vec{u}_1 = \langle 1, 0, -1\rangle$ and $\vec{u}_2 = \langle 0, 1, 0\rangle$ are orthogonal as $\vec{u}_1 \bullet \vec{u}_2 = 0$ and they are both on S as $\vec{n} \bullet \vec{u}_1 = 0$ and $\vec{n} \bullet \vec{u}_2 = 0$. Normalize the vectors,*

$$\widehat{u}_1 = \frac{1}{\sqrt{2}}\langle 1, 0, -1\rangle, \qquad \widehat{u}_2 = \langle 0, 1, 0\rangle \qquad \widehat{n} = \frac{1}{\sqrt{2}}\langle 1, 0, 1\rangle$$

Let $\vec{v} = \langle a, b, c\rangle$. Calculate,

$$\begin{aligned} Proj_S(\vec{v}) &= \big(\vec{v} \bullet \widehat{u}_1\big)\widehat{u}_1 + \big(\vec{v} \bullet \widehat{u}_2\big)\widehat{u}_2 \\ &= \frac{1}{2}(a - c)\langle 1, 0, -1\rangle + b\langle 0, 1, 0\rangle \\ &= \langle\ \frac{1}{2}(a - c),\ b,\ \frac{1}{2}(c - a)\ \rangle. \end{aligned}$$

On the other hand, the normal formula says,

$$\begin{aligned} Proj_S(\vec{v}) &= \vec{v} - (\vec{v} \bullet \widehat{n})\widehat{n} \\ &= \langle a, b, c\rangle - \frac{1}{2}(a+c)\langle 1, 0, 1\rangle \\ &= \langle a - \frac{1}{2}(a+c), b, c - \frac{1}{2}(a+c)\rangle \\ &= \langle\ \frac{1}{2}(a-c),\ b,\ \frac{1}{2}(c-a)\ \rangle. \end{aligned}$$

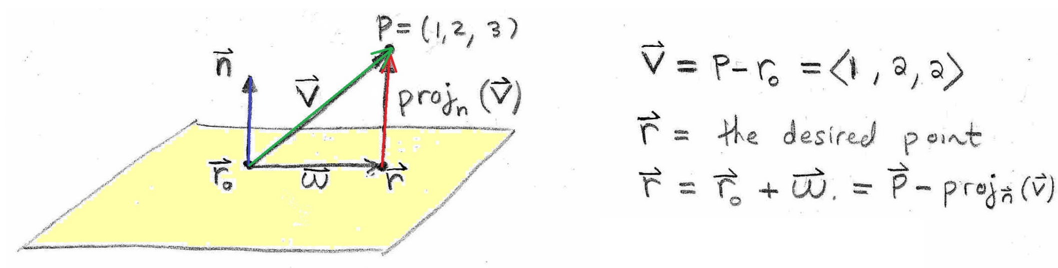
We typically use the normal formulation of the projection since it's easier to find a normal than it is to find a pair of orthonormal vectors on the plane. That said, the orthonormal projection formula naturally generalizes to higher dimensional studies. We discuss applications of such formulas in linear algebra. It is the math behind least squares data fitting and Fourier analysis.

Example 1.3.10. *Consider $S : x - y + 10z = 10$. Find the point (x, y, z) in S which is closest to the point $P = (1, 2, 3)$.*

Solution: *to begin we find a point on S by setting $x = y = 0$ for convenience. We find $\vec{r}_o = (0, 0, 1)$ is a point on S as $0 - 0 + 10(1) = 10$. Consider $\vec{v} = P - \vec{r}_o$. Explicitly,*

$$\vec{v} = P - \vec{r}_o = (1, 2, 3) - (0, 0, 1) = \langle 1, 2, 2\rangle.$$

The projection of this vector onto the normal direction is $Proj_{\vec{n}}(\vec{v})$. Geometrically, it is clear that $P - Proj_{\vec{n}}(\vec{v})$ gives the point $\vec{r} = (x, y, z)$ which is closest to P in S. See the diagram below:



Observe $\vec{n} = \langle 1, -1, 10\rangle$ is manifest from the equation $x - y + 10z = 10$ for the plane S. The projection of $\vec{v} = \langle 1, 2, 2\rangle$ onto $\vec{n}$ is calculated as we studied earlier in this chapter,

$$Proj_{\vec{n}}(\vec{v}) = \frac{\vec{n} \bullet \vec{v}}{\vec{n} \bullet \vec{n}}\,\vec{n} = \frac{19}{102}\langle 1, -1, 10\rangle$$

Therefore,

$$\vec{r} = P - Proj_{\vec{n}}(\vec{v}) = (1, 2, 3) - \frac{19}{102}\langle 1, -1, 10\rangle = \langle \tfrac{83}{102}, \tfrac{223}{102}, \tfrac{116}{102}\rangle.$$

Indeed, I invite the reader to verify that $(\frac{83}{102}, \frac{223}{102}, \frac{116}{102}) \in S$.

The fact that this point in S is closest to P can be shown algebraically as follows: let $\vec{r}\,' \in S$. We wish to argue that the distance from P to $\vec{r}\,'$ is shortest when $\vec{r}\,' = \vec{r}$. Observe $\vec{r}\,' - \vec{r}$ is tangent to S and thus perpendicular to $Proj_{\vec{n}}(\vec{v})$. Also, $P = \vec{r} + Proj_{\vec{n}}(\vec{v})$ therefore $P - \vec{r} = Proj_{\vec{n}}(\vec{v})$. Hence:

$$||P - \vec{r}\,'||^2 = ||P - \vec{r} + \vec{r} - \vec{r}\,'||^2 = ||P - \vec{r} + \vec{r} - \vec{r}\,'||^2 = ||Proj_{\vec{n}}(\vec{v})||^2 + ||\vec{r} - \vec{r}\,'||^2$$

The squared-distance $||P - \vec{r}\,'||^2$ is smallest when $\vec{r} = \vec{r}\,'$ in which case we see $||P - \vec{r}|| = ||Proj_{\vec{n}}(\vec{v})||$.

1.3.6 additional examples

Example 1.3.11. *Find the parametrization of the line from point P to point Q.*

Solution: *there is a simple formula to use here:*

$$\vec{r}(t) = P + t(Q - P)$$

Observe that $\vec{r}(0) = P$ and $\vec{r}(1) = P + Q - P = Q$. It follows the parametrization is that of a line from P to Q. Moreover, you can see $Q - P$ is the direction-vector of the parametrization.

Example 1.3.12. *Let L be a line with direction-vector $\vec{v} = \langle 1, 0, 3\rangle$ and point $\vec{r}_o = (\pi, \pi, \pi)$. Find the vector and scalar parametric equations for L.*

Solution: *The vector parametric equations for L are simply:*

$$\vec{r}(t) = (\pi, \pi, \pi) + t\langle 1, 0, 3\rangle.$$

We find the scalar parametric equations by writing $\vec{r}(t) = \langle x(t), y(t), z(t)\rangle$. It follows:

$$x(t) = \pi + t, \quad y(t) = \pi, \quad z(t) = \pi + 3t$$

Sometimes we omit the explicit function notation and write $x = \pi + t, y = \pi, z = \pi + 3t$.

I prefer to refer to the calculation above as giving a parametrization to L. To find a parametrization is to provide a mapping whose image covers the given object. In contrast, an equation is better associated with an object which is described as a solution set. That said, the terminology *parametric equations* is common and my complaint here is not too serious.

Example 1.3.13. *Suppose L is parametrized by $\vec{r}(t) = (3 - t, t + 5, 2t + 8)$ for all $t \in \mathbb{R}$. Find the direction-vector and natural base-point for L. Also, derive the Cartesian equations for L.*

Solution: *Observe*

$$\vec{r}(t) = \langle 3 - t, t + 5, 2t + 8\rangle = \langle 3, 5, 8\rangle + t\langle -1, 1, 2\rangle$$

thus, by inspection, $\vec{r}_o = (3, 5, 8)$ is a base-point for L and the direction vector is $\vec{v} = \langle -1, 1, 2\rangle$. The Cartesian equations can be derived by eliminating t. Algebraically, this is accomplished by solving $x = 3 - t, y = t + 5$ and $z = 2t + 8$ for t:

$$t = \underbrace{3 - x = y - 5 = \frac{z - 8}{2}}_{\textit{Cartesian eqns. for } L}.$$

Geometrically, these equations correspond to supposing (x, y, z) is found on two planes. The intersection of which is the line L.

Generally, a line L with base point (x_o, y_o, z_o) and direction-vector $\vec{v} = \langle a, b, c\rangle$ (with $abc \neq 0$) has Cartesian equations $\frac{x - x_o}{a} = \frac{y - y_o}{b} = \frac{z - z_o}{c}$. These are known as the **symmetric equations for the line**. However, if $abc = 0$ then the symmetric equations cannot be written due to division by zero. In contrast, the parametric formulation of the line suffices for all possible lines. Moreover, the parametric formulation works as well on one-dimension as it does in n-dimensions; $\vec{r}(t) = \vec{r}_o + t\vec{v}$ parametrizes a line in $\mathbb{R}^n$ for $n \geq 1$. However, generally, it takes $(n - 1)$ Cartesian equations for a line in n-dimensions. In the case of $\mathbb{R}^2$ we just need one equation $y - mx = b$, in $n = 3$ it takes two independent equations, in $n = 4$ it would take 3.

Example 1.3.14. *Suppose L_1 is a line with $\vec{r}_1(t) = \langle 3+t, 2+t, 1+t\rangle$ and L_2 is a second line parametrized by $\vec{r}_2(t) = \langle 3t, 3t, 3t+6\rangle$. What can we say about L_1 and L_2?*

Solution: *Observe the direction vector of $\vec{r}_1(t)$ is $\vec{v}_1 = \langle 1,1,1\rangle$ whereas the direction vector of $\vec{r}_2(t)$ is $\vec{v}_2 = \langle 3,3,3\rangle$. Clearly $\vec{v}_2 = 3\vec{v}_1$ hence these lines share the same direction. We say L_1 and L_2 are parallel lines.*

Example 1.3.15. *Find a parametrization of a line which goes through $P = (-2,4,10)$ and is parallel to the line through $R = (2,2,2)$ and $S = (5,3,-6)$.*

Solution: *First, note the direction of the line through R and S can be taken to be $S - R = \langle 3,1,-8\rangle$. Second, we're given base-point P. Therefore, the desired line has parametrization:*

$$\vec{r}(t) = (-2,4,10) + t\langle 3,1,-8\rangle.$$

Or, if you prefer, $x = 3t-2, y = 4+t$ and $z = 10-8t$.

I should warn, the answers in this section are far from unique. We should appreciate there are infinitely many parametrizations for a given set of points. That said, the solutions I've given here are fairly standard.

Example 1.3.16. *Let $L_1 : \vec{r}_1(t) = \langle 1+t, 3t-2, 4-t\rangle$ and $L_2 : \vec{r}_2(t) = \langle 2t, 3+t, 4t-3\rangle$. Show that L_1 and L_2 do not intersect.*

Solution: *to be fair we ought not use t for both parametrizations as they may meet for differing parameters. Thus, we search for solutions of $\vec{r}_1(t) = \vec{r}_2(s)$:*

$$\langle 1+t, 3t-2, 4-t\rangle = \langle 2s, 3+s, 4s-3\rangle \quad \Rightarrow \quad \begin{array}{c} 1+t = 2s \\ 3t-2 = 3+s \\ 4-t = 4s-3 \end{array}$$

Adding the first and third equations gives $5 = 6s-3$ hence $6s = 8$ hence $s = 4/3$. However, then from the first equation $t = 2s-1 = 8/3-1 = 5/3$. Now, check the second equation for consistency: $3t-2 = 3(5/3)-2 = 9/3 = 3$ whereas $3+s = 3+8/3 = 14/3$. Clearly equation two is false. We cannot simultaneously solve this set of 3 equations and 2 unknowns hence there is no intersection point for L_1 and L_2.

One possible modification of the preceding example is given as follows: when t is time and $\vec{r}(t)$ is position at time t then find a solution of $\vec{r}_1(t) = \vec{r}_2(t)$ indicates a **collision** of particles 1 and 2.

Example 1.3.17. *Let $P = (2,1,0)$ and $Q = (3,4,1)$ and $R = (4,5,6)$ be in a plane S. Find a parametrization of the plane, a normal to the plane and its Cartesian equation.*

Solution: *one method to write the parametric equations is to extend the technique of Example 1.3.11 to three points:*

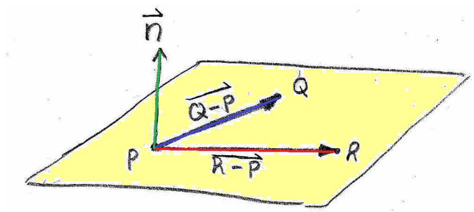
$$\vec{r}(s,t) = P + s(Q-P) + t(R-P) = (2,1,0) + s\langle 1,3,1\rangle + t\langle 2,4,6\rangle.$$

you can check $\vec{r}(0,0) = P$, $\vec{r}(1,0) = Q$ and $\vec{r}(0,1) = R$ hence $\vec{r}(s,t)$ yields the desired parametrization of the plane.

Notice that vectors connecting the given points are tangent to S. Let us define:

$$\vec{A} = Q - P = \langle 1, 3, 1\rangle \qquad \vec{B} = R - P = \langle 2, 4, 6\rangle.$$

The normal to the plane must be perpendicular to both $\vec{A}$ and $\vec{B}$. Recall, this was exactly what we designed $\vec{A}\times\vec{B}$ to accomplish.



Calculate,

$$\begin{aligned}\vec{n} = \vec{A}\times\vec{B} = det &\begin{bmatrix} \widehat{x} & \widehat{y} & \widehat{z} \\ 1 & 3 & 1 \\ 2 & 4 & 6 \end{bmatrix} \\ &= \widehat{x}(18-4) - \widehat{y}(6-2) + \widehat{z}(4-6) \\ &= \langle 14, -4, -2\rangle.\end{aligned}$$

Check that $\vec{A}$ and $\vec{B}$ are orthogonal to $\vec{A}\times\vec{B}$. Yes. Thus, we find normal to the plane $\vec{n} = \langle 14, -4, -2\rangle$ and the Cartesian equation of the plane as follows:

$$14(x-2) - 4(y-1) - 2z = 0.$$

I used P as the base-point for the equation above. You can use either P, Q or R, they all will yield an equivalent equation.

Note: we say two equations are equivalent if they share the same solution set.

Example 1.3.18. *Find the plane through $(1, 2, 3)$ with normal parallel to the intersection line of the planes $P_1 : x + y + z = 10$ and $P_2 : 2x + 3y + z = 20$.*

Solution: *I begin by finding the line of intersection. A general principle you ought to think over: the intersection points are where both equations hold true. Furthermore, a convenient method to find the line is simply to locate two points on the line of intersection. Begin by hoping for a point with $x = 0$ hence $P_1(x = 0) : y + z = 10$ whereas $P_2(x = 0) : 3y + z = 20$. Subtracting these yields $2y = 10$ thus $y = 5$ and it follows $z = 5$. This gives us $Q_1 = (0, 5, 5)$ on the line of intersection. Next, I search for the point on the line of intersection where $y = 0$. Observe, $P_1(y = 0) : x + z = 10$ and $P_2(y = 0) : 2x + z = 20$. Again, subtracting equations yields $x = 10$ and thus $z = 0$. We find $Q_2 = (10, 0, 0)$ on the line of intersection. The direction of the line of intersection is given by $\vec{v} = Q_2 - Q_1 = \langle 10, -5, -5\rangle$. Note, we can use $\vec{n} = \langle 2, -1, -1\rangle$ as it points in the same direction. Therefore, the plane we seek is simply:*

$$2(x-1) - (y-2) - (z-3) = 0 \quad\Rightarrow\quad \boxed{2x - y - z = -3.}$$

There are probably a half-dozen other popular methods to solve the preceding example. I merely try to give you some ideas about possible methods. Sometimes we just need to find a point or two and setting a coordinate or two to zero simplifies the algebra considerably.

Example 1.3.19. *Find the line through $(2,1,0)$ and perpendicular in direction to both $\vec{A} = \widehat{x} + \widehat{y}$ and $\vec{B} = \widehat{y} + \widehat{z}$.*

Solution: *if the direction of the line is perpendicular to both $\vec{A}$ and $\vec{B}$ then (by geometry) it must be colinear to $\vec{A} \times \vec{B}$. Thus calculate,*

$$\vec{A} \times \vec{B} = (\widehat{x} + \widehat{y}) \times (\widehat{y} + \widehat{z}) = \widehat{z} - \widehat{y} + \widehat{x}.$$

Therefpre, the desired line has parametrization $\vec{r}(t) = (2,1,0) + t\langle 1,-1,1\rangle$.

Example 1.3.20. *Find the line through $(1,0,6)$ and perpendicular to the plane $x + 3y + z = 5$.*

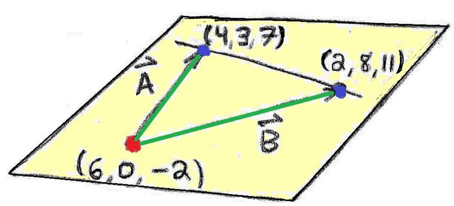
Solution: *the line is perpendicular to the plane if it is parallel to the normal of the plane. The normal to the plane is $\langle 1,3,1\rangle$ by inspection of the given equation. Thus, $\vec{r}(t) = (1,0,6) + t\langle 1,3,1\rangle$ parametrizes the desired line. In other words, the line is given by $x = 1 + t, y = 3t$ and $z = 6 + t$.*

Example 1.3.21. *Given a plane contains the point $P = (6,0,-2)$ and a line with parametric equations $x = 4 - 2t$, $y = 3 + 5t$ and $z = 7 + 4t$. Find the Cartesian equation for the plane.*

Solution: *there are many ways to solve this. Note, choosing $t = 0$ and $t = 1$ give us two points on the plane:*

$$Q = \vec{r}(0) = (4,3,7) \qquad \& \qquad R = \vec{r}(1) = (2,8,11).$$

Next construct tangent vectors to the plane by the directed line segments:



We connect P with Q and P with R:

$$\vec{A} = Q - P = \langle -2,3,9\rangle \qquad \& \qquad \vec{B} = R - P = \langle -4,8,13\rangle$$

We know $\vec{A} \times \vec{B}$ makes a good normal as it is perpendicular to both $\vec{A}$ and $\vec{B}$:

$$\begin{aligned}
\vec{n} = \vec{A} \times \vec{B} = det \begin{bmatrix} \widehat{x} & \widehat{y} & \widehat{z} \\ -2 & 3 & 9 \\ -4 & 8 & 13 \end{bmatrix} \\
&= \widehat{x}(39 - 72) - \widehat{y}(-26 + 36) + \widehat{z}(-16 + 12) \\
&= \langle -33, -10, -4\rangle.
\end{aligned}$$

Therefore, using P as the base-point, we find the plane is $-33(x-6) - 10y - 4(z+2) = 0$.

1.4 curves

A curve is a one-dimensional subset of some space. There are at least three common, but distinct, ways to frame the mathematics of a curve. These viewpoints were already explored in the previous section but I list them once more: we can describe a curve:

1. as a path, that is as a parametrized curve.
2. as a level curve, also known as a solution set.
3. as a graph.

I expect you master all three viewpoints in the two-dimensional context. However, for three or more dimensions we primarily use the parametric viewpoint in this course. Exceptions to this rule are fairly rare: the occasional homework problem where you are asked to find the curve of intersection for two surfaces, or the symmetric equations for a line. In contrast, the parametric decription of a curve in three dimensions is far more natural. Do you want to describe a curve as where two surfaces intersect or would you rather describe a curve as a set of points formed by pasting a copy of the real line through your space? I much prefer the parametric view[18].

Definition 1.4.1. *vector-valued functions, curves and paths.*

A vector valued function of a real variable is an assignment of a vector for each real number in some domain. It's a mapping $t \mapsto \vec{f}(t) = \langle f_1(t), f_2(t), \dots, f_n(t) \rangle$ for each $t \in J \subset \mathbb{R}$. We say $f_j : J \subseteq \mathbb{R} \to \mathbb{R}$ is the j-th component function of $\vec{f}$. Let $C = \vec{f}(J)$ then C is said to be a **curve which is parametrized by** $\vec{f}$. We can also say that $t \mapsto \vec{f}(t)$ is a **path** in $\mathbb{R}^n$. Equivalently, but not so usefully, we can write the scalar parametric equations for C above as

$$x_1 = f_1(t), \quad x_2 = f_2(t), \ \dots, \ x_n = f_n(t)$$

for all $t \in J$.

When we define a parametrization of a curve it is important to give the formula for the path **and** the domain of the parameter[19]. Note that when I say the word *curve* I mean for us to think about some set of points, whereas when I say the word *path* I mean to refer to the particular mapping whose image is a curve. We may cover a particular curve with infinitely many different paths.

1.4.1 curves in two-dimensional space

We have several viewpoints to consider. Graphs, parametrized curves and level sets.

graphs in the plane

Let's begin by reminding ourselves of the definition of a graph:

Definition 1.4.2. *Graph of a function.*

Let $f : dom(f) \to \mathbb{R}$ be a function then

$$graph(f) = \{(x, f(x)) \mid x \in dom(f)\}.$$

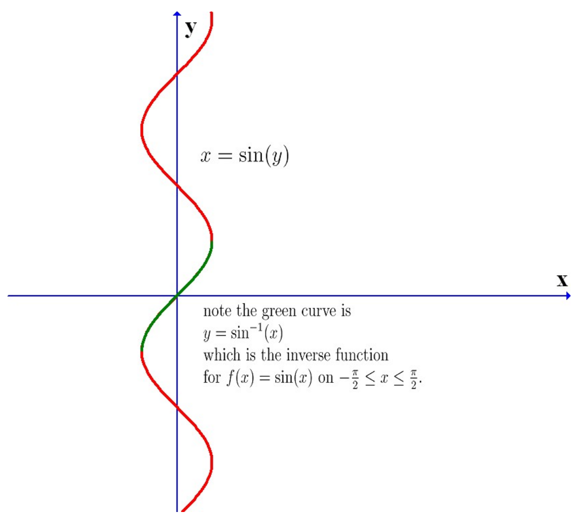
[18] algebraic geometers take a different view...

[19] this was not an issue for the lines considered in the last section as the domain was simply $\mathbb{R}$ in that context

We know this is quite restrictive. We must satisfy the vertical line test if we say our curve is the graph of a function.

Example 1.4.3. *To form a circle centered at the origin of radius R we need to glue together two graphs. In particular we solve the equation $x^2 + y^2 = R^2$ for $y = \sqrt{R^2 - x^2}$ or $y = -\sqrt{R^2 - x^2}$. Let $f(x) = \sqrt{R^2 - x^2}$ and $g(x) = -\sqrt{R^2 - x^2}$ then we find $graph(f) \cup graph(g)$ gives us the whole circle.*

Example 1.4.4. *On the other hand, if we wish to describe the set of all points such that $\sin(y) = x$ we also face a similar difficulty if we insist on functions having independent variable x. Naturally, if we allow for functions with y as the independent variable then $f(y) = \sin(y)$ has graph $graph(f) = \{(f(y), y) \mid y \in dom(f)\}$. You might wonder, is this correct? I would say a better question is, "is this allowed?". Different books are more or less imaginative about what is permissible as a function. This much we can say, if a shape fails both the vertical and horizontal line tests then it is not the graph of a single function of x or y.*



Example 1.4.5. *Let $f(x) = mx + b$ for some constants m, b then $y = f(x)$ is the line with slope m and y-intercept b.*

level curves in two-dimensions

Level curves are amazing. The full calculus of level curves is only partially appreciated even in calculus III, but trust me, this viewpoint has many advantages as you learn more. For now it's simple enough:

Definition 1.4.6. *Level Curve.*

A level curve is given by a function of two variables $F : dom(F) \subseteq \mathbb{R}^2 \to \mathbb{R}$ and a constant k. In particular, the set of all $(x, y) \in \mathbb{R}^2$ such that $F(x, y) = k$ is called the level-set of F, but more commonly we just say $F(x, y) = k$ is a level curve.

In an algebra class you might have called this the "graph of an equation", but that terminology is dead to us now. For us, it is a level curve. Moreover, for a particular set of points $C \subseteq \mathbb{R}^2$ we can find more than one function F which produces C as a level set. Unlike functions, for a particular curve there is not just one function which returns that curve. This means that it might be important to give both the level-function F and the level k to specify a level curve $F(x, y) = k$.

Example 1.4.7. *A circle of radius R centered at the origin is a level curve $F(x,y) = R^2$ where $F(x,y) = x^2 + y^2$. We call F the level function (of two variables).*

Example 1.4.8. *To describe $\sin(y) = x$ as a level curve we simply write $\sin(y) - x = 0$ and identify the level function is $F(x,y) = \sin(y) - x$ and in this case $k = 0$. Notice, we could just as well say it is the level curve $G(x,y) = 1$ where $G(x,y) = x - \sin(y) + 1$.*

Note once more this type of ambiguity is one distinction of the level curve language, in contrast, the graph *graph*(f) of a function $y = f(x)$ and the function f are interchangeable. Some mathematicians insist the rule $x \mapsto f(x)$ defines a function whereas others insist that a function is a set of pairs $(x, f(x))$. I prefer the mapping rule because it's how I think about functions in general whereas the idea of a graph is much less useful in general.

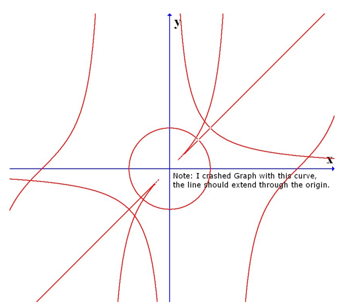
Example 1.4.9. *A line with slope m and y-intercept b can be described by $F(x,y) = mx+b-y = 0$. Alternatively, a line with x-intercept x_o and y-intercept y_o can be described as the level curve $G(x,y) = \frac{x}{x_o} + \frac{y}{y_o} = 1$.*

Example 1.4.10. *Level curves need not be simple things. They can be lots of simple things glued together in one grand equation:*

$$F(x,y) = (x-y)(x^2+y^2-1)(xy-1)(y-\tan(x)) = 0.$$

Solutions to the equation above include the line $y = x$, the unit circle $x^2+y^2 = 1$, the tilted-hyperbola known more commonly as the reciprocal function $y = \frac{1}{x}$ and finally the graph of the tangent. Some of these intersect, others are disconnected from each other.

It is sometimes helpful to use software to plot equations. However, we must be careful since they are not as reliable as you might suppose. The example above is not too complicated but look what happens with Graph:



Theorem 1.4.11. *any graph of a function can be wriiten as a level curve.*

> If $y = f(x)$ is the graph of a function then we can write $F(x,y) = f(x) - y = 0$ hence the graph $y = f(x)$ is also a level curve.

Not much of a theorem. But, it's true. The converse is not true without a lot of qualification. I'll state that theorem (it's called the implicit function theorem) in a future chapter after we've studied partial differentiation.

parametrized curves in two-dimensions

Example 1.4.12. *Suppose $a, b > 0$ and $h, k \in \mathbb{R}$. The parametrization*

$$\vec{r}(t) = \langle h + a\cos(t), k + b\sin(t)\rangle$$

for $t \in [0, 2\pi]$ coveres and the ellipse

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

Example 1.4.13. *Suppose $a, b > 0$ and $h, k \in \mathbb{R}$. The parametrization*

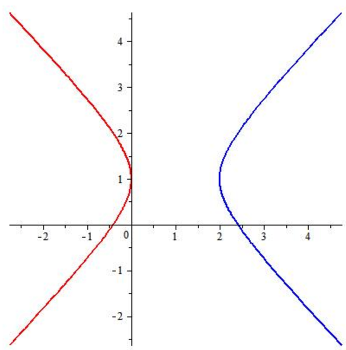
$$\vec{r}_1(t) = \langle h + a\cosh(t), k + b\sinh(t)\rangle$$

for $t \in \mathbb{R}$ coveres one branch of the hyperbola

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

Note $x = h + a\cosh(t)$ implies $\frac{x-h}{a} = \cosh(t) \geq 1$ therefore it follows $x \geq h + a$. We've covered the right branch. If we wish to cover the left branch of this hyperbola then use:

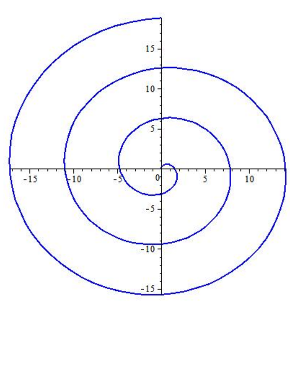
$$\vec{r}_2(t) = \langle h - a\cosh(t), k + b\sinh(t)\rangle.$$



Example 1.4.14. *A spiral can be thought of as a sort of circle with a variable radius. With that in mind I write: for $t \geq 0$,*

$$\vec{r}(t) = \langle t\cos(t), t\sin(t)\rangle$$

to give a spiral whose "radius" is proportional to the angle t subtended from $t = 0$.

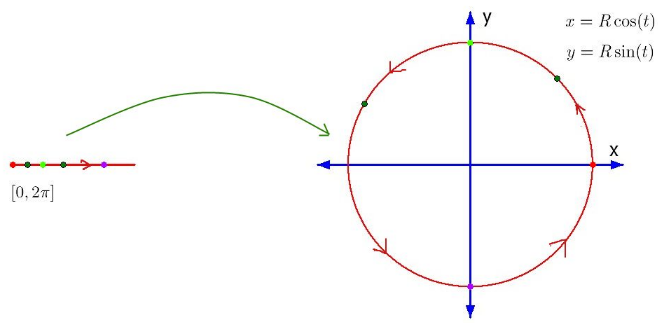


Finding the parametric equations for a curve does require a certain amount of creativity. However, it's almost always some slight twist on the examples I give in this section. The remaining examples I also give in calculus II, I add some detail to emphasize how the parametrization matches the already known identities of certain curves and I add pictures which emphasize the idea that the parametrization pastes a line into $\mathbb{R}^2$.

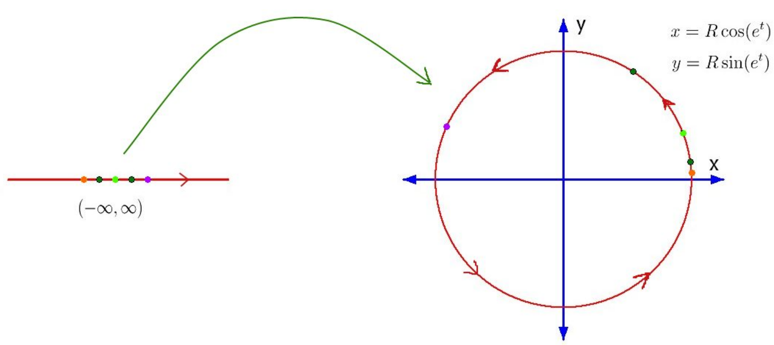
Example 1.4.15. *Let $x = R\cos(t)$ and $y = R\sin(t)$ for $t \in [0, 2\pi]$. This is a parametrization of the circle of radius R centered at the origin. We can check this by substituting the equations back into our standard Cartesian equation for the circle:*

$$x^2 + y^2 = (R\cos(t))^2 + (R\sin(t))^2 = R^2(\cos^2(t) + \sin^2(t))$$

Recall that $\cos^2(t) + \sin^2(t) = 1$ therefore, $x(t)^2 + y(t)^2 = R^2$ for each $t \in [0, 2\pi]$. This shows that the parametric equations do return the set of points which we call a circle of radius R. Moreover, we can identify the parameter in this case as the standard angle from standard polar coordinates.



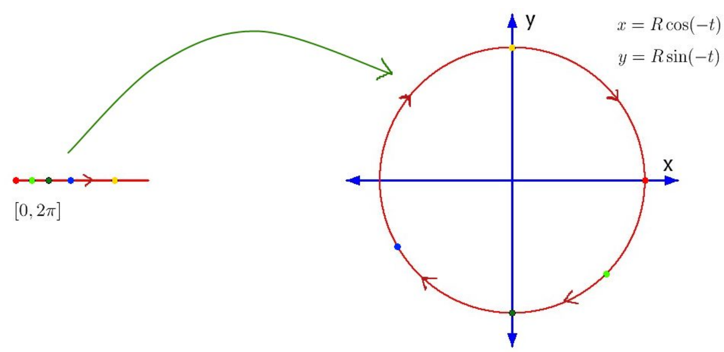
Example 1.4.16. *Let $x = R\cos(e^t)$ and $y = R\sin(e^t)$ for $t \in \mathbb{R}$. We again cover the circle at t varies since it is still true that $(R\cos(e^t))^2 + (R\sin(e^t))^2 = R^2(\cos^2(e^t) + \sin^2(e^t)) = R^2$. However, since range$(e^t) = [1, \infty)$ it is clear that we will actually wrap around the circle infinitely many times. The parametrizations from this example and the last do cover the same set, but they are radically different parametrizations: the last example winds around the circle just once whereas this example winds around the circle ∞-ly many times.*



Example 1.4.17. *Let $x = R\cos(-t)$ and $y = R\sin(-t)$ for $t \in [0, 2\pi]$. This is a parametrization of the circle of radius R centered at the origin. We can check this by substituting the equations back into our standard Cartesian equation for the circle:*

$$x^2 + y^2 = (R\cos(-t))^2 + (R\sin(-t))^2 = R^2(\cos^2(-t) + \sin^2(-t))$$

Recall that $\cos^2(-t)+\sin^2(-t) = 1$ *therefore,* $x(t)^2+y(t)^2 = R^2$ *for each* $t \in [0, 2\pi]$. *This shows that the parametric equations do return the set of points which we call a circle of radius* R. *Moreover, we can identify the parameter an angle measured CW*[20] *from the positive* x*-axis. In contrast, the standard polar coordinate angle is measured CCW from the postive* x*-axis. Note that in this example we cover the circle just once, but the direction of the curve is opposite that of Example 1.4.15.*



The idea of directionality is not at all evident from Cartesian equations for a curve. Given a graph $y = f(x)$ or a level-curve $F(x, y) = k$ there is no intrinsic concept of direction ascribed to the curve. For example, if I ask you whether $x^2 + y^2 = R^2$ goes CW or CCW then you ought not have an answer. I suppose you could ad-hoc pick a direction, but it wouldn't be natural. This means that if we care about giving a direction to a curve we need the concept of the parametrized curve. We can use the ordering of the real line to induce an ordering on the curve.

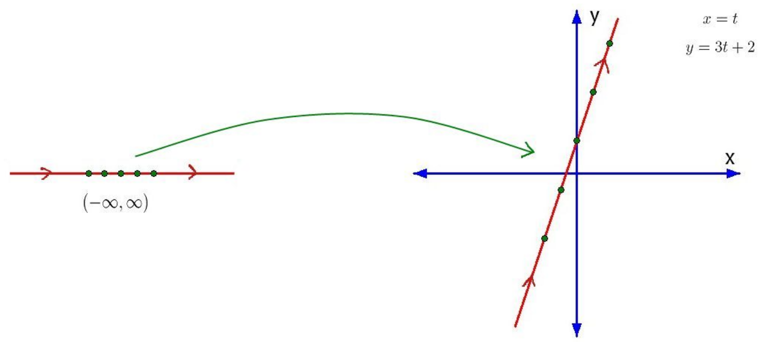
Definition 1.4.18. *oriented curve.*

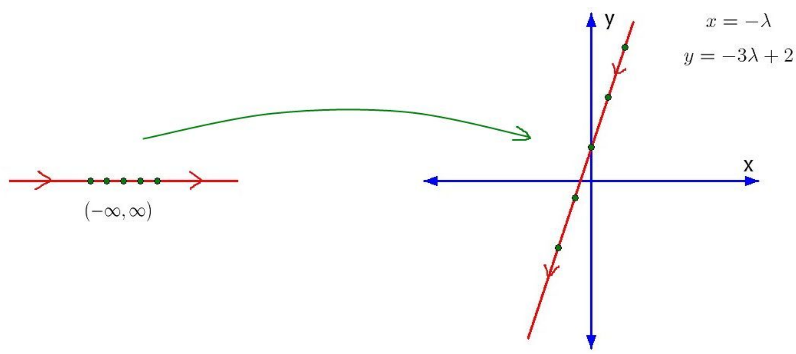
Suppose $f, g : J \subseteq \mathbb{R} \to \mathbb{R}$ are $1-1$ functions. We say the set $\{(f(t), g(t)) \mid t \in J\}$ is an **oriented curve** and say $t \to (f(t), g(t))$ is a consistently oriented **path** which covers C. If $J = [a, b]$ and $(f(a), g(a)) = p$ and $(f(b), g(b)) = q$ then we can say that C is a curve from to p to q.

I often illustrate the orientation of a curve by drawing little arrows along the curve to indicate the direction. Furthermore, in my previous definition of parametrization I did not insist the parametric functions were $1-1$, this means that those parametrizations could reverse direction and go back and forth along a given curve. What is meant by the terms "path", "curve" and "parametric equations" may differ from text to text so you have to keep a bit of an open mind and try to let context be your guide when ambguity occurs. I will try to be uniform in my language within this course.

Example 1.4.19. *The line* $y = 3x + 2$ *can be parametrized by* $x = t$ *and* $y = 3t + 2$ *for* $t \in \mathbb{R}$. *This induces an orientation which goes from left to right for the line. On the other hand, if we use* $x = -\lambda$ *and* $y = -3\lambda + 2$ *then as* λ *increases we travel from right to left on the curve. So the* λ*-equations give the line the opposite orientation.*

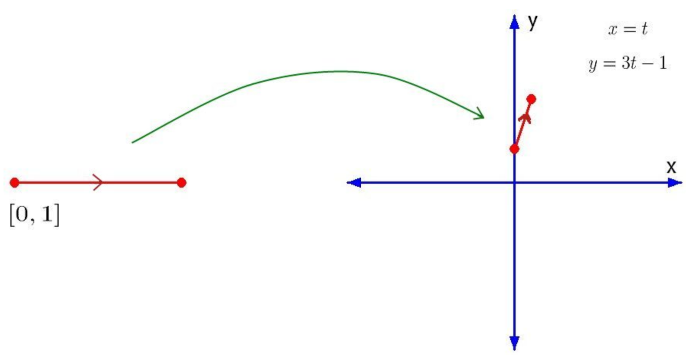
[20] CW is an abbreviation for ClockWise,whereas CCW is an abbreviation for CounterClockWise.

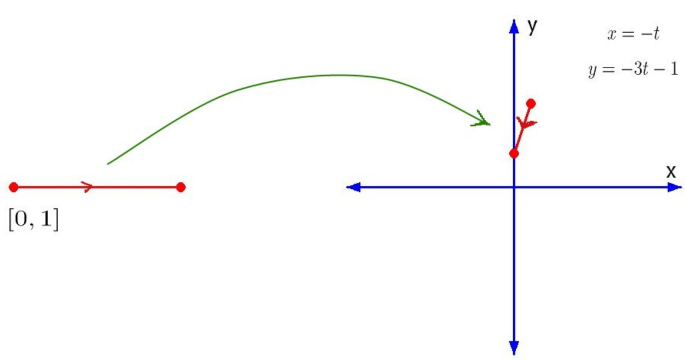




To reverse orientation for $x = f(t), y = g(t)$ for $t \in J = [a, b]$ one may simply replace t by $-t$ in the parametric equations, this gives new equations which will cover the same curve via $x = f(-t), y = g(-t)$ for $t \in [-a, -b]$.

Example 1.4.20. *The line-segment from $(0, -1)$ to $(1, 2)$ can be parametrized by $x = t$ and $y = 3t - 1$ for $0 \leq t \leq 1$. On the other hand, the line-segment from $(1, 2)$ to $(0, -1)$ is parametrized by $x = -t, y = -3t - 1$ for $-1 \leq t \leq 0$.*





The other method to graph parametric curves is simply to start plugging in values for the parameter and assemble a table of values to plot. I have illustrated that in part by plotting the green dots in the domain of the parameter together with their images on the curve. Those dots are the results of plugging in the parameter to find corresponding values for x, y. I don't find that is a very reliable approach in the same way I find plugging in values to $f(x)$ provides a very good plot of $y = f(x)$. That sort of brute-force approach is more appropriate for a CAS system. There are many excellent tools for plotting parametric curves, hopefully I will have some posted on the course website. In addition, the possibility of animation gives us an even more exciting method for visualization of the time-evolution of a parametric curve. In the next chapter we connect the parametric viewpoint with physics and such an animation actually represents the physical motion of some object. My focus in the remainder of this chapter is almost purely algebraic, I could draw pictures to explain, but I wanted the notes to show you that the pictures are not necessary when you understand the algebraic process. That said, the best approach is to do some combination of algebraic manipulation/figuring and graphical reasoning.

converting to and from the parametric viewpoint in 2D

Let's change gears a bit, we've seen that parametric equations for curves give us a new method to describe particular geometric concepts such as orientability or multiple covering. Without the introduction of the parametric concept these geometric ideas are not so easy to describe. That said, I now turn to the question of how to connect parametric descriptions with Cartesian descriptions of a curve. We'd like to understand how to go both ways if possible:

1. how can we find the Cartesian form for a given parametric curve?
2. how can we find a parametrization of a given Cartesian curve?

In case (2.) we mean to include the ideas of level curves and graphs. It turns out that both questions can be quite challenging for certain examples. However, in other cases, not so much: for example any graph $y = f(x)$ is easily recast as the set of parametric equations $x = t$ and $y = f(t)$ for $t \in dom(f)$. For the standard graph of a function we use x as the parameter.

1.4.2 how can we find the Cartesian form for a given parametric curve?

Example 1.4.21. *What curve has parametric equations $x = t$ for $y = t^2$ for $t \in \mathbb{R}$? To find Cartesian equation we eliminate the parameter (when possible)*

$$t^2 = x^2 = y \quad \Rightarrow \quad y = x^2$$

Thus the Cartesian form of the given parametrized curve is simply $y = x^2$.

Example 1.4.22. *Example 15.2.2: Find parametric equations to describe the graph $y = \sqrt{x+3}$ for $0 \leq x < \infty$. We can use $x = t^2$ and $y = \sqrt{t^2+3}$ for $t \in \mathbb{R}$. Or, we could use $x = \lambda$ and $y = \sqrt{\lambda+3}$ for $\lambda \in [0, \infty)$.*

Example 1.4.23. *What curve has parametric equations $x = t$ for $y = t^2$ for $t \in [0, 1]$? To find Cartesian equation we eliminate the parameter (when possible)*

$$t^2 = x^2 = y \quad \Rightarrow \quad y = x^2$$

Thus the Cartesian form of the given parametrized curve is simply $y = x^2$, however, given that $0 \leq t \leq 1$ and $x = t$ it follows we do not have the whole parabola, instead just $y = x^2$ for $0 \leq x \leq 1$.

Example 1.4.24. *Identify what curve has parametric equations $x = \tan^{-1}(t)$ and $y = \tan^{-1}(t)$ for $t \in \mathbb{R}$. Recall that range$(\tan^{-1}(t)) = (-\pi/2, \pi/2)$. It follows that $-\pi/2 < x < \pi/2$. Naturally we just equate inverse tangent to obtain $\tan^{-1}(t) = y = x$. The curve is the open line-segment with equation $y = x$ for $-\pi/2 < x < \pi/2$. This is an interesting parameterization, notice that as $t \to \infty$ we approach the point $(\pi/2, \pi/2)$, but we never quite get there.*

Example 1.4.25. *Consider $x = \ln(t)$ and $y = e^t - 1$ for $t \geq 1$. We can solve both for t to obtain*

$$t = e^x = \ln(y+1) \quad \Rightarrow \quad y = -1 + exp(exp(x)).$$

The domain for the expression above is revealed by analyzing $x = \ln(t)$ for $t \geq 1$, the image of $[1, \infty)$ under natural log is precisely $[0, \infty)$; $\ln[1, \infty) = [0, \infty)$.

Example 1.4.26. *Suppose $x = \cosh(t) - 1$ and $y = 2\sinh(t) + 3$ for $t \in \mathbb{R}$. To eliminate t it helps to take an indirect approach. We recall the most important identity for the hyperbolic sine and cosine: $\cosh^2(t) - \sinh^2(t) = 1$. Solve for hyperbolic cosine; $\cosh(t) = x + 1$. Solve for hyperbolic sine; $\sinh(t) = \frac{y-3}{2}$. Now put these together via the identity:*

$$\cosh^2(t) - \sinh^2(t) = 1 \quad \Rightarrow \quad (x+1)^2 - \frac{(y-3)^2}{4} = 1.$$

Note that $\cosh(t) \geq 1$ hence $x + 1 \geq 1$ thus $x \geq 0$ for the curve described above. On the other hand y is free to range over all of $\mathbb{R}$ since hyperbolic sine has range $\mathbb{R}$. You should[21] *recognize the equation as a hyperbola centered at $(-1, 3)$.*

how can we find a parametrization of a given Cartesian curve?

I like this topic more, the preceding bunch of examples, while needed, are boring. The art of parameterizing level curves is much more fun.

Example 1.4.27. Find parametric equations for the circle centered at (h, k) with radius R.

To begin recall the equation for such a circle is $(x-h)^2 + (y-k)^2 = R^2$. Our inspiration is the identity $\cos^2(t) + \sin^2(t) = 1$. Let $x - h = R\cos(t)$ and $y - k = R\sin(t)$ thus

$$\boxed{x = h + R\cos(t)} \qquad \textit{and} \qquad \boxed{y = k + R\sin(t).}$$

I invite the reader to verify these do indeed parametrize the circle by explicitly plugging in the equations into the circle equation. Notice, if we want the whole circle then we simply choose any interval for t of length 2π or longer. On the other hand, if you want to select just a part of the circle you need to think about where sine and cosine are positive and negative. For example, if I want to parametrize just the part of the circle for which $x > h$ then I would choose $t \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

The reason I choose that intuitively is that the parametrization given for the circle above is basically built from polar coordinates[22] centered at (h, k). That said, to be sure about my choice of parameter domain I like to actually plug in some of my proposed domain and make sure it matches the desired criteria. I think about the graphs of sine and cosine to double check my logic. I know that

[21] many students need to review these at this point, we use circles, ellipses and hyperbolas as examples in this course. I'll give examples of each in this chapter.

[22] we will discuss further in a later section, but this should have been covered in at least your precalculus course.

$\cos(-\frac{\pi}{2}, \frac{\pi}{2}) = (0, 1]$ whereas $\sin(-\frac{\pi}{2}, \frac{\pi}{2}) = (-1, 1)$, I see it in my mind. Then I think about the parametric equations in view of those facts,

$$x = h + R\cos(t) \qquad \text{and} \qquad y = k + R\sin(t).$$

I see that x will range over $(h, h + R]$ and y will range over $(k - R, k + R)$. This is exactly what I should expect geometrically for half of the circle. Visualize that $x = h$ is a vertical line which cuts our circle in half. These are the thoughts I think to make certain my creative leaps are correct. I would encourage you to think about these matters. Don't try to just memorize everything, it will not work for you, there are simply too many cases. It's actually way easier to just understand these as a consequence of trigonometry, algebra and analytic geometry.

Example 1.4.28. Find parametric equations for the level curve $x^2 + 2x + \frac{1}{4}y^2 = 0$ which give the ellipse a CW orientation.

To begin we complete the square to understand the equation:

$$x^2 + 2x + \frac{1}{4}y^2 = 0 \quad \Rightarrow \quad (x + 1)^2 + \frac{1}{4}y^2 = 1.$$

We identify this is an ellipse centered at $(-1, 0)$. Again, I use the pythagorean trig. identity as my guide: I want $(x + 1)^2 = \cos^2(t)$ and $\frac{1}{4}y^2 = \sin^2(t)$ because that will force the parametric equations to solve the ellipse equation. However, I would like for the equations to describe CW direction so I replace the t with $-t$ and propose:

$$\boxed{x = -1 + \cos(-t)} \qquad \textit{and} \qquad \boxed{y = 2\sin(-t)}$$

If we choose $t \in [0, 2\pi)$ then the whole ellipse will be covered. I could simplify $\cos(-t) = \cos(t)$ and $\sin(-t) = -\sin(t)$ but I have left the minus to emphasize the idea about reversing the orientation. In the preceding example we gave the circle a CCW orientation.

Example 1.4.29. Find parametric equations for the part of the level curve $x^2 - y^2 = 1$ which is found in the first quadrant.

We recognize this is a hyperbola which opens horizontally since $x = 0$ gives us $-y^2 = 1$ which has no real solutions. Hyperbolic trig. functions are built for a problem just such as this: recall $\cosh^2(t) - \sinh^2(t) = 1$ thus we choose $x = \cosh(t)$ and $y = \sinh(t)$. Furthermore, the hyperbolic sine function $\sinh(t) = \frac{1}{2}\left(e^t - e^{-t}\right)$ is everywhere increasing since it has derivative $\cosh(t)$ which is everywhere positive. Moreover, since $\sinh(0) = 0$ we see that $\sinh(t) \geq 0$ for $t \geq 0$. Choose non-negative t for the domain of the parametrization:

$$\boxed{x = \cosh(t), \qquad y = \sinh(t), \qquad t \in [0, \infty).}$$

Example 1.4.30. Find parametric equations for the part of the level curve $x^2 - y^2 = 1$ which is found in the third quadrant.

Based on our thinking from the last example we just need to modify the solution a bit:

$$\boxed{x = -\cosh(t), \qquad y = \sinh(t), \qquad t \in (-\infty, 0].}$$

Note that if $t \in (-\infty, 0]$ then $-\cosh(t) \leq -1$ and $\sinh(t) \leq 0$, this puts us in the third quadrant. It is also clear that these parametric equations solve the hyperbola equation since

$$(-\cosh(t))^2 - (\sinh(t))^2 = \cosh^2(t) - \sinh^2(t) = 1.$$

The examples thus far are rather specialized, and in general there is no method to find parametric equations. This is why I said it is an art.

Example 1.4.31. Find parametric equations for the level curve $x^2y^2 = x - 2$.

This example is actually pretty easy because we can solve for $y^2 = \frac{x-2}{x^2}$ *hence* $y = \pm\sqrt{\frac{x-2}{x^2}}$. *We can choose* x *as parameter so the parametric equations are just*

$$x = t \quad \text{and } y = \sqrt{\frac{t-2}{t^2}}$$

for $t \geq 2$. *Or, we could give parametric equations*

$$x = t \quad \text{and } y = -\sqrt{\frac{t-2}{t^2}}$$

for $t \geq 2$. *These parametrizations simply cover different parts of the same level curve.*

Remark 1.4.32. *but... what is* t *?*

If you are at all like me when I first learned about parametric curves you're probably wondering what is t ? You probably, like me, suppose incorrectly that t should be just like x or y. There is a crucial difference between x and y and t. The notations x and y are actually shorthands for the Cartesian coordinate maps $x : \mathbb{R}^2 \to \mathbb{R}$ and $y : \mathbb{R}^2 \to \mathbb{R}$ where $x(a, b) = a$ and $y(a, b) = b$. When I use the notaion $x = 3$ then you know what I mean, you know that I'm focusing on the vertical line with first coordinate 3. On the other hand, if I say $t = 3$ and ask where is it? Then you should say, your question doesn't make sense. The concept of t is tied to the curve for which it is the parameter. There are infinitely many geometric meanings for t. In other words, if you try to find t in the xy-plane without regard to a curve then you'll never find an answer. It's a meaningless question.

On the other hand if we are given a curve and ask what the meaning of t is for that curve then we ask a meaningful question. There are two popular meanings.

1. the parameter $t = s$ measures the arclength from some base point on the given curve.
2. the parameter t gives the time along the curve.

In case (1.) for an oriented curve this actually is uniquely specified if we have a starting point. Such a parameterization is called the **arclength parametrization** or **unit-speed** parametrization of a curve. These play a fundamental role in the study of the differential geometry of curves. In case (2.) we have in mind that the curve represents the physical trajectory of some object, as t increases, time goes on and the object moves. I tend to use (2.) as my conceptual backdrop. But, keep in mind that these are just applications of parametric curves. In general, the parameter need not be time or arclength. It might just be what is suggested by algebraic convenience.

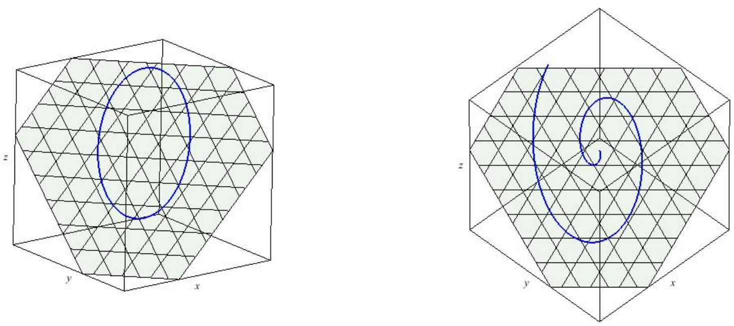
1.4.3 curves in three dimensional space

Other interesting curves can be obtained by feeding a simple curve like a circle into the parameterization of a plane.

Example 1.4.33. *Suppose $\vec{R}(u,v) = \vec{r}_o + u\vec{A} + v\vec{B}$ is the parametrization of a plane S then if we compose $\vec{R}$ with the path $t \mapsto \vec{\gamma}(t) = \langle R\cos(t), R\sin(t)\rangle$ we obtain an ellipse on the plane:*

$$\vec{r}(t) = (\vec{R} \circ \vec{\gamma})(t) = \vec{r}_o + R\cos(t)\vec{A} + R\sin(t)\vec{B}$$

Of course, we could also put a spiral on a plane through much the same device:



The idea of the last example can be used to create many interesting examples. These should suffice for our purposes here. I really just want you to think about what a parametrization does. Moreover, I hope you can find it in your heart to regard the parametric viewpoint as primary. Notice that any curves in three dimensions would require two independent equations in x, y, z. We saw how much of a hassel this was for something as simple as a line. I'd rather not attempt a general treatment of the purely cartesian description of the curves in this section[23] I instead offer a pair of examples to give you a flavor:

Example 1.4.34. *Suppose $x^2 + y^2 + z^2 = 4$ and $x = \sqrt{2}$ then the solution set of this pair of equations defines a curve in $\mathbb{R}^3$. Substutiting $x = \sqrt{3}$ into $x^2 + y^2 + z^2 = 4$ gives $y^2 + z^2 = 1$. The solution set is just a unit-circle in the yz-coordinates placed at $x = \sqrt{3}$. We can parametrize it via:*

$$\vec{r}(t) = \langle \sqrt{3}, \cos t, \sin t\rangle.$$

Example 1.4.35. *Suppose $z = x^2 - y^2$ and $z = 2x$. The solution set is once more a curve in $\mathbb{R}^3$. We can substitute $z = 2x$ into $z = x^2 - y^2$ to obtain $x^2 - y^2 = 2x$ hence $x^2 - 2x - y^2 = 0$ and completing the square reveals $(x-1)^2 - y^2 = 1$. This is the equation of a hyperbola. A natural parametrization is given by $x = 1 + \cosh t$ and $y = \sinh t$ then since $z = 2x$ we have $z = 2 + 2\cosh t$. In total,*

$$\vec{r}(t) = \langle 1 + \cosh t, \sinh t, 2 + 2\cosh t\rangle$$

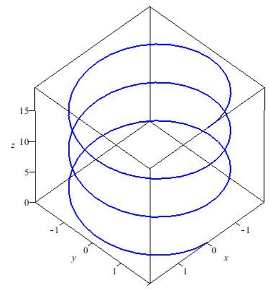
We'll explain the geometry of these calculations in the next section. Basically the idea is just that when two surfaces intersect in $\mathbb{R}^3$ we may obtain a curve.

Example 1.4.36. *A helix of radius R which wraps around the z-axis and has a slope of m is given by:*

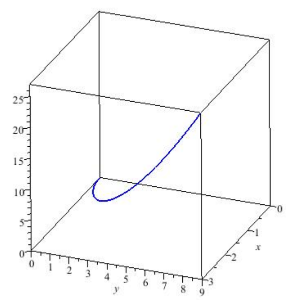
$$\vec{r}(t) = \langle R\cos(t), R\sin(t), mt\rangle$$

for $t \in [0, \infty)$.

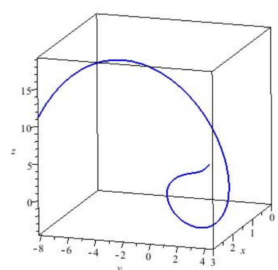
[23] which is not to say it hasn't been done, in fact, viewing curves as solutions to equations is also a powerful technique, but we focus our efforts in the parametric setting.



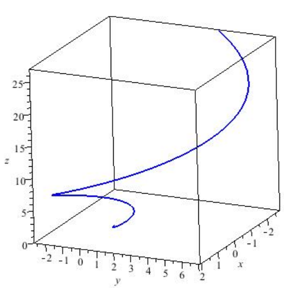
Example 1.4.37. *The curve parametrized by* $\vec{r}(t) = \langle t, t^2, t^3 \rangle$ *for* $t \geq 0$ *has scalar parametric equations* $x = t, y = t^2, z = t^3$ *and a graph*



Example 1.4.38. *The curve parametrized by* $\vec{r}(t) = \langle t, t^2\cos(3t), t^3\sin(3t) \rangle$ *for* $t \geq 0$ *has scalar parametric equations* $x = t, y = t^2\cos(3t), z = t^3\sin(3t)$ *and a graph*



Example 1.4.39. *The curve parametrized by* $\vec{r}(t) = \langle t\cos(3t), t^2\sin(3t), t^3 \rangle$ *for* $t \geq 0$ *has scalar parametric equations* $x = t\cos(3t), y = t^2\sin(3t), z = t^3$ *and a graph*



We will explore the geometry of curves in the next chapter. We'll find relatively simple calculations which allow us to test how a curve bends within its plane of motion and bend off its plane of motion.

In other words, we'll find a way to test if a curve lies in a plane and also how it curves away from it's tangential direction. These quantities are called *torsion* and *curvature*. It turns out that these two quantities often classify a curve up to congruence in the sense of high-school geometry. In other words, there is just one circle of radius 1 and we can rotate it and translate it throughout $\mathbb{R}^3$. In this sense all circles in $\mathbb{R}^3$ are the same. We've already seen in this section that parametrization alone does not capture this concept. Why? Well there are many parametrizations of a circle. Are those different circles? I would say not. I say there is a circle and there are many pictures of the circle, some CW, some CCW, but so long as those pictures cover the same curve then they are merely differing perspectives on the same object. That said, these differing pictures are different. They are unique in their assignments of points to parameter values. The problem of the differential geometry of curves is to extract from this infinity of parametrizations some universal data. One seeks a few constants which invariantly characterize the curve independent of the perspective a particular geometer has used to capture it. More generally this is the problem of geometry. How can we classify spaces? What constants can we label a space with unambiguously?

1.5 surfaces

A surface in $\mathbb{R}^3$ is a subset which usually looks two dimensional. There are three main viewpoints; graphs, parametrizations or patches, and level-surfaces. As usual, the parametric view naturally generalizes to surfaces in $\mathbb{R}^n$ for $n > 3$ with relatively little qualification. That said, we almost without exception focus on surfaces in $\mathbb{R}^3$ in this course so I focus our efforts in that direction. This section is introductory in nature, basically this is just show and tell with a little algebra. Your goal should be to learn the names of these surfaces and more importantly to gain a conceptual foothold on the different ways to look at two-dimensional subsets of $\mathbb{R}^3$. Many of the diagrams in this section were created with Maple, others perhaps Mathematica. Ask if interested.

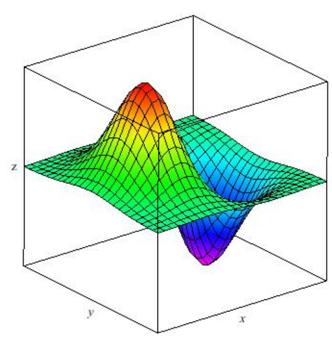
1.5.1 surfaces as graphs

Given a function of two variables it is natural to graph such a function in three-dimensions. In particular, we define:

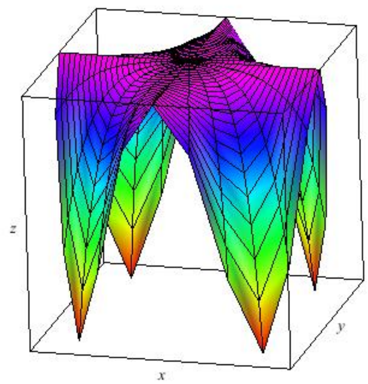
Definition 1.5.1. *graph of a function of two variables.*

Suppose $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ is a function then the set of all (x, y, z) such that $z = f(x, y)$ for some $(x, y) \in dom(f)$ is called the **graph** of f. Moreover, we denote $graph(f) = \{(x, y, f(x, y)) \mid (x, y) \in dom(f)\}$.

Example 1.5.2. *Let $f(x, y) = xe^{-x^2-y^2}$. The graph looks something like:*



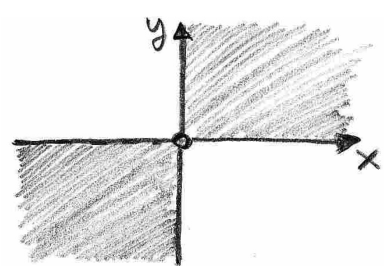
Example 1.5.3. *Let $f(x,y) = -\cosh(xy)$. The graph looks something like:*



What is f? Well, many interpretations exist. For example, f could represent the temperature at (x,y). Or, f could represent the mass per unit area close to (x,y), this would make f a mass density function. More generally, if you have a variable which depends by some single-valued rule to another pair of variables then you can find a function in that application. Sometimes college algebra students will ask, but what *is* a function really? With a little imagination the answer is most anything. It could be that f is the cost for making x widgets and y gadgets. Or, perhaps $f(x,y)$ is the grade of a class as a function of x males and y females. Enough. Let's get back to the math, I'll generally avoid cluttering these notes with these silly comments, however, you are free to ask in office hours. Not all such discussion is without merit. Application is important, but is not at all necessary for good mathematics.

We can add, multiple and divide functions of two variables in the same way we did for functions of one variable. Natural domains are also implicit within formulas and points are excluded for much the same reason as in single-variable calculus; we cannot divide by zero, take an even root of a negative number or take a logarithm of a non-positive quantity if we wish to maintain a real output[24]. A typical example is given below:

Example 1.5.4. *Let $f(x,y) = \frac{\sqrt{xy}}{x^2+y^2}$. We can ask: what is the largest domain compatible with the given formula? We need $xy \geq 0$ for the numerator to be real. However, we cannot allow $x = y = 0$ due to division by 0. Therefore, (x,y) needs $xy > 0$ which is solved by $x, y > 0$ or $x, y < 0$. Thus, $dom(f) = (0,\infty)^2 \cup (-\infty, 0)^2$. Perhaps the picture below helps:*



[24] complex variables do give meaning to even roots and logarithms of negative numbers however, division by zero and logarithm of zero continue to lack an arithmetical interpretation.

1.5.2 parametrized surfaces

Definition 1.5.5. *vector-valued functions of two real variables, parametrized surfaces.*

A vector valued function of a two real variables is an assignment of a vector for each pair of real numbers in some domain D of $\mathbb{R}^2$. It's a mapping $(u, v) \mapsto \vec{F}(u, v) = \langle F_1(u, v), F_2(u, v), \dots, F_n(u, v)\rangle$ for each $(u, v) \in D \subset \mathbb{R}^2$. We say $F_j : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ is the j-th component function of $\vec{F}$. Let $S = \vec{F}(D)$ then S is said to be a **surface parametrized by** $\vec{F}$. Equivalently, but not so usefully, we can write the scalar parametric equations for S above as

$$x_1 = F_1(u, v), \quad x_2 = F_2(u, v), \; \dots, \; x_n = F_n(u, v)$$

for all $(u, v) \in D$. We call $\vec{F}$ a **patch** on S.

When we define a parametrization of a surface it is important to give the formula for the patch **and** the domain D of the parameters. We call D the **parameter space**. Usually we are interested in the case of a surface which is embedded in $\mathbb{R}^3$ so I will focus the examples in that direction. Note however that the parametric equation for a plane actually embeds the plane in $\mathbb{R}^n$ for whatever n you wish, there is nothing particular to three dimensions for the construction of the line or plane parametrizations.

Example 1.5.6. *Suppose $S = \{(x, y, z) \mid z = f(x, y)\}$ where f is a function. Naturally we parametrize this graph via the choice $x = u, y = v$,*

$$\vec{r}(u, v) = \langle u, v, f(u, v)\rangle$$

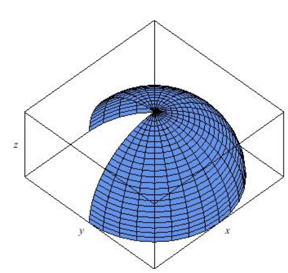
for $(u, v) \in dom(f)$.

As I discussed in the plane section, a graph is given in terms of cartesian coordinates. In the case of surfaces in $\mathbb{R}^3$ you'll often encounter the presentation $z = f(x, y)$ for some function f. This is an important class of examples, however, the criteria that f be a function is quite limiting.

Example 1.5.7. *Let $\vec{r}(u, v) = \langle R\cos(u)\sin(v), R\sin(u)\sin(v), R\cos(v)\rangle$ for $(u, v) \in [0, 2\pi] \times [0, \pi]$. In this case we have scalar equations:*

$$x = R\cos(u)\sin(v), \quad y = R\sin(u)\sin(v), \quad z = R\cos(v).$$

It's easy to show $x^2 + y^2 + z^2 = R^2$ and we should recognize that these are the parametric equations which force $\vec{r}(u, v)$ to land a distance of R away from the origin for each choice of (u, v). Let $S = \vec{r}(D)$ and recognize S is a **sphere** *of radius R centered at the origin. If we restrict the domain of $\vec{r}$ to $0 \leq u \leq \frac{3\pi}{2}$ and $0 \leq v \leq \frac{\pi}{2}$ then we select just a portion of the sphere:*

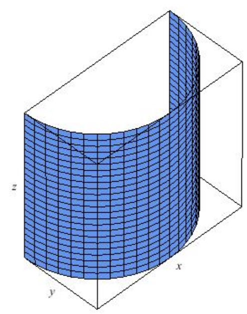


Notice that we could cover the whole sphere with a single patch. We cannot do that with a graph. This is the same story we saw in the two-dimensional case in calculus II. Parametrized curves are not limited by the vertical line test. Graphs are terribly boring in comparison to the geometrical richness of the parametric curve. As an exotic example from 1890, Peano constructed a continuous[25] path from $[0, 1]$ which covers all of $[0, 1] \times [0, 1]$. Think about that. Such curves are called *space filling curves.* There are textbooks devoted to the study of just those curves. For example, see Hans Sagan's *Space Filling Curves.*

Example 1.5.8. *Let $\vec{r}(u, v) = \langle R\cos(u), R\sin(u), v\rangle$ for $(u, v) \in [0, 2\pi] \times \mathbb{R}$. In this case we have scalar equations:*

$$x = R\cos(u), \quad y = R\sin(u), \quad z = v.$$

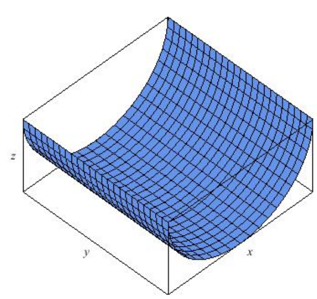
It's easy to show $x^2 + y^2 = R^2$ and z is free to range over all values. This surface is a circle at each possible z. We should recognize that these are the parametric equations which force $\vec{r}(u, v)$ to land on a cylinder of raduius R centered on the z-axis. If we restrict the domain of $\vec{r}$ to $0 \leq u \leq \pi$ and $0 \leq v \leq 2$ then we select a finite half-cylinder:



Example 1.5.9. *Let $\vec{r}(u, v) = \langle a\cos(u), v, b\sin(u)\rangle$ for $(u, v) \in [0, 2\pi] \times \mathbb{R}$. In this case we have scalar equations:*

$$x = a\cos(u), \quad y = v, \quad z = b\sin(u).$$

It's easy to show $x^2/a^2 + z^2/b^2 = 1$ and y is free to range over all values. This surface is a ellipse at each possible y. We should recognize that these are the parametric equations which force $\vec{r}(u, v)$ to land on an elliptical cylinder centered on the y-axis. If we restrict the domain of $\vec{r}$ to $0 \leq u \leq \pi$ and $0 \leq v \leq 2$ then we select a finite half-cylinder:

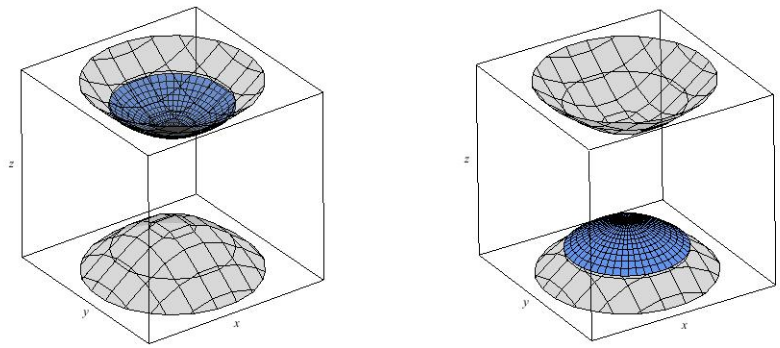


[25] we will define this carefully in a future chapter

Example 1.5.10. *Let $\vec{r}(u,v) = \langle R\cos(u)\sinh(v), R\sin(u)\sinh(v), R\cosh(v)\rangle$ for $(u,v) \in [0,2\pi] \times \mathbb{R}$. In this case we have scalar equations:*

$$x = R\cos(u)\sinh(v), \quad y = R\sin(u)\sinh(v), \quad z = R\cosh(v).$$

It's easy to show $-x^2 - y^2 + z^2 = R^2$. If we restrict the domain of $\vec{r}$ to $0 \leq u \leq 2\pi$ and $-2 \leq v \leq 2$ then we select a portion of the upper branch:

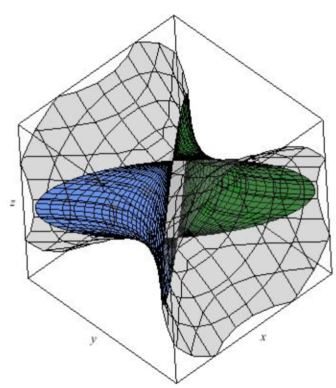


The part of the lower branch which is graphed above is covered by the mapping Let $\vec{r}(u,v) = \langle R\cos(u)\sinh(v), R\sin(u)\sinh(v), -R\cosh(v)\rangle$ for $(u,v) \in [0,2\pi] \times [-2,2]$. The grey shape is where the parametrization will cover if we enlarge the domain of the parameterizations.

Example 1.5.11. *Let $\vec{r}(u,v) = \langle R\cosh(u)\sin(v), R\sinh(u)\sin(v), R\cos(v)\rangle$ for $(u,v) \in \mathbb{R} \times [0,2\pi]$. In this case we have scalar equations:*

$$x = R\cosh(u)\sin(v), \quad y = R\sinh(u)\sin(v), \quad z = R\cos(v).$$

It's easy to show $x^2 - y^2 + z^2 = R^2$. I've plotted $\vec{r}$ with domain restricted to $dom(\vec{r}) = [-1.3, 1.3] \times [0,\pi]$ in blue and $dom(\vec{r}) = [-1.3, 1.3] \times [\pi, 2\pi]$ in green. The grey shape is where the parametrization will go if we enlarge the domain.



1.5.3 surfaces as level sets

Unlike curves, we do not need two equations to fix a surface in $\mathbb{R}^3$. In three dimensional space[26] if we have just one equation in x, y, z that should suffice to leave just two free variables. In a nutshell

[26] to pick out a two-dimensional surface in $\mathbb{R}^4$ it would take two equations in t, x, y, z, but, we really only care about $\mathbb{R}^3$ so, I'll behave and stick with that case.

that is what a surface is. It is a space which has two degrees of freedom. In the parametric set-up we declare those freedoms explicitly from the outset by the construction of the patch in terms of the parameters. In the level set formulation we focus the attention on an equation which defines the surface of interest. We already saw this for a plane; the solutions of $ax + by + cz = d$ fill out a plane with normal $\langle a, b, c\rangle$.

Definition 1.5.12. *level surface in three dimensional space*

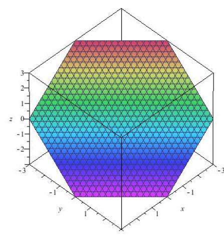
Suppose $F : dom(F) \subseteq \mathbb{R}^3 \to \mathbb{R}$ is a function. Let

$$S = \{(x, y, z) \mid F(x, y, z) = k\}.$$

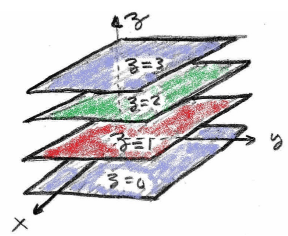
We say that S is a **level surface** of level k with level function F. In other words, $S = F^{-1}\{k\}$ is a level surface.

A level surface is a *fiber* of a real-valued function on $\mathbb{R}^3$.

Example 1.5.13. *Let $F(x, y, z) = a(x - x_o) + b(y - y_o) + c(z - z_o)$. Recognize that the solution set of $F(x, y, z) = 0$ is the plane with base-point (x_o, y_o, z_o) and normal $\langle a, b, c\rangle$.*

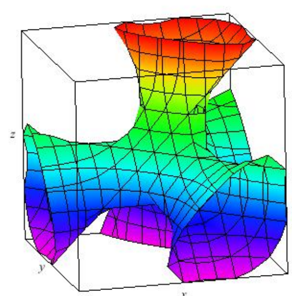


Example 1.5.14. *some level surfaces I can plot without fancy CAS programs: let $f(x, y, z) = z$ I plot $f(x, y, z) = 0, 1, 2, 3$ below:*



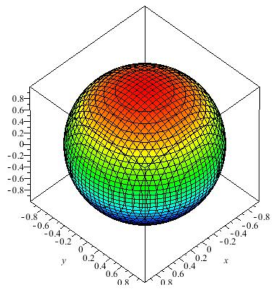
The example below is not such a case:

Example 1.5.15. *This surface has four holes. I have an animation on my website, check it out.*

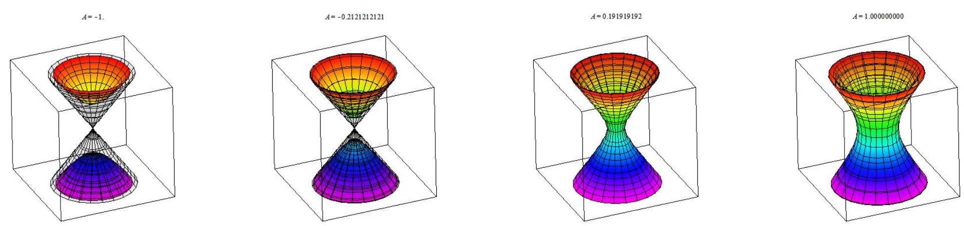


In fact, I don't think I want to parametrize this beast[27].

Example 1.5.16. *Let $F(x, y, z) = x^2/a^2 + y^2/b^2 + z^2/c^2$. The solution set of $F(x, y, z) = 1$ is called an* **ellipsoid** *centered at the origin. In the special case $a = b = c = R$ the ellipsoid is a sphere of radius R. Here's a special case, $a = b = c = 1$ the unit-sphere:*



Example 1.5.17. *Let $F(x, y, z) = x^2 + y^2 - z^2$. The solution set of $F(x, y, z) = 0$ is called an* **cone** *through the origin. However, the solution set of $F(x, y, z) = k$ for $k \neq 0$ forms a* **hyperboloid of one-sheet** *for $k > 0$ and a* **hyperboloid of two-sheets** *for $k < 0$. The hyperboloids approach the cone as the distance from the origin grows. I plot a few representative cases:*



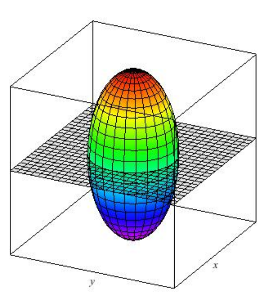
There is an animation on my webage, take a look.

Some of the examples above fall under the general category of a **quadratic surface**. Suppose

$$Q(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz.$$

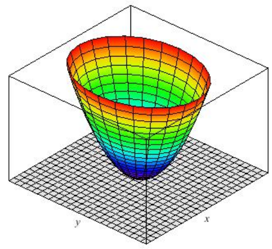
For any particular nontrivial selection of constants $a, b, \ldots, h, i$ we say the solution of $Q(x, y, z) = k$ is a **quadratic surface**. For future reference let me list the proper terminology. We'd like to get comfortable with these terms.

1. a standard **ellipsoid** is the solution set of $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$. If $a = b = c$ then we say the ellipsoid is a **sphere**.

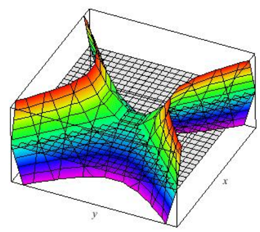


[27] Wait, I have students, isn't this what homework is for?

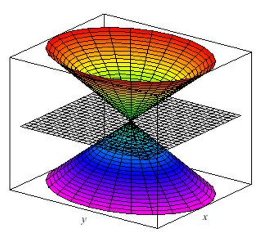
2. a standard **elliptic paraboloid** is the solution set of $z/c = x^2/a^2 + y^2/b^2$. If $a = b$ then we say the paraboloid is **circular**.



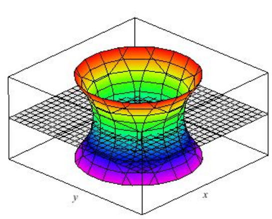
3. a standard **hyperbolic paraboloids** is the solution set of $z/c = y^2/b^2 - x^2/a^2$.



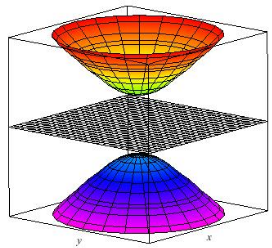
4. a standard **elliptic cone** is the solution set of $z^2/c^2 = x^2/a^2 + y^2/b^2$. If $a = b$ then we say the cone is **circular**.



5. a standard **hyperboloid of one sheet** is the solution set of $x^2/a^2 + y^2/b^2 - z^2/c^2 = 1$.



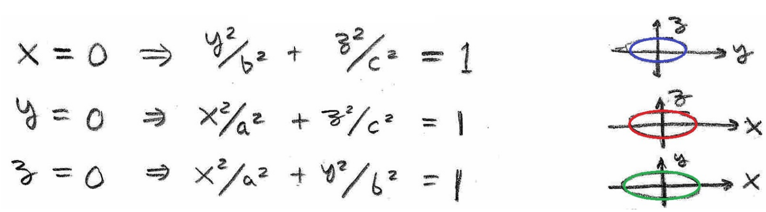
6. a standard **hyperboloid of two sheets** is the solution set of $x^2/a^2 + y^2/b^2 - z^2/c^2 = -1$.



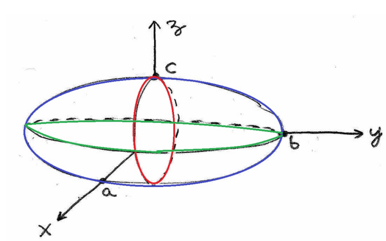
If you study the formulas above you'll notice the absense of certain terms in the general quadratic form: terms such as dxy, exz, fyz, gx, hy are absent. Inclusion of these terms will either shift or rotate the standard equations. However, we need linear algebra to properly construct the rotations from the eigenvectors of the quadratic form. I leave that to linear algebra where we have more toys to play with. You'll have to be content with the standard examples for the majority of this course.

I've inserted the term *standard* because I don't mean to say that every elliptic cone has the same equation as I give. I expect you can translate the standard examples up to an interchange of coordinates, that ought not be too hard to understand. For example, $y = x^2 + 2z^2$ is clearly an elliptical paraboloid. Or $y = x^2 - z^2$ is clearly a hyperbolic paraboloid. Or $x^2 + z^2 - y^2 = 1$ is clearly a hyperboloid of one sheet whereas $-x^2 - z^2 + y^2 = 1$ is a hyperboloid of two sheets. These are the possibilities we ought to anticipate when faced with the level set of some quadratic form. I don't try to memorize all of these, I use the method sketched in the next pair of examples. Basically the idea is simply to slice the graph into planes where we find either circles, hyperbolas, lines, ellipses or perhaps nothing at all. Then we take a few such slices and extrapolate the graph. Often the slices $x = 0$, $y = 0$ and $z = 0$ are very helpful.

Example 1.5.18. *To understand the structure of the ellipsoid* $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ *by considering the intersection of the surface with the coordinate planes:*



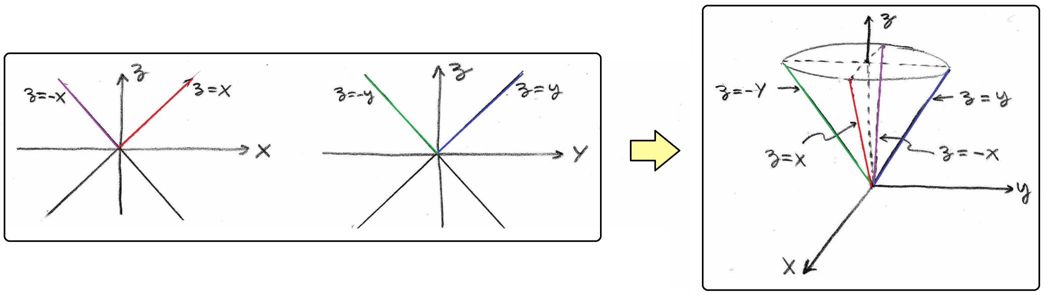
It follows the picture looks something like:



Example 1.5.19. *To understand the shape of the cone $z^2 = x^2+y^2$ we can consider its intersection with the three coordinate planes:*

$$\begin{array}{lllll} z=0 & \Rightarrow & 0=x^2+y^2 & \Rightarrow & x=y=z=0. \\ x=0 & \Rightarrow & z^2=y^2 & \Rightarrow & x=0,\ z=\pm y. \\ y=0 & \Rightarrow & z^2=x^2 & \Rightarrow & y=0,\ z=\pm x. \end{array}$$

This data helps us envision the cone: (I only plot the $z>0$ part of the cone)



Example 1.5.20. *Let $F(x,y,z) = x^2+y^2-z$. The solution set of $F(x,y,z) = k$ sometimes called a paraboloid. Notice that if we fix a value for z say $z=c$ then $x^2+y^2-c=k$ reduces to $x^2+y^2=k+c$. If $k+c\geq 0$ then the solution in the $z=c$ plane is a circle or a point. In other words, all horizontal cross-sections of this shape are circles and if $z<-k$ there is no solution. This surface opens up from the vertex at $(0,0,-k)$.*

What about the geometry of surfaces? How can we classify surfaces? What constants can we label a surface with unambiguously? Here's a simple example: $x^2+y^2+z^2=1$ and $(x-1)^2+(y-2)^2+(z-3)^2=1$ define the same shape at different points in $\mathbb{R}^3$. The problem of the differential geometry of surfaces is to find general invariants which discover this equivalence through mathematical calculation. This is a more difficult problem and we will not treat it in this course. It turns out this geometry begins to provide the concepts needed for Einstein's General Relativity. If you are interested in learning differential geometry please stop by sometime so we can chat.

1.5.4 combined concept examples

Example 1.5.21. *Find where the path $\vec{r}(t) = \langle t, 0, 2t-t^2\rangle$ interects the paraboloid $z=x^2+y^2$.*

Solution: *to find points of intersection we assume both equations hold: $x=t, y=0$ and $z=2t-t^2$ and $z=x^2+y^2$ hence*

$$2t-t^2=t^2 \quad\Rightarrow\quad 2t-2t^2=2t(1-t)=0$$

Thus $t=0$ and $t=1$ give points of intersection. In particular, $\vec{r}(0)=(0,0,0)$ and $\vec{r}(1)=(1,0,1)$ are the points at which the path punctures the paraboloid.

Example 1.5.22. *Find a parametrization of the curve of intersection formed by the cylinder $x^2+y^2=4$ and $z=xy$.*

Solution: *if $y\geq 0$ then $y=\sqrt{4-x^2}$ thus $z=xy=x\sqrt{4-x^2}$ and we may use x as the parameter of the curve:*

$$\vec{r}_+(x) = \langle x, \sqrt{4-x^2}, x\sqrt{4-x^2}\rangle.$$

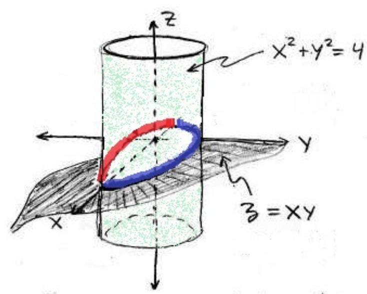
Likewise, for the other part of the intersection curve with $y \leq 0$ we may use

$$\vec{r}_-(x) = \langle x, -\sqrt{4-x^2}, -x\sqrt{4-x^2}\rangle.$$

The domain for $\vec{r}_\pm$ is given by $[-2,2]$. These formulas are fairly easy to find, but, if you do calculus on this curve later you'll regret using these formulas. There is a better way: use trigonometric insight! Let $x = 2\cos t$ and $y = 2\sin t$ this puts us on the cylinder $x^2+y^2=4$. Next, impose $z = xy$, this says $z = 4\sin t\cos t = 2\sin 2t$. In total,

$$\vec{r}(t) = \langle 2\cos t, 2\sin t, 2\sin 2t\rangle$$

where $t \in [0, 2\pi]$. The trigonometry-inspired path is far easier for future applications. Here is a horrible picture where I illustrate the $\pm$ curves in red $(-)$ and blue $(+)$.



Example 1.5.23. *Find the curve of intersection for $z = \sqrt{x^2+y^2}$ and the plane $z = 1+y$.*

Solution: *We seek a parametrization of the intersection curve. A simple way to impose both surface equations is to equate z:*

$$\sqrt{x^2+y^2} = 1+y.$$

Square both sides to obtain

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

Thus, $x^2 = 1+2y$ and we obtain $y = \frac{1}{2}(x^2-1)$ which suggests x is a good choice of parameter here. Let $x = t$ then $y = \frac{1}{2}(t^2-1)$ and $z = 1+y = 1+\frac{1}{2}(t^2-1)$. Hence,

$$\vec{r}(t) = \langle\ t, \frac{1}{2}(t^2-1),\ 1+\frac{1}{2}(t^2-1)\ \rangle.$$

Example 1.5.24. *Consider the trajectories $\vec{r}_1(t) = \langle t^2,\ 7t-12,\ t^2\rangle$ and $\vec{r}_2(t) = \langle 4t-3,\ t^2,\ 5t-6\rangle$ for $t \geq 0$. Suppose $\vec{r}_1$ is the position of Madara and $\vec{r}_2$ is the position of Jack Bauer. Do they collide?*

Solution: *We seek t for which $\vec{r}_1(t) = \vec{r}_2(t)$ hence:*

$$t^2 = 4t-3, \quad 7t-12 = t^2, \quad t^2 = 5t-6$$

From which we obtain,

$$t^2-4t+3 = (t-1)(t-3) = 0, \quad t^2-7t+12 = (t-3)(t-4) = 0, \quad t^2-5t+6 = (t-3)(t-2) = 0.$$

The solution which is common to all three equations is $t = 3$. At time $t = 3$ it's on. Who will win? The collision occurs at $t = 3$ at the point $(9, 9, 9)$.

Example 1.5.25. *Suppose $\vec{r}_1(t) = \langle t, t^2, t^3\rangle$ and $\vec{r}_2(s) = \langle 1+2s, 1+6s, 1+14s\rangle$. If $\vec{r}_1$ is the position of a mouse at time t and $\vec{r}_2$ is the position of a cat at time s then does the cat run across the path of the mouse?*

Solution: *We seek points of intersection $\vec{r}_1(t) = \vec{r}_2(s)$ for which $t < s$. Consider,*

$$t = 1+2s, \quad t^2 = 1+6s, \quad t^3 = 1+14s$$

Note we may solve for 1,

$$1 = t - 2s = t^2 - 6s = t^3 - 14s$$

Note, $s = 0$ implies $1 = t = t^2 = t^3$ hence $t = 1$. One solution is given by $s = 0$ and $t = 1$. However, $1 \not< 0$ so the mouse is safe thus far. Are there any other solutions? In fact, yes, $s = 1/2$ and $t = 2$ also solves the given system. Again, the mouse is fine as he intersects the path of the cat after the cat is already gone.

You might wonder how I found the second solution (or the first) in the preceding example. To be more systematic, we could solve each equation for s and equate the resulting polynomials in t.

1.6 curvelinear coordinates

Cartesian coordinates are a nice starting point, but they make many simple problems needlessly complex. If a two-dimensional problem has a quantity which only depends on distance from the central point then probably polar coordinates will simplify the equations of the problem. Similarly, if a three dimensional problem possesses a cylindrical symmetry then use cylindrical coordinates. If a three dimensional problem has spherical symmetry then use spherical coordinates.

A coordinate system is called **right-handed** if the unit-vectors which point in the direction of increasing coordinates at each point are related to each other by the right-hand-rule just like the xyz-coordinates. We call this set of unit-vectors the **frame** of the coordinate system. Generally a frame in $\mathbb{R}^n$ is just an assignment of n-vectors at each point in $\mathbb{R}^n$. In linear-algebra language, a frame is an assignment of a basis at each point of $\mathbb{R}^n$. Dimensions $n = 2$ and $n = 3$ suffice for our purposes. If y_1, y_2, y_3 denote coordinates with unit-vectors $\widehat{u}_1, \widehat{u}_2, \widehat{u}_3$ in the direction of increasing y_1, y_2, y_3 respective then we say the coordinate system is **right-handed** iff

$$\widehat{u}_1 \times \widehat{u}_2 = \widehat{u}_3, \quad \widehat{u}_2 \times \widehat{u}_3 = \widehat{u}_1, \quad \widehat{u}_3 \times \widehat{u}_1 = \widehat{u}_2.$$

In contrast with the constant frame $\{\widehat{x}, \widehat{y}, \widehat{z}\}$ the frame $\{\widehat{u}_1, \widehat{u}_2, \widehat{u}_3\}$ is usually point-dependent. An assignment of a vector to each point in some space is called a **vector field**. A frame is actually a triple of vector fields which is given over a space. Enough terminology, the equations speak for themselves soon enough.

1.6.1 polar coordinates

Polar coordinates (r, θ) are defined by

$$x = r\cos\theta, \qquad y = r\sin\theta$$

In quadrants I and IV (regions with $x > 0$) we also have

$$r^2 = x^2 + y^2, \qquad \theta = \tan^{-1}\left[\frac{y}{x}\right].$$

In quadrants II and III (regions with $x < 0$) we have

$$r^2 = x^2 + y^2, \qquad \theta = \pi + \tan^{-1}\left[\frac{y}{x}\right].$$

Geometrically it is clear that we can label any point in $\mathbb{R}^2$ either by cartesian coordinates (x, y) or by polar coordinates (r, θ). We may view equations in cartesian or polar form.

Example 1.6.1. *The circle $x^2 + y^2 = R^2$ has polar equation $r = R$.*

Typically in the polar context the angle plays the role of the independent variable. In the same way it is usually customary to write $y = f(x)$ for a graph we try to write $r = f(\theta)$.

Example 1.6.2. *The line $y = mx + b$ has polar equation $r \sin\theta = mr\cos\theta + b$ hence*

$$r = \frac{b}{\sin\theta - m\cos\theta}.$$

Example 1.6.3. *The polar equation $\theta = \pi/4$ translates to $y = x$ for $x > 0$. The reason is that*

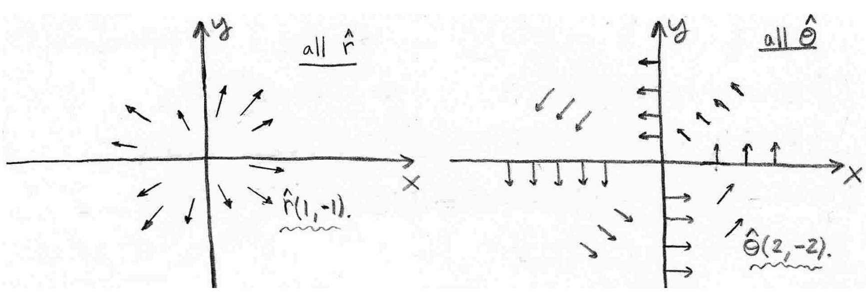
$$\frac{\pi}{4} = \tan^{-1}\left[\frac{y}{x}\right] \quad \Rightarrow \quad \tan\frac{\pi}{4} = \frac{y}{x} \quad \Rightarrow \quad 1 = \frac{y}{x} \quad \Rightarrow \quad y = x$$

and the ray with $\theta = \pi/4$ is found in quadrant I where $x > 0$.

Let us denote unit vectors in the direction of increasing r, θ by $\widehat{r}$, $\widehat{\theta}$ respective. You can derive by geometry alone that

$$\begin{aligned} \widehat{r} &= \cos(\theta)\,\widehat{x} + \sin(\theta)\,\widehat{y} \\ \widehat{\theta} &= -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}. \end{aligned} \tag{1.3}$$

We call $\{\widehat{r},\ \widehat{\theta}\}$ the **frame** of polar coordinates. Notice that these are perpendicular at each point; $\widehat{r} \bullet \widehat{\theta} = 0$.



Example 1.6.4. *If we want to assign a vector to each point on the unit circle such that the vector is tangent and pointing in the counter-clockwise (CCW) direction then a natural choice is $\widehat{\theta}$.*

Example 1.6.5. *If we want to assign a vector to each point on the unit circle such that the vector is pointing radially out from the center then a natural choice is $\widehat{r}$.*

Suppose you have a perfectly flat floor and you pour paint slowly in a perfect even stream then in principle you'd expect it would spread out on the floor in the $\widehat{r}$ direction if we take the spill spot as the origin and the floor as the xy-plane.

1.6.2 cylindrical coordinates

Cylindrical coordinates (r, θ, z) are defined by

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad z = z$$

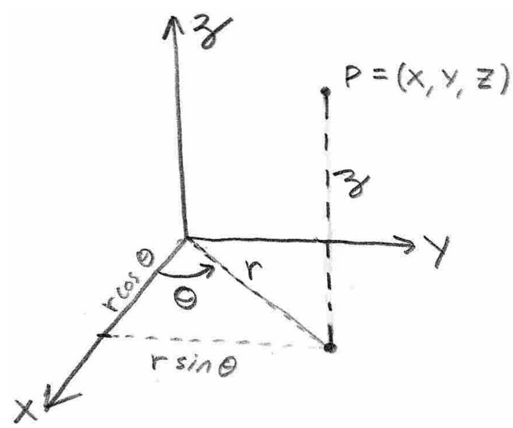
In quadrants I and IV (regions with $x > 0$) we also have

$$r^2 = x^2 + y^2, \qquad \theta = \tan^{-1}\left[\frac{y}{x}\right].$$

In quadrants II and III (regions with $x < 0$) we have

$$r^2 = x^2 + y^2, \qquad \theta = \pi + \tan^{-1}\left[\frac{y}{x}\right].$$

Geometrically it is clear that we can label any point in $\mathbb{R}^3$ either by cartesian coordinates (x, y, z) or by cylindrical coordinates (r, θ, z).



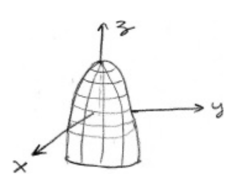
We may view equations in cartesian or cylindrical form.

Example 1.6.6. *In cylindrical coordinates the equation $r = 1$ is a cylinder since the z-variable is free. If we denote the unit-circle by $S_1 = \{(x, y) \mid x^2 + y^2 = 1\}$ then the solution set of $r = 1$ has the form $S_1 \times \mathbb{R}$. At each z we get a copy of the circle S_1.*

Example 1.6.7. *The equation $\theta = \pi/4$ is a half-plane which has equation $y = x$ subject to the condition $x > 0$.*

Example 1.6.8. *Consider the cylindrical equation $z = 4 - r^2$. Characterize the solution set.*

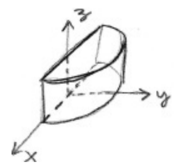
Solution: *notice $z = 4 - x^2 - y^2$ thus $x = 0$ yields $z = 4 - y^2$. Also, when $y = 0$ we have $z = 4 - x^2$. The intersection of the surface with coordinate planes are parabolas and as there is no θ-dependence we find a parabola in all other vertical planes through the origin. This is a paraboloid that opens down and has top point (0,0,4). It may also be helpful to note that fixing $z = z_o < 4$ we have $4 - z_0 = r^2$, that is, a circle in the plane $z = z_o$. Here is a crude sketch:*



Example 1.6.9. *We may sketch the region given by cylindrical coordinates r, θ, z subject the constraints:*

$$0 \leq r \leq 2, \quad 0 \leq \theta \leq \pi, \quad 0 \leq z \leq 1.$$

as follows:

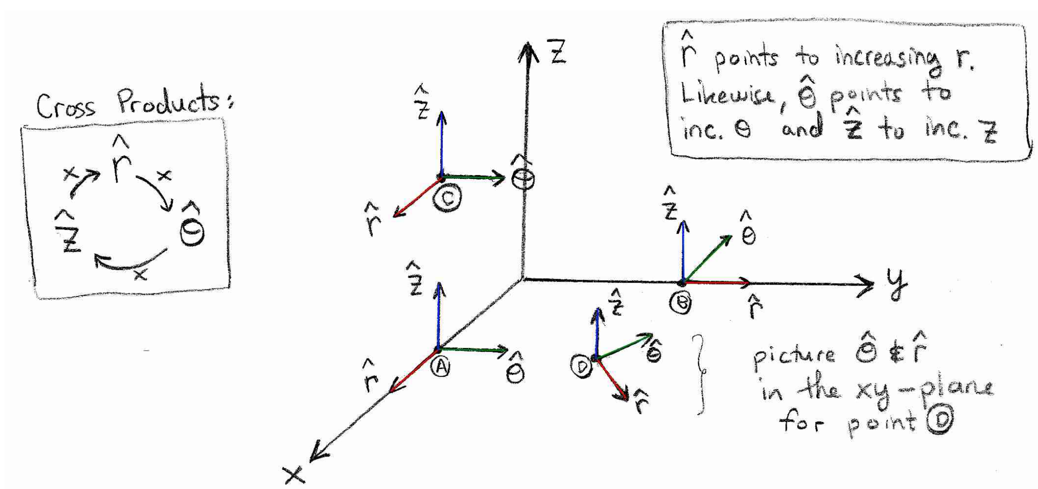


This is a height one, half-circular-cylinder resting on the xy-plane.

Let us denote unit vectors in the direction of increasing $r,\ \theta,\ z$ by $\widehat{r},\ \widehat{\theta},\ \widehat{z}$ respectively. You can derive by geometry alone that

$$\begin{aligned}\widehat{r} &= \cos(\theta)\,\widehat{x} + \sin(\theta)\,\widehat{y} \\ \widehat{\theta} &= -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y} \\ \widehat{z} &= \langle 0, 0, 1\rangle.\end{aligned}$$

We call $\{\widehat{r},\ \widehat{\theta},\ \widehat{z}\}$ the **unit-frame** of cylindrical coordinates.



Example 1.6.10. *Suppose we have a line of electric charge smeared along the z-axis with charge density λ. One can easily derive from Gauss' Law that the electric field has the form:*

$$\vec{E} = \frac{\lambda}{2\pi\epsilon_o r}\widehat{r}.$$

Example 1.6.11. *If we have a uniform current $I\,\widehat{z}$ flowing along the z-axis then the magnetic field can be derived from Ampere's Law and has the simple form:*

$$\vec{B} = \frac{\mu_o I}{2\pi r}\widehat{\theta}$$

Trust me when I tell you that the formulas in terms of cartesian coordinates are not nearly as clean.

If we fix our attention to a particular point the cylindrical frame has the same structure as the cartesian frame $\{\widehat{x}, \widehat{y}, \widehat{z}\}$. In particular, we can show that

$$\widehat{r} \bullet \widehat{r} = 1, \quad \widehat{\theta} \bullet \widehat{\theta} = 1, \quad \widehat{z} \bullet \widehat{z} = 1$$

$$\widehat{\theta} \bullet \widehat{r} = 0, \quad \widehat{\theta} \bullet \widehat{z} = 0, \quad \widehat{z} \bullet \widehat{r} = 0.$$

We can also calculate either algebraically or geometrically that:

$$\widehat{r} \times \widehat{\theta} = \widehat{z}, \quad \widehat{\theta} \times \widehat{z} = \widehat{r}, \quad \widehat{z} \times \widehat{r} = \widehat{\theta}$$

Therefore, the cylindrical coordinate system (r, θ, z) is a **right-handed** coordinate system since it provides a right-handed basis of unit-vectors at each point. We can summarize these relations compactly with the notation $\widehat{u}_1 = \widehat{r}$, $\widehat{u}_2 = \widehat{\theta}$, $\widehat{u}_3 = \widehat{z}$ whence:

$$\widehat{u}_i \bullet \widehat{u}_j = \delta_{ij}, \qquad \widehat{u}_i \times \widehat{u}_j = \sum_{k=1}^{3} \epsilon_{ijk} \widehat{u}_k$$

this is the same pattern we saw for the cartesian unit vectors.

1.6.3 spherical coordinates

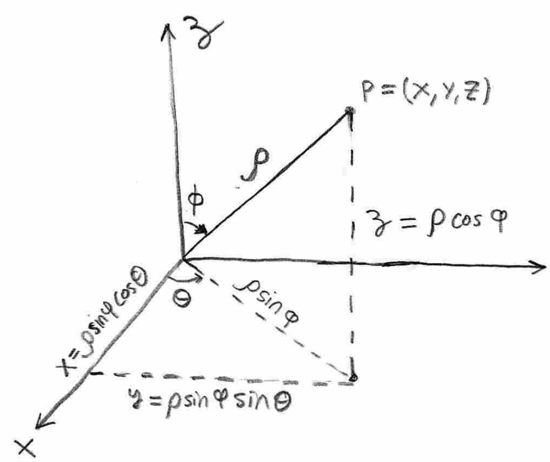
Spherical coordinates[28] (ρ, ϕ, θ) relate to Cartesian coordinates as follows

$$\begin{aligned} x &= \rho \cos(\theta) \sin(\phi) \\ y &= \rho \sin(\theta) \sin(\phi) \\ z &= \rho \cos(\phi) \end{aligned} \tag{1.4}$$

where $\rho > 0$ and $0 \leq \phi \leq \pi$ and $0 \leq \theta \leq 2\pi$. We can derive,

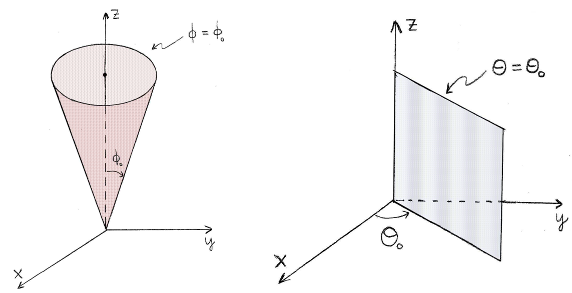
$$\begin{aligned} \rho^2 &= x^2 + y^2 + z^2 \\ \tan(\phi) &= \sqrt{x^2 + y^2}/z \\ \tan(\theta) &= y/x. \end{aligned} \tag{1.5}$$

It is clear that any point in $\mathbb{R}^3$ is labeled both by cartesian coordinates (x, y, z) or spherical coordinates (ρ, ϕ, θ).



[28] I'll use notation which is consistent with typical calculus texts, but beware there is a better notation used in physics where the meaning of ϕ and θ are switched and the spherical radius ρ is instead denoted by r.

Also, it is important to distinguish[29] between the geometry of the **azmuthial angle** θ and the **polar angle** ϕ.



Example 1.6.12. *The equation* $\sqrt{x^2+y^2+z^2} = R$ *is written as* $\rho = R$ *in spherical coordinates.*

Example 1.6.13. *The plane* $a(x-1)+b(y-2)+c(z-3) = 0$ *has a much uglier form in spherical coordinates. Its:* $a(\rho\cos(\theta)\sin(\phi)-1)+b(\rho\sin(\theta)\sin(\phi)-2)+c(\rho\cos(\phi)-3) = 0$ *hence*

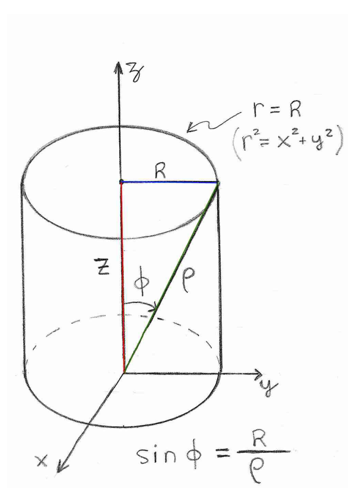
$$\rho = \frac{a+2b+3c}{a\cos(\theta)\sin(\phi)+b\sin(\theta)\sin(\phi)+c\cos(\phi)}.$$

Example 1.6.14. *The equation of a cylinder is* $r = R$ *in cylindrical coordinates. In spherical coordinates it is not as pretty. Note that* $r = R$ *gives* $x^2+y^2 = R^2$ *and*

$$x^2+y^2 = \rho^2\cos^2(\theta)\sin^2(\phi)+\rho^2\cos^2(\theta)\sin^2(\phi) = \rho^2\sin^2(\phi)$$

Thus, the equation of a cylinder in spherical coordinates is $R = \rho\sin(\phi)$.

You might notice that the formula above is easily derived geometrically. If you picture a cylinder and draw a rectangle as shown below it is clear that $\sin(\phi) = \frac{R}{\rho}$.



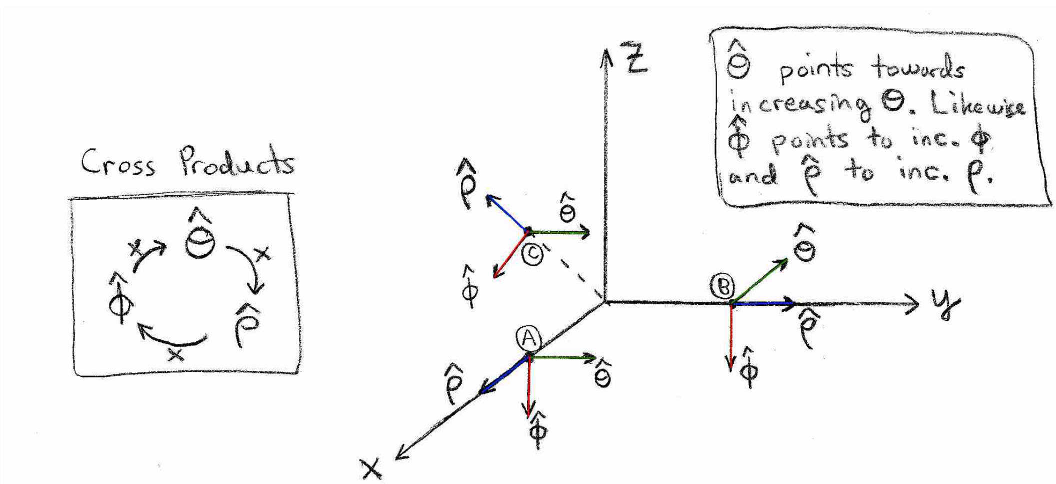
It is important to be proficient in both visualization and calculation. They work together to solve problems in this course, if you get stuck in one direction sometimes the other will help you get free.

[29] Wikipedia currently has an lengthy article on the distinction between math and physics notation for spherical angles and some explanation concerning the terminology.

Let us denote unit vectors in the direction of increasing $\rho,\ \phi,\ \theta$ by $\widehat{\rho},\ \widehat{\phi},\ \widehat{\theta}$ respectively. You can derive by geometry alone[30] that

$$\begin{aligned}\widehat{\rho} &= \sin(\phi)\cos(\theta)\,\widehat{x} + \sin(\phi)\sin(\theta)\,\widehat{y} + \cos(\phi)\,\widehat{z} \\ \widehat{\phi} &= \cos(\phi)\cos(\theta)\,\widehat{x} + \cos(\phi)\sin(\theta)\,\widehat{y} - \sin(\phi)\,\widehat{z} \\ \widehat{\theta} &= -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}.\end{aligned} \tag{1.6}$$

We call $\{\widehat{\rho},\ \widehat{\phi},\ \widehat{\theta}\}$ the **frame** of spherical coordinates. At each point these unit-vectors point in particular direction.



In contrast to the cartesian frame which is constant[31] over all of $\mathbb{R}^3$.

Example 1.6.15. *Suppose we a point charge q is placed at the origin then by Gauss' Law we can derive*

$$\vec{E} = \frac{q}{4\pi\epsilon_o\rho^2}\widehat{\rho}.$$

This formula makes manifest the spherical direction of the electric field, the absence of the angular unit-vectors says the field has no angular dependence and hence its values depend only on the spherical radius ρ. This is called the **Coulomb field** *or monopole field. Almost the same math applies to gravity. If M is placed at the origin then*

$$\vec{F} = \frac{GmM}{\rho^2}(-\widehat{\rho}).$$

gives the gravitational force $\vec{F}$ of M on m at distance ρ from the origin. The direction of the gravitational field is $-\widehat{\rho}$ which simply says the field points radially inward.

The spherical frame gives us a basis of vectors to build vectors at each point in $\mathbb{R}^3$. More than that, the spherical frame is an orthonormal frame since at any particular point the frame provides an orthonormal set of vectors. In particular, we can show that

$$\widehat{\rho}\bullet\widehat{\rho} = 1, \quad \widehat{\phi}\bullet\widehat{\phi} = 1, \quad \widehat{\theta}\bullet\widehat{\theta} = 1$$

[30] but, we'll find an easier way with calculus a bit later

[31] How do I know the cartesian frame is unchanging? It's not complicated really; $\widehat{x} = \langle 1,0,0\rangle$, $\widehat{y} = \langle 0,1,0\rangle$ and $\widehat{z} = \langle 0,0,1\rangle$.

$$\widehat{\phi} \bullet \widehat{\rho} = 0, \quad \widehat{\theta} \bullet \widehat{\rho} = 0, \quad \widehat{\phi} \bullet \widehat{\theta} = 0.$$

We can also calculate either algebraically or geometrically that:

$$\widehat{\theta} \times \widehat{\rho} = \widehat{\phi}, \quad \widehat{\rho} \times \widehat{\phi} = \widehat{\theta}, \quad \widehat{\phi} \times \widehat{\theta} = \widehat{\rho}$$

Therefore, the spherical coordinate system (ρ, ϕ, θ) is a **right-handed** coordinate system since it provides a right-handed basis of unit-vectors at each point. We can summarize these relations compactly with the notation $\widehat{u}_1 = \widehat{\rho}$, $\widehat{u}_2 = \widehat{\phi}$, $\widehat{u}_3 = \widehat{\theta}$ whence:

$$\widehat{u}_i \bullet \widehat{u}_j = \delta_{ij}, \qquad \widehat{u}_i \times \widehat{u}_j = \sum_{k=1}^{3} \epsilon_{ijk} \widehat{u}_k$$

this is the same pattern we saw for the cartesian unit vectors.

We will return to the polar, cylindrical and spherical coordinate systems as the course progresses. Even now we could consider a multitude of problems based on the combination of the material covered thus-far and it's intersection with curvelinear coordinates. There are other curved coordinate systems beyond these standard three, but I leave those to your imagination for the time being. I do discuss a more general concept of coordinate system in the advanced calculus course. In manifold theory the concept of a coordinate system is refined in considerable depth. We have no need of such abstraction here so I'll behave[32].

The examples which follow show how we convert from spherical to Cartesian and vice-versa:

Example 1.6.16. *Suppose we are given a cone centered around the z-axis which is defined by $\phi = \frac{\pi}{3}$ in spherical coordinates. Find the equation of the cone in Cartesian coordinates.*

Solution: *To convert to cartesian coordinates we can substitute the given $\phi = \frac{\pi}{3}$ into the spherical coordinate formulas:*

$$z = \rho \cos \phi = \rho \cos(\pi/3) = \frac{\rho}{2}$$
$$x = \rho \cos \theta \sin \phi = \frac{\sqrt{3}}{2} \rho \cos \theta$$
$$y = \rho \sin \theta \sin \phi = \frac{\sqrt{3}}{2} \rho \sin \theta$$

Thus $x^2+y^2 = \frac{3}{4}\rho^2(\cos^2\theta+\sin^2\theta) = \frac{3}{4}\rho^2$ or $3\rho^2 = 4(x^2+y^2)$. But, $\rho^2 = x^2+y^2+z^2$ hence $4x^2+4y^2 = 3(x^2+y^2+z^2)$ and we conclude with the Cartesian equation for the cone: $\boxed{z^2 = \frac{1}{3}(x^2+y^2).}$

Was the method used in the example above the most efficient? Are the other ways to derive $z^2 = \frac{1}{3}(x^2+y^2)$ from the given $\phi = \pi/3$? Certainly. I merely illustrate one solution. Another solution based on $r = \rho \sin \phi$ would be a bit easier.

Example 1.6.17. *Write the expression $z^2 = x^2 + y^2$ in spherical coordinates.*

[32] I'd guess most calculus text editors would say this whole paragraph is misbehaving, but I have no editor so ha.

Solution: *replace x, y, z with the spherical coordinate equivalents:*

$$\begin{aligned}(\rho\cos\phi)^2 &= (\rho\cos\theta\sin\phi)^2 + (\rho\sin\theta\sin\phi)^2 \\ &= \rho^2\cos^2\theta\sin^2\phi + \rho^2\sin^2\theta\sin^2\phi \\ &= \rho^2(\cos^2\theta + \sin^2\theta)\sin^2\phi \\ &= \rho^2\sin^2\phi\end{aligned}$$

Thus $\cos^2\phi = \sin^2\phi$ which yields $(\cos\phi - \sin\phi)(\cos\phi + \sin\phi) = 0$ or $\cos\phi = \pm\sin\phi$. Note, $0 \leq \phi \leq \pi$ thus we find solutions $\phi = \pi/4$ and $\phi = 3\pi/4$. I encourage the reader to think about why this is geometrically clear.

Remark 1.6.18.

I'm not particularly enamored with these math-conventions. In my undergraduate years, as I was raised by physicists, it was often the case that the meaning of θ and ϕ were switched. In the **physics** notation:

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

Hence $r^2 = x^2 + y^2 + z^2 = s^2 + z^2$ and $0 \leq \phi \leq 2\pi$ whereas $0 \leq \theta \leq \pi$. I mention this to make the reader aware that the math conventions for spherical coordinates are not universally accepted. That said, we will use math conventions unless otherwise explicitly stated. You can look at pages 382-384 of my handwritten calculus III notes where I contrast the conventions from Thomas' *Calculus* and those of Griffith's *Electrodynamics*.

1.7 Problems

Problem 1 The Cartesian product of three sets A, B and C is defined as follows:

$$A \times B \times C = \{(a, b, c) \mid a \in A,\ b \in B,\ c \in C\}.$$

If $(x, y, z) \in [0, 1] \times \mathbb{R} \times \mathbb{Z}$ then what can you tell me about x, y and z?

Problem 2 Suppose $P = (1, 2)$, $Q = (-1, -2)$ and $R = (0, 3)$. What point is in the middle of these three points?
Hint: the "middle" is found by the vector-average of these points; the middle point M is simply given by $M = \frac{1}{3}(P + Q + R)$. This idea also works for 4,5 or 500 points.

Problem 3 *A tetrahedron is an example of a boundary to a Platonic solid. It's formed by gluing four equilateral triangles together. Place a unit tetrahedron with vertices at $(0, 0, 0)$ and $(1, 0, 0)$. Assume it is placed so that $z \geq 0$ for points on the tetrahedron.

(a.) find the height of the tetrahedron (this is maximum z-value in our set-up)

(b.) Find the angle between the line-segments which connect the center of a tetrahedron and two vertices of a unit tetrahedron via a vector argument. (chemistry sees this shape in CH_4)

Problem 4 The displacement from point P to point Q is defined to be the vector which points from P to Q, in particular $\overrightarrow{PQ} = Q - P$. Suppose $A = (1, 2, 3)$, $B = (1, 1, -2)$ and $C = (4, 4, 4)$.

(a) calculate the vectors $\overrightarrow{AB}$, $\overrightarrow{BC}$ and $\overrightarrow{CA}$.

(b) calculate $\overrightarrow{AB} + \overrightarrow{BC}$

(c) calculate $\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CA}$

(d) find the angle between $\overrightarrow{AB}$ and $\overrightarrow{AC}$, call it θ

(e) find the angle between $\overrightarrow{CA}$ and $\overrightarrow{CB}$, call it β

(f) find the angle between $\overrightarrow{BC}$ and $\overrightarrow{BA}$, call it α

(g) calculate $\theta + \beta + \alpha$. Does your result make sense? Comment on the geometric meaning of this problem.

Problem 5 Are the vectors $\vec{v} = \langle 1, 0, 4\rangle$ and $\vec{w} = \langle 0, 2, 0\rangle$ orthogonal, colinear or neither? (give a calculation and explain)

Problem 6 Calculate the projection of $\vec{v} = \langle 1, 1, 1\rangle$ onto the vector $\vec{w} = 2\,\widehat{y} - \widehat{z}$. Write $\vec{v}$ as a sum of a vector which is colinear to $\vec{w}$ and another vector which is orthogonal to $\vec{w}$.

Problem 7 Suppose $\vec{A} = \widehat{x} + \widehat{y}$ and $\vec{B} = \widehat{z}$ and $\vec{C} = \widehat{y}$.

(a) calculate $\vec{A} \bullet (\vec{B} \times \vec{C})$

(b) calculate $\vec{B} \bullet (\vec{A} \times \vec{C})$

(c) which of the calculations above gives the volume of the parallel piped spanned by the given vectors ?

Problem 8 Suppose $\vec{A}, \vec{B}$ and $\vec{C}$ are vectors which do not all lie in a common plane. When does $\vec{A} \bullet (\vec{B} \times \vec{C})$ yield the volume of the parallel piped with edges $\vec{A}, \vec{B}$ and $\vec{C}$?

Problem 9 Find the equation of a plane which contains the line parametrized by $\vec{r}(t) = \langle 1+t, 2-t, 3\rangle$ and the vector $\vec{w} = \langle 1, 2, 3\rangle$.

Problem 10 (this problem is bonus!) I explain in the notes that a convenient compact notation for the cross-product is provided by the antisymmetric symbol ϵ_{ijk} in particular:

$$\vec{A} \times \vec{B} = \sum_{i,j,k=1}^{3} \epsilon_{ijk} A_i B_j \,\widehat{x}_k.$$

The identity below can be checked case-by-case, it is true for all m, n, i, k in $\{1, 2, 3\}$,

$$\boxed{\sum_{j=1}^{3} \epsilon_{mnj}\epsilon_{ikj} = \delta_{mi}\delta_{nk} - \delta_{mk}\delta_{ni}. \quad \star}$$

I give you two options: (choose just one please)

(a.) show the $\star$ identity is true for all $3^4 = 81$ cases.

(b.) show that $\vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \bullet \vec{C})\vec{B} - (\vec{A} \bullet \vec{B})\vec{C}$ by implementing the $\star$-identity in an index-calculation.

Problem 11 Suppose a force $\vec{F}$ with magnitude of 100 in the $\hat{x}$-direction is applied to a mass displacing it from the point $(1, 2, 3)$ to the point $(4, 4, 4)$. Calculate the work done by $\vec{F}$ (assume $\vec{F}$ is constant).

Problem 12 Suppose a rigid body rotates with angular frequency $\vec{\omega}$. It is customary for the direction of $\vec{\omega}$ to point along the axis of rotation while the magnitude ω gives the number of radians rotated per time period. It is known that the velocity of a particle on such a rigid body at $\vec{r}$ from the axis of rotation is simply given by $\vec{v} = \vec{\omega} \times \vec{r}$. Here the vector $\vec{r}$ points from the axis to the point on the body which has velocity $\vec{v}$. See the picture below, the choice of $\vec{r}$ is not unique. Why is the formula $\vec{v} = \vec{\omega} \times \vec{r}$ unambiguous? Why is $\vec{\omega} \times \vec{r}_1 = \vec{\omega} \times \vec{r}_2$ where $\vec{r}_1, \vec{r}_2$ are related as indicated in the diagram below? (please remind me to draw this in class, hopefully this will allow all involved to get the right picture in mind)

Problem 13 Find the direction vector of the line of intersection of the planes $x + y + z = 3$ and $2x - 3y - 4z = 7$.

Problem 14 Explain what the parametric equations $x = u + v, y = u - v$ and $z = 1 + u$ describe. Find the corresponding cartesian equations (either give answer as a graph or as a level surface, please say which your answer is)

Problem 15 Suppose a plane S contains the points $(1, 0, 2), (3, 4, 1)$ and $(0, 0, 1)$. Find the point on S which is closest to the origin.

Problem 16 Refer once more to the plane S from the previous problem. Find a parametrization of the triangular region bounded by the points given in the previous problem.

Problem 17 Suppose $\vec{a} \bullet \vec{b} = \vec{a} \bullet \vec{c}$ for some vector $\vec{a}$. Does it follow that $\vec{b} = \vec{c}$?

Problem 18 Suppose $\vec{a} \bullet \vec{b} = \vec{a} \bullet \vec{c}$ for **all** vectors $\vec{a}$. Does it follow that $\vec{b} = \vec{c}$?

Problem 19 Suppose $\vec{a} \times \vec{b} = \vec{a} \times \vec{c}$ for some vector $\vec{a}$. Does it follow that $\vec{b} = \vec{c}$?

Problem 20 Suppose $\vec{a} \times \vec{b} = \vec{a} \times \vec{c}$ for **all** vectors $\vec{a}$. Does it follow that $\vec{b} = \vec{c}$?

Problem 21 * **Claim:** If $\vec{v} \bullet \vec{x} = c$ and $\vec{v} \times \vec{x} = \vec{b}$ for particular nonzero vectors $\vec{v}, \vec{b}$ and some constant c then we can solve for $\vec{x}$ uniquely in terms of the given vectors $\vec{v}, \vec{b}$ and c. **Show the claim is true**.

Problem 22 Suppose that $\vec{A}, \vec{B}, \vec{C}, \vec{D}$ are coplanar vectors in $\mathbb{R}^3$. Show that $(\vec{A} \times \vec{B}) \times (\vec{C} \times \vec{D}) = 0$.

Problem 23 Use vectors to show that the diagonals of a parallelogram are orthogonal iff the parallelogram is a rhombus.

Problem 24 * Suppose that A, B, C, D are points in $\mathbb{R}^3$ where no three of these points lie on the same line. Furthermore, suppose S is the quadrilateral $ABCD$ formed from these points. Note S need not lie in a plane. Let M_1, M_2, M_3, M_4 be the midpoints of of the edges of S. Use a vector argument to show $M_1M_2M_3M_4$ is a parallelogram!

Problem 25 This problem explores the concept of **direction cosines**, these gadgets are still popular in certain applications, we don't use them much in this course, but here it is for breadth and also practice on your dot-product skilz. Find the **cosine of the angle** between $\vec{v} = \langle 2, 4, \sqrt{5} \rangle$ and the

(a) positive x-axis (usually denoted $\cos(\alpha)$)

(b) positive y-axis (usually denoted $\cos(\beta)$)

(c) positive z-axis (usually denoted $\cos(\gamma)$)

(d) how are the answers to parts a, b, c related to $\vec{v}$?

Problem 26 Show that a triangle inscribed in a semicircle must be a right triangle. To say the triangle is inscribed in a semicircle means one leg of the triangle lies on the diameter of the circle.

Problem 27 Find the point where the line $\vec{r}(t) = \langle 1+t, 2-3t, 3+4t\rangle$ intersects the plane $x + y + z = 8$.

Problem 28 What is the distance between the line $\vec{r}_1(t) = \langle 1+t, 2-3t, 3+4t\rangle$ and $\vec{r}_2(t) = \langle 1+2t, 2+t, 3t\rangle$? Note: when discussing the distance between extended objects it is customary to define the distance between objects as the closest possible distance.

Problem 29 Simplify: $[\vec{A} + \vec{B}] \bullet (\vec{A} \times \vec{B})$.

Problem 30 Simplify: $[\vec{A} - \vec{B}] \bullet [\vec{A} + \vec{B}]$. If you are given that $\vec{A} - \vec{B}, \vec{A} + \vec{B}$ are orthogonal then what condition must this pair of vectors satisfy?

Problem 31 Use a computer graphing system to plot each of the curves below and include a printout in your homework. Comment on the identity of each curve (as in: "this is a circle, ellipse, hyperbola, spiral etc...")

(a.) let $x = \cos(t)$, $y = \sin(t)$ and $z = \cos(t)$ for $t \in [0, 2\pi]$

(b.) let $x = \cosh(t), y = 4\sinh(t)$ for $t \in \mathbb{R}$

(c.) let $x = e^{-t}\cos(50t)$ and $y = e^{-t}\sin(50t)$ and $z = e^{-t}\cos(50t)$ for $t \in [0, 1]$

(d.) let $x = e^t$, $y = 2e^t$ and $z = 3e^t$ for $t \in [0, \ln(3)]$

Problem 32 Identify each surface below. Either use results from my notes and simply state the name of the surface, or use a computer to plot the surface and print your result to justify your claim. (at least one of the surfaces below avoids explicit tabulation by my notes)

(a.) $z = x^2 + y^2$

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

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

(d.) $x^2 + 2y^2 = 1$

(e.) $x^2 + y^2 + z^2 + 2xy + 2xz = 1$

Problem 33 Provide parametrizations of the surfaces (a,b,c,d) given in the preceding problem. (or, provide a patch for part of the surface and explain where your patch covers, you must explicitly state the domain of your parameters for full credit, thanks!)

Problem 34 Parametrize the subset of the plane $x + 3y - z = 10$ for which $1 \leq x \leq 3$ and $2 \leq y \leq 4$.

Problem 35 Parametrize the subset of the plane $x + 3y - z = 10$ for which $1 \leq y^2 + z^2 \leq 4$.

Problem 36 Convert the following inequalities/equations to either spherical or cylindrical coordinates as appropriate and comment on the result.

(a.) $1 \leq x^2 + y^2 + z^2 \leq 3$

(b.) $0 \leq x^2 + y^2 \leq 4$

Problem 37 Consider the point $P = (\sqrt{3}, 1, 2)$. Find the

(a.) cylindrical coordinates of P

(b.) spherical coordinates of P

Problem 38 Write the vector $\vec{v} = \langle 1, 2, 3 \rangle$ in terms of the spherical frame at an arbitrary point with spherical coordinates ρ, ϕ, θ. In other words, find functions a, b, c of spherical coordinates such that $\vec{v} = a\widehat{\rho} + b\widehat{\phi} + c\widehat{\theta}$.
Hint: dot-products with respect to $\widehat{\rho}, \widehat{\phi}, \widehat{\theta}$ nicely isolate a, b and c if you make use of the fact that $\widehat{\rho}, \widehat{\phi}, \widehat{\theta}$ forms an orthonormal frame.

Problem 39 Find and parametrize the curve of intersection of $x^2 + y^2 = 4$ and $z = x^2 - y^2$.

Problem 40 Find and parametrize the curve of intersection of $x + y + z = 10$ and $z = x^2 + y^2$.

Problem 41 Convert the equation $4 = \rho \sin \phi$ to

(a.) cylindrical coordinates

(b.) cartesian coordinates

Problem 42 Find the intersection of $\phi = \pi/3$ and $z = 4$ and provide a parametrization which covers this curve of intersection.

Problem 43 Find the intersection of $\theta = \pi/4$ and $\rho = 4$ and provide a parametrization which covers this curve of intersection.

Problem 44 Find the polar coordinate equations for

(a.) the ellipse $x^2/a^2 + y^2/b^2 = 1$,

(b.) the line $y = 1 - 2x$

Problem 45 I recommend the use of vector arguments ($||\vec{c}||^2 = \vec{c} \bullet \vec{c}$ etc. . .) to analyze the following:

(a.) let $\vec{a} \neq 0$, characterize all vectors $\vec{r} = \langle x, y \rangle$ such that $||\vec{r} - \vec{a}|| = ||\vec{r} + \vec{a}||$

(b.) let $\vec{a} \neq 0$, characterize all vectors $\vec{r} = \langle x, y, z \rangle$ such that $||\vec{r} - \vec{a}|| = ||\vec{r} + \vec{a}||$

Chapter 2

calculus and geometry of curves

In this chapter we study mappings from $\mathbb{R}$ into $\mathbb{R}^2$ or $\mathbb{R}^3$ primarily. However, in general, a function $\phi : I \subseteq \mathbb{R} \to S$ is called a **path** in S. The set of points $\phi(I)$ gives a **curve** in S. This is my default terminology, a given curve admits many paths which parametrize the curve. That said, it is often the case that these concepts are interchanged or even identified. I cannot protect the reader from these abuses and at times I may even be the instigator.

The calculus of paths is simple enough; we differentiate and integrate one component at a time. Many of the results we know and love from single variable calculus have a natural extension. In particular, the Fundamental Theorem of Calculus part I and II apply to paths in the natural manner. In addition, the new vector operations we learned in the earlier chapter all obtain simple product rules. In the calculus of paths we have a product rule for scalar multiplication, dot-products and cross-products.

Geometrically, I at times refer to the path as a set of points. Other times, we think of the path as a vector given for each t in the domain. This is reasonable due to identification of points and vectors based at the origin as we introduced in the previous chapter. However, we also discuss vectors based at points other than the origin in some depth in this chapter. In particular, the Frenet-Frame is detailed. This triple of orthonormal vectors is attached at each point along a non-stop, non-linear regular curve. We also study the application of the calculus of paths to the physics to three dimensional motion. We derive beautiful formulas which decompose the acceleration of a path into tangent and normal directions. Our vector-based language allows treatment of essentially arbitrary three-dimensional motion. Of course, to be fair, we do not explain how to solve Newton's Equations in any depth. We do show how to find the equation of a tangent line to a path. Also, we show how to find the equation of the osculating plane and we relate the radius of the osculating circle to the curvature of the path. We can briefly describe curvature and torsion of a path:

> The curvature tells us how the path lifts away from the tangent directions whereas the torsion tells us when the curve lifts off the osculating plane.

Let me briefly outline the sections. The first section lays out the basic differentiation and integration rules for paths as well as a number of useful propositions. In the second section we study geometry of paths. In particular, arclength is described and the tangent, normal and binormal are introduced to study the length and shape of a curve. We introduce curvature and torsion and derive the Frenet-Serret equations. These are a prototype of a whole family of related theorems in differential geometry as powerfully promoted by Cartan in the early twentieth century. Of course, the study of curves is also a powerful example of how calculus and analytic geometry bring profound insight into problems of classical geometry. In short, we argue, for reasonable curves, two curves are congruent in the sense of high-school geometry if they share the same curvature and torsion. The problem of classifying curves in $\mathbb{R}^3$ is a useful prelude to deeper studies of surfaces or curves in exotic spaces which is an **active field** of mathematics. In the third section we define standard physical terms such as speed, velocity and acceleration. A number of interesting examples are given. For example, we calculate the arclength of the motion of a projectile near the surface of the earth. If time permits, we'll see how the calculus of paths allows a beautiful derivation of Kepler's Laws for planetary motion. Finally, in our concluding section, we explain how to integrate a scalar quantity along a curve. This integral allows us to calculate the mass of a wire, or the balancing point of an irregular curve, or the average temperature along a road which winds up and down through mountains and deserts. Pedagogically, the final section is here to help ease a certain conceptual difficulty which appears later in this course[1].

2.1 calculus for curves

In this section we describe the calculus for functions with a domain of real numbers and a range of vectors. It is possible to define the derivative in terms of a limiting process, but, little is gained by doing so in this section so I make a more pragmatic definition[2]. We'll begin with $\mathbb{R}^3$,

Definition 2.1.1. *calculus of 3-vector-valued functions.*

Suppose $\vec{F}(t) = \langle F_1(t), F_2(t), F_3(t)\rangle$ then

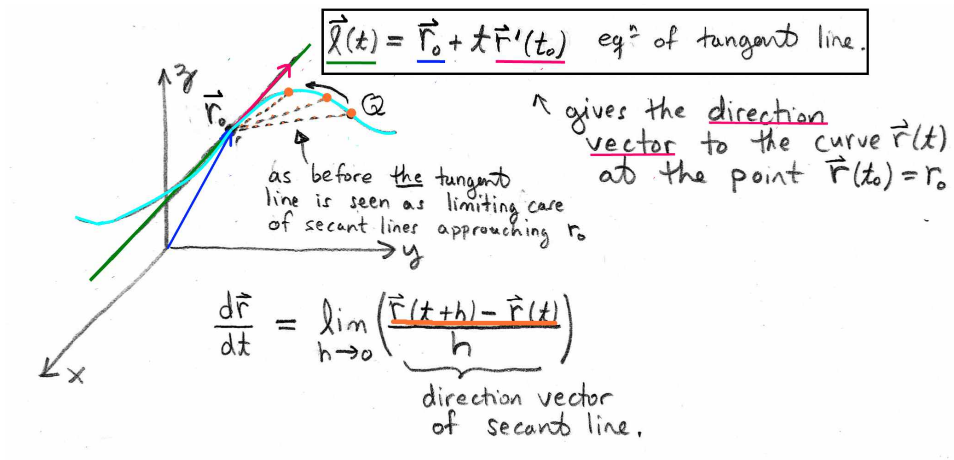
1. If F_1, F_2 and F_3 are differentiable functions near t we define
$$\frac{d\vec{F}}{dt} = \frac{d}{dt}\langle F_1, F_2, F_3\rangle = \left\langle \frac{dF_1}{dt}, \frac{dF_2}{dt}, \frac{dF_3}{dt}\right\rangle.$$
2. If F_1, F_2 and F_3 are integrable functions on $[a, b]$ then we define
$$\int_a^b \vec{F}(t)dt = \int_a^b \langle F_1, F_2, F_3\rangle dt = \left\langle \int_a^b F_1(t)dt, \int_a^b F_2(t)dt, \int_a^b F_3(t)dt\right\rangle.$$
3. We write $\int \vec{f}(t)dt = \vec{F}(t) + \vec{c}$ iff $\frac{d\vec{F}}{dt} = \vec{f}(t)$ and $\vec{c} = \langle c_1, c_2, c_3\rangle$ is a constant vector. Equivalently,
$$\int \vec{f}(t)dt = \left\langle \int f_1(t)dt, \int f_2(t)dt, \int f_3(t)dt\right\rangle.$$

We also use the prime notation for differentiation of vector valued functions if it is convenient; this means $\vec{A}\,'(t) = d\vec{A}/dt = \frac{d\vec{A}}{dt}$. Higher derivatives are also denoted in the same manner as previously;

[1] this course of action was recommended to me by my brother who has had good success with the inclusion of this topic for the past several years of his teaching multivariate calculus

[2] for the purist you can skip ahead to the chapter on differentiation where I describe how to differentiate a general function from $\mathbb{R}^n$ to $\mathbb{R}^m$, this is definition can be derived from that definition with a few basic theorems of advanced calculus.

for example, $\frac{d^2\vec{A}}{dt^2} = \frac{d}{dt}\left[\frac{d\vec{A}}{dt}\right]$. The geometric meaning of the definition is encapsulated in the picture below:



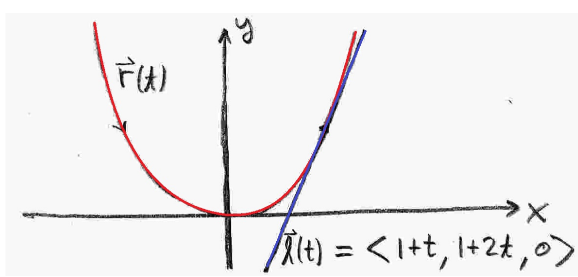
If $t \mapsto \vec{r}(t)$ is some parametrized curve and $t_o \in dom(\vec{r})$ such that $\vec{r}\,'(t_o) \neq 0$ defines the tangent vector to the curve at $\vec{r}(t_o)$. Moreover, a natural parametrization of the tangent line is given by $\vec{l}(s) = \vec{r}_o + s\vec{r}\,'(t_o)$. Recall that the parametric view is natural one in this context. Hopefully we learned this already in the first chapter of these notes.

Example 2.1.2. *Let $C : \vec{r}(t) = (t, t^2, 0)$ find the tangent line to C at $(1, 1, 0)$.*

Solution: *observe $\langle t, t^2, 0\rangle = \langle 1, 1, 0\rangle$ for $t = 1$. Furthermore,*

$$\frac{d\vec{r}}{dt} = \langle 1, 2t, 0\rangle$$

Setting $t = 1$ give the direction-vector $\langle 1, 2, 0\rangle$ for the tangent line. We find $\vec{l}(t) = \langle 1, 1, 0\rangle + t\langle 1, 2, 0\rangle$ is a parametrization of the tangent line to C at $(1, 1, 0)$. The curve and tangent line are both in the $z = 0$ plane which is plotted below:



Example 2.1.3. *Let a, b, c, x_o, y_o, z_o be constants and suppose $\vec{r}(t) = \langle x_o + ta, y_o + tb, z_o + tc\rangle$. Differentiate $\vec{r}(t)$ as follows:*

$$\begin{aligned}\frac{d\vec{r}}{dt} &= \frac{d}{dt}\Big\langle x_o + ta, y_o + tb, z_o + tc\Big\rangle \\ &= \Big\langle \frac{d}{dt}(x_o + ta),\ \frac{d}{dt}(y_o + tb),\ \frac{d}{dt}(z_o + tc)\Big\rangle \\ &= \langle a, b, c\rangle.\end{aligned}$$

In other words, if $\vec{r}(t)$ is a line with base-point (x_o, y_o, z_o) and direction-vector $\vec{v} = \langle a, b, c\rangle$ then the derivative of the line is the direction-vector $\frac{d\vec{r}}{dt} = \vec{v}$.

Let's pause from our geometric inspirations to simply study the calculation. It is nothing more than your previous calculus applied one component at a time. All the rules and techniques we developed in previous courses still apply here.

Example 2.1.4. *Let $\vec{F}(t) = \langle 1,\ t,\ \cos(t)\rangle$.*

$$\frac{d\vec{F}}{dt} = \left\langle \frac{d}{dt}(1), \frac{d}{dt}(t), \frac{d}{dt}(\cos(t)) \right\rangle = \langle 0,\ 1,\ -\sin(t)\rangle.$$

$$\int \vec{F}(t)dt = \left\langle \int dt, \int tdt, \int \cos(t)dt \right\rangle = \left\langle t + c_1,\ \frac{1}{2}t^2 + c_2,\ \sin(t) + c_3 \right\rangle.$$

$$\int_0^1 \vec{F}(t)dt = \left\langle \int_0^1 dt, \int_0^1 tdt, \int_0^1 \cos(t)dt \right\rangle = \left\langle 1,\ \frac{1}{2},\ \sin(1) \right\rangle.$$

Example 2.1.5. *Let $\vec{F}(t) = \langle e^t f(t), f(t^2)\rangle$ where f is a differentiable function on $\mathbb{R}$. We calculate the derivative as follows:*

$$\vec{F}\,(t) = \left\langle \frac{d}{dt}(e^t f(t)),\ \frac{d}{dt}(f(t^2)) \right\rangle = \left\langle e^t f(t) + e^t f'(t),\ 2tf'(t^2) \right\rangle.$$

We used the product-rule for the x-component and the chain-rule for the y-component.

Example 2.1.6. *Let $\vec{G}(t) = \langle \tan(t), \cosh(t), \tan^{-1}(t)\rangle$. Differentiating $\vec{G}$:*

$$\vec{G}'(t) = \left\langle \sec^2(t), \sinh(t), \frac{1}{t^2+1} \right\rangle.$$

Integration requires a bit more thought. Recall,

$$\int \tan t\, dt = \int \frac{\sin t dt}{\cos t} = \int \frac{-d(\cos t)}{\cos t} = -\ln|\cos t| + c_1.$$

On the other hand, we define[3] $\cosh t = \frac{1}{2}(e^t + e^{-t})$ *and* $\sinh t = \frac{1}{2}(e^t - e^{-t})$

$$\int \cosh t\, dt = \int \frac{1}{2}(e^t + e^{-t})dt = \frac{1}{2}(e^t - e^{-t}) + c_2 = \sinh t + c_2.$$

Next, the integral of an inverse function is often nicely revealed by the technique of integration by parts. Consider:

$$\int \underbrace{\tan^{-1}(t)}_{u} \underbrace{dt}_{dv} = uv - \int vdu = t\tan^{-1}(t) - \int \frac{tdt}{1+t^2} = t\tan^{-1}(t) + \frac{1}{2}\ln|1+t^2| + c_3$$

the last step was a basic $u = 1 + t^2$ substitution. Let us collect our work,

$$\int \vec{G}(t)\, dt = \left\langle -\ln|\cos t| + c_1,\ \sinh t + c_2,\ t\tan^{-1}(t) + \ln\sqrt{1+t^2} + c_3 \right\rangle.$$

I include the previous example is to ask a few questions:

[3] if you have not studied these previously then you ought to read over my materials from calculus I and II to see why they are worth your attention, also, I usually will just write $\int \cosh t dt = \sinh t + c$ in future work for this course.

1. Do you remember the basic definitions of the functions to which I refer? Do you need to review the precalculus chapter of my calculus I notes or some similar resource?

2. Are you ready to differentiate with the full force of the product, quotient and chain rules? If not, perhaps a review of materials from the previous course is in order.

3. Do you know how to integrate without breaking a sweat? Are you comfortable with the major techniques from previous semesters? Is a review of integration techniques a good idea at this point?

One more calculational example before we return to the task of visualization and tangent lines in two and more dimensions.

Example 2.1.7. *Let* $\vec{F}(t) = \Big\langle \cos^2 t, \frac{1}{1+3t}, t^2 \sin\big(t^3\big) \Big\rangle$. *The chain and product rules reveal:*

$$\frac{d\vec{F}}{dt} = \Big\langle -2\cos t \sin t, \frac{-3}{(1+3t)^2}, 2t\sin\big(t^3\big) + 3t^4\cos\big(t^3\big) \Big\rangle.$$

Integration is not too hard if we know $\cos^2(t) = \frac{1}{2}\big(1+\cos 2t\big)$.

$$\int \vec{F}(t)\,dt = \Big\langle \frac{1}{2}\big(t + \frac{1}{2}\sin 2t\big) + c_1,\ \frac{1}{3}\ln|1+3t| + c_2,\ -\frac{1}{3}\cos\big(t^3\big) + c_3 \Big\rangle.$$

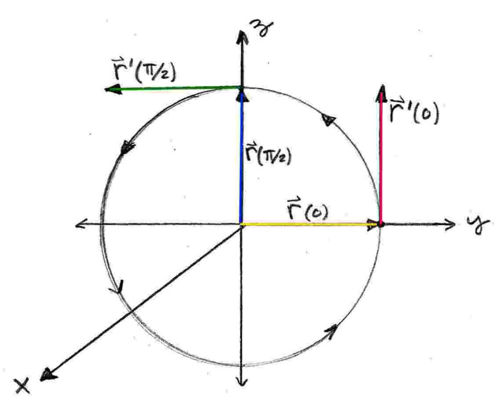
If in doubt about the integration then I invite the reader to check the details with explicit u-substitutions. We often omit such details when they are routine calculus.

Example 2.1.8. *Let* $\vec{r}(t) = \langle 0, \cos t, \sin t\rangle$. *Find* $\vec{r}\,'(t)$ *and plot the derivative at* $t = 0$ *and* $t = \pi/2$.

Solution: *differentiate component-wise to find:*

$$\vec{r}\,'(t) = \langle 0, -\sin t, \cos t\rangle$$

thus $\vec{r}\,'(0) = \langle 0, 0, 1\rangle$ *and* $\vec{r}\,'(\pi/2) = \langle 0, -1, 0\rangle$. *I plot these below where I attach the derivative vectors at the point on the curve where they are tangent.*



Admittably the example above is a bit silly. It is a two dimensional example masquerading as a three dimensional problem.

Example 2.1.9. *Let $\vec{r}_1(t) = \langle t, t^2, t^3\rangle$ and $\vec{r}_2(t) = \langle \sin t, \sin 2t, t\rangle$. Find the angle at which these curves intersect.*

Solution: *the angle of intersection for two curves is defined to be the angle between their tangent vectors at the point of intersection. Calculate,*

$$\frac{d\vec{r}_1}{dt} = \langle 1, 2t, 3t^2\rangle, \qquad \frac{d\vec{r}_2}{dt} = \langle \cos t, 2\cos 2t, 1\rangle.$$

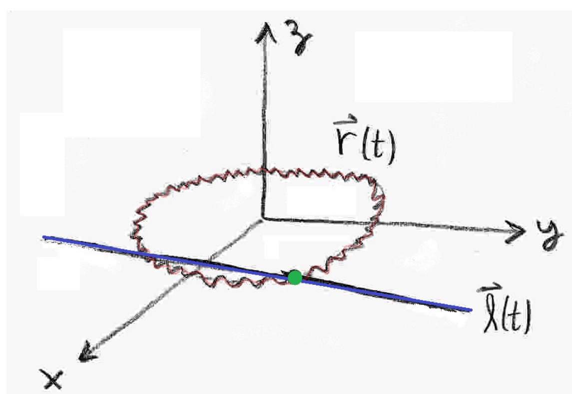
The point at which these curves intersect is given by $t = 0$ thus we need to find the angle between $\frac{d\vec{r}_1}{dt}(0) = \langle 1, 0, 0\rangle$ and $\frac{d\vec{r}_2}{dt}(0) = \langle 1, 2, 1\rangle$. Observe, $\frac{d\vec{r}_1}{dt}(0) \bullet \frac{d\vec{r}_2}{dt}(0) = 1$ while $||\frac{d\vec{r}_1}{dt}(0)|| = 1$ and $||\frac{d\vec{r}_2}{dt}(0)|| = \sqrt{6}$. It follows that $1 = \sqrt{6}\cos\theta$ hence $\theta = \cos^{-1}(1/\sqrt{6}) = 66^o$.

Example 2.1.10. *Consider the curve $\vec{r}(t) = \langle \cos t,\ \sin t,\ \frac{1}{100}\sin(100t)\rangle$ for $0 \leq t \leq 2\pi$. Sketch the curve and its tangent to the point at $t = \pi/4$.*

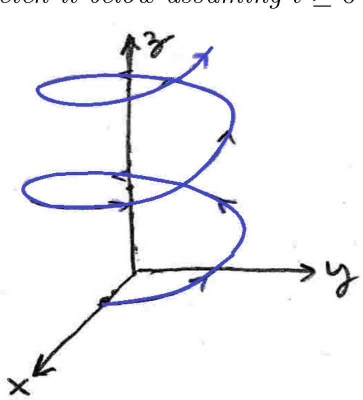
Solution: *Differentiating yields $\vec{r}\,'(t) = \langle -\sin t,\ \cos t,\ \cos(100t)\rangle$. Evaluate at $t = \pi/4$ to find $\vec{r}\,'(\pi/4) = \langle -1/\sqrt{2},\ 1/\sqrt{2},\ -1\rangle$. Thus the tangent line $\vec{l}(t) = \vec{r}(\pi/4) + t\vec{r}\,'(\pi/4)$ yields:*

$$\vec{l}(t) = \langle 1/\sqrt{2},\ 1/\sqrt{2},\ 0\rangle + t\langle -1/\sqrt{2},\ 1/\sqrt{2},\ -1\rangle.$$

Technically, t in $\vec{l}(t)$ should be distinguished from t in $\vec{r}(t)$. However, it is common practice to simply use t for many curves. Furthermore, the bulk motion of this curve in in the x, y-directions; to a good approximation the curve is $x^2 + y^2 = 1$ with $z = 0$. In truth, the curve is a sinusoid waving transversally to the unit-circle in the xy-plane. Below is an attempt at a sketch:



Example 2.1.11. *Consider $\vec{r}(t) = \langle \cos t, \sin t, t\rangle$. This curve falls on the cylinder $x^2 + y^2 = 1$ as is easily verified by noting $\cos^2 t + \sin^2 t = 1$. Furthermore, the derivative is given by $\vec{r}\,'(t) = \langle -\sin t, \cos t, 1\rangle$ and this shows that the curve is always increasing in the z-component. This curve is known as a* **circular helix***. I sketch it below assuming $t \geq 0$*



The derivative of an n-vector valued functions of a real variable is likewise calculated component by component.

Definition 2.1.12. *calculus of n-vector-valued functions.*

Suppose $\vec{F}(t) = \langle F_1(t), F_2(t), \dots, F_n(t)\rangle$ then

1. If $F_1, F_2, \dots, F_n$ are differentiable functions near t we define
$$\frac{d\vec{F}}{dt} = \frac{d}{dt}\langle F_1, F_2, \dots, F_n\rangle = \left\langle \frac{dF_1}{dt}, \frac{dF_2}{dt}, \dots, \frac{dF_n}{dt}\right\rangle.$$
2. If $F_1, F_2, \dots, F_n$ are integrable functions on $[a,b]$ then we define
$$\int_a^b \vec{F}(t)dt = \int_a^b \langle F_1, F_2, \dots, F_n\rangle dt = \left\langle \int_a^b F_1(t)dt, \int_a^b F_2(t)dt, \dots, \int_a^b F_n(t)dt\right\rangle.$$
3. We write $\int \vec{f}(t)dt = \vec{F}(t) + \vec{c}$ iff $\frac{d\vec{F}}{dt} = \vec{f}(t)$ and $\vec{c} = \langle c_1, c_2, \dots, c_n\rangle$ is a constant vector. Equivalently,
$$\int \vec{f}(t)dt = \left\langle \int f_1(t)dt, \int f_2(t)dt, \dots, \int f_n(t)dt\right\rangle.$$

In summation notation the definitions translate to:

$$\frac{d}{dt}\left[\sum_{j=1}^n F_j\,\widehat{x}_j\right] = \sum_{j=1}^n \frac{dF_j}{dt}\,\widehat{x}_j \qquad \text{and} \qquad \int_a^b\left[\sum_{j=1}^n F_j\,\widehat{x}_j\right]dt = \sum_{j=1}^n\left[\int_a^b F_j\ dt\right]\widehat{x}_j$$

We differentiate and integrate componentwise.

Example 2.1.13. *Let $\vec{F}(t) = \langle t,\ t^2,\ \dots,\ t^n\rangle$. It follows that,*

$$\frac{d\vec{F}}{dt} = \langle\ 1, 2t, \dots, nt^{n-1}\ \rangle$$

and, for constants $c_1, c_2, \dots, c_n$,

$$\int \vec{F}(t)dt = \langle\ \frac{1}{2}t^2 + c_1, \frac{1}{3}t^3 + c_2, \dots, \frac{1}{n+1}t^{n+1} + c_n\ \rangle.$$

Or we could calculate via summation notation, note that $\vec{F}(t) = \sum_{j=1}^n t^j\,\widehat{x}_j$ hence,

$$\frac{d\vec{F}}{dt} = \frac{d}{dt}\sum_{j=1}^n t^j\,\widehat{x}_j = \sum_{j=1}^n \frac{d}{dt}(t^j)\,\widehat{x}_j = \sum_{j=1}^n jt^{j-1}\,\widehat{x}_j.$$

Likewise,

$$\int \vec{F}(t)dt = \int\sum_{j=1}^n t^j\,\widehat{x}_j dt = \sum_{j=1}^n\int t^j\,dt\,\widehat{x}_j = \sum_{j=1}^n\left(\frac{1}{j+1}t^{j+1} + c_j\right)\widehat{x}_j.$$

We usually find ourselves working problems with $n = 1, 2$ or 3 in this course. Many of the theorems known to us from calculus I apply equally well to vector-valued functions of a real variable. The key is that the differentiation concerns the domain whereas the range just rides along. If there was somehow a time-dependence for $\widehat{x}$, $\widehat{y}$, $\widehat{z}$ then the story would change, but we insist that $\widehat{x}$, $\widehat{y}$, $\widehat{z}$ are the unit-vectors of a fixed x, y, z-coodinate system[4].

Theorem 2.1.14. *fundamental theorems of calculus for space curves.*

$$(I.) \quad \frac{d}{dt}\int_a^t \vec{F}(\tau)d\tau = \vec{F}(t)$$

$$(II.) \quad \int_a^b \frac{d\vec{G}}{dt}dt = \vec{G}(b) - \vec{G}(a)$$

Proof: Apply the FTC part I componentwise as shown below:

$$\begin{aligned}
\frac{d}{dt}\int_a^t \vec{F}(\tau)d\tau = \frac{d}{dt}\int_a^t \left[\sum_{j=1}^n F_j(\tau)\,\widehat{x}_j\right] d\tau &= \frac{d}{dt}\sum_{j=1}^n \left[\int_a^t F_j(\tau)\,d\tau\right]\widehat{x}_j \\
&= \sum_{j=1}^n \left[\frac{d}{dt}\int_a^t F_j(\tau)\,d\tau\right]\widehat{x}_j \\
&= \sum_{j=1}^n F_j(t)\,\widehat{x}_j = \vec{F}(t).
\end{aligned}$$

thus (I.) holds true. Next apply FTC part II componentwise as shown below:

$$\begin{aligned}
\int_a^b \frac{d\vec{G}}{dt}dt = \int_a^b \frac{d}{dt}\left[\sum_{j=1}^n G_j(t)\,\widehat{x}_j\right]dt &= \int_a^b \left[\sum_{j=1}^n \frac{dG_j}{dt}\,\widehat{x}_j\right]dt \\
&= \sum_{j=1}^n \left[\int_a^b \frac{dG_j}{dt}dt\right]\widehat{x}_j \\
&= \sum_{j=1}^n \left[G_j(b) - G_j(a)\right]\widehat{x}_j \\
&= \sum_{j=1}^n G_j(b)\,\widehat{x}_j - \sum_{j=1}^n G_j(a)\,\widehat{x}_j \\
&= \vec{G}(b) - \vec{G}(a).
\end{aligned}$$

Therefore, part (II.) is true. $\square$

[4] In physics one might consider moving coordinate systems and in such a context the rules are a bit more interesting.

Example 2.1.15. *Note* $\int \frac{dt}{1+t^2} = \tan^{-1}(t) + c$ *and* $\int \frac{2t\,dt}{1+t^2} = \ln(1+t^2) + c$ *hence,*

$$\begin{aligned}
\int_0^1 \left\langle 4t^3, \ \frac{4}{1+t^2}, \ \frac{2t\,dt}{1+t^2} \right\rangle dt &= \left\langle \int_0^1 4t^3\,dt, \ \int_0^1 \frac{4}{1+t^2}\,dt, \ \int_0^1 \frac{2t\,dt}{1+t^2} \right\rangle \\
&= \left\langle t^4 \Big|_0^1, \ 4\tan^{-1}(t)\Big|_0^1, \ \ln(1+t^2)\Big|_0^1 \right\rangle \\
&= \langle 1, \ \pi, \ \ln(2) \ \rangle.
\end{aligned}$$

Many other properties of differentiation and integration hold for vector-valued functions.

Theorem 2.1.16. *rules of calculus for space curves.*

Let $\vec{A}, \vec{B} : J \subseteq \mathbb{R} \to \mathbb{R}^n$ and $f : J \subseteq \mathbb{R} \to \mathbb{R}$ are differentiable functions and $c \in \mathbb{R}$,

$$(1.) \ \ \frac{d}{dt}\left[\vec{A} + \vec{B}\right] = \frac{d\vec{A}}{dt} + \frac{d\vec{B}}{dt} \qquad (3.) \ \ \int [\vec{A} + \vec{B}]\,dt = \int \vec{A}dt + \int \vec{B}dt$$

$$(2.) \ \ \frac{d}{dt}\left[c\vec{A}\right] = c\frac{d\vec{A}}{dt} \qquad (4.) \ \ \int c\vec{A}dt = c\int \vec{A}dt$$

Proof: The proof of the theorem above is easily derived by simply expanding what the vector notation means and borrowing the corresponding theorems from calculus I to simplify the component expressions. I might ask for this in homework so I'll not offer details here. $\square$.

There are several types of products we can consider for vector-valued function. Each has a natural product rule.

Theorem 2.1.17. *product rules of calculus for space curves.*

Let $\vec{A}, \vec{B} : J \subseteq \mathbb{R} \to \mathbb{R}^n$ and $f : J \subseteq \mathbb{R} \to \mathbb{R}$ are differentiable functions and $c \in \mathbb{R}$,

$$(1.) \ \ \frac{d}{dt}\left[f\vec{A}\right] = \frac{df}{dt}\vec{A} + f\frac{d\vec{A}}{dt}$$

$$(2.) \ \ \frac{d}{dt}\left[\vec{A} \bullet \vec{B}\right] = \frac{d\vec{A}}{dt} \bullet \vec{B} + \vec{A} \bullet \frac{d\vec{B}}{dt}$$

$$(3.) \ \ \text{for } n = 3, \quad \frac{d}{dt}\left[\vec{A} \times \vec{B}\right] = \frac{d\vec{A}}{dt} \times \vec{B} + \vec{A} \times \frac{d\vec{B}}{dt}$$

Proof: let $\vec{A}$ and f be differentiable near t and suppose $\vec{A} = \sum_{j=1}^n A_j\,\widehat{x}_j$. Note $f\vec{A} = \sum_{j=1}^n fA_j\,\widehat{x}_j$

and calculate

$$
\begin{aligned}
\frac{d}{dt}[\, f\vec{A}\,] &= \frac{d}{dt}\left[\sum_{j=1}^{n} fA_j\, \widehat{x}_j\right] \\
&= \sum_{j=1}^{n} \frac{d}{dt}[\, fA_j\,]\, \widehat{x}_j \\
&= \sum_{j=1}^{n} \left[\, \frac{df}{dt}A_j + f\frac{dA_j}{dt}\, \right] \widehat{x}_j \\
&= \frac{df}{dt}\sum_{j=1}^{n} A_j\, \widehat{x}_j + f\sum_{j=1}^{n} \frac{dA_j}{dt}\, \widehat{x}_j \\
&= \frac{df}{dt}\vec{A} + f\frac{d\vec{A}}{dt}.
\end{aligned}
$$

The proof of (1.) is complete. Now consider the dot-product of $\vec{A}$ with $\vec{B}$,

$$
\begin{aligned}
\frac{d}{dt}[\, \vec{A}\bullet\vec{B}\,] &= \frac{d}{dt}\left[\sum_{j=1}^{n} A_jB_j\right] \\
&= \sum_{j=1}^{n} \frac{d}{dt}[\, A_jB_j\,] \\
&= \sum_{j=1}^{n} \left[\frac{dA_j}{dt}B_j + A_j\frac{dB_j}{dt}\right] \\
&= \sum_{j=1}^{n} \frac{dA_j}{dt}B_j + \sum_{j=1}^{n} A_j\frac{dB_j}{dt} \\
&= \frac{d\vec{A}}{dt}\bullet\vec{B} + \vec{A}\bullet\frac{d\vec{B}}{dt}.
\end{aligned}
$$

The proof of (2.) is complete. Only in $n = 3$ do we have a binary operation which is a cross-product, fortunately we have an easy notation so this will not be much harder than (2.).

$$\begin{aligned}
\frac{d}{dt}\big[\ \vec{A}\times\vec{B}\ \big] &= \frac{d}{dt}\Big[\sum_{i,j,k=1}^{n} A_iB_j\epsilon_{ijk}\,\widehat{x}_k\Big] \\
&= \sum_{k=1}^{n}\frac{d}{dt}\Big[\sum_{i,j=1}^{3} A_iB_j\epsilon_{ijk}\Big]\widehat{x}_k \\
&= \sum_{k=1}^{n}\Big[\sum_{i,j=1}^{3}\epsilon_{ijk}\frac{d}{dt}\big[\ A_iB_j\ \big]\Big]\widehat{x}_k \\
&= \sum_{k=1}^{n}\Big[\sum_{i,j=1}^{3}\epsilon_{ijk}\Big[\frac{dA_i}{dt}B_j + A_i\frac{dB_j}{dt}\Big]\Big]\widehat{x}_k \\
&= \sum_{i,j,k=1}^{3}\epsilon_{ijk}\frac{dA_i}{dt}B_j\,\widehat{x}_k + \sum_{i,j,k=1}^{3}\epsilon_{ijk}A_i\frac{dB_j}{dt}\,\widehat{x}_k \\
&= \frac{d\vec{A}}{dt}\times\vec{B} + \vec{A}\times\frac{d\vec{B}}{dt}.
\end{aligned}$$

The proof of (3.) is complete. I know some students don't care for the use of summations in calculus , but I would encourage such students to work this out without sums and reconsider your thinking. In the proof above you would have 6 products which would yield 12 terms and then those have to be rearranged to see the cross-products. Not that its impossible, or even too difficult, it's just that the summation notation is much cleaner. $\square$

In the prime notation the product rules for vector products are

$$(\vec{A}\bullet\vec{B})' = \vec{A}'\bullet\vec{B} + \vec{A}\bullet\vec{B}' \qquad (\vec{A}\times\vec{B})' = \vec{A}'\times\vec{B} + \vec{A}\times\vec{B}'.$$

We use these in future section to help uncover the geometry of curves.

Example 2.1.18. *The* **angular momentum** *of a mass m at $\vec{r}$ with velocity $\vec{v} = \frac{d\vec{r}}{dt}$ is defined by $\vec{L} = m\vec{r}\times\vec{v}$. If the net-force on m is $\vec{F}$ then $\vec{F} = m\vec{a}$ where $\vec{a} = \frac{d\vec{v}}{dt}$. If we define the* **torque** *produced by $\vec{F}$ on m as $\vec{\tau} = \vec{r}\times\vec{F}$. We study how torque relates to the change in angular momentum:*

$$\begin{aligned}
\frac{d\vec{L}}{dt} &= \frac{d}{dt}\left[m\vec{r}\times\vec{v}\right] \\
&= m\left[\frac{d\vec{r}}{dt}\times\vec{v} + \vec{r}\times\frac{d\vec{v}}{dt}\right] \\
&= m\vec{v}\times\vec{v} + \vec{r}\times m\vec{a} \\
&= \vec{r}\times\vec{F} \\
&= \vec{\tau}.
\end{aligned}$$

In particular, if $\vec{\tau} = 0$ then the angular momentum of m is constant. This happens in gravitational motion where the net-force is **central** *which means the direction of the force is colinear with $\vec{r}$ hence the torque is zero.*

Example 2.1.19. *Suppose $\vec{A}$ is a time-varying vector for which $A = 1$. We study how $\vec{A}$ and $\frac{d\vec{A}}{dt}$ relate geometrically. To accomplish this, we write $A^2 = 1$ in terms of dot-products; $1 = \vec{A}\bullet\vec{A}$ thus,*

$$\frac{d}{dt}(1) = \frac{d}{dt}[\vec{A}\bullet\vec{A}] \quad\Rightarrow\quad 0 = \frac{d\vec{A}}{dt}\bullet\vec{A} + \vec{A}\bullet\frac{d\vec{A}}{dt}.$$

Commutativity of the dot-product yields $\vec{A} \bullet \frac{d\vec{A}}{dt} = 0$. *Therefore, a unit-vector field will be orthogonal to its derivative. Or perhaps a more interesting point: to find a vector orthonormal to a given unit-vector-valued map* $t \mapsto \vec{A}(t)$ *simply normalize* $t \mapsto \frac{d\vec{A}}{dt}$.

The calculation in the above example is important to the study of the geometry of curves. We will encounter it in multiple instances as we study the Frenet apparatus for curves in $\mathbb{R}^3$.

Theorem 2.1.20. *chain rule of calculus for space curves.*

Let $g : K \subseteq \mathbb{R} \to \mathbb{R}$ be differentiable near t and $\vec{F} : J \subseteq \mathbb{R} \to \mathbb{R}^n$ be differentiable near $g(t)$ then near t we have

$$\frac{d}{dt}\big[\vec{F}(g(t))\big] = \frac{d\vec{F}}{dt}(g(t))\frac{dg}{dt}.$$

Proof: let $\vec{F} = \langle F_1, F_2, \dots, F_n\rangle$ and calculate,

$$\begin{aligned}
\frac{d}{dt}\big[\vec{F}(g(t))\big] &= \frac{d}{dt}\Big\langle\ F_1(g(t)),\ F_2(g(t)),\ \dots,\ F_n(g(t))\ \Big\rangle \\
&= \Big\langle\ \frac{d}{dt}F_1(g(t)),\ \frac{d}{dt}F_2(g(t)),\ \dots,\ \frac{d}{dt}F_n(g(t))\ \Big\rangle \\
&= \Big\langle\ \frac{dF_1}{dg}\frac{dg}{dt},\ \frac{dF_2}{dg}\frac{dg}{dt},\ \dots,\ \frac{dF_n}{dg}\frac{dg}{dt}\ \Big\rangle \\
&= \frac{dg}{dt}\Big\langle\ \frac{dF_1}{dg},\ \frac{dF_2}{dg},\ \dots,\ \frac{dF_n}{dg}\ \Big\rangle \\
&= \frac{dg}{dt}\frac{d\vec{F}}{dt}(g(t))
\end{aligned}$$

where I have used the notation $\frac{dF_j}{dg} = \frac{dF_j}{dt}(g(t))$ which you might recall from calculus I. As usual, the proof amounts to sorting through a little notation and quoting the basic result from calculus I. $\square$

We can use the notation $\frac{d\vec{F}}{dg}$ in place of the clumsy, but more technically accurate, $\frac{d\vec{F}}{dt}(g(t))$. With this notation the chain-rule looks nice:

$$\boxed{\frac{d}{dt}\big[\vec{F}(g(t))\big] = \frac{d\vec{F}}{dg}\frac{dg}{dt}.}$$

Example 2.1.21. *Let* $\vec{F} = \langle t^3, t^2, t\rangle$ *and* $g(t) = \cosh t$. *Note that:*

$$\vec{F}'(t) = \langle 3t^2, 2t, 1\rangle \qquad \& \qquad g'(t) = \sinh t$$

Thus,

$$\frac{d}{dt}\big[(\vec{F} \circ g)(t)\big] = \vec{F}'(g(t))g'(t) = \langle 3\sinh t\cosh^2 t, 2\sinh t\cosh t, \sinh t\rangle.$$

Of course, we would obtain the same result by noting $\vec{F}(g(t)) = \langle \cosh^3 t, \cosh^2 t, \cosh t\rangle$ *and differentiating component-wise.*

Some problems do not allow to take both paths. It is important to know the chain-rules both computationally and symbollically.

Example 2.1.22. *Suppose $\vec{F} : \mathbb{R} \to \mathbb{R}^3$ is differentiable g is also a differentiable function. If at $t = 1$ we have $g(1) = 3$ and $g'(1) = 7$ and $\vec{F}'(3) = \langle 1, 2, 3\rangle$ then find $(\vec{F} \circ g)'(1)$.*

Solution: *simply apply the chain-rule:*

$$(\vec{F} \circ g)'(1) = \vec{F}'(g(1))g'(1) = \vec{F}'(3)g'(1) = 7\langle 1, 2, 3\rangle = \langle 7, 14, 21\rangle.$$

We could not just plug in the formulas and differentiate because only partial information was given. Often in real world scenarios you have incomplete data as in this example. It is important to know how to relate changes even in the absence of complete formulaic relations.

2.2 geometry of smooth oriented curves

If the curve[5] is assigned a sense of direction then we call it an **oriented curve**. A particular curve can be parametrized by many different paths. You can think of a parametrization of a curve as a process of pasting a flexible numberline onto the curve.

Definition 2.2.1.

Let $C \subseteq \mathbb{R}^n$ be an oriented curve which starts at P and ends at Q. We say that $\vec{\gamma} : [a, b] \to \mathbb{R}^n$ is a **smooth non-stop parametrization** of C if $\vec{\gamma}([a, b]) = C$, $\vec{\gamma}(a) = P$, $\vec{\gamma}(b) = Q$, and $\vec{\gamma}$ is smooth with $\vec{\gamma}\,'(t) \neq 0$ for all $t \in [a, b]$. We will typically call $\vec{\gamma}$ a **path** from P to Q which covers the curve C.

I have limited the definition to curves with endpoints however the definition for curves which go on without end is very similar. You can just drop one or both of the endpoint conditions.

2.2.1 arclength

Let's begin by analyzing the tangent vector to a path in three dimensional space. Denote $\vec{\gamma} = (x, y, z)$ where $x, y, z \in C^\infty([a, b], \mathbb{R})$ and calculate that

$$\vec{\gamma}\,'(t) = \tfrac{d\vec{\gamma}}{dt} = \langle \tfrac{dx}{dt}, \tfrac{dy}{dt}, \tfrac{dz}{dt}\rangle.$$

Multiplying by dt yields

$$\vec{\gamma}\,'(t)dt = \tfrac{d\vec{\gamma}}{dt}dt = \langle \tfrac{dx}{dt}, \tfrac{dy}{dt}, \tfrac{dz}{dt}\rangle dt.$$

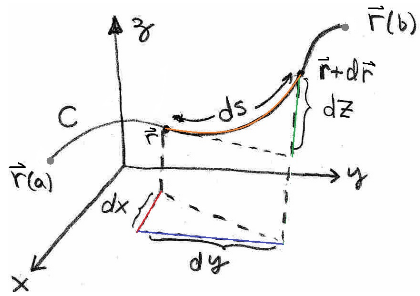
The arclength ds subtended from time t to time $t + dt$ is simply the length of the vector $\vec{\gamma}\,'(t)dt$ which yields,

$$ds = ||\vec{\gamma}\,'(t)dt|| = \sqrt{\tfrac{dx}{dt}^2 + \tfrac{dy}{dt}^2 + \tfrac{dz}{dt}^2}\,dt$$

You can think of this as the length of a tiny bit of string that is laid out along the curve from the point $\vec{\gamma}(t)$ to the point $\vec{\gamma}(t + dt)$.

---

[5] you will notice that I begin using $\vec{\gamma}$ to denote the studied curve in this section, if it helps you feel free to just set $\vec{\gamma} = \vec{r}$. I oscillate between those notations in the remainder of this chapter and course.



Of course this infinitesimal notation is just shorthand for an explicit limiting processes. If we sum together all the little bits of arclength we will arrive at the total arclength of the curve. In fact, this is how we define the arclength of a curve. The preceding discussion was in 3 dimensions but the formulas stated in terms of the norm generalizes naturally to $\mathbb{R}^n$.

Definition 2.2.2.

Let $\vec{\gamma} : [a, b] \to \mathbb{R}^n$ be a smooth, non-stop path which covers the oriented curve C. The **arclength function** of $\vec{\gamma}$ is a function $s_{\vec{\gamma}} : [a, b] \to \mathbb{R}$ where

$$s_{\vec{\gamma}} = \int_a^t ||\vec{\gamma}\,'(u)||\, du$$

for each $t \in [a, b]$. If $\tilde{\gamma}$ is a smooth non-stop path such that $||\tilde{\gamma}'(t)|| = 1$ then we say that $\tilde{\gamma}$ is a unit-speed curve. Moreover, we say $\tilde{\gamma}$ is parametrized with respect to arclength.

The examples below illustrate how we calculate arclength and also, for reasonable arclength functions, how we can explicitly reparametrize the path with respect to arclength. Sorry the notation in the examples below does not match the definition above. The connection is simple though, just think $\vec{r} = \vec{\gamma}$. This notational divide continues throughout my work, I sometimes use $\vec{r}$ for a path and sometimes $\vec{\gamma}$. Sometimes, I'll use another letter. Context is important and this is one of the reasons it is important to declare the domain and range for functions in this course. If we declare the domain and target spaces then the letter need not confuse us.

Example 2.2.3. *Consider $\vec{r}(t) = \langle R\cos t, R\sin t\rangle$ for $R > 0$. Note $\frac{d\vec{r}}{dt} = \langle -R\sin t, R\cos t\rangle$ hence $||\frac{d\vec{r}}{dt}|| = R$. If we wish to parametrize a circle then $0 \le t \le 2\pi$ is a natural choice. Hopefully this arclength result is familar:*

$$s = \int_0^{2\pi} R\, dt = Rt\Big|_0^{2\pi} = 2\pi R.$$

The arclength function based at $t = 0$ is given by $s(t) = Rt$.

Example 2.2.4. *A helix with slope b and radius a is given by:*

$$\vec{r}(t) = \langle a\cos t, a\sin t, bt\rangle.$$

If we restrict $0 \le t \le 2\pi$ then we obtain one complete spiral. Note,

$$\vec{r}\,'(t) = \langle -a\sin t, a\cos t, b\rangle.$$

hence

$$||\vec{r}\,'(t)|| = \sqrt{a^2\sin^2 t + a^2\cos^2 t + b^2} = \sqrt{a^2 + b^2}.$$

thus the arclength function based at $t = 0$ is given by:

$$s(t) = \int_0^t \sqrt{a^2 + b^2}du = t\sqrt{a^2 + b^2}.$$

We may note that the example above reverts to a circle in the plane if we set $b = 0$. It is often informative to consider the case $b \to 0$ as we study such a helix.

Example 2.2.5. *Find the* **arc-length** *parametrization for the helix with slope b and radius a twisting about the z-axis. This is accomplished by solving the arclength function $s(t)$ for t and thus replacing t with formulas involving the arclength. In particular, we found in the previous example*[6] $s = t\sqrt{a^2+b^2}$ *thus* $t = \frac{s}{\sqrt{a^2+b^2}}$ *hence:*

$$\vec{r}(s) = \left\langle a\cos\left(\frac{s}{\sqrt{a^2+b^2}}\right), a\sin\left(\frac{s}{\sqrt{a^2+b^2}}\right), \frac{sb}{\sqrt{a^2+b^2}}\right\rangle.$$

Technically, we ought to write $\vec{r}(t(s))$ but it is a common abuse of notation to use the same symbol to denote the curve after a **reparametrization**. *The* **parametrization** *we considered here is called the* **arclength** *or* **unit-speed** *parametrization.*

Explicit reparametrization of the curve below with respect to arclength is not a simple task. You'd likely need to break into cases.

Example 2.2.6. *Let* $\vec{r}(t) = \langle t, \frac{\sqrt{2}}{2}t^2, \frac{1}{3}t^3\rangle$ *for* $t \geq 0$. *We find* $\vec{r}\,'(t) = \langle 1, t\sqrt{2}, t^2\rangle$ *hence*

$$||\vec{r}\,'(t)|| = \sqrt{1+2t^2+t^4} = \sqrt{(1+t^2)^2} = 1+t^2.$$

Therefore, the arclength function is derived:

$$s(t) = \int_0^t (1+u^2)du = t + \tfrac{1}{3}t^3.$$

It is possible to solve $s = t + t^3/3$ for t, but, perhaps this example helps you appreciate that there exist curves where the reparametrization with respect to arclength will be impossible to implement with elementary functions. There simply is no guarantee that the arclength integral is tractable. On the other hand, the existence of the arclength function is not in jeapordy. This is somewhat like the situtation we faced in integral calculus. The FTC showed that mere continuity sufficed to give the exitence of an antiderivative, yet, the examples where we can explicitly calculate the antiderivative in terms of elementary functions are precious and somewhat rare in the space of all possible examples[7]

The arclength function has many special properties. Notice that item (1.) reduces to the statement that the speed is the magnitude of the velocity vector if we set $w = t$.

Proposition 2.2.7.

Let $\vec{\gamma} : [a,b] \to \mathbb{R}^n$ be a smooth, non-stop path which covers the oriented curve C. The **arclength function** of $\vec{\gamma}$ denoted by $s_{\vec{\gamma}} : [a,b] \to \mathbb{R}$ has the following properties:

1. $\frac{d}{dt}(s_{\vec{\gamma}}(w)) = ||\vec{\gamma}\,'(w)||\frac{dw}{dt}$,
2. $\frac{ds_{\vec{\gamma}}}{dt} > 0$ for all $t \in (a,b)$,
3. $s_{\vec{\gamma}}$ is a 1-1 function,
4. $s_{\vec{\gamma}}$ has inverse $s_{\vec{\gamma}}^{-1} : s_{\vec{\gamma}}([a,b]) \to [a,b]$.

[6] I wrote $s(t)$ in the previous example to emphasize that s is a function of t, that is **not** multiplication by t.

[7] you can read about Robert Risch's algorithm concerning the existence of nice antiderivatives. See http://math.stackexchange.com/a/5723/36530 for some nice pointers on where to read. I may have some discussion of this in Math 422 of Spring 2017 when we discuss differential Galois theory briefly.

Proof: We begin with (1.). We apply the fundamental theorem of calculus:

$$\frac{d}{dt}(s_{\vec{\gamma}}(w)) = \frac{d}{dt}\int_a^w ||\vec{\gamma}\,'(u)||\,du = ||\vec{\gamma}\,'(w)||\frac{dw}{dt}$$

for all $w \in (a,b)$. For (2.), set $w = t$ and recall that $||\vec{\gamma}\,'(t)|| = 0$ iff $\vec{\gamma}\,'(t) = 0$ however we were given that $\vec{\gamma}$ is non-stop so $\vec{\gamma}\,'(t) \neq 0$. We find $\frac{ds_{\vec{\gamma}}}{dt} > 0$ for all $t \in (a,b)$ and consequently the arclength function is an increasing function on (a,b). For (3.), suppose (towards a contradiction) that $s_{\vec{\gamma}}(x) = s_{\vec{\gamma}}(y)$ where $a < x < y < b$. Note that $\vec{\gamma}$ smooth implies $s_{\vec{\gamma}}$ is differentiable with continuous derivative on (a,b) therefore the mean value theorem applies and we can deduce that there is some point on $c \in (x,y)$ such that $s'_{\vec{\gamma}}(c) = 0$, which is impossible, therefore (3.) follows. If a function is 1-1 then we can construct the inverse pointwise by simply going backwards for each point mapped to in the range; $s_{\vec{\gamma}}^{-1}(x) = y$ iff $s_{\vec{\gamma}}(y) = x$. The fact that $s_{\vec{\gamma}}$ is single-valued follows from (3.). $\square$

If we are given a curve C covered by a path $\vec{\gamma}$ (which is smooth and non-stop but may not be unit-speed) then we can reparametrize the curve C with a unit-speed path $\tilde{\gamma}$ as follows:

$$\boxed{\tilde{\gamma}(s) = \vec{\gamma}(s_{\vec{\gamma}}^{-1}(s))}$$

where $s_{\vec{\gamma}}^{-1}$ is the inverse of the arclength function.

Proposition 2.2.8.

If $\vec{\gamma}$ is a smooth non-stop path then the path $\tilde{\gamma}$ defined by $\tilde{\gamma}(s) = \vec{\gamma}(s_{\vec{\gamma}}^{-1}(s))$ is unit-speed.

Proof: Differentiate $\tilde{\gamma}(t)$ with respect to t, we use the chain-rule,

$$\tilde{\gamma}'(t) = \tfrac{d}{dt}(\vec{\gamma}(s_{\vec{\gamma}}^{-1}(t))) = \vec{\gamma}\,'(s_{\vec{\gamma}}^{-1}(t))\tfrac{d}{dt}(s_{\vec{\gamma}}^{-1}(t)).$$

Hence $\tilde{\gamma}'(t) = \vec{\gamma}\,'(s_{\vec{\gamma}}^{-1}(t))\frac{d}{dt}(s_{\vec{\gamma}}^{-1}(t))$. Recall that if a function is increasing on an interval then its inverse is likewise increasing hence, by (2.) of the previous proposition, we can pull the positive constant $\frac{d}{dt}(s_{\vec{\gamma}}^{-1}(t))$ out of the norm. We find, using item (1.) in the previous proposition,

$$||\tilde{\gamma}'(t)|| = ||\vec{\gamma}\,'(s_{\vec{\gamma}}^{-1}(t))||\tfrac{d}{dt}(s_{\vec{\gamma}}^{-1}(t)) = \tfrac{d}{dt}(s_{\vec{\gamma}}(s_{\vec{\gamma}}^{-1}(t))) = \tfrac{d}{dt}(t) = 1.$$

Therefore, the curve $\tilde{\gamma}$ is unit-speed. $\square$

The notation relating speed and arclength is consistent with the above proposition. Generally the **speed** of the curve is $\frac{ds}{dt}$ so when $t = s$ we have $\frac{ds}{ds} = 1$.

Remark 2.2.9.

While there are many paths which cover a particular oriented curve the unit-speed path is unique and we'll see that the Frenet-Serret equations for unit-speed curves are particularly simple.

Example 2.2.10. *Let $\vec{r}(t) = \langle 1 + ae^t, 2 + be^t, 3 + ce^t\rangle$ for $0 \leq t \leq \ln 2$. Find the unit-speed parametrization of this curve. Also, identify the curve.*

Solution: *Note $\vec{r}\,'(t) = \langle ae^t, be^t, ce^t\rangle = e^t\langle a, b, c\rangle$ thus we find speed:*

$$||\vec{r}\,'(t)|| = e^t\sqrt{a^2 + b^2 + c^2}$$

Therefore, $s(t) = \int_0^t e^u\sqrt{a^2 + b^2 + c^2}\,du = (e^t - 1)\sqrt{a^2 + b^2 + c^2}$. We should solve $s = (e^t - 1)\sqrt{a^2 + b^2 + c^2}$ for t:

$$\frac{s}{\sqrt{a^2 + b^2 + c^2}} + 1 = e^t \quad \Rightarrow \quad t = \ln\left(\frac{s}{\sqrt{a^2 + b^2 + c^2}} + 1\right).$$

We find (after a bit of algebra)

$$\vec{r}(s) = \left\langle 1 + \frac{as}{\sqrt{a^2 + b^2 + c^2}} + a, 2 + \frac{bs}{\sqrt{a^2 + b^2 + c^2}} + b, 3 + \frac{cs}{\sqrt{a^2 + b^2 + c^2}} + c\right\rangle$$

which we could write as:

$$\vec{r}(s) = (1 + a, 2 + b, 3 + c) + \frac{s}{\sqrt{a^2 + b^2 + c^2}}\langle a, b, c\rangle.$$

This is a line with base-point $(1 + a, 2 + b, 3 + c)$ and direction $\frac{1}{\sqrt{a^2+b^2+c^2}}\langle a, b, c\rangle$.

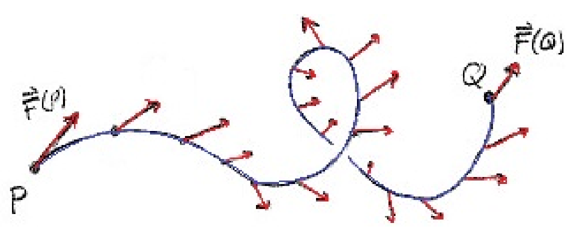
If we are free of the unit-speed goal there are simple ways to deal with examples like the one above. I would just set $\lambda = e^t$ as to reparametrize the curve to $\vec{r}(\lambda) = (1, 2, 3) + \lambda\langle a, b, c\rangle$ for $1 \leq \lambda \leq 2$. Again, it is clear that we have a line in the $\langle a, b, c\rangle$ direction.

2.2.2 vector fields along a path

Definition 2.2.11.

Let $C \subseteq \mathbb{R}^3$ be an oriented curve which starts at P and ends at Q. A **vector field along the curve** C is a function which attaches a vector to each point on C.

There are many different vector fields we could attach to a given curve. You can easily imagine attaching different vectors than the ones pictured below:



The tangent ($\vec{T}$), normal($\vec{N}$) and binormal ($\vec{B}$) vector fields defined below will allow us to identify when two oriented curves have the same shape. Note, we say a path is **regular** if it has a derivative which is nonzero everywhere. We say a path is **non-linear** if the path is not a line. Regularity is needed to give a well-defined tangent vector field along a curve. Likewise, lines must be avoided since $\vec{N}$ is not well-defined for $\vec{T}\,' = 0$. Lines are a special case and we don't need the machine described in this section to analyze lines. We've already seen vector algebra is quite sufficient to understand the geometry of lines.

Definition 2.2.12.

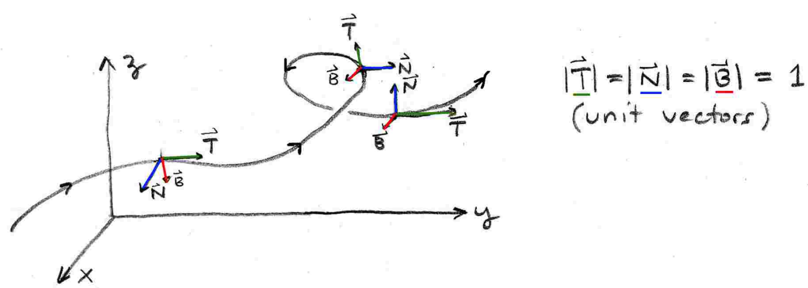
Let $\vec{\gamma} : [a, b] \rightarrow \mathbb{R}^3$ be a path from P to Q in $\mathbb{R}^3$ which is non-linear and regular. The **tangent vector field** of $\vec{\gamma}$ is given by

$$\vec{T}(t) = \frac{1}{||\vec{\gamma}\,'(t)||}\vec{\gamma}\,'(t)$$

for each $t \in [a, b]$. Likewise, if $\vec{T}\,'(t) \neq 0$ for all $t \in [a, b]$ then the **normal vector field** of $\vec{\gamma}$ is defined by

$$\vec{N}(t) = \frac{1}{||\vec{T}\,'(t)||}\vec{T}\,'(t)$$

for each $t \in [a, b]$. Finally, if $\vec{T}_{,}'(t) \neq 0$ for all $t \in [a, b]$ then the **binormal vector field** of $\vec{\gamma}$ is defined by $\vec{B}(t) = \vec{T}(t) \times \vec{N}(t)$ for all $t \in [a, b]$



Example 2.2.13. *Let $R > 0$ and suppose $\vec{\gamma}(t) = (R\cos(t), R\sin(t), 0)$ for $0 \leq t \leq 2\pi$. We can calculate*

$$\vec{\gamma}\,'(t) = \langle -R\sin(t), R\cos(t), 0\rangle \quad \Rightarrow \quad ||\vec{\gamma}\,'(t)|| = R.$$

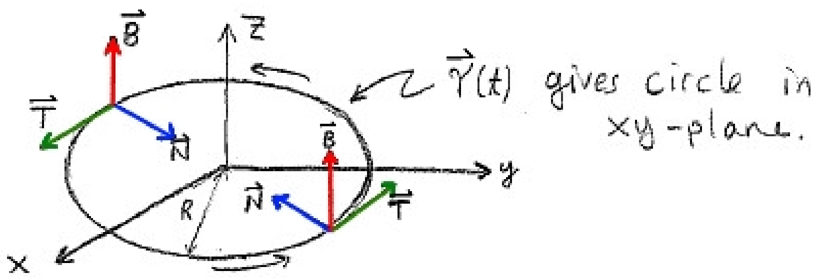
Hence $\vec{T}(t) = \langle -\sin(t), \cos(t), 0\rangle$ and we can calculate,

$$\vec{T}\,'(t) = \langle -\cos(t), -\sin(t), 0\rangle \quad \Rightarrow \quad ||\vec{T}\,'(t)|| = 1.$$

Thus $\vec{N}(t) = \langle -\cos(t), -\sin(t), 0\rangle$. Finally we calculate the binormal vector field,

$$\begin{aligned}
\vec{B}(t) = \vec{T}(t) \times \vec{N}(t) \quad &= [-\sin(t)e_1 + \cos(t)e_2] \times [-\cos(t)e_1 - \sin(t)e_2] \\
&= [\sin^2(t)e_1 \times e_2 - \cos^2(t)e_2 \times e_1 \\
&= [\sin^2(t) + \cos^2(t)]e_1 \times e_2 \\
&= e_3 = \langle 0, 0, 1\rangle
\end{aligned}$$

Notice that $\vec{T} \bullet \vec{N} = \vec{N} \bullet \vec{B} = \vec{T} \bullet \vec{B} = 0$. For a particular value of t the vectors $\{\vec{T}(t), \vec{N}(t), \vec{B}(t)\}$ give an orthogonal set of unit vectors, they provide a comoving frame for $\vec{\gamma}$. It can be shown that the tangent and normal vectors span the plane in which the path travels for times infintesimally close to t. This plane is called the **osculating plane**. *The binormal vector gives the normal to the osculating plane. The curve considered in this example has a rather boring osculating plane since $\vec{B}$ is constant. This curve is just a circle in the xy-plane which is traversed at constant speed.*



Example 2.2.14. *Notice that $s_{\vec{\gamma}}(t) = Rt$ in the preceding example. It follows that $\widetilde{\gamma}(s) = (R\cos(s/R), R\sin(s/R), 0)$ for $0 \leq s \leq 2\pi R$ is the unit-speed path for curve. We can calculate*

$$\widetilde{\gamma}'(s) = \langle -\sin(s/R), \cos(s/R), 0\rangle \quad \Rightarrow \quad ||\widetilde{\gamma}'(s)|| = 1.$$

Hence $\widetilde{T}(s) = \langle -\sin(s/R), \cos(s/R), 0\rangle$ and we can also calculate,

$$\widetilde{T}'(s) = \tfrac{1}{R}\langle -\cos(s/R), -\sin(s/R), 0\rangle \quad \Rightarrow \quad ||\widetilde{T}'(s)|| = 1/R.$$

Thus $\widetilde{N}(s) = \langle -\cos(s/R), -\sin(s/R), 0\rangle$. Note $\widetilde{B} = \widetilde{T} \times \widetilde{N} = \langle 0, 0, 1\rangle$ as before.

Example 2.2.15. *Let $m, R > 0$ and suppose $\vec{\gamma}(t) = (R\cos(t), R\sin(t), mt)$ for $0 \leq t \leq 2\pi$. We can calculate*

$$\vec{\gamma}\,'(t) = \langle -R\sin(t), R\cos(t), m\rangle \quad \Rightarrow \quad ||\vec{\gamma}\,'(t)|| = \sqrt{R^2 + m^2}.$$

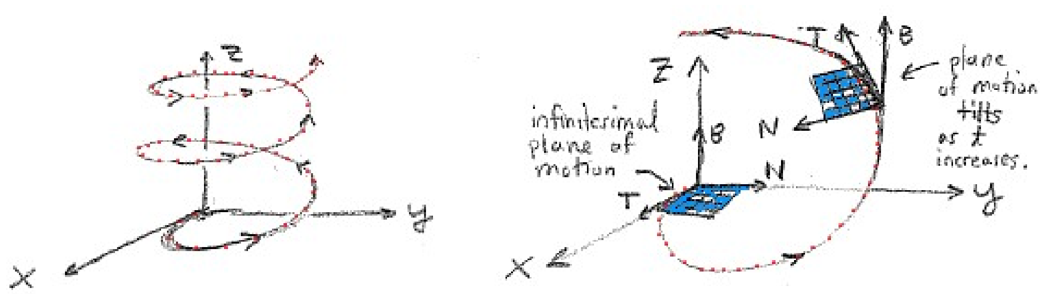
Hence $\vec{T}(t) = \frac{1}{\sqrt{R^2+m^2}}\langle -R\sin(t), R\cos(t), m\rangle$ and we can calculate,

$$\vec{T}\,'(t) = \tfrac{1}{\sqrt{R^2+m^2}}\langle -R\cos(t), -R\sin(t), 0\rangle \quad \Rightarrow \quad ||\vec{T}\,'(t)|| = \tfrac{R}{\sqrt{R^2+m^2}}.$$

Thus $\vec{N}(t) = \langle -\cos(t), -\sin(t), 0\rangle$. Finally we calculate the binormal vector field,

$$\begin{aligned} \vec{B}(t) = \vec{T}(t) \times \vec{N}(t) &= \tfrac{1}{\sqrt{R^2+m^2}}[-R\sin(t)e_1 + R\cos(t)e_2 + me_3] \times [-\cos(t)e_1 - \sin(t)e_2] \\ &= \tfrac{1}{\sqrt{R^2+m^2}}\langle m\sin(t), -m\cos(t), R\rangle \end{aligned}$$

We again observe that $\vec{T} \bullet \vec{N} = \vec{N} \bullet \vec{B} = \vec{T} \bullet \vec{B} = 0$. The **osculating plane** *is moving for this curve, note the t-dependence. This curve does not stay in a single plane, it is not a planar curve. In fact this is a circular helix with radius R and slope m.*



Example 2.2.16. *Lets reparametrize the helix as a unit-speed path. Notice that* $s_{\vec{\gamma}}(t) = t\sqrt{R^2+m^2}$ *thus we should replace* t *with* $s/\sqrt{R^2+m^2}$ *to obtain* $\widetilde{\gamma}(s)$*. Let* $a = 1/\sqrt{R^2+m^2}$ *and* $\widetilde{\gamma}(s) = (R\cos(as), R\sin(as), ams)$ *for* $0 \leq s \leq 2\pi\sqrt{R^2+m^2}$*. We can calculate*

$$\widetilde{\gamma}'(s) = \langle -Ra\sin(as), Ra\cos(as), am\rangle \quad \Rightarrow \quad ||\widetilde{\gamma}'(s)|| = a\sqrt{R^2+m^2} = 1.$$

Hence $\widetilde{T}(s) = a\langle -R\sin(as), R\cos(as), m\rangle$ *and we can calculate,*

$$\widetilde{T}'(s) = Ra^2\langle -\cos(as), -\sin(as), 0\rangle \quad \Rightarrow \quad ||\widetilde{T}'(s)|| = Ra^2 = \tfrac{R}{R^2+m^2}.$$

Thus $\widetilde{N}(s) = \langle -\cos(as), -\sin(as), 0\rangle$*. Next, calculate the binormal vector field,*

$$\begin{aligned} \widetilde{B}(s) = \widetilde{T}(s) \times \widetilde{N}(s) &= a\langle -R\sin(as), R\cos(as), m\rangle \times \langle -\cos(as), -\sin(as), 0\rangle \\ &= \tfrac{1}{\sqrt{R^2+m^2}}\langle m\sin(as), -m\cos(as), R\rangle \end{aligned}$$

Hopefully you can start to see that the unit-speed path shares the same $\vec{T}, \vec{N}, \vec{B}$ *frame at arclength* s *as the previous example with* $t = s/\sqrt{R^2+m^2}$*.*

2.2.3 Frenet Serret equations

We now prepare to prove the Frenet Serret formulas for the $\vec{T}, \vec{N}, \vec{B}$ frame fields. It turns out that for nonlinear curves the $\vec{T}, \vec{N}, \vec{B}$ vector fields always provide an orthonormal frame. Moreover, for nonlinear curves, we'll see that the **torsion** and **curvature** capture the geometry of the curve.

Proposition 2.2.17.

If $\vec{\gamma}$ is a path with tangent, normal and binormal vector fields $\vec{T}, \vec{N}$ and $\vec{B}$ then $\{\vec{T}(t), \vec{N}(t), \vec{B}(t)\}$ is an orthonormal set of vectors for each $t \in dom(\vec{\gamma})$.

Proof: It is clear from $\vec{B}(t) = \vec{T}(t) \times \vec{N}(t)$ that $\vec{T}(t) \bullet \vec{B}(t) = \vec{N}(t) \bullet \vec{B}(t) = 0$. Furthermore, it is also clear that these vectors have length one due to their construction as unit vectors. In particular this means that $\vec{T}(t) \bullet \vec{T}(t) = 1$. We can differentiate this to obtain (by the product rule for dot-products)

$$\vec{T}\,'(t) \bullet \vec{T}(t) + \vec{T}(t) \bullet \vec{T}\,'(t) = 0 \quad \Rightarrow \quad 2\vec{T}(t) \bullet \vec{T}\,'(t) = 0$$

Divide by $||\vec{T}\,'(t)||$ to obtain $\vec{T}(t) \bullet \vec{N}(t) = 0$. $\square$

We omit the explicit t-dependence for the dicussion to follow here, also you should assume the vector fields are all derived from a particular path $\vec{\gamma}$. Since $\vec{T}, \vec{N}, \vec{B}$ are nonzero and point in three mutually distinct directions it follows that any other vector can be written as a linear combination of $\vec{T}, \vec{N}, \vec{B}$. This means[8] if $\vec{v}$ is a vector then there exist c_1, c_2, c_3 such that $v = c_1T + c_2N + c_3B$. The orthonormality is very nice because it tells us we can calculate the coefficients in terms of dot-products with $\vec{T}, \vec{N}$ and $\vec{B}$:

$$\boxed{\vec{v} = c_1\vec{T} + c_2\vec{N} + c_3\vec{B} \quad \Rightarrow \quad c_1 = \vec{v} \bullet \vec{T},\ c_2 = \vec{v} \bullet \vec{N},\ c_3 = \vec{v} \bullet \vec{B}}$$

We will make much use of the observations above in the calculations that follow. Suppose that

$$\begin{aligned} \vec{T}\,' &= c_{11}\vec{T} + c_{12}\vec{N} + c_{13}\vec{B} \\ \vec{N}\,' &= c_{21}\vec{T} + c_{22}\vec{N} + c_{23}\vec{B} \\ \vec{B}\,' &= c_{31}\vec{T} + c_{32}\vec{N} + c_{33}\vec{B}. \end{aligned}$$

[8] You might recognize $[v]_\beta = [c_1, c_2, c_3]^T$ as the coordinate vector with respect to the basis $\beta = \{\vec{T}, \vec{N}, \vec{B}\}$

We observed previously that $\vec{T}\,' \bullet \vec{T} = 0$ thus $c_{11} = 0$. It is easy to show $\vec{N}\,' \bullet \vec{N} = 0$ and $\vec{B}\,' \bullet \vec{B} = 0$ thus $c_{22} = 0$ and c_{33}. Furthermore, we defined $\vec{N} = \frac{1}{||\vec{T}\,'||}\vec{T}\,'$ hence $c_{13} = 0$. Note that

$$\vec{T}\,' = c_{12}\vec{N} = \tfrac{c_{12}}{||\vec{T}\,'||}\vec{T}\,' \quad \Rightarrow \quad c_{12} = ||\vec{T}\,'||.$$

To summarize what we've learned so far:

$$\begin{aligned} \vec{T}\,' &= c_{12}\vec{N} \\ \vec{N}\,' &= c_{21}\vec{T} + c_{23}\vec{B} \\ \vec{B}\,' &= c_{31}\vec{T} + c_{32}\vec{N}. \end{aligned}$$

We'd like to find some condition on the remaining coefficients. Consider that:

$$\begin{array}{lll} \vec{B} = \vec{T} \times \vec{N} & \Rightarrow \quad \vec{B}\,' = \vec{T}\,' \times \vec{N} + \vec{T} \times \vec{N}\,' & \text{a product rule} \\ & \Rightarrow \quad \vec{B}\,' = [c_{12}\vec{N}] \times \vec{N} + \vec{T} \times [c_{21}\vec{T} + c_{23}\vec{B}] & \text{using previous eqn.} \\ & \Rightarrow \quad \vec{B}\,' = c_{23}\vec{T} \times \vec{B} & \text{noted } \vec{N} \times \vec{N} = \vec{T} \times \vec{T} = 0 \\ & \Rightarrow \quad \vec{B}\,' = -c_{23}\vec{N} & \text{you can show } \vec{N} = \vec{B} \times \vec{T}. \\ & \Rightarrow \quad c_{31}\vec{T} + c_{32}\vec{N} = -c_{23}\vec{N} & \text{refer to previous eqn.} \\ & \Rightarrow \quad c_{31} = 0 \text{ and } c_{32} = -c_{23}. & \text{using LI of } \{T, N\} \end{array}$$

The "LI" is linear independence. The fact that $\vec{T}, \vec{N}$ are LI follows from the fact that they form a nonzero and orthogonal set of vectors. We can equate coefficients of LI sums of vectors. This is the principle I'm using. Alternatively, you can just take the dot-product of the next to last equation with $\vec{N}$ and then $\vec{T}$ and use $\vec{T} \bullet \vec{N} = 0$ to obtain the final line. We have reduced the initial set of equations to the following:

$$\begin{aligned} \vec{T}\,' &= c_{12}\vec{N} \\ \vec{N}\,' &= c_{21}\vec{T} + c_{23}\vec{B} \\ \vec{B}\,' &= -c_{23}\vec{N}. \end{aligned}$$

The equations above encourage us to define the **curvature** and **torsion** as follows:

Definition 2.2.18.

Let C be a curve which is covered by the unit-speed path $\widetilde{\gamma}$ then we define the curvature κ and torsion τ as follows:

$$\boxed{\kappa(s) = \left|\left|\frac{d\widetilde{T}}{ds}\right|\right| \qquad \tau(s) = -\frac{d\widetilde{B}}{ds} \bullet \widetilde{N}(s)}$$

I leave it to the reader[9] to show that $c_{21} = -c_{12}$. Given that result we find the famous **Frenet-Serret** equations:

$$\boxed{\tfrac{d\widetilde{T}}{ds} = \kappa\widetilde{N} \qquad \tfrac{d\widetilde{N}}{ds} = -\kappa\widetilde{T} + \tau\widetilde{B} \qquad \tfrac{d\widetilde{B}}{ds} = -\tau\widetilde{N}.}$$

We had to use the arclength parameterization to insure that the formulas above unambiguously define the curvature and the torsion. In fact, if we take a particular (unoriented) curve then there are two choices for orienting the curve. You can show that that the torsion and curvature are independent of the choice of orientation. Naturally the total arclength is also independent of the orientation of a given curve.

[9]there is much more to read about this and the generalization to surfaces on my website. See the materials I posted on Elementary Differential Geometry. Alternatively, get a copy of Barrett Oneiil's excellent text. All you need is calculus III to get started on classical elementary differential geometry.

Example 2.2.19. *We found in Example 2.2.14 a unit-speed parametrization of the circle of radius R is $\widetilde{\gamma}(s) = (R\cos(s/R), R\sin(s/R), 0)$ for $0 \leq s \leq 2\pi R$. Moreover, the Frenet frame in terms of arclength was derived:*

$$\widetilde{T}(s) = \langle -\sin(s/R), \cos(s/R), 0\rangle \qquad \widetilde{N}(s) = \langle -\cos(s/R), -\sin(s/R), 0\rangle, \qquad \widetilde{B} = \langle 0, 0, 1\rangle$$

The curvature is the magnitude the derivative of the tangent:

$$\frac{d\widetilde{T}}{ds} = \frac{1}{R}\langle -\cos(s/R), -\sin(s/R), 0\rangle \quad \Rightarrow \quad \kappa = \frac{1}{R}.$$

In contrast, the torsion $\tau = 0$ as $\frac{d\widetilde{B}}{ds} = 0$.

Curvature, torsion can also be calculated in terms of a path which is not unit speed. We simply replace s with the arclength function $s_{\vec{\gamma}}(t)$ and make use of the chain rule. Notice that $d\vec{F}/dt = (ds/dt)(d\widetilde{F}/ds)$ hence,

$$\tfrac{d\vec{T}}{dt} = \tfrac{ds}{dt}\tfrac{d\widetilde{T}}{ds}, \quad \tfrac{d\vec{N}}{dt} = \tfrac{ds}{dt}\tfrac{d\widetilde{N}}{ds}, \quad \tfrac{d\vec{B}}{dt} = \tfrac{ds}{dt}\tfrac{d\widetilde{B}}{ds}$$

Or if you prefer, use the dot-notation $ds/dt = \dot{s}$ to write:

$$\tfrac{1}{\dot{s}}\tfrac{d\vec{T}}{dt} = \tfrac{d\widetilde{T}}{ds}, \quad \tfrac{1}{\dot{s}}\tfrac{d\vec{N}}{dt} = \tfrac{d\widetilde{N}}{ds}, \quad \tfrac{1}{\dot{s}}\tfrac{d\vec{B}}{dt} = \tfrac{d\widetilde{B}}{ds}$$

Substituting these into the unit-speed Frenet Serret formulas yield:

$$\boxed{\tfrac{d\vec{T}}{dt} = \dot{s}\kappa\vec{N} \qquad \tfrac{d\vec{N}}{dt} = -\dot{s}\kappa\vec{T} + \dot{s}\tau\vec{B} \qquad \tfrac{d\vec{B}}{dt} = -\dot{s}\tau\vec{N}.}$$

where $\widetilde{T}(s_{\vec{\gamma}}(t)) = \vec{T}(t), \widetilde{N}(s_{\vec{\gamma}}(t)) = \vec{N}(t)$ and $\widetilde{B}(s_{\vec{\gamma}}(t)) = \vec{B}(t)$. Likewise deduce[10] that

$$\kappa(t) = \frac{1}{\dot{s}}\left|\left|\frac{d\vec{T}}{dt}\right|\right| \qquad \tau(t) = -\frac{1}{\dot{s}}\left(\frac{d\vec{B}}{dt} \bullet \vec{N}(t)\right)$$

Let's see how these formulas are useful in an example or two.

Example 2.2.20. *Returning once more to Example 2.2.15 where we found the helix with radius R and slope m centered about the z-axis as parametrized by:*

$$\vec{\gamma}(t) = (R\cos(t), R\sin(t), mt)$$

for $t \geq 0$ has speed $\dot{s} = \sqrt{R^2 + m^2}$ and the Frenet frame:

$$\begin{aligned}
\vec{T}(t) &= \tfrac{1}{\sqrt{R^2+m^2}}\langle -R\sin(t), R\cos(t), m\rangle, \\
\vec{N}(t) &= \langle -\cos(t), -\sin(t), 0\rangle, \\
\vec{B}(t) &= \tfrac{1}{\sqrt{R^2+m^2}}\langle m\sin(t), -m\cos(t), R\rangle
\end{aligned}$$

Calculate,

$$\frac{d\vec{T}}{dt} = \frac{1}{\sqrt{R^2+m^2}}\langle -R\cos(t), -R\sin(t), 0\rangle$$

[10] I'm using the somewhat ambiguous notation $\kappa(t) = \kappa(s_\gamma(t))$ and $\tau(t) = \tau(s_\gamma(t))$. We do this often in applications of calculus. Ask me if you'd like further clarification on this point.

Consequently,

$$\left|\left|\frac{d\vec{T}}{dt}\right|\right| = \frac{R}{\sqrt{R^2+m^2}} \quad \Rightarrow \quad \kappa(t) = \frac{1}{\dot{s}}\left|\left|\frac{d\vec{T}}{dt}\right|\right| = \frac{1}{\sqrt{R^2+m^2}}\frac{R}{\sqrt{R^2+m^2}} \quad \Rightarrow \quad \boxed{\kappa = \frac{R}{R^2+m^2}}$$

The calculation of torsion is similar, the formula $\tau(t) = (-1/\dot{s})\left(\frac{d\vec{B}}{dt} \bullet \vec{N}\right)$ *guides us:*

$$\frac{d\vec{B}}{dt} = \frac{1}{\sqrt{R^2+m^2}}\langle m\cos(t), m\sin(t), 0\rangle \quad \Rightarrow \quad \frac{d\vec{B}}{dt} \bullet \vec{N} = \frac{-m}{\sqrt{R^2+m^2}} \quad \Rightarrow \quad \boxed{\tau = \frac{m}{R^2+m^2}.}$$

Example 2.2.21. *Let* $\vec{r}(t) = \langle e^t\cos t, e^t\sin t, e^t\rangle$*. Find the Frenet frame as well as the curvature and torsion for the given curve.*

Solution: *differentiation requires some care, remember the product rule. Also, to reduce clutter I factor out* e^t *from the outset* $\vec{r}(t) = e^t\langle\cos t, \sin t, 1\rangle$*,*

$$\frac{d\vec{r}}{dt} = e^t\langle\cos t, \sin t, 1\rangle + e^t\langle-\sin t, \cos t, 0\rangle = e^t\langle\cos t - \sin t, \sin t + \cos t, 1\rangle$$

Observe, $(\cos t + \sin t)^2 + (\cos t - \sin t)^2 = 2$ *hence the speed of this path is* $||\frac{d\vec{r}}{dt}|| = \sqrt{3}e^t$*. Thus,*

$$\vec{T} = \frac{1}{\sqrt{3}}\Big\langle \cos t - \sin t, \sin t + \cos t, 1\Big\rangle.$$

Differentiate again to obtain

$$\frac{d\vec{T}}{dt} = \frac{1}{\sqrt{3}}\Big\langle -\sin t - \cos t, \cos t - \sin t, 0\Big\rangle.$$

Once again, after absorbing a minus-sign, $(\cos t + \sin t)^2 + (\cos t - \sin t)^2 = 2$ *shows the length of* $\frac{d\vec{T}}{dt}$ *is* $\sqrt{2/3}$ *hence we find* $\boxed{\kappa = \frac{\sqrt{2}}{3e^t}}$ *(I used* $\kappa = (1/\dot{s})||d\vec{T}/dt||$ *for path with speed* $\dot{s} = e^t\sqrt{3}$*)*

$$\vec{N} = \frac{1}{\sqrt{2}}\Big\langle -(\cos t + \sin t), \cos t - \sin t, 0\Big\rangle.$$

The binormal follows from the cross product $\vec{B} = \vec{T} \times \vec{N}$*,*

$$\vec{B} = \frac{1}{\sqrt{6}}\Big\langle \sin t - \cos t, -\sin t - \cos t, 2\Big\rangle.$$

To find torsion we need to study the derivative of $\vec{B}$*,*

$$\frac{d\vec{B}}{dt} = \frac{1}{\sqrt{6}}\Big\langle \cos t + \sin t, -\cos t + \sin t, 0\Big\rangle.$$

We calculate $\frac{d\vec{B}}{dt} \bullet \vec{N} = \frac{-1}{\sqrt{12}}[(\cos t + \sin t)^2 + (\cos t - \sin t)^2] = \frac{-2}{\sqrt{12}} = \frac{-1}{\sqrt{3}}$*. Thus,*

$$\tau(t) = (-1/\dot{s})\left(\frac{d\vec{B}}{dt} \bullet \vec{N}\right) = \left(\frac{-1}{e^t\sqrt{3}}\right)\left(\frac{-1}{\sqrt{3}}\right) \quad \Rightarrow \quad \boxed{\tau = \frac{1}{3e^t}.}$$

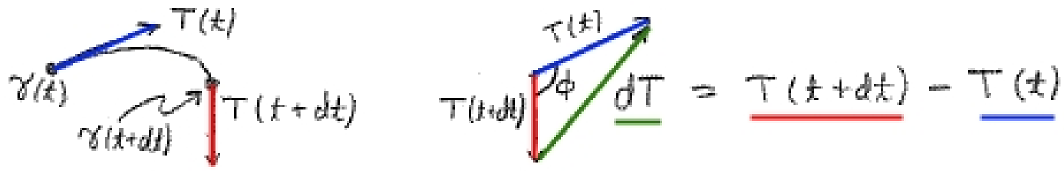
We've seen in this section how calculus and vector algebra encourage us to define curvature and torsion. It remains to examine the geometric significance of those definitions. We pursue that geometry in the remainder of this section.

2.2.4 curvature

Let use begin with the curvature. Assume $\vec{\gamma}$ is a non-stop smooth path,

$$\kappa = \frac{1}{\dot{s}}\left|\left|\frac{d\vec{T}}{dt}\right|\right|$$

Infinitesimally this equation gives $||d\vec{T}|| = \kappa\dot{s}dt = \kappa\frac{ds}{dt}dt = \kappa\, ds$. But this is a strange equation since $||\vec{T}|| = 1$. So what does this mean? Perhaps we should add some more detail to resolve this puzzle; let $d\vec{T} = \vec{T}(t+dt) - \vec{T}(t)$.



Notice that

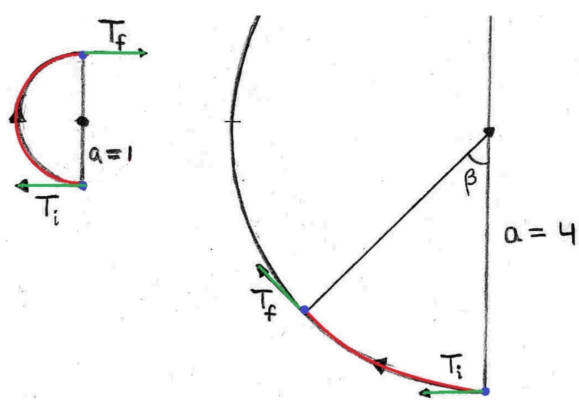
$$\begin{aligned} ||d\vec{T}||^2 &= [\vec{T}(t+dt) - \vec{T}(t)] \bullet [\vec{T}(t+dt) - \vec{T}(t)] \\ &= \vec{T}(t+dt) \bullet \vec{T}(t+dt) + \vec{T}(t) \bullet \vec{T}(t) - 2\vec{T}(t) \bullet \vec{T}(t+dt) \\ &= \vec{T}(t+dt) \bullet \vec{T}(t+dt) + \vec{T}(t) \bullet \vec{T}(t) - 2\vec{T}(t) \bullet \vec{T}(t+dt) \\ &= 2(1 - \cos(\phi))) \end{aligned}$$

where we define ϕ to be the angle between $\vec{T}(t)$ and $\vec{T}(t+dt)$. This angle measures the change in direction of the tangent vector at t goes to $t + dt$. Since this is a small change in time it is reasonable to expect the angle ϕ is small thus $\cos(\phi) \approx 1 - \frac{1}{2}\phi^2$ and we find that

$$||d\vec{T}|| = \sqrt{2(1-\cos(\phi)} = \sqrt{2(1-1+\tfrac{1}{2}\phi^2)} = \sqrt{\phi^2} = |\phi|$$

Therefore, $||d\vec{T}|| = \kappa\, ds = |\phi|$ and we find[11] $\boxed{\kappa = \pm\frac{ds}{d\phi}}$.

Example 2.2.22. *A circle of radius a in the xy-plane has curvature $\kappa = 1/a$. If we study the motion of the unit-tangent to a circle of radius $a = 1$ versus $a = 4$ for a given length of arc we find the smaller circle bends the direction of the unit-tangent more. This is an example of how the magnitude of curvature informs us how fast the direction of the unit-tangent vector is changing: in the small circle, the tangent undergoes a change in direction of π radians whereas in the large circle the change in direction is only $\pi/4$ radians. To be fair, these are compared for the same arclength.*

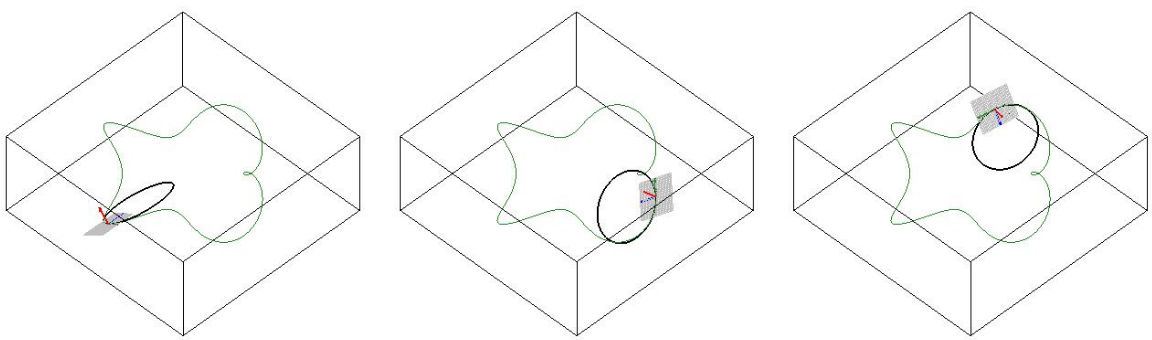


[11] If you study *geodesic curvature* on a surface then the sign takes a particular significance.

Example 2.2.23. *Technically, a line falls outside the scope of the Frenet Serret equations. However, we can gain some appreciation for how lines fit by examining limiting cases. For example, a circle with radius $R \to \infty$ has $\kappa = 1/R \to 0$. Or, for a helix, if we allow the slope $m \to \infty$ while holding R fixed then $\kappa = \frac{m}{m^2+R^2} \to 0$. Both of these provide intuition that $\kappa = 0$ for a line. Also, $\kappa = ||\vec{T}'(t)|| = 0$ for $\vec{r}(t) = \vec{r}_o + t\vec{v}_o$ as $\vec{r}''(t) = 0$ hence $\vec{T}' = 0$.*

Remark 2.2.24.

The curvature measures the infinitesimal change in the direction of the unit-tangent vector to the curve. We say the the reciprocal of the curvature is the **radius of curvature** $r = \frac{1}{\kappa}$. This makes sense as $ds = |1/\kappa| d\phi$ suggests that a circle of radius $1/\kappa$ fits snuggly against the path at time t. We form the **osculating circle at each point along the path by placing a circle of radius $1/\kappa$ tangent to the unit-tangent vector in the plane with normal $\vec{B}(t)$**. Here's a picture, the red-vector is the tangent, the blue the binormal, the green the normal and the black circle is in the grey osculating plane. I have an animated version on my webpage, go take a look.



2.2.5 osculating plane and circle

It was claimed that the "infinitesimal" motion of the path resides in a plane with normal $\vec{B}$. Suppose that at some time t_o the path reaches the point $\vec{\gamma}(t_o) = P_o$. Infinitesimally the tangent line matches the path and we can write the parametric equation for the tangent line as follows:

$$\vec{l}(t) = \vec{\gamma}(t_o) + t\vec{\gamma}\,'(t_o) = P_o + tv_o\vec{T}_o$$

where we used that $\vec{\gamma}\,'(t) = \dot{s}T(t)$ and we evaluated at $t = t_o$ to define $\dot{s}(t_o) = v_o$ and $\vec{T}(t_o) = \vec{T}_o$. The normal line through P_o has parametric equations (using $\vec{N}_o = \vec{N}(t_o)$):

$$\vec{n}(\lambda) = P_o + \lambda\vec{N}_o$$

We learned in the last section that the path bends away from the tangent line along a circle whose radius is $1/\kappa_o$. We find the infinitesimal motion resides in the plane spanned by $\vec{T}_o$ and $\vec{N}_o$ which has normal $\vec{T}_o \times \vec{N}_o = \vec{B}(t_o)$. The tangent line and the normal line are perpendicular and could be thought of as a xy-coordinate axes in the osculating plane. The osculating circle is found with its center on the normal line a distance of $1/\kappa_o$ from P_o. Thus the center of the circle is at:

$$Q_o = P_o + \frac{1}{\kappa_o}\vec{N}_o$$

I'll think of constructing x, y, z coordinates based at P_o with respect to the $\vec{T}_o, \vec{N}_o, \vec{B}_o$ frame. We suppose $\vec{r}$ be a point on the osculating circle and x, y, z to be the coefficients in $\vec{r} = P_o + x\vec{T}_o + y\vec{N}_o + z\vec{B}_o$. Since the circle is in the plane based at P_o with normal $\vec{B}_o$ we should set $z = 0$ for our circle thus $\vec{r} = P_o + x\vec{T}_o + y\vec{N}_o$. Observe,

$$\vec{r} - Q_o = P_o + x\vec{T}_o + y\vec{N}_o - (P_o + \tfrac{1}{\kappa_o}\vec{N}_o) = x\vec{T}_o + (y - \tfrac{1}{\kappa_o})\vec{N}_o.$$

Therefore,

$$||\vec{r} - Q_o||^2 = \tfrac{1}{\kappa_o^2} \quad \Rightarrow \quad ||x\vec{T}_o + (y - \tfrac{1}{\kappa_o})\vec{N}_o]||^2 = \tfrac{1}{\kappa_o^2}.$$

Therefore, by the pythagorean theorem for orthogonal vectors, the x, y, z equations for the osculating circle are simply[12] :

$$\boxed{x^2 + (y - \tfrac{1}{\kappa_o})^2 = \tfrac{1}{\kappa_o^2}, \quad z = 0.}$$

Example 2.2.25. *Let $\vec{r}(t) = \langle 2\sin(3t), t, 2\cos(3t)\rangle$ find the equation of the normal plane and the equation of the osculating plane to the curve at $(0, \pi, -2)$.*

Solution: *straight-forward calculations yield: for $t = \pi$,*

$$\vec{T} = \frac{1}{\sqrt{37}}\langle -6, 1, 0\rangle, \quad \vec{N} = \langle 0, 0, 1\rangle, \quad \vec{B} = \frac{1}{\sqrt{37}}\langle 1, 6, 0\rangle.$$

The osculating plane takes $\vec{B}$ as its normal thus the equation of the osculating plane at $(0, \pi, -2)$ is $x + 6(y - \pi) = 0$. The normal plane has $\vec{N}$ and $\vec{B}$ as tangents which indicates that $\vec{N} \times \vec{B} = \vec{T}$ is the normal to the normal plane. Therefore, the equation of the normal plane is $-6x + y - \pi = 0$.

Finally, notice that if the torsion is zero then the Frenet Serret formulas simplify to:

$$\boxed{\frac{d\vec{T}}{dt} = \dot{s}\kappa\vec{N} \qquad \frac{d\vec{N}}{dt} = -\dot{s}\kappa\vec{T} \qquad \frac{d\vec{B}}{dt} = 0.}$$

we see that $\vec{B}$ is a constant vector field and motion will remain in the osculating plane. The change in the normal vector causes a change in the tangent vector and vice-versa however the binormal vector is not coupled to $\vec{T}$ or $\vec{N}$.

Remark 2.2.26.

The torsion measures the infinitesimal change in the direction of the binormal vector relative to the normal vector of the curve. Because the normal vector is in the plane of infinitesimal motion and the binormal is perpendicular to that plane we can say that the torsion measures how the path lifts or twists up off the plane of infinitesimal motion. Furthermore, we can expect path which is trapped in a particular plane (these are called **planar** curves) will have torsion which is identically zero. We should also expect that the torsion for something like a helix will be nonzero everywhere since the motion is always twisting up off the plane of infinitesimal motion.

Further study of the differential geometry of curves is an excellent topic for future course work. Moreover, the method explored in the past few sections is the quintessential example of the *method of moving frames*. The basic idea is to attach a frame of vectors at each point to the curve. Then,

[12] Of course if we already use x, y, z in a different context then we should use other symbols for the equation of the osculating circle.

because the way they are attached naturally reveals the intrinsic geometry of the curve we derive the invariants of curvature and torsion. A well-known theorem reveals that if we have two curves with matching curvature and torsion and arclength then the curves are **congruent** in the sense of high-school geometry. What that means is there exists an **isometry** of Euclidean three-space which moves one curve to other by a rotation and translation; that is, by a **rigid motion**. Thus, in a sense, we have solved the problem:

> When are two curves congruent in three dimensional space?

Well, in honesty, we are missing some finer points in this discussion so you ought to read a text on classical differential geometry to see a more nuanced presentation. That said, what is the next question? How about:

> When are two surfaces congruent in three dimensional space?

This turns out to be a harder question. However, shortly after Frenet and Serret worked out the case for curves around 1860, Darboux began adapting the method to surfaces around 1880. The theory required[13] deeper insight into what are known as **differential forms**. Cartan's method of moving frames (1900-1920) completed the thought of Darboux and ultimately gave rise to the modern topic of **Exterior Differential Systems**. If you want to *really* understand what a differential equation is then you must study exterior differential systems. There are a few courses you need to complete before that is a reasonable topic to dive into, but I comment here to give you a view ahead. The next text which is immediately accessible (with a little linear algebraic work) is Barret Oneiil's *Elementary Differential Geometry*. See my website for some solutions from that text as well as some crude notes which parallel the text. A more elementary source at essentially the same level as these notes is Colley's *Vector Calculus*. She also discusses the differential geometry of curves and has a nice basic introductory chapter on differential forms.

2.3 physics of motion

In this section we study kinematics. That is, we study how position, velocity and acceleration are related for physical motions. We do not ask where the force comes from, that is a question for physics. Our starting point is the equation of motion $\vec{F} = m\vec{A}$ which is called Newton's Second Law. Given the force and some initial conditions we can in principle integrate the equations of motion and derive the resulting kinematics. We have already, in the usual calculus sequence, twice studied kinematics. In calculus I for one-dimensional motion, in calculus II for two-dimensional motion. I recycle some examples for our current discussion. However, some comments are added since we now have the proper machinery to break-down vectors along a physical path. Let's see how the preceding section is useful in the analysis of the motion of physical objects. The solution of Newton's equation $\vec{F} = m\vec{A}$ is a path $t \mapsto \vec{r}(t)$. It follows we can analyze the velocity and acceleration of the physical path in terms of the **Frenet Frame** $\{T, N, B\}$. To keep it interesting we'll assume the motion is non-stop and smooth so that the analysis of the last section applies.

In this section the notations $\vec{r}$, $\vec{v}$ and $\vec{a}$ are special and set-apart. I don't use these as abstract variables here with no set meaning. Instead, these are connected as is described in the definition that follows:

[13] ok, perhaps this is my bias, but surely differential forms allow one of the most elegant solutions for this problem

Definition 2.3.1. *position, velocity and acceleration.*

The position, velocity and acceleration of an object are vector-valued functions of time and we define them as follows:

1. $\vec{r}(t)$ is the position at time t. (we insist physical paths are parametrized by time)
2. $\vec{v}(t) = \frac{d\vec{r}}{dt}$ is the velocity at time t.
3. $\vec{a}(t) = \frac{d\vec{v}}{dt}$ is the acceleration at time t.

We also define the tangential and normal accelerations of the motion by

$$a_T = \vec{a} \bullet \vec{T} \qquad a_N = \vec{a} \bullet \vec{N} \qquad \text{note:} \qquad \vec{a} = a_T\vec{T} + a_N\vec{N}.$$

We know from our study of the geometry of curves that the binormal component of the acceleration is trivial. Acceleration must lie in the osculating plane and as such is perpendicular to the binormal vector which is the normal to the osculating plane. If you're curious, the position vector itself can have nontrivial components in each direction of the Frenet frame whereas the velocity vector clearly has only a tangential component; $\vec{v} = v\vec{T}$.

If we are given the position vector as a function of time then we need only differentiate to find the velocity and acceleration. On the other hand, if we are given the acceleration then we need to integrate and apply initial conditions to obtain the velocity and position.

Example 2.3.2. *Suppose R and ω are positive constants and the motion of an object is observed to follow the path $\vec{r}(t) = \langle R\cos(\omega t), R\sin(\omega t)\rangle = R\langle \cos(\omega t), \sin(\omega t)\rangle$. We wish to calculate the velocity and acceleration as functions of time.*

Differentiate to obtain the velocity

$$\boxed{\vec{v}(t) = R\omega\langle -\sin(\omega t), \cos(\omega t)\rangle.}$$

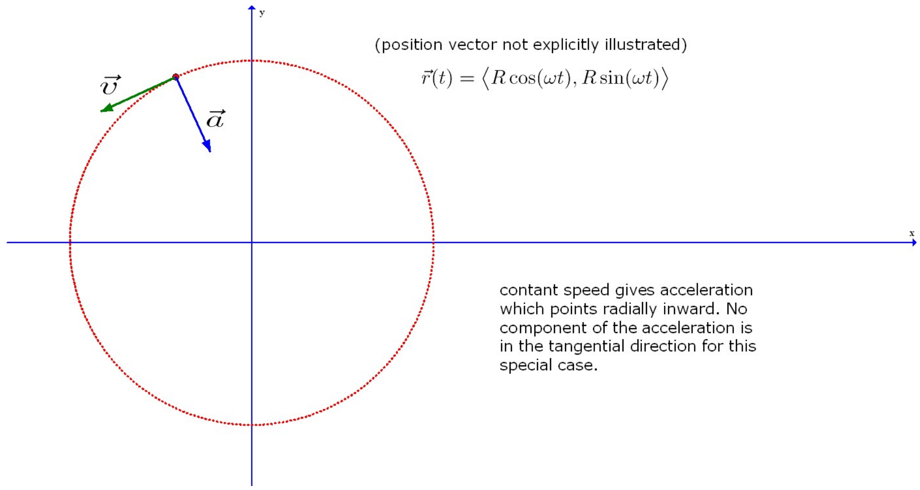
Differentiate once more to obtain the acceleration:

$$\vec{a}(t) = R\omega\langle -\omega\cos(\omega t), -\omega\sin(\omega t)\rangle = \boxed{-R\omega^2\langle \cos(\omega t), \sin(\omega t)\rangle.}$$

Notice we can write that $\vec{a}(t) = -\omega^2\vec{r}(t) = R\omega^2\vec{N}$ in this very special example. This means the acceleration is opposite the direction of the position and it is purely normal. Furthermore, we can calculate

$$r = R, \qquad v = R\omega, \qquad a = R\omega^2$$

Thus the magnitudes of the position, velocity and acceleration are all constant. However, their directions are always changing. Perhaps you recognize these equations as the foundational equations describing constant speed circular motion. This acceleration is called the **centripetal** *or center-seeking acceleration since it points towards the center. Here we imagine attaching the acceleration vector to the object which is traveling in the circle.*



Incidentally, you might wonder how the binormal should be thought of in the example above. We should adjoin a zero to make the vectors three-dimensional and then the cross-product of $\vec{T} \times \vec{N}$ points in the direction given by the right-hand-rule for circles. Curl your right hand around the circle following the motion and your thumb will point in the binormal direction. You can calculate that the binormal is constant:

$$\vec{B} = \vec{T} \times \vec{N} = \langle -\sin(\omega t), \cos(\omega t), 0\rangle \times \langle -\cos(\omega t), -\sin(\omega t), 0\rangle = \langle 0, 0, 1\rangle$$

Often when we consider planar motion we omit the third dimension in the vectors since those components are zero throughout the whole discussion. That said, if we wish to properly employ the Frenet Frame analysis then we must think in three dimensions. The next example is also two-dimensional[14].

Example 2.3.3. *Suppose that the acceleration of an object is known to be $\vec{a} = \langle 0, -g\rangle$ where g is a positive constant. Furthermore, suppose that initially the object is at $\vec{r}_o$ and has velocity $\vec{v}_o$. We wish to calculate the position and velocity as functions of time.*

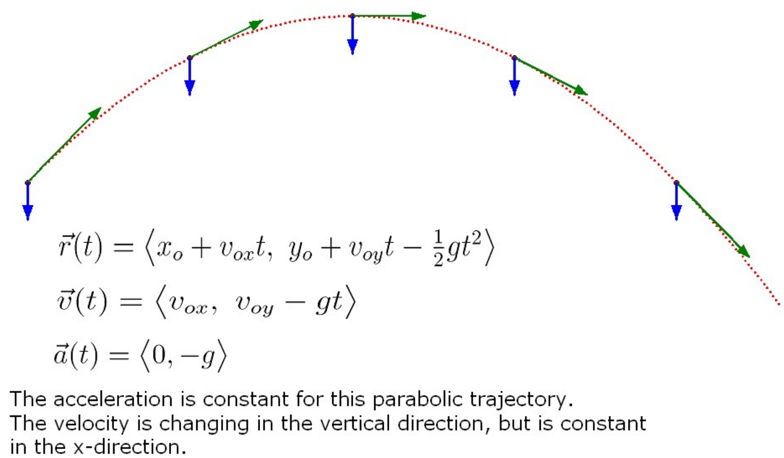
Integrate the acceleration from 0 to t,

$$\int_0^t \frac{d\vec{v}}{d\tau} d\tau = \int_0^t \vec{a}(\tau)d\tau \quad \Rightarrow \quad \vec{v}(t) - \vec{v}(0) = \int_0^t \langle 0, -g\rangle d\tau \quad \Rightarrow \quad \boxed{\vec{v}(t) = \vec{v}_o + \langle 0, -gt\rangle}$$

Integrate the velocity from 0 to t,

$$\int_0^t \frac{d\vec{r}}{d\tau} d\tau = \int_0^t \vec{v}(\tau)d\tau \quad \Rightarrow \quad \vec{r}(t) - \vec{r}(0) = \int_0^t \big(\vec{v}_o + \langle 0, -gt\rangle\big)d\tau \quad \Rightarrow \quad \boxed{\vec{r}(t) = \vec{r}_o + t\vec{v}_o + \langle 0, -\frac{1}{2}gt^2\rangle}$$

[14] all motion generated from Newtonian gravity alone is planar. A more general result states all central force motion lies in a plane, probably a homework of yours



I'm curious how the decomposition of the acceleration into normal and tangential components works out for the example above. It might make an interesting exercise.

The best understanding of Newtonian Mechanics is given by a combination of both vectors and calculus. We need vectors to phrase the geometry of force addition whereas we need calculus to understand how the position, velocity and acceleration variables change in concert.

2.3.1 position vs. displacement vs. distance traveled

The position of an object is simply the (x, y, z) coordinates of the object. Usually it is convenient to think of the position as a vector-valued function of time which we denote $\vec{r}(t)$. The displacement is also a vector, however it compares two possibly distinct positions:

Definition 2.3.4. *displacement and distance traveled.*

Suppose $\vec{r}(t)$ is the position at time t of some object.

1. The **displacement** from position $\vec{r}_1$ to position $\vec{r}_2$ is the vector $\Delta\vec{r} = \vec{r}_2 - \vec{r}_1$.
2. The **distance travelled** during the interval $[t_1, t_2]$ along the curve $t \mapsto \vec{r}(t)$ is given by
$$s_{12} = \int_{t_1}^{t_2} v(t)dt = \int_{t_1}^{t_2} \sqrt{\frac{dx}{dt}^2 + \frac{dy}{dt}^2 + \frac{dz}{dt}^2}\ dt$$
where $v(t) = ||d\vec{r}/dt||$.

Note that the position is the displacement from the origin. Distance travelled is a **scalar** quantity which means it is just a number or if we think of an endpoint as variable it could be a function.

Definition 2.3.5. *arclength function and speed.*

We define
$$s(t) = \int_{t_1}^{t} v(\tau)d\tau = \int_{t_1}^{t} \sqrt{\frac{dx}{d\tau}^2 + \frac{dy}{d\tau}^2 + \frac{dz}{d\tau}^2}\ d\tau$$
to be the arclength travelled from time t_1 to t along the parametrized curve $t \mapsto \vec{r}(t)$. Futhermore, we define the **speed** to be the instantaneous rate of change in the arclength; speed is ds/dt.

Notice it is simple to show that the speed is also equal to the magnitude of the velocity; $ds/dt = v$. Also, note that we drop the z-terms for a typical two-dimensional problem. If you insist on being three dimensional you can just adjoin a bunch of zeros in the examples below. There are unavoidably three dimensional examples a little later in the section.

Example 2.3.6. *Let $\omega, R > 0$. Suppose $\vec{r}(t) = \langle R\cos(\omega t), R\sin(\omega t)\rangle$ for $t \geq 0$. We can calculate that*

$$\frac{d\vec{r}}{dt} = \langle -R\omega\sin(\omega t), R\omega\cos(\omega t)\rangle \quad \Rightarrow \quad v(t) = \sqrt{(-R\omega\sin(\omega t))^2 + (R\omega\cos(\omega t))^2} = \sqrt{R^2\omega^2} = R\omega.$$

Now use this to help calculate the distance travelled during the interval $[0, t]$

$$s(t) = \int_0^t v(\tau)d\tau = \int_0^t R\omega d\tau = R\omega\tau\Big|_0^t = R\omega t.$$

In other words, $\Delta s = R\omega\Delta t$. On a circle the arclength subtended Δs divided by the radius R is defined to be the radian measure of that arc which we typically denote $\Delta\theta$. We find that $\Delta\theta = \omega\Delta t$ or if you prefer $\omega = \Delta\theta/\Delta t$.

Circular motion which is not at a constant speed can be obtained mathematically by replacing the constant ω with a function of time. Let's examine such an example.

Example 2.3.7. *Suppose $\vec{r}(t) = \langle R\cos(\theta), R\sin(\theta)\rangle$ for $t \geq 0$ where $\theta_o, \omega_o, \alpha$ are constants and $\theta = \theta_o + \omega_o t + \frac{1}{2}\alpha t^2$. To calculate the distance travelled it helps to first calculate the velocity:*

$$\frac{d\vec{r}}{dt} = \langle -R(\omega_o + \alpha t)\sin(\theta), R(\omega_o + \alpha t)\cos(\theta)\rangle$$

Next, the speed is the length of the velocity vector,

$$v = \sqrt{[-R(\omega_o + \alpha t)\sin(\theta)]^2 + [R(\omega_o + \alpha t)\cos(\theta)]^2} = R\sqrt{(\omega_o + \alpha t)^2} = R|\omega_o + \alpha t|.$$

Therefore, the distance travelled is given by the integral below:

$$s(t) = \int_0^t R|\omega_o + \alpha\tau|d\tau$$

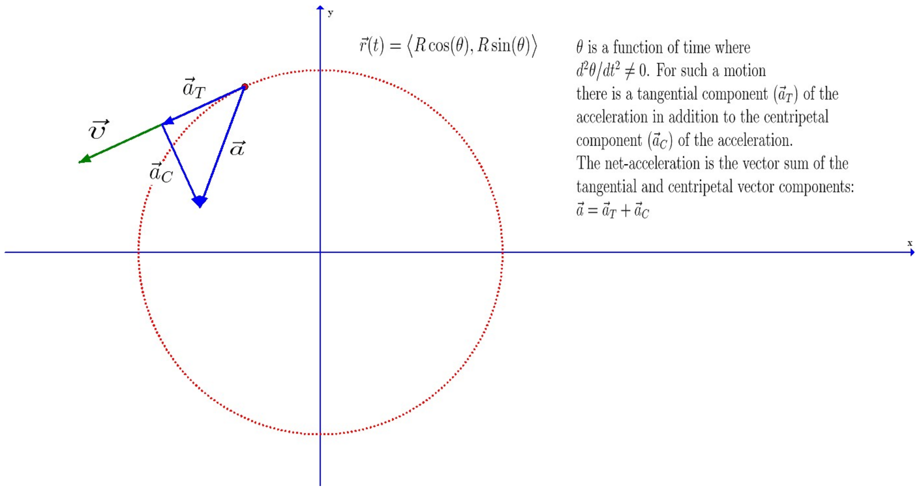
To keep things simple, let's suppose that ω_o, α are given such that $\omega_o + \alpha t \geq 0$ hence $v = R\omega_o + R\alpha t$. To suppose otherwise would indicate the motion came to a stopping point and reversed direction, which is interesting, just not to us here.

$$s(t) = R\int_0^t (\omega_o + \alpha\tau)d\tau = R\omega_o t + \frac{1}{2}R\alpha t^2.$$

Observe that $\theta(t) - \theta_o = (s(t) - s(0))/R$ thus we find that $\Delta\theta = \omega_o t + \frac{1}{2}\alpha t^2$ which is the formula for the angle subtended due to motion at a constant **angular acceleration** *α. In invite the reader to differentiate the position twice and show that*

$$\vec{a}(t) = \frac{d^2\vec{r}}{dt^2} = -\underbrace{R\omega^2\Big\langle\cos(\theta(t)), \sin(\theta(t))\Big\rangle}_{centripetal} + \underbrace{R\alpha\Big\langle -\sin(\theta(t)), \cos(\theta(t))\Big\rangle}_{tangential}$$

where $\omega = \omega_o + \alpha t$. Notice, the term **centripetal** *could be replaced with* **normal** *in the sense of the Frenet-Serret frames. Recall the normal pointed towards the center of the osculating circle thus the center-seeking acceleration is precisely the normal acceleration.*



Distance travelled is not always something we can calculate in closed form. Sometimes we need to relegate the calculation of the arclength integral to a numerical method. However, the example that follows is still calculable without numerical assistance. It did require some thought.

Example 2.3.8. *We found that $\vec{a} = \langle 0, -g\rangle$ twice integrated yields a position of $\vec{r}(t) = \vec{r}_o + t\vec{v}_o + \langle 0, -\frac{1}{2}gt^2\rangle$ for some constant vectors $\vec{r}_o = \langle x_o, y_o\rangle$ and $\vec{v}_o = \langle v_{ox}, v_{oy}\rangle$. Thus,*

$$\vec{r}(t) = \left\langle x_o + v_{ox}t, y_o + v_{oy}t - \frac{1}{2}gt^2\right\rangle$$

From which we can differentiate to derive the velocity,

$$\vec{v}(t) = \langle v_{ox}, v_{oy} - gt\rangle.$$

If you've had any course in physics, or just a proper science education, you should be happy to observe that the zero-acceleration in the x-direction gives rise to constant-velocity motion in the x-direction whereas the gravitational acceleration in the y-direction makes the object fall back down as a consequence of gravity. If you think about $v_{oy} - gt$ it will be negative for some $t > 0$ whatever the initial velocity v_{oy} happens to be, this point where $v_{oy} - gt = 0$ is the turning point in the flight of the object and it gives the top of the parabolic[15] trajectory which is parametrized by $t \to \vec{r}(t)$. Suppose $x_o = y_o = 0$ and calculate the distance travelled from time $t = 0$ to time $t_1 = v_{oy}/g$. Additionally, let us assume $v_{ox}, v_{oy} \geq 0$.

$$\begin{aligned} s = \int_0^{t_1} v(t)dt &= \int_0^{t_1} \sqrt{(v_{ox})^2 + (v_{oy} - gt)^2}dt \\ &= \int_{v_{oy}}^0 \sqrt{(v_{ox})^2 + (u)^2}\left(\frac{du}{-g}\right) \qquad u = v_{oy} - gt \\ &= \frac{1}{g}\int_0^{v_{oy}} \sqrt{(v_{ox})^2 + (u)^2}du \end{aligned}$$

Recall that a nice substitution for an integral such as this is provided by the $\sinh(z)$ *since* $1 + \sinh^2(z) = \cosh^2(z)$ *hence a* $u = v_{ox}\sinh(z)$ *subsitution will give*

$$(v_{ox})^2 + (u)^2 = (v_{ox})^2 + (v_{ox}\sinh(z))^2 = v_{ox}^2\cosh^2(z)$$

[15] no, we have not shown this is a parabola, I invite the reader to verify this claim. That is find A, B, C such that the graph $y = Ax^2 + Bx + C$ is the same set of points as $\vec{r}(\mathbb{R})$.

and $du = v_{ox}\cosh(z)dz$ *thus,* $\int\sqrt{(v_{ox})^2+(u)^2}du = \int\sqrt{v_{ox}^2\cosh^2(z)}v_{ox}\cosh(z)dz = \int v_{ox}^2\cosh^2(z)dz$. *Furthermore,* $\cosh^2(z) = \frac{1}{2}(1+\cosh(2z))$ *hence*

$$\int\sqrt{(v_{ox})^2+(u)^2}du = \frac{v_{ox}^2}{2}\left[z+\frac{1}{2}\sinh(2z)\right]+c = \frac{v_{ox}^2}{2}\left[z+\sinh(z)\cosh(z)\right]+c$$

Note $u = v_{ox}\sinh(z)$ *and* $v_{ox}\cosh(z) = \sqrt{(v_{ox})^2+(u)^2}$ *hence substituting,*

$$\int\sqrt{(v_{ox})^2+(u)^2}du = \frac{1}{2}\left[v_{ox}^2\sinh^{-1}\left(\frac{u}{v_{ox}}\right)+u\sqrt{v_{ox}^2+u^2}\right]+c$$

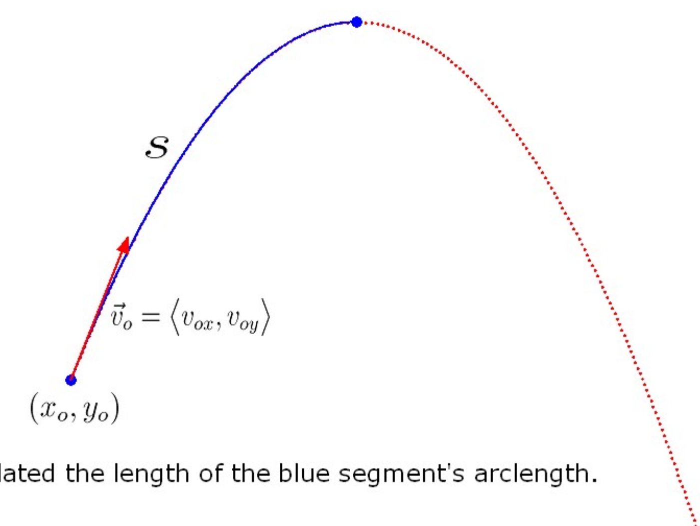
Well, I didn't think that was actually solvable, but there it is. Returning to the definite integral to calculate s *we can use the antiderivative just calculated together with FTC part II to conclude: (provided* $v_{ox}\neq 0$*)*

$$\boxed{s = \frac{1}{2g}\left[v_{ox}^2\sinh^{-1}\left(\frac{v_{oy}}{v_{ox}}\right)+v_{oy}\sqrt{v_{ox}^2+v_{oy}^2}\right]}$$

If $v_{ox}=0$ *then the problem is much easier since* $v(t) = |v_{oy}-gt| = v_{oy}-gt$ *for* $0\leq t\leq t_1 = v_{oy}/g$ *hence*

$$s = \int_0^{t_1}v(t)dt = \int_0^{t_1}(v_{oy}-gt)dt = \left[v_{oy}t-\frac{1}{2}gt^2\right]\Bigg|_0^{v_{oy}/g} = \boxed{\frac{v_{oy}^2}{2g}}$$

Interestingly, this is the formula for the height of the parabola even if $v_{ox}\neq 0$. *The initial* x*-velocity simply determines the horizontal displacement as the object is accelerated vertically by gravity.*



We calculated the length of the blue segment's arclength.

Example 2.3.9. *Suppose $\vec{r}(t) = \langle t^2+1,\ t^3,\ t^2-1\rangle$ for $t \geq 0$. Find the velocity, acceleration and speed relative the given position.*

Solution: *From the definitions of velocity ($\vec{v}$) , acceleration ($\vec{a}$) and speed (v):*

$$\begin{aligned}
\vec{v}(t) &= \frac{d\vec{r}}{dt} = \frac{d}{dt}\langle t^2+1,\ t^3,\ t^2-1\rangle = \langle 2t,\ 3t^2,\ 2t\rangle. \\
\vec{a}(t) &= \frac{d\vec{v}}{dt} = \frac{d}{dt}\langle 2t,\ 3t^2,\ 2t\rangle = \langle 2,\ 6t,\ 2\rangle. \\
v(t) &= ||\vec{v}|| = \sqrt{(2t)^2+(3t^2)^2+(2t)^2} = \sqrt{8t^2+9t^4}.
\end{aligned}$$

Notice the notations (t) indicate the function in question is evaluated at time t. In some instances, this notation is omitted.

Example 2.3.10. *Suppose that $\vec{v}(t) = \langle e^t, t^2, \cos t\rangle$ is the velocity of Dwight. Calculate the acceleration of Dwight. If the initial position of Dwight is the origin then find the position of Dwight at time t.*

Solution: *Acceleration is found by differentiation; $\vec{a}(t) = \frac{d\vec{v}}{dt} = \langle e^t, 2t, -\sin t\rangle$. On the other hand, the position is found by integration:*

$$\int_0^t \frac{d\vec{r}}{du}du = \int_0^t \vec{v}(u)du \quad \Rightarrow \quad \vec{r}(t) - \vec{r}(0) = \langle e^t - 1, \tfrac{1}{3}t^3, \sin t\rangle$$

We know $\vec{r}(0) = \langle 0,0,0\rangle$ thus the position of Dwight at time t is $\vec{r}(t) = \langle e^t - 1, \frac{1}{3}t^3, \sin t\rangle$.

Example 2.3.11. *Suppose $\vec{a}(t) = \langle 2, 6t, 12t^2\rangle$ and we know the initial velocity is $\vec{v}(0) = \langle 1,0,0\rangle$ and the initial position is $\vec{r}(0) = \langle 0,1,-1\rangle$. Find the velocity and position at time t.*

Solution: *note $\int_0^t \frac{d\vec{B}}{du}du = \vec{B}(t) - \vec{B}(0)$ hence $\vec{C}(t) = \frac{d\vec{B}}{dt}$ implies $\vec{B}(t) = \vec{B}(0) + \int_0^t \vec{C}(u)\,du$. Therefore, as $\vec{a} = \frac{d\vec{v}}{dt}$ and $\vec{v}(0) = \langle 1,0,0\rangle$,*

$$\vec{v}(t) = \langle 1,0,0\rangle + \int_0^t \langle 2, 6u, 12u^2\rangle\,du \quad \Rightarrow \quad \boxed{\vec{v}(t) = \langle 1+2t, 3t^2, 4t^3\rangle.}$$

Likewise, as $\vec{v} = \frac{d\vec{r}}{dt}$ and $\vec{r}(0) = \langle 0,1,-1\rangle$,

$$\vec{r}(t) = \langle 0,1,-1\rangle + \int_0^t \langle 1+2u, 3u^2, 4u^3\rangle\,du \quad \Rightarrow \quad \boxed{\vec{r}(t) = \langle t+t^2, 1+t^3, -1+t^4\rangle.}$$

Example 2.3.12. *This example is closely related to Example 2.3.3 where a picture is given. Suppose a projectile is fired on earth and is subject to the acceleration $\langle 0, -g\rangle$ where $g = 9.8m/s^2$. If the projectile has an initial speed of v_o as it is fired at an angle of θ above the horizontal. Assume the projectile is fired on a level plane and find the maximum hieght and range. Also, find the position and velocity as a function of time t. Ignore effects other than gravity and assume the motion is close to the surface of the earth so gravity is essentially a constant force giving rise to a constant acceleration.*

Solution: *we omit units to reduce clutter. To begin, note that if $\theta = 30^o$ then $\vec{v}(0) = \langle v_o\cos\theta, v_o\sin\theta\rangle$. Also, we set the origin as the location at which the projectile is fired*[16]*. Integrate and apply $\vec{v}(0) = \langle v_o\cos\theta, v_o\sin\theta\rangle$ to obtain*

$$\vec{v}(t) = \langle v_o\cos\theta, v_o\sin\theta - gt\rangle$$

However, by our definition $\vec{r}(0) = \langle 0, 0\rangle$ hence integrating again yields:

$$\vec{r}(t) = \big\langle (v_o\cos\theta)t,\ (v_o\sin\theta)t - \tfrac{1}{2}gt^2\big\rangle$$

Notice, the position and velocity functions above implicitly contain the following **scalar** *functions:*

$$v_x(t) = v_o\cos\theta, \quad v_y(t) = v_o\sin\theta - gt$$

and

$$x(t) = (v_o\cos\theta)t, \quad y(t) = (v_o\sin\theta)t - \tfrac{1}{2}gt^2.$$

To find the maximum y value we find critical times for y. Consider $v_y(t) = 0$ gives the only critical time which is interesting. Algebra yields $t = \frac{v_o\sin\theta}{g}$ and $y'' = -g < 0$ hence by the second derivative test we find the maximum y is attained at $t_c = \frac{v_o\sin\theta}{g}$. This is called the height h of the trajectory,

$$h = (v_o\sin\theta)\left(\frac{v_o\sin\theta}{g}\right) - \tfrac{1}{2}g\left(\frac{v_o\sin\theta}{g}\right)^2 \quad \Rightarrow \quad \boxed{h = \frac{v_o^2\sin^2\theta}{2g}.}$$

The range is found by determining the time for which the equation $y(t) = 0$

$$(v_o\sin\theta)t - \tfrac{1}{2}gt^2 = t(v_o\sin\theta - \tfrac{1}{2}gt) = 0.$$

Thus $t = 0$ or $t = \frac{2v_o\sin\theta}{g}$. Notice the nonzero solution is precisely $2t_c$ as it expected by the symmetry of the problem. Finally, set $t = \frac{2v_o\sin\theta}{g}$ in $x(t)$ to obtain the range R:

$$R = (v_o\cos\theta)\frac{2v_o\sin\theta}{g} \quad \Rightarrow \quad \boxed{R = \frac{v_o^2\sin 2\theta}{g}.}$$

Perhaps you used the boxed equations in your highschool physics course.

[16]This freedom is implicitly granted to us by the structure of the problem, beware, when a coordinate system is more completely described at the outset of the problem this freedom may not be available.

Example 2.3.13. *Find the tangential and normal components of the acceleration of the motion with position* $\vec{r}(t) = \langle 3t - t^3, 3t^2, 0\rangle$.

Solution: *Note* $\vec{v}(t) = \langle 3 - 3t^2, 6t, 0\rangle = 3\langle 1 - t^2, 2t, 0\rangle$ *thus noting* $(1-t^2)^2 + 4t^2 = (1+t^2)^2$,

$$\vec{T}(t) = \frac{1}{1+t^2}\Big\langle 1-t^2,\ 2t,\ 0\Big\rangle.$$

Differentiating, use product rule and simplify to obtain:

$$\vec{T}\,'(t) = \frac{2}{(1+t^2)^2}\Big\langle -2t,\ 1-t^2,\ 0\Big\rangle.$$

Normalizing $d\vec{T}/dt$ *gives (by the same algebra as was used in normalizing* $\vec{v}$ *to give* $\vec{T}$*):*

$$\vec{N}(t) = \frac{1}{1+t^2}\Big\langle -2t,\ 1-t^2,\ 0\Big\rangle$$

Finally, we can select $\vec{T}$ *and* $\vec{N}$ *components from* $\vec{a}(t) = 6\langle -t,\ 1,\ 0\rangle$ *by taking dot-products with the orthonormal vectors* $\vec{T}(t), \vec{N}(t)$

$$a_T = \vec{a} \bullet \vec{T} = \frac{6}{1+t^2}\big(-t(1-t^2) + 2t\big) = \frac{6}{1+t^2}\big(t^3 + t\big) \quad \Rightarrow \quad \boxed{a_T = 6t.}$$

and,

$$a_N = \vec{a} \bullet \vec{N} = \frac{6}{1+t^2}\big(-t(-2t) + 1 - t^2\big) = \frac{6}{1+t^2}\big(t^2 + 1\big) \quad \Rightarrow \quad \boxed{a_N = 6.}.$$

This curve is always increasing speed and circling around.

One last connection we ought to explore: what do the Frenet-Serret equations say about the normal and tangent components of acceleration. In other words, can we cipher some geometry from a_T and a_N directly? I'll begin by expressing the unit-tangent explicitly in terms of velocity and speed

$$\vec{T} = \frac{\vec{v}}{v}$$

Differentiate with respect to time, use the product and chain rules as usual:

$$\frac{d\vec{T}}{dt} = \frac{1}{v}\frac{d\vec{v}}{dt} - \frac{\vec{v}}{v^2}\frac{dv}{dt} = \frac{1}{v}\vec{a} - \frac{1}{v}\frac{dv}{dt}\vec{T}$$

Recall, in terms of the curvature κ the unit-tangent evolves into the normal direction $\frac{d\vec{T}}{dt} = v\kappa\vec{N}$ (we cannot assume unit-speed). Thus,

$$v\kappa\vec{N} = \frac{1}{v}\vec{a} - \frac{1}{v}\frac{dv}{dt}\vec{T} \quad \Rightarrow \quad \boxed{\vec{a}(t) = \kappa v^2\vec{N} + \frac{dv}{dt}\vec{T}.}$$

which show us that $a_N = \kappa v^2$ and $a_T = \frac{dv}{dt}$. In other words, the normal acceleration is proportional to the curvture whereas the tangential acceleration reveals how the speed is changing with time.

Example 2.3.14. *Find the curvture of the osculating circle to* $\vec{r}(t) = \langle 3t - t^3, 3t^2, 0\rangle$ *at time* t.

Solution: *The calculations of Example 2.3.13 show that* $v = 3(1+t^2)$ *and* $a_T(t) = 6t$ *and* $a_N(t) = 6$. *However, in view of the boxed equation above this example we have:*

$$6 = \kappa \cdot 9(1+t^2)^2 \quad \Rightarrow \quad \kappa(t) = \frac{2}{3(1+t^2)^2}.$$

2.3.2 Keplers' laws of planetary motion

I realize these notes are a little stale, but the equations are as true now as they were when Newton derived them without the aid of vectors and modern notation. I don't always find time to lecture on these, but I will try.

KEPLER'S LAWS OF PLANETARY MOTION (284)

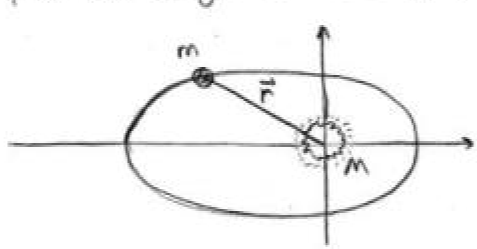
In antiquity there have been radically different views of the universe at large and the motion or lack of motion of the earth through it. At the time of Kepler the heliocentric view of Copernicus (1473–1543) had taken hold, but astronomers insisted that planets traveled in circles, then circles on top of circles on top of circles... This system of "perfect" circles were known as epicycles. Epicycles worked quite well but Kepler (1571–1630) found them unnatural. Kepler instead thought he could explain the motion of planets by a few simple rules. He found these rules emperically by studying the exquisite data taken by Tycho Brahe. These laws were chosen simply to fit the data. Only later were these laws derived from basic physical law. By the way, much of modern physics are still like Kepler's Laws, it is always the dream/goal/aspiration to derive known phenominological law from basic principles. There is some controversy as to who first derived Kepler's Laws, many credit Newton himself others credit Johann Bernoulli in 1710. The incredible thing is that we can derive the laws in a few short pages. Our notation and understanding of vector calculus is several hundred years in advance, so ordinary folks like myself can grasp the proof.

Set-up

Keplers laws for the Sun and a single planet are:

1.) The orbit of the planet is elliptical with the sun at a focus.
2.) During equal times the planet sweeps out equal areas in the ellipse.
3.) $T^2 \propto a^3$ where T = period of planet's orbit, a = length of semimajor axis of ellipse.

We place the origin at the sun. We expect that



• My proof of Kepler's Laws follows Colley's of §3.1 fairly closely. (285)

Proposition: The motion of the planet lies in a plane which also contains the sun if we assume Newton's Universal Law of Gravitation governs the motion through Newton's Laws.

Proof: our goal is to show that $\vec{r} \times \vec{V} = \vec{C}$ for some constant vector $\vec{C}$. This will show that planet moves in a plane with normal $\vec{C}$. Note,

$$\frac{d}{dt}\left(\vec{r} \times \vec{V}\right) = \underbrace{\frac{d\vec{r}}{dt} \times \vec{V}}_{\vec{V}\times\vec{V}=0.} + \vec{r} \times \frac{d\vec{V}}{dt} = \vec{r} \times \vec{a}.$$

Recall in our current notation that $\vec{r} = r\hat{r}$ and Newton tells us that,

$$\vec{F} = m\vec{a} = -\frac{GmM}{r^2}\hat{r} = -\frac{GmM}{r^3}\vec{r}$$

m = mass of planet
M = mass of sun
G = Gravitational Constant.

$$\therefore \quad \vec{a} = \frac{-GM}{r^3}\vec{r} \quad \text{thus } \vec{a} \parallel \vec{r}$$

$$\Rightarrow \vec{a} \times \vec{r} = 0 \quad \Rightarrow \quad \frac{d}{dt}\left(\vec{r} \times \vec{V}\right) = \vec{r} \times \vec{a} = 0 \quad \therefore \quad \vec{r} \times \vec{V} = \vec{C}$$

Th^m/ Kepler's 1^st Law: The planets orbit is an ellipse with sun at one focus

Proof: this will take a little work so be patient, lets get a better hold on $\vec{C}$,

$$\vec{V} = \frac{d\vec{r}}{dt} = \frac{d}{dt}\left(r\hat{r}\right) = \frac{dr}{dt}\hat{r} + r\frac{d\hat{r}}{dt} = \dot{r}\hat{r} + r\frac{d\hat{r}}{dt}$$

Apply this to the following,

$$\vec{C} = \vec{r} \times \vec{V} = r\hat{r} \times [\cancel{\dot{r}\hat{r}}^{\,0} + r\frac{d\hat{r}}{dt}] = \underbrace{r^2\,\hat{r} \times \frac{d\hat{r}}{dt} = \vec{C}}_{\mathrm{I}}$$

Calculate then, using I

$$\begin{aligned}
\vec{a} \times \vec{C} &= \left(\frac{-GM}{r^2}\hat{r}\right) \times \left(r^2\,\hat{r} \times \frac{d\hat{r}}{dt}\right) \\
&= -GM\left[\hat{r} \times \left(\hat{r} \times \frac{d\hat{r}}{dt}\right)\right] \\
&= GM\left[\left(\hat{r} \times \frac{d\hat{r}}{dt}\right) \times \hat{r}\right] \quad : \text{recall } \overbrace{A\times(B\times C) = (A\cdot C)B - (A\cdot B)C}^{\text{see §9.4 \#30}} \\
&= GM\left[(\hat{r}\cdot\hat{r})\frac{d\hat{r}}{dt} - \underbrace{\left(\hat{r}\cdot\frac{d\hat{r}}{dt}\right)}_{\to 0}\hat{r}\right] \quad : \hat{r}\cdot\hat{r} = 1 \Rightarrow 2\hat{r}\cdot\dot{\hat{r}} = 0 \Rightarrow \hat{r}\cdot\dot{\hat{r}} = 0. \\
&= GM\frac{d\hat{r}}{dt} \\
&= \underbrace{\frac{d}{dt}\left(GM\hat{r}\right) = \vec{a} \times \vec{C}}_{\mathrm{II}}
\end{aligned}$$

(286)

Proof of Kepler's 1st Law continued

We may derive another identity for $\vec{a}\times\vec{c}$,

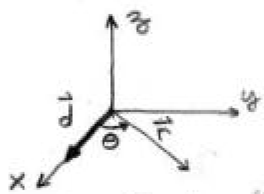
$$\vec{a}\times\vec{c} = \frac{d\vec{V}}{dt}\times\vec{c} + \vec{V}\times\frac{d\vec{c}}{dt} \quad : \text{added zero since } \frac{d\vec{c}}{dt} = 0.$$

$$= \underbrace{\frac{d}{dt}\left[\vec{V}\times\vec{c}\right]}_{\text{III}} \quad : \text{using identity (V.) on (265)}$$

Thus comparing (II.) & (III.) we find

$$\frac{d}{dt}\left(GM\hat{r}\right) = \frac{d}{dt}\left(\vec{V}\times\vec{c}\right) \quad \therefore \quad \underbrace{\vec{V}\times\vec{c} = GM\hat{r} + \vec{d}}_{\text{IV}}$$

where $\vec{d}$ is a constant vector, it lies in the orbital plane since $\vec{V}\times\vec{c}$ and $\hat{r}$ do. Now choose coordinates in the orbital plane so that $\vec{d}$ lines up with the x-axis. Let θ be the usual θ in the xy-plane,



$$\hat{r}\cdot\vec{d} = |\hat{r}||\vec{d}|\cos\theta = d\cos\theta$$

where $|\vec{d}| = d$ in our notation here.

Now consider the length of $\vec{c}$ squared,

$$\begin{aligned} c^2 &= \vec{c}\cdot\vec{c} \\ &= (\vec{r}\times\vec{V})\cdot\vec{c} \\ &= \vec{r}\cdot(\vec{V}\times\vec{c}) \quad : \text{using identity (V.) of (248)} \\ &= r\hat{r}\cdot\left[GM\hat{r} + \vec{d}\right] \quad : \text{using IV. we found just above.} \\ &= GMr + r\hat{r}\cdot\vec{d} \\ &= GMr + rd\cos\theta \\ &= r(GM + d\cos\theta) \end{aligned}$$

Therefore we solve for $r = \sqrt{x^2+y^2+z^2} = \sqrt{x^2+y^2}$ (we're in $z=0$) and obtain the eq$^{\underline{n}}$ of an ellipse (or parabola or hyperbola)

$$r = \frac{c^2}{GM + d\cos\theta} = \frac{c^2/GM}{1 + (d/GM)\cos\theta} = \boxed{\frac{p}{1 + e\cos\theta} = r}$$

where we define $p = c^2/GM$ and the eccentricity $e = d/GM$. This is an ellipse in polar coordinates. Since you've likely not seen that recently (or maybe never) we'll connect to

Proof of Kepler's 1st Law continued (287)

the usual Cartesian eq^n's for the ellipse. The details will be of use to us in proving the 3rd Law of Kepler later on.

$$r = \frac{p}{1+e\cos\theta} \Rightarrow r = p - er\cos\theta$$

Trying to convert the polar coordinates (r,θ) to (x,y) where $x = r\cos\theta$ and $y = r\sin\theta$. We see, using $x = r\cos\theta$

$$r = p - ex$$

$$r^2 = x^2 + y^2 = p^2 - 2epx + e^2x^2$$

$$x^2 - e^2x^2 + y^2 + 2epx = p^2$$

$$x^2(1-e^2) + 2epx + y^2 = p^2$$

$$x^2 + \frac{2ep}{1-e^2}x + \frac{y^2}{1-e^2} = \frac{p^2}{1-e^2} \quad : \quad \underline{\text{assume } e \neq \pm 1}$$

$$\left(x - \frac{ep}{1-e^2}\right)^2 + \frac{y^2}{(1-e^2)} = \frac{p^2}{1-e^2} + \frac{e^2p^2}{(1-e^2)^2} = \frac{p^2 - e^2p^2 + e^2p^2}{(1-e^2)^2} = \frac{p^2}{(1-e^2)^2}$$

$$\therefore \boxed{\frac{\left(x - \frac{ep}{1-e^2}\right)^2}{p^2/(1-e^2)^2} + \frac{y^2}{p^2/(1-e^2)} = 1} \quad \underset{(0<e<1)}{\underline{\text{ellipse}}} \text{ OR } \underset{(e>1)}{\underline{\text{hyperbola}}}$$

This is an ellipse with center $(ep/(1-e^2), 0)$ and it has semimajor axis length $a = p/(1-e^2)$ and semiminor axis $b = p/\sqrt{1-e^2}$.

Remark: recall that we defined $p = c^2/GM$ so $p>0$ and we need not worry about ÷ by p. Now $e = d/GM > 0$ so we can rule out $e = -1$ as a problem. Notice we have division by $\sqrt{1-e^2}$ as part of our sol^n, this only makes sense if $0 < e < 1$. The case $e = 1$ needs seperate treatment. Motion in the case $0 < e < 1$ is that of planets.

$\underline{e = 1}$) $r = p - r\cos\theta \quad \therefore \quad r^2 = (p-x)^2 = p^2 - 2xp + x^2$

that is $x^2 + y^2 = p^2 - 2xp + x^2 \Rightarrow 2xp = p^2 - y^2$

$$\therefore \boxed{x = p/2 - y^2/2p} \quad \underline{\text{parabola}}$$

Remark: One nice resource for background on conic-sections and polar coordinates is "Precalculus, Concepts through functions" Sullivan & Sullivan. There is just about all the cases you can imagine, rotated ellipses for example.

288

Th^m/ KEPLER'S 2nd LAW: During equal times a planet sweeps through equal areas.

Proof: Pick a point P_o at angle θ_o. The later in this course we will learn that the area in polar coordinates swept by the region from θ_o to θ is simply

$$A(\theta) = \int_{\theta_o}^{\theta} \frac{1}{2} r^2 \, d\beta$$

We seek to show that $\frac{dA}{dt}$ = constant. Consider then

$$\frac{dA}{d\theta} = \frac{d}{d\theta}\int_{\theta_o}^{\theta} \frac{1}{2} r^2 \, d\beta = \frac{1}{2} r^2 \qquad \text{by F.T.C.}$$

Then the chain rule tells us

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

Notice that $\hat{r} = \langle \cos\theta, \sin\theta \rangle$ thus diff. implicitly, remember $\theta = \theta(t)$.

$$\frac{d\hat{r}}{dt} = \langle -\sin\theta, \cos\theta \rangle \frac{d\theta}{dt} = \langle -\sin\theta, \cos\theta, 0 \rangle \frac{d\theta}{dt}$$

(we've been suppressing the z-comp.)

(189), ① $\Rightarrow$ $\vec{C} = r^2\left(\hat{r} \times \frac{d\hat{r}}{dt}\right) = r^2 \frac{d\theta}{dt} \langle \cos\theta, \sin\theta, 0 \rangle \times \langle -\sin\theta, \cos\theta, 0 \rangle$

$$\vec{C} = r^2 \frac{d\theta}{dt}\langle 0, 0, 1 \rangle \qquad \therefore \; C = r^2 \frac{d\theta}{dt}$$

Hence $\frac{dA}{dt} = \frac{1}{2} r^2 \frac{d\theta}{dt} = \frac{C}{2}$ = constant. //

Th^m/ KEPLER'S 3rd Law: $T^2 = Ka^3$ where T is the orbital period and a is the length of the semimajor axis, K = some constant

Proof: I proved back on pg. 138 in E7 that the area of an ellipse is $A = \pi ab$. On the other hand we could say that $dA = \frac{dA}{dt} dt$ and integrate over a whole orbit to find

$$\pi ab = \int_0^T \frac{dA}{dt} dt = \int_0^T \frac{C}{2} dt = \frac{CT}{2} \qquad \therefore \; T = \frac{2\pi ab}{C} \quad \therefore \; T^2 = \frac{4\pi^2 a^2 b^2}{C^2}$$

notice that $a^2 = p^2/(1-e^2)^2$ and $b^2 = p^2/(1-e^2)$, also $C^2 = GMp$.

$$T^2 = \frac{4\pi^2}{GMp} \frac{p^2}{(1-e^2)^2} \cdot \frac{p^2}{(1-e^2)} = \frac{4\pi^2}{GM}\left(\frac{p}{1-e^2}\right)^3 = \boxed{\frac{4\pi^2 a^3}{GM} = T^2} \; //.$$

It is interesting that $K = \frac{4\pi^2}{GM}$ is independent of the planets mass: all the planets orbit under the same K-value.

Remark: There is another method of proving Kepler's Laws that begins with the two-body Lagrangian for a central potential (well force really but $\vec{F} = f(r)\hat{r} \Rightarrow U = U(r) \dots$). In that derivation one need not assume the sun is at the origin. Instead you consider the center of mass to be at the origin and work out how the reduced mass μ orbits. Anyway its very beautiful, take Mechanics at the Junior/Senior level to see the more general derivation. Also they will actually find $r(t)$ explicitly as opposed to the indirect arguments we have offered (or rather stolen from Colley ☺.)

2.4 integration of scalar function along a curve

In this section we learn how to sum a quantity along some curve. Let's begin by reviewing some terminology. A **path** in $\mathbb{R}^3$ is a continuous function $\vec{\gamma}$ with connected domain I such that $\vec{\gamma} : I \subset \mathbb{R} \to \mathbb{R}^3$. If $\partial I = \{a, b\}$ then we say that $\vec{\gamma}(a)$ and $\vec{\gamma}(b)$ are the endpoints of the path $\vec{\gamma}$. When $\vec{\gamma}$ has continuous derivatives of all orders we say it is a smooth path (of class C^∞), if it has at least one continuous derivative we say it is a differentiable path(of class C^1). When $I = [a, b]$ then the path is said to go from $\vec{\gamma}(a) = P$ to $\vec{\gamma}(b) = Q$ and the image $C = \vec{\gamma}([a, b])$ is said to be an **oriented curve** C from P to Q. The same curve from Q to P is denoted $-C$. We say C and $-C$ have opposite orientations.

Hopefully most of this is already familar from our earlier work on parametrizations. I give another example just in case.

Example 2.4.1. *The line-segment L from $(1, 2, 3)$ to $(5, 5, 5)$ has parametric equations $x = 1 + 4t, y = 2 + 3t, z = 3 + 2t$ for $0 \leq t \leq 1$. In other words, the path $\vec{\gamma}(t) = \langle 1 + 4t, 2 + 3t, 3 + 2t\rangle$ covers the line-segment L. In contrast $-L$ goes from $(5, 5, 5)$ to $(1, 2, 3)$ and we can parametrize it by $x = 5-4u, y = 5-3u, z = 5-2u$ or in terms of a vector-formula $\vec{\gamma}_{reverse}(u) = \langle 5-4u, 5-3u, 5-2u\rangle$. How are these related? Observe:*

$$\vec{\gamma}_{reverse}(0) = \vec{\gamma}(1) \qquad \& \qquad \vec{\gamma}_{reverse}(1) = \vec{\gamma}(0)$$

Generally, $\vec{\gamma}_{reverse}(t) = \vec{\gamma}(1 - t)$.

We can generalize this construction to other curves. If we are given C from P to Q parametrized by $\vec{\gamma} : [a, b] \to \mathbb{R}^3$ then we can parametrize $-C$ by $\vec{\gamma}_{reverse} : [a, b] \to \mathbb{R}^3$ defined by $\vec{\gamma}_{reverse}(t) = \vec{\gamma}(a + b - t)$. Clearly we have $\vec{\gamma}_{reverse}(a) = \vec{\gamma}(b) = Q$ whereas $\vec{\gamma}_{reverse}(b) = \vec{\gamma}(a) = Q$. Perhaps it is interesting to compare these paths at a common point,

$$\vec{\gamma}(t) = \vec{\gamma}_{reverse}(a + b - t)$$

The velocity vectors naturally point in opposite directions, (by the chain-rule)

$$\frac{d\vec{\gamma}}{dt}(t) = -\frac{d\vec{\gamma}_{reverse}}{dt}(a + b - t).$$

Example 2.4.2. *Suppose* $\vec{\gamma}(t) = \langle \cos(t), \sin(t) \rangle$ *for* $\pi \leq t \leq 2\pi$ *covers the oriented curve* C*. If we wish to parametrize* $-C$ *by* $\vec{\beta}$ *then we can use*

$$\vec{\beta}(t) = \vec{\gamma}(3\pi - t) = \langle \cos(3\pi - t), \sin(3\pi - t) \rangle$$

Simplifying via trigonometry yields $\vec{\beta}(t) = \langle -\cos(t), -\sin(t) \rangle$ *for* $\pi \leq t \leq 2\pi$*. You can easily verify that* $\vec{\beta}$ *covers the lower half of the unit-circle in a CW-fashion, it goes from* (1.0) *to* $(-1, 0)$

What I have just described is a general method to reverse a path whilst keeping the same domain for the new path. Naturally, you might want to use a different domain after you change the parametrization of a given curve. Let's settle the general idea with a definition. This definition describes what we allow as a reasonable reparametrization of a curve.

Definition 2.4.3.

Let $\vec{\gamma}_1 : [a_1, b_1] \to \mathbb{R}^3$ be a path. We say another path $\vec{\gamma}_2 : [a_2, b_2] \to \mathbb{R}^3$ is a **reparametrization** of $\vec{\gamma}_1$ if there exists a bijective (one-one and onto), continuous function $u : [a_1, b_1] \to [a_2, b_2]$ with continuous inverse $u^{-1} : [a_2, b_2] \to [a_1, b_1]$ such that $\vec{\gamma}_1(t) = \vec{\gamma}_2(u(t))$ for all $t \in [a_1, b_1]$. If the given curve is smooth or k-times differentiable then we also insist that the transition function u and its inverse be likewise smooth or k-times differentiable.

In short, we want the allowed reparametrizations to capture the same curve without adding any artificial stops, starts or multiple coverings. If the original path wound around a circle 10 times then we insist that the allowed reparametrizations also wind 10 times around the circle. Finally, let's compare the a path and its reparametrization's velocity vectors, by the chain rule we find:

$$\vec{\gamma}_1(t) = \vec{\gamma}_2(u(t)) \qquad \Rightarrow \qquad \frac{d\vec{\gamma}_1}{dt}(t) = \frac{du}{dt}\frac{d\vec{\gamma}_2}{dt}(u(t)).$$

This calculation is important in the section that follows. Observe that:

1. if $du/dt > 0$ then the paths progress in the same direction and are **consistently oriented**
2. if $du/dt < 0$ then the paths go in opposite directions and are **oppositely oriented**

Reparametrizations with $du/dt > 0$ are said to be **orientation preserving**.

2.4.1 line-integral of scalar function

These are also commonly called the **integral with respect to arclength**. In lecture we framed the need for this definition by posing the question of finding the area of a curved fence with height $f(x, y)$. It stood to reason that the infinitesimal area dA of the curved fence over the arclength ds would simply be $dA = f(x, y)ds$. Then integration is used to sum all the little areas up. Moreover, the natural calculation to accomplish this is clearly as given below:

Definition 2.4.4.

Let $\vec{\gamma} : [a, b] \to C \subset \mathbb{R}^n$ be a differentiable path and suppose that $C \subset dom(f)$ for a continuous function $f : dom(f) \to \mathbb{R}$ then the **scalar line integral of f along** C is

$$\int_C f\ ds \equiv \int_a^b f(\vec{\gamma}(t))\ ||\vec{\gamma}\,'(t)||\ dt.$$

We should check to make sure there is no dependence on the choice of parametrization above. If there was then this would not be a reasonable definition. Suppose $\vec{\gamma}_1(t) = \vec{\gamma}_2(u(t))$ for $a_1 \leq t \leq b_1$ where $u : [a_1, b_1] \to [a_2, b_2]$ is differentiable and strictly monotonic. Note

$$\begin{aligned}\int_{a_1}^{b_1} f(\vec{\gamma}_1(t))\ \left|\left|\frac{d\vec{\gamma}_1}{dt}\right|\right|\ dt &= \int_{a_1}^{b_1} f(\vec{\gamma}_2(u(t)))\ \left|\left|\frac{du}{dt}\frac{d\vec{\gamma}_2}{dt}(u(t))\right|\right|\ dt \\ &= \int_{a_1}^{b_1} f(\vec{\gamma}_2(u(t)))\ \left|\left|\frac{d\vec{\gamma}_2}{dt}(u(t))\right|\right| \cdot \left|\frac{du}{dt}\right|\ dt\end{aligned}$$

If u is orientation preserving then $du/dt > 0$ hence $u(a_1) = a_2$ and $u(b_1) = b_2$ and thus

$$\begin{aligned}\int_{a_1}^{b_1} f(\vec{\gamma}_1(t))\ \left|\left|\frac{d\vec{\gamma}_1}{dt}\right|\right|\ dt &= \int_{a_1}^{b_1} f(\vec{\gamma}_2(u(t)))\ \left|\left|\frac{d\vec{\gamma}_2}{dt}(u(t))\right|\right|\frac{du}{dt}\ dt \\ &= \int_{a_2}^{b_2} f(\vec{\gamma}_2(u))\ \left|\left|\frac{d\vec{\gamma}_2}{du}\right|\right| du.\end{aligned}$$

On the other hand, if $du/dt < 0$ then $|du/dt| = -du/dt$ and the bounds flip since $u(a_1) = b_2$ and $u(b_1) = a_2$

$$\begin{aligned}\int_{a_1}^{b_1} f(\vec{\gamma}_1(t))\ \left|\left|\frac{d\vec{\gamma}_1}{dt}\right|\right|\ dt &= -\int_{a_1}^{b_1} f(\vec{\gamma}_2(u(t)))\ \left|\left|\frac{d\vec{\gamma}_2}{dt}(u(t))\right|\right|\frac{du}{dt}\ dt \\ &= -\int_{b_2}^{a_2} f(\vec{\gamma}_2(u))\ \left|\left|\frac{d\vec{\gamma}_2}{du}\right|\right| du. \\ &= \int_{a_2}^{b_2} f(\vec{\gamma}_2(u))\ \left|\left|\frac{d\vec{\gamma}_2}{du}\right|\right| du.\end{aligned}$$

Note, the definition requires me to flip the bounds before I judge if we have the same result. This is implicit in the statement in the definition that $dom(\vec{\gamma}) = [a, b]$ this forces $a < b$ and hence the integral in turn. Technical details aside we have derived the following important fact:

$$\boxed{\int_C f ds = \int_{-C} f ds}$$

The scalar-line integral of function with no attachment to C is independent of the orientation of the curve. Given our original motivation for calculating the area of a curved fence this is not surprising.

One convenient notation calculation of the scalar-line integral is given by the dot-notation of Newton. Recall that $dx/dt = \dot{x}$ hence $\vec{\gamma} = \langle x, y, z\rangle$ has $\vec{\gamma}\,'(t) = \langle \dot{x}, \dot{y}, \dot{z}\rangle$. Thus, for a space curve,

$$\int_C f\ ds \equiv \int_a^b f(x,y,z)\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}\,dt.$$

We can also calculate the scalar line integral of f along some curve which is made of finitely many differentiable segments, we simply calculate each segment's contribution and sum them together. Just like calculating the integral of a piecewise continuous function with a finite number of jump-discontinuities, you break it into pieces.

Furthermore, notice that if we calculate the scalar line integral of the constant function $f = 1$ then we will obtain the arclength of the curve. More generally the scalar line integral calculates the weighted sum of the values that the function f takes over the curve C. If we divide the result by the length of C then we would have the average of f over C.

Example 2.4.5. *Suppose the linear mass density of a helix $x = R\cos(t), y = R\sin(t), z = t$ is given by $dm/ds = z$. Calculate the total mass around the two twists of the helix given by $0 \le t \le 4\pi$.*

$$\begin{aligned} m_{total\ on\ C} &= \int_C z ds = \int_0^{4\pi} z\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}\ dt \qquad (2.1)\\ &= \int_0^{4\pi} t\sqrt{R^2+1}\ dt \\ &= \frac{t^2\sqrt{R^2+1}}{2}\bigg|_0^{4\pi} \\ &= \boxed{8\pi^2\sqrt{R^2+1}.} \end{aligned}$$

In contrast to total mass we could find the arclength by simply adding up ds, the total length L of C is given by

$$\begin{aligned} L = \int_C ds &= \int_0^{4\pi} \sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}\ dt \\ &= \int_0^{4\pi} \sqrt{R^2+1}\ dt \\ &= \boxed{4\pi\sqrt{R^2+1}.} \end{aligned}$$

Definition 2.4.6.

Let C be a curve with length L then the average of f over C is given by

$$f_{avg} = \frac{1}{L}\int_C f\ ds.$$

Example 2.4.7. *The average mass per unit length of the helix with $dm/ds = z$ as studied above is given by*

$$m_{avg} = \frac{1}{L}\int_C f\ ds = \frac{1}{4\pi\sqrt{R^2+1}}8\pi^2\sqrt{R^2+1} = \boxed{2\pi.}$$

Since $z = t$ and $0 \le t \le 4\pi$ over C this result is hardly surprising.

Another important application of the scalar line integral is to find the center of mass of a wire. The idea here is nearly the same as we discussed for volumes, the difference is that the mass is distributed over a one-dimensional space so the integration is one-dimensional as opposed to two-dimensional to find the center of mass for a planar laminate or three-dimensional to find the center of mass for a volume.

Definition 2.4.8.

Let C be a curve with length L and suppose $dM/ds = \delta$ is the mass-density of C. The total mass of the curve found by $M = \int_c \delta ds$. The **centroid** or **center of mass** for C is found at $(\bar{x}, \bar{y}, \bar{z})$ where

$$\bar{x} = \frac{1}{M}\int_C x\delta\ ds, \qquad \bar{y} = \frac{1}{M}\int_C y\delta\ ds, \qquad \bar{z} = \frac{1}{M}\int_C z\delta\ ds.$$

Often the centroid is found off the curve.

Example 2.4.9. *Suppose $x = R\cos(t), y = R\sin(t), z = h$ for $0 \leq t \leq \pi$ for a curve with $\delta = 1$. Clearly $ds = Rdt$ and thus $M = \int_C \delta ds = \int_0^\pi Rdt = \pi R$. Consider,*

$$\bar{x} = \frac{1}{\pi R}\int_C xds = \frac{1}{\pi R}\int_0^\pi R^2\cos(t)dt = 0$$

whereas,

$$\bar{y} = \frac{1}{\pi R}\int_C yds = \frac{1}{\pi R}\int_0^\pi R^2\sin(t)dt = \frac{1}{\pi R}(-R^2\cos(t)\bigg|_0^\pi = \frac{2R}{\pi}$$

The reader can easily verify that $\bar{z} = h$ hence the centroid is at $(0, \frac{2R}{\pi}, h)$.

Of course there are many other applications, but I believe these should suffice for our current purposes. We will eventually learn that $\int_C \vec{F} \bullet \vec{T} ds$ and $\int_C \vec{F} \bullet \vec{N} ds$ are also of interest, but we should cover other topics before returning to these. Incidentally, it is pretty obvious that we have the following properties for the scalar-line integral:

$$\int_C (f + cg)ds = \int_C fds + c\int_C gds \qquad \& \qquad \int_{C\cup\bar{C}} fds = \int_C fds + \int_{\bar{C}} fds$$

in addition if $f \leq g$ on C then $\int_C fds \leq \int_C gds$. I leave the proof to the reader.

Remark 2.4.10.

I have a few solved problems on integrals along a curve and centroids. They are attached to a later Chapter. See Problems 187, 188, 189.

2.5 Problems

Problem 46 Calculate the following:

(a.) $\frac{d}{dt}\langle t^2, e^t, \ln(t)\rangle$

(b.) $\frac{d}{dt}\langle \cosh(t^2), \sinh(\ln(t))\rangle$

(c.) $\int \langle 1, t, \sin(t)\rangle dt$

Problem 47 Let $\vec{g}, \vec{r}_o, \vec{v}_o$ be constant vectors. Let $\vec{r}(t) = \vec{r}_o + t\vec{v}_o + \frac{1}{2}t^2\vec{g}$ and calculate:

(a.) $\frac{d}{dt}\big[\,\vec{r}\,\big]$

(b.) $\frac{d^2}{dt^2}\big[\,\vec{r}\,\big]$

(c.) $\frac{d^3}{dt^3}\big[\,\vec{r}\,\big]$

Problem 48 Suppose $\vec{r}(t) = \langle \cos(\pi t), \sin(\pi t), t/8 \rangle$ for $t \geq 0$.

(a.) How many revolutions does this helix make at it travels from $z = 0$ to $z = 1$.

(b.) Calculate $d\vec{r}/dt$

(c.) Find the arclength of this curve from $z = 0$ to $z = 1$.

Problem 49 Find the total arclength of $\vec{r}(t) = 2\cos(t)\,\widehat{x} + t\,\widehat{y} + 2\sin(t)\,\widehat{z}$ for $0 \leq t \leq 4\pi$. Also, find the arclength function and reparametrize this helix in terms of the arclength.

Problem 50 Find the arclength functions for $t \geq 0$ for:

(a.) $\vec{r}(t) = \langle e^{-t}, 1 - e^{-t} \rangle$

(b.) $\vec{r}(t) = (2 - 3t)\,\widehat{x} + (1 + t)\,\widehat{y} - 4t\,\widehat{z}$

Problem 51 Let R be a fixed positive constant. Suppose $\vec{r}(t) = \langle R\cos(t), 42, R\sin(t) \rangle$ for $0 \leq t \leq 2\pi$. Calculate, and simplify, the tangent, normal and binormal vector fields for the given path.

Problem 52 Calculate the curvature of the curve given in Problem 51.

Problem 53 Calculate the torsion of the curve given in Problem 51.

Problem 54 Suppose . $x = e^{-t}\cos(t)$ and $y = e^{-t}\sin(t)$ and $z = e^{-t}$ for $0 \leq t \leq 4\pi$. Calculate and simplify the tangent, normal and binormal vector fields for the curve parametrized by the given scalar parametric equations.

Problem 55 Calculate the curvature of the curve given in Problem 54.

Problem 56 Calculate the torsion of the curve given in Problem 54.

Problem 57 Find the point on the curve $y = 1/x$ for which the curvature is maximized. You may focus your efforts on the part of the curve with $x > 0$. (you can use the formula from Stewart for curvature if you wish... I'll probably attack it from my notes, it is likely the standard $y = f(x)$ curvature formula makes this easier)

Problem 58 * Suppose it is given that the three dimensional vectors $\vec{A}, \vec{B}$ are orthogonal vectors with $A = B = 1$. Show that if $\vec{A} \times \vec{B} = \vec{C}$ then $\vec{B} \times \vec{C} = \vec{A}$ and $\vec{C} \times \vec{A} = \vec{B}$.

Problem 59 * Show that $c_{12} = -c_{21}$ as defined in the discussion on pages 112-113 of my notes.

Problem 60 Suppose that you are given a path with torsion which is identically zero at all points on the path. Show that this path parametrizes a curve in a plane.

Problem 61 * A force is said to be central if it is directed along the line connecting a central point and has a magnitude which depends on on the distance from the center. For convenience put this force center at the origin thus we have $\vec{F}(x, y, z) = F(\rho)\widehat{\rho}$. If a mass m is subject to this central force alone then show that $\vec{F} = m\vec{a}$ implies the motion is planar.

Problem 62 By the previous problem we find the orbital motion of a particular planet revolving around the sun must lie in a plane since Newton's universal law of gravitation is a central force. To a good approximation we can take the sun as motionless at the center of the solar system. Place the center of the sun at the origin and use xy-coordinates to label the plane of orbital motion. Also, use polar coordinates in the same plane. We can write the force of gravity on a planetat position $\vec{r}$ as follows:

$$\vec{F} = \frac{GmM}{r^2}\widehat{r} = -\frac{GmM}{r^3}\vec{r}.$$

Show that the angular momentum of the planet is conserved. Recall that the angular momentum of the planet is simply given by $\vec{L} = m\vec{r} \times \vec{v}$. Also, recall that Newton's second law for the planet states $\vec{F} = m\frac{d\vec{v}}{dt}$. You'll need all these facts together with the formula for gravity above to show that $\frac{d\vec{L}}{dt} = 0$.

Problem 63 Suppose two ninja begin travelling the paths given below. To begin, at $t = 0$, a relatively slow genin level ninja sets off in a NE direction given by

$$\vec{r}_1(t) = \langle -10 + t, 1 + t \rangle.$$

However, at the same time $t = 0$, an enemy Jonin sets off in a NW direction given by

$$\vec{r}_2(t) = \langle 20 - 4t, 6 + t \rangle.$$

Both of these paths are placed in a forest thick with a mist which lowers visability to near zero. Suppose the Jonin level ninja has advanced tracking skills that allow his to pick up on the faintest of scents. If he crosses the path of an enemy he can smell it and then alter his path to pursue and attack the enemy genin. Should the genin worry? Is he in danger? (a Jonin is no match for a typical genin in a usual battle, if the Jonin catches the genin it's game over for the lowly genin)

Problem 64 Suppose the velocity is given by $\vec{v}(t) = \langle t, 3, t\cosh(t^2) \rangle$ for some particle which has initial position $(1, 2, 3)$. Find the acceleration and position at time $t \geq 0$.

Problem 65 Suppose $\vec{a} = 3\,\widehat{x}$. Given this constant acceleration, derive the velocity and position as functions of time. Please state your answers in terms of arbitrary initial position $\vec{r}(0) = \vec{r}_o$ and initial velocity $\vec{v}(0) = \vec{v}_o$.

Problem 66 Suppose $\vec{r}(t) = \langle 2^t, \ln(t), \sqrt{t^2+1} \rangle$. Find the parametric equations of the tangent line to the given curve at the point $(8, \ln(8), \sqrt{10})$.

Problem 67 Flashback: calculate: $\int \sin^n(x)\ ,dx$ for $n = 2$ and $n = 3$. (by hand, no Mathematica here please, at least not in your solution)

Problem 68 Flashback: calculate: (simplify answer please, assume $t > 0$)

$$\frac{d}{dt}\Big[2t\cosh(2\ln(t))\Big]$$

Problem 69 Suppose $\vec{r}(t) = \langle 2^{-t}, 3\sin(t), 4t \rangle$ for $t \geq 0$ describes the path travelled by a hover car. Set-up, but do not evaluate, the distance travelled by time t for this path. Also, if this path was given in units of miles per second and the hover craft police placed a speed limit of 6mps on hover cars then is there any chance an honest hover cop will give you a ticket? Try to estimate a reasonable upper bound on the speed of the given hover car path.

Chapter 3

multivariate limits

Calculus for vector valued functions of one real variable was easy. Essentially, we just did first semester calculus in each component. The truly new material in some sense begins here. The problem of differentiation for functions of several variables requires we develop a new formulation of the *derivative*. However, before we can even attempt such a process we must first settle the more basic question:

What is the concept of a limit for functions of several variables ?

Suppose $f : \mathbb{R}^n \to \mathbb{R}$. What should it mean for $f(\vec{r}) \to L$ as $\vec{r} \to \vec{a}$? This is carefully defined in Section 3.2. In short, we just take the $\epsilon - \delta$-definition from first semester calculus and replace absolute values with norm[1].

We begin the chapter with a brief discussion of the euclidean topology of $\mathbb{R}^n$ in Section 3.1. This gives precise meaning for a set being an *open set* or a *closed set*. Also, we give discuss the meaning for a point to be an *interior point* or *boundary point* and most relevant to this chapter we define a **limit point**. The terminology of this section is used from time to time throughout the remainder of this course.

The results of Section 3.2 are easily summarized: if the formula for a function of several variables is free of undefined or non-real quantities then we can just plug in the limit point to find the value of the limit. We do not supply proof of these assertions in this course. The interested student is refered to advanced calculus[2]

Section 3.3 deals with the less clear cases. What should we do if we cannot just substitute the limit point into the limited expression ? Such cases are called **indeterminant**. Once more, the troubles we faced in first semester calculus are still with us. We face indeterminant forms $0/0, 0 \cdot \infty, \infty/\infty, \infty - \infty$ as well as $0^0, 1^\infty, \infty^0$. Let us summarize:

(i.) **methods for calculation of convergent limits:** Direct computation by algebra is still at times possible. There are examples where we may do algebra to remove the indeterminancy of the limit. There is a multivariate squeeze theorem which allows solution for many problems. Conveniently, for two-dimensional limits, we have a new tool at our disposal; substitution

[1]or vector length if you prefer not to use the term "norm", although, norms are normal.

[2]but, admittedly, even there I don't belabor these results for long, the proper, long-winded, study of continuity of functions is left to our topology course

of polar coordinates. It is sometimes the case that when we change the limit to polar coordinates the limit simplifies to a single-variate limit.

(ii.) **methods for ascertaining divergence of a given limit:** if a function is not bounded near the limit point then divergence to $\pm\infty$ are possible. We still say the limit does not exist when the limit tends toward ∞ on one-side and $-\infty$ on the another side. However, this is the essential difference between single and multivariate limits: in the case of two or more variables, the number of paths to approach a limit point is **infinite**. In contrast, for one-dimensional domains there is just left and rightor if you are more vertically-minded up and down. In particular, you may recall: $\lim_{x\to a^+} f(x) = L$ and $\lim_{x\to a^-} f(x) = L$ if and only if $\lim_{x\to a} f(x) = L$. Therefore, in the one-dimensional case we could break into two cases and work each separately. This is no longer possible. You might expect, if we have limits from two independent directions that ought to be enough to conclude convergence. But, that is not enough. Ok, what if we know **all** linear paths yield the same limit, is that enough? No. I illustrate how this happens and try to give some sense of just how subtle the problem is here. This issue arises again in the next chapter as we study the subtle distinction between mere differentiability and continuous differentiability. Subtle points aside, often we simply want the reader to realize that if just **two** particular paths disagree on the limiting value then we may conclude the limit does not exist. Once more, for two-dimensional problems, polar coordinate substitution can be used to give some guidance to pick the paths of differing limiting value.

The verity of the substitution concept is implicit within the results of Section 3.2. In fact, the concept of limit calculation via substitution is a topic which could be expanded in far greater generality than we attempt in this meager chapter.

The first two sections of this chapter will not be covered in lecture in their entirety. However, I would like you to read through it and try to get the general idea. You should get a good sense of what parts of the first two sections will be tested from the discussion in lecture. The third section contains the problems that are typically tested on this material in a generic calculus III course.

3.1 basic euclidean topology

In this section we describe the *euclidean topology* for $\mathbb{R}^n$. In the study of functions of one real variable we often need to refer to open or closed intervals. The definition that follows generalizes those concepts to n-dimensions.

Definition 3.1.1.

An **open ball** of radius ϵ centered at $\vec{a} \in \mathbb{R}^n$ is the subset of all points in $\mathbb{R}^n$ which are less than ϵ units from $\vec{a}$, we denote this open ball by

$$B_\epsilon(\vec{a}) = \{\vec{x} \in \mathbb{R}^n \mid ||\vec{x} - \vec{a}|| < \epsilon\}$$

The **closed ball** of radius ϵ centered at $\vec{a} \in \mathbb{R}^n$ is likewise defined

$$\overline{B}_\epsilon(\vec{a}) = \{\vec{x} \in \mathbb{R}^n \mid ||\vec{x} - \vec{a}|| \leq \epsilon\}$$

Notice that in the $n = 1$ case we observe an open ball is an open interval: let $a \in \mathbb{R}$,

$$B_\epsilon(a) = \{x \in \mathbb{R} \mid ||x - a|| < \epsilon\} = \{x \in \mathbb{R} \mid |x - a| < \epsilon\} = (a - \epsilon, a + \epsilon)$$

In the $n = 2$ case we observe that an open ball is an open disk: let $(a, b) \in \mathbb{R}^2$,

$$B_\epsilon((a, b)) = \big\{(x, y) \in \mathbb{R}^2 \mid ||\,(x, y) - (a, b)\,|| < \epsilon\big\} = \big\{(x, y) \in \mathbb{R}^2 \mid \sqrt{(x - a)^2 + (y - b)^2} < \epsilon\big\}$$

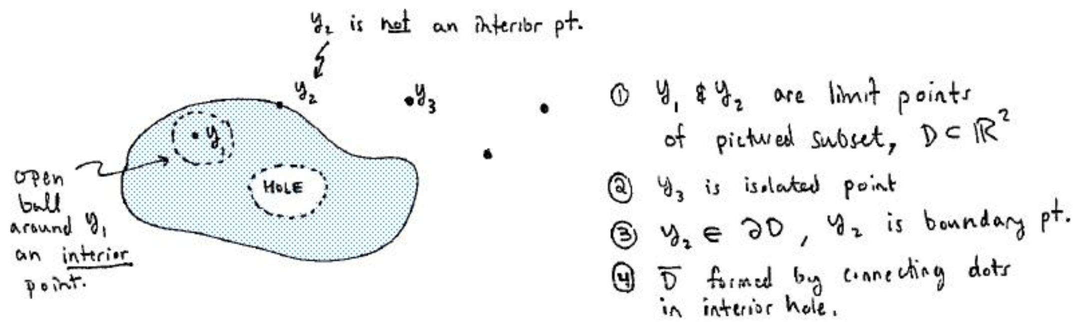
For $n = 3$ an open-ball is a sphere without the outer shell. In contrast, a closed ball in $n = 3$ is a solid sphere which includes the outer shell of the sphere.

$$B_\epsilon((a, b, c)) = \big\{(x, y, z) \in \mathbb{R}^3 \mid \sqrt{(x - a)^2 + (y - b)^2 + (z - c)^2} < \epsilon\big\}$$

Definition 3.1.2.

Let $D \subseteq \mathbb{R}^n$. We say $\vec{y} \in D$ is an **interior point** of D iff there exists some open ball centered at $\vec{y}$ which is completely contained in D. We say $\vec{y} \in \mathbb{R}^n$ is a **limit point** of D iff every open ball centered at $\vec{y}$ contains points in $D - \{\vec{y}\}$. We say $\vec{y} \in \mathbb{R}^n$ is a **boundary point**[3] of D iff every open ball centered at $\vec{y}$ contains points not in D and other points which are in $D - \{\vec{y}\}$. We say $\vec{y} \in D$ is an **isolated point** of D if there exist open balls about $\vec{y}$ which do not contain other points in D. The set of all interior points of D is called the **interior of** D. Likewise the set of all boundary points[4] for D is denoted ∂D. The **closure** of D is defined to be $\overline{D} = D \cup \{\vec{y} \in \mathbb{R}^n \mid \vec{y} \text{ a limit point}\}$

If you're like me the paragraph above doesn't help much until I see the picture below. All the terms are aptly named. The term "limit point" is given because those points are the ones for which it is natural to define a limit.



Definition 3.1.3.

Let $A \subseteq \mathbb{R}^n$ is an **open set** iff for each $\vec{x} \in A$ there exists $\epsilon > 0$ such that $\vec{x} \in B_\epsilon(\vec{x})$ and $B_\epsilon(\vec{x}) \subseteq A$. Let $B \subseteq \mathbb{R}^n$ is an **closed set** iff its complement $\mathbb{R}^n - B = \{\vec{x} \in \mathbb{R}^n \mid \vec{x} \notin B\}$ is an open set.

Notice that $\mathbb{R} - [a, b] = (\infty, a) \cup (b, \infty)$. It is not hard to prove that open intervals are open hence we find that a closed interval is a closed set. Likewise it is not hard to prove that open balls are open sets and closed balls are closed sets.

3.2 the multivariate limit and continuity

The definition of the limit here is the natural generalization of the $\epsilon\delta$-defn. we studied in before.

Definition 3.2.1.

Let $\vec{f} : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$ be a mapping. We say that $\vec{f}$ has limit $\vec{b} \in \mathbb{R}^m$ at limit point $\vec{a}$ of U iff for each $\epsilon > 0$ there exists a $\delta > 0$ such that $\vec{x} \in \mathbb{R}^n$ with $0 < ||\vec{x} - \vec{a}|| < \delta$ implies $||\vec{f}(\vec{x}) - \vec{b}|| < \epsilon$. In such a case we can denote the above by stating that

$$\lim_{\vec{x} \to \vec{a}} \vec{f}(\,\vec{x}\,) = \vec{b}.$$

In single variable calculus the limit of a function is defined in terms of deleted open intervals centered about the limit point. We just defined the limit of a mapping in terms of deleted open balls centered at the limit point. The term "deleted" refers to the fact that we assume $0 < ||\vec{x} - \vec{a}||$ which means we do not consider $\vec{x} = \vec{a}$ in the limiting process. In other words, the limit of a mapping considers values close to the limit point but not necessarily the limit point itself. The case that the function is defined at the limit point is special, when the limit and the mapping agree then we say the mapping is continuous at that point.

Definition 3.2.2.

Let $\vec{f} : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$ be a mapping. If $\vec{a} \in U$ is a limit point of $\vec{f}$ then we say that $\vec{f}$ is **continuous at** $\vec{a}$ iff

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

If $\vec{a} \in U$ is an isolated point then we also say that $\vec{f}$ is continous at $\vec{a}$. The mapping $\vec{f}$ is **continous on** S iff it is continous at each point in S. The **mapping** $\vec{f}$ **is continuous** iff it is continuous on its domain.

Notice that in the $m = n = 1$ case we recover the definition of continuous functions on $\mathbb{R}$. In practice we seldom calculate a multivariate limit in terms of an $\epsilon - \delta$ argument. I include these to better illustrate just how the definition is directly implemented. In fact, to those of you who are not math majors feel free to skip the proofs in the remainder of this section. I do think it is worthwhile for everyone to at least read the results which are known about continuity of multivariate functions. The section that follows contains problems which I expect you to be able to work yourself by the next test. This section is for breadth and depth of concept.

Example 3.2.3. Claim: the identity function $Id : \mathbb{R}^n \to \mathbb{R}^n$ defined by $Id(\vec{x}) = \vec{x}$ is continuous on $\mathbb{R}^n$.

Proof: *Let $\epsilon > 0$ and choose $\delta = \epsilon$. If $\vec{x} \in \mathbb{R}^n$ such that $0 < ||\vec{x} - \vec{a}|| < \delta$ then it follows that $||\vec{x} - \vec{a}|| < \epsilon$.. Therefore, $\lim_{\vec{x} \to \vec{a}} \vec{x} = \vec{a}$ which means that $\lim_{\vec{x} \to \vec{a}} Id(\vec{x}) = Id(\vec{a})$ for all $\vec{a} \in \mathbb{R}^n$. Hence Id is continuous on $\mathbb{R}^n$ which means Id is continuous.*

Example 3.2.4. Claim: the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(\vec{x}) = ||\vec{x}||^2$ is continuous on $\mathbb{R}^n$. *To prepare for the proof consider we must show that for an appropriately chosen δ the* *condition* $||\vec{x}-\vec{a}|| < \delta$ *implies the difference* $|f(\vec{x})-f(\vec{a})| < \epsilon$. *Suppose* $||\vec{x}-\vec{a}|| < \delta$ *and observe:*

$$\begin{aligned}
|f(\vec{x})-f(\vec{a})| &= |\vec{x}\bullet\vec{x}-\vec{a}\bullet\vec{a}| \\
&= |\vec{x}\bullet\vec{x}-\vec{x}\bullet\vec{a}+\vec{a}\bullet\vec{x}-\vec{a}\bullet\vec{a}| \\
&= |\vec{x}\bullet(\vec{x}-\vec{a})+\vec{a}\bullet(\vec{x}-\vec{a})| \\
&= |(\vec{x}+\vec{a})\bullet(\vec{x}-\vec{a})| \\
&\leq ||\vec{x}+\vec{a}||\,||\vec{x}-\vec{a}|| \\
&< \delta\,||\vec{x}+\vec{a}|| \\
&= \delta\,||\vec{x}-\vec{a}+2\vec{a}|| \\
&\leq \delta\left[||\vec{x}-\vec{a}||+2||\vec{a}||\right] \\
&< \delta\left[\delta+2||\vec{a}||\right] \qquad \star
\end{aligned}$$

This suggests we choose δ *such that* $\delta^2+2\delta||\vec{a}|| \leq \epsilon$. *Let's go for equality, let* $||\vec{a}|| = a$ *and solve* $\delta^2+2a\delta-\epsilon = 0$ *to find* $\delta = -a\pm\sqrt{a^2+\epsilon}$. *The solution* $\delta = -a+\sqrt{a^2+\epsilon}$ *is clearly positive for* $\epsilon > 0$.

Proof: *Let* $\epsilon > 0$ *and let* $\vec{a} \in \mathbb{R}^n$ *with* $||\vec{a}|| = a$ *choose* $\delta = -a+\sqrt{a^2+\epsilon}$ *which is clearly positive. Suppose* $\vec{x} \in \mathbb{R}^n$ *and* $0 < ||\vec{x}-\vec{a}|| < \delta$ *and calculate, following the* $\star$ *calculation,*

$$|f(\vec{x})-f(\vec{a})| = \delta^2+2a\delta = \epsilon.$$

Therefore, by the definition of the limit, $\lim_{\vec{x}\to\vec{a}} ||\vec{x}||^2 = ||\vec{a}||^2$. *and since* $\vec{a}$ *is arbitrary this shows* f *is continuous on all of* $\mathbb{R}^n$. $\square$

The examples that follow are somewhat abstract, but their use is astounding once they're paired with a couple basic theorems about the multivariate limit.

Proposition 3.2.5.

Let $\vec{f} : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$ be a mapping with component functions $f_1, f_2, \dots, f_m$ hence $\vec{f} = (f_1, f_2, \dots, f_m)$. If $\vec{a} \in U$ is a limit point of $\vec{f}$ then

$$\lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x}) = \vec{b} \qquad \Leftrightarrow \qquad \lim_{\vec{x}\to\vec{a}} f_j(\vec{x}) = b_j \text{ for each } j = 1, 2, \dots, m.$$

.

Proof: ($\Rightarrow$) Suppose $\lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x}) = \vec{b}$. Then for each $\epsilon > 0$ choose $\delta > 0$ such that $0 < ||\vec{x}-\vec{a}|| < \delta$ implies $||\vec{f}(\vec{x})-\vec{b}|| < \epsilon$. This choice of δ suffices for our purposes as:

$$|f_j(\vec{x})-b_j| = \sqrt{(f_j(\vec{x})-b_j)^2} \leq \sqrt{\sum_{j=1}^{m}(f_j(\vec{x})-b_j)^2} = ||\vec{f}(\vec{x})-\vec{b}|| < \epsilon.$$

Hence we have shown that $\lim_{\vec{x}\to\vec{a}} f_j(\vec{x}) = b_j$ for all $j = 1, 2, \dots m$.

($\Leftarrow$) Suppose $\lim_{\vec{x}\to\vec{a}} f_j(\vec{x}) = b_j$ for all $j = 1, 2, \dots m$. Let $\epsilon > 0$. Note that $\epsilon/m > 0$ and therefore by the given limits we can choose $\delta_j > 0$ such that $0 < ||\vec{x}-\vec{a}|| < \delta$ implies $||f_j(\vec{x})-b_j|| < \sqrt{\epsilon/m}$. Choose $\delta = min\{\delta_1, \delta_2, \dots, \delta_m\}$ clearly $\delta > 0$. Moreoever, notice $0 < ||\vec{x}-\vec{a}|| < \delta \leq \delta_j$ hence requiring $0 < ||\vec{x}-\vec{a}|| < \delta$ automatically induces $0 < ||\vec{x}-\vec{a}|| < \delta_j$ for all j. Suppose that $\vec{x} \in \mathbb{R}^n$ and $0 < ||\vec{x}-\vec{a}|| < \delta$ it follows that

$$||\vec{f}(\vec{x})-\vec{b}|| = ||\sum_{j=1}^{m}(f_j(\vec{x})-b_j)e_j|| = \sqrt{\sum_{j=1}^{m}|f_j(\vec{x})-b_j|^2} \leq \sum_{j=1}^{m}(\sqrt{\epsilon/m})^2 < \sum_{j=1}^{m}\epsilon/m = \epsilon.$$

Therefore, $\lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x}) = \vec{b}$ and the proposition follows. $\square$

Remark 3.2.6.

Notice, one simple application of the Proposition above is that $\lim_{h\to 0}(\lim_{h\to 0} f(h), \lim_{h\to 0} g(h))$ provided the limits exist. This vector limit law is what is behind my glib assertion that the difference quotient of a path descends naturally to a vector of difference quotients of the components: using $\vec{r} = \langle x, y\rangle : \mathbb{R} \to \mathbb{R}^2$,

$$\lim_{h\to 0} \frac{\vec{r}(t+h) - \vec{r}(t)}{h} = \left(\lim_{h\to 0} \frac{x(t+h) - x(t)}{h}, \lim_{h\to 0} \frac{y(t+h) - y(t)}{h} \right)$$

Thus $\frac{d\vec{r}}{dt} = \left(\frac{dx}{dt}, \frac{dy}{dt}\right)$. In other words, my lazy Definition 2.1.1 naturally derives from the more natural limiting definition.

We can analyze the limit of a mapping by analyzing the limits of the component functions:

Example 3.2.7. *Let* $\vec{F}(x) = \langle \sqrt{\tan^2(x) + 1},\ \cos x,\ \frac{\sin x}{x}\rangle$ *then we calculate,*

$$\begin{aligned}
\lim_{x\to 0} \vec{F}(x) &= \lim_{x\to 0} \left\langle \sqrt{\tan^2(x) + 1}, \cos x, \frac{\sin x}{x} \right\rangle \\
&= \left\langle \lim_{x\to 0} \sqrt{\tan^2(x) + 1},\ \lim_{x\to 0} \cos x,\ \lim_{x\to 0} \frac{\sin x}{x} \right\rangle \\
&= \langle 1, 1, 1\rangle.
\end{aligned}$$

The following follows immediately from the preceding proposition.

Proposition 3.2.8.

Suppose that $\vec{f} : U \subseteq \mathbb{R}^n \to V \subseteq \mathbb{R}^m$ is a mapping with component functions $f_1, f_2, \dots, f_m$. Let $\vec{a} \in U$ be a limit point of $\vec{f}$ then $\vec{f}$ is continous at $\vec{a}$ iff f_j is continuous at $\vec{a}$ for $j = 1, 2, \dots, m$. Moreover, $\vec{f}$ is continuous on S iff all the component functions of $\vec{f}$ are continuous on S. Finally, a mapping $\vec{f}$ is continous iff all of its component functions are continuous. .

Proposition 3.2.9.

The cartesian coordinate functions are continuous. The identity mapping is continuous.

Proof: Since the cartesian coordinate functions are component functions of the identity mapping it follows that the coordinate functions are also continuous (using the previous proposition and the fact we showed the identity function was continuous earlier in this chapter). $\square$

Many things you saw for functions on $\mathbb{R}$ are true for mappings from $\mathbb{R}^m$ to $\mathbb{R}^n$. You can pull limits inside the arguments of appropriately continuous maps[5]. The composite of continuous mappings is continuous. Continuing this investigation we'd eventually reach the same intuition as was found in first semester calculus: when the formula for the function has no ambiguity such as division by zero or unboundedness then the map in question is continuous at the limit point and the value of the limit is simply obtained by evaluation. In contrast, the next section deals with the problem of indeterminant limit types. In the multivariate theory, there are genuinely new features.

[5] see C.H. Edward's *Advanced Calculus of Several Variables* pages 46-47 for example

3.3 multivariate indeterminants

Explicit calculations which show multivariate limits do not exist are important to think about. They bring added understanding to the constructions we've thus far endured. Also, from the perspective of the student, they are important since they are common test[6] questions.

I can summarize the results of the last section with a simple slogan: *the limit of a multivariate function is given by evaluation at the limit point provided the evaluation does not violate the laws of arithmetic*. In other words, you can solve the limit by just plugging in the limit point if there is no division by zero, even root of a negative number and/or inputs outside the domain of the elementary functions. We offer some methods to evaluate indeterminant limits and we will see how to show the limit does not exist.

In single variable calculus we learned that the double-sided limit exists iff both the left and right limits exist and are equal. However, consider $f : dom(f) \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$. If we look at a limit point (a, b) then there are infinitely many paths in $dom(f)$ which approach the limit point. Suppose we have a path $t \mapsto \vec{r}(t)$ with $\vec{r}(0) = (a, b)$ and $\lim_{t\rightarrow 0} \vec{r}(t) = (a, b)$. Then, for a real number L,

$$\boxed{\lim_{(x,y)\rightarrow(a,b)} f(x, y) = L \quad \Rightarrow \quad \lim_{t\rightarrow 0} f(\vec{r}(t)) = L.}$$

A direct consequence of the implication boxed above is that if we obtain different limits for two different paths through (a, b) then the limit of $f(x, y)$ as $(x, y) \rightarrow (a, b)$ does not exist. The examples below show how to pragmatically and orderly use this boxed equation to show limits fail to exist. I found these examples in Thomas' Calculus, which, depending on which edition you look at, can be an excellent text.

Example 3.3.1. *Suppose* $f(x, y) = \begin{cases} \frac{2xy}{x^2+y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0) \end{cases}$*. Notice that we can calculate the limit for* $(a, b) \neq (0, 0)$ *with ease:*

$$\lim_{(x,y)\rightarrow(a,b)} = \frac{2ab}{a^2 + b^2}.$$

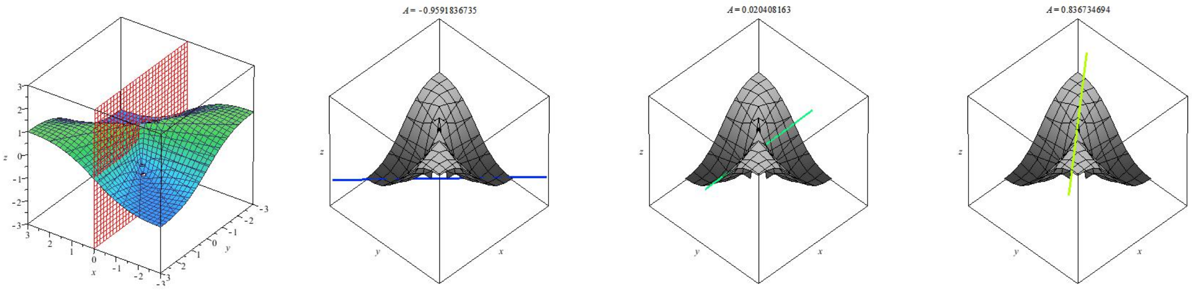
However, if we consider the limit at $(0, 0)$ *it is indeterminant since we have an expression of type* $0/0$. *Other calculation is required. Consider the path* $\vec{r}(t) = (t, mt)$ *then clearly this is continuous at* $t = 0$ *and* $\vec{r}(0) = (0, 0)$*; in-fact, this is just the parametric equation of a line* $y = mx$. *Consider,*

$$\lim_{t\rightarrow 0} f(\vec{r}(t)) = \lim_{t\rightarrow 0} \frac{2mt^2}{t^2 + m^2t^2} = \lim_{t\rightarrow 0} \frac{2m}{1 + m^2} = \frac{2m}{1 + m^2}.$$

The proposed limit $L = \frac{2m}{1+m^2}$ *depends nontrivially on* m *which means that paths with different* m *yield different limits. For example,* $m = 1$ *suggests* $L = 1$ *whereas* $m = -1$ *yields* $L = -1$. *It follows that the limit does not exist. Here's the graph of this function, maybe you can see the problem at the origin. The red plane is vertical through the origin. The three pictures on the right illustrate how differing linear paths yield differing limits*[7]

[6] I do talk about things which are not on the test because this course is more than the tests or the homework, those are just tools to get you to start thinking, those are not the end, there is no end, this is the essence of what university education should be, an invitation to think, not just a path to a degree.

[7] I'm using the function in Maple called "zhue" which colors an object corresponding to it's z-values.



Curious, what if all linear-paths through the limit point yield the same limiting value? Is that a sufficient criteria to obtain a converse to the boxed implication? It seems plausible, if we look at all lines through the limit point then we've covered a neighborhood of the limit point with our analysis. Seems like we have a real chance. Well, until you study the next example:

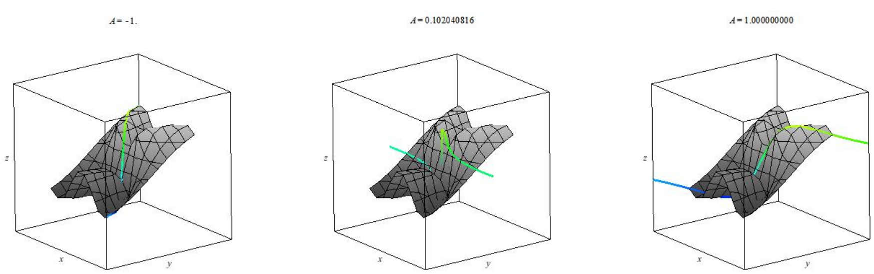
Example 3.3.2. *Suppose* $f(x,y) = \begin{cases} \frac{2x^2y}{x^4+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. *Notice that we can calculate the limit for* $(a,b) \neq (0,0)$ *with ease:*

$$\lim_{(x,y)\to(a,b)} = \frac{2a^2b}{a^4+b^2}.$$

However, if we consider the limit at $(0,0)$ *it is indeterminant since we have an expression of type* $0/0$. *Other calculation is required. Consider the path* $\vec{r}(t) = (t, mt)$ *then clearly this is continuous at* $t = 0$ *and* $\vec{r}(0) = (0,0)$*; in-fact, this is just the parametric equation of a line* $y = mx$. *Consider, for* $m \neq 0$,

$$\lim_{t\to 0} f(\vec{r}(t)) = \lim_{t\to 0} \frac{2mt^3}{t^4+m^2t^2} = \lim_{t\to 0} \frac{2mt}{t^2+m^2} = \frac{2m(0)}{0+m^2} = 0.$$

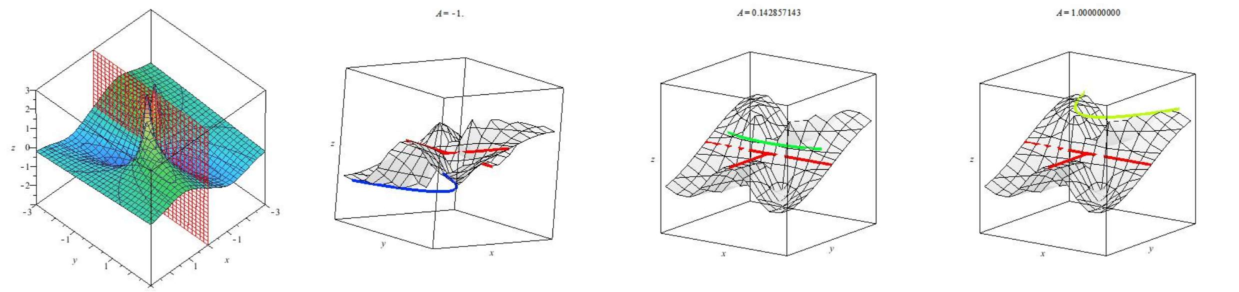
If $\vec{r}(t) = (t,0)$ *then for* $t \neq 0$ *we have* $f(\vec{r}(t)) = f(t,0) = 0$ *thus the limit of the function restricted to any linear path is just zero. The three pictures on the right illustrate how differing linear paths yield the same limits. The red lines are the* x, y *axes.*



What about parabolic paths? Those are easily constructed via $\vec{r}_2(t) = (t, kt^2)$ *again* $\vec{r}_2(0) = (0,0)$ *and* $\lim_{t\to 0} \vec{r}_2(t) = (0,0)$. *Calculate, for* $k \neq 0$,

$$\lim_{t\to 0} f(\vec{r}_2(t)) = \lim_{t\to 0} \frac{2kt^4}{t^4+k^2t^4} = \lim_{t\to 0} \frac{2k}{1+k^2} = \frac{2k}{1+k^2}.$$

Clearly if we choose differing values for k *we obtain different values for the limit hence the limit of* f *does not exist as* $(x,y) \to (0,0)$. *Here's the graph of this function, maybe you can see the problem at the origin. The red plane is vertical through the origin. The three pictures on the right illustrate how differing parabolic paths yield differing limits. The red lines are the* x, y *axes.*



I believe if we knew that the limit of $f \circ \vec{r}$ existed and were equal to L for all possible continuous paths then the multivariate limit of f would exist. However, I have not included proof in these notes at the present time. If you can prove or disprove this claim I'd be interested. It turns out that a multivariate function which is differentiable is consqeuently continuous. Thus, in the next chapter we find another tool for indirectly analyzing continuity. In any event, I hope this pair of examples gives you the idea. Moreover, while I have illustrated the concept for $f : \mathbb{R}^2 \to \mathbb{R}$ the techniques above equally well apply to indeterminant limits of $f : \mathbb{R}^3 \to \mathbb{R}$ or with a bit more imagination $f : \mathbb{R}^n \to \mathbb{R}$.

If you find the calculations of this last pair of examples a bit disheartening then you have my sympathy. Fortunately, there is another way to look at these examples and the technique will bring us to examples which are both indeterminant in their initial formulation and finite once the indeterminant form is resolved. The trick is coordinate change. Can we trade x, y for new coordinates which simplify the expression? Polar coordinates suffice for the examples above.

Example 3.3.3. *Suppose* $f(x,y) = \begin{cases} \frac{2xy}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. *I argue it is intuitively clear the substitution below is valid:*

$$\lim_{(x,y)\to(0,0)} \frac{2xy}{x^2+y^2} = \lim_{r\to 0} \frac{2r^2\cos\theta\sin\theta}{r^2} = \lim_{r\to 0}\bigg(2\cos\theta\sin\theta\bigg).$$

Therefore, the limit as $(x,y) \to (0,0)$ *of* f *does not exist since as* $r \to 0$ *the function tends to* $2\cos\theta\sin\theta$ *which is not single-valued at the origin. What value of* θ *is assigned to the origin in polar coordinates?*

You might complain that I am using polar coordinates precisely where they fail to be defined. However, I would argue it's reasonable since the limiting process considers points near the limit point but not the limit point itself. Polar coordinates are uniquely defined for points near the origin, just not the origin itself because the angle is infinitely-many valued at the origin. In practice we don't face trouble from this for a variety of reasons, but it is a fact that the origin is not labled in the same way as all the other points in the plane by polar coordinates.

Example 3.3.4. *Suppose* $f(x,y) = \begin{cases} \frac{2x^2y}{x^4+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. *Again, we use polar coordinate substitution,*

$$\lim_{(x,y)\to(0,0)} \frac{2x^2y}{x^4+y^2} = \lim_{r\to 0} \frac{2r^3\cos^2\theta\sin\theta}{r^4\cos^4\theta + r^2\sin^2\theta} = \lim_{r\to 0} \frac{2r\cos^2\theta\sin\theta}{r^2\cos^4\theta + \sin^2\theta}.$$

If we choose $\theta = \pi/4$ *then the limit above clearly tends to zero however if we consider the spiral path* $\theta = r$ *then we have* $\frac{2r\cos^2 r\sin r}{r^2\cos^4 r + \sin^2 r} \to \frac{2rr}{r^2+r^2} = 1$. *Therefore, the limit does not exist. This example is a* *bit subtle in any coordinate system, notice that looking at various choices for $\theta = const$ corresponds to sorting through lines of various slope. All of these linear paths, or rays in our current context, lead to the apparent triviality of the limit.*

If there is no θ-dependence after changing to polar coordinates then the analysis simplifies.

Example 3.3.5. *Suppose* $f(x,y) = \begin{cases} \frac{x^2+y^2}{x^4+2x^2y^2+y^4} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. *Again, we use polar coordinate substitution,*

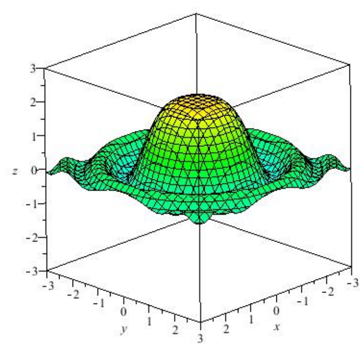
$$\lim_{(x,y)\to(0,0)} \frac{x^2+y^2}{x^4+2x^2y^2+y^4} = \lim_{r\to 0} \frac{r^2}{r^4} = \lim_{r\to 0} \frac{1}{r^2} = \infty.$$

this limit **diverges** *to ∞. We could see evidence of this in the graph, although infinities in graphs should be treated with great caution.*

Example 3.3.6. *Suppose* $f(x,y) = \begin{cases} \frac{\sin(x^2+y^2)}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$. *Again, we use polar coordinate substitution,*

$$\lim_{(x,y)\to(0,0)} \frac{\sin(x^2+y^2)}{x^2+y^2} = \underbrace{\lim_{r\to 0} \frac{\sin(r^2)}{r^2} = \lim_{r\to 0} \frac{2r\cos(r^2)}{2r}}_{L'Hospital's\ Rule\ on\ 0/0} = \lim_{r\to 0} \cos(r^2) = 1.$$

The graph agrees with our result. (rescaled for ease of viewing)



Naturally, we can also apply this technique to functions which admit a simplification in terms of spherical coordinates.

Example 3.3.7. *Suppose* $f(x,y,z) = \begin{cases} \sqrt{x^2+y^2+z^2}\ln(x^2+y^2+z^2) & (x,y) \neq (0,0,0) \\ 0 & (x,y,z) = (0,0,0) \end{cases}$.

Use spherical coordinate substitution,

$$\begin{aligned}
\lim_{(x,y,z)\to(0,0,0)} \left[\sqrt{x^2+y^2+z^2}\ln(x^2+y^2+z^2)\right] &= \lim_{\rho\to 0}\left[\ \rho\ln(\rho^2)\ \right] \\
&= \lim_{\rho\to 0}\left[\ \frac{2\ln(\rho)}{1/\rho}\ \right] \qquad L'Hospital's\ Rule\ on\ \frac{\infty}{\infty}. \\
&= \lim_{\rho\to 0}\left[\ \frac{2/\rho}{-1/\rho^2}\ \right] \\
&= \lim_{\rho\to 0}\left[-2\rho\right] \\
&= 0.
\end{aligned}$$

The idea of substitution need not be limited to standard coordinate systems. Perhaps you'll find a challenge problem in the homework.

3.3.1 additional examples

More of the same here, but perhaps these help.

Example 3.3.8. *Let* $f(x,y) = x^2 + \sqrt{y} + \tan^{-1}(x) + 3$*. To calculate the limit as* $(x,y) \to (0,1)$ *we may simply evaluate* f *at* $(0,1)$ *since all the expressions involved at continuous near the values of* $x = 0$ *and* $y = 1$*:*

$$\lim_{(x,y)\to(0,1)} f(x,y) = 0^2 + \sqrt{1} + \tan^{-1}(0) + 3 = 4.$$

Example 3.3.9. *Once more, this limit is simple since polynomials in one, two or* n *variables are continuous everywhere:*

$$\lim_{(x,y)\to(2,-1)} \left(x^5 + 4x^3y + 5xy^2\right) = 32 + 4(8)(-1) + 5(2)(-1)^2 = 10.$$

Example 3.3.10. *Let* $f(x,y) = \frac{xy\cos y}{3x^2+y^2}$ *and calculate the limit as* $(x,y) \to (0,0)$*. Notice the limit in question is indeterminant as the expression of* f *has the form* $0/0$ *at* $(0,0)$*. We suspect the limit does not exist, hence, we seek to show that different values are obtained by approaching* $(0,0)$ *along different paths.*

1. $(x,0) \to (0,0)$ *has* $f(x,0) = \frac{0}{3x^2+0} = 0$ *for* $x \neq 0$ *hence* $f(x,0) \to 0$ *along the* x*-axis.*
2. $(x,x) \to (0,0)$ *has* $f(x,x) = \frac{x^2\cos x}{3x^2+x^2} = \frac{1}{4}\cos x$ *for* $x \neq 0$ *hence* $f(x,x) \to \frac{1}{4}$ *along* $y = x$*-path.*

Clearly the path-limits differ hence $\lim_{(x,y)\to(0,0)} f(x,y)$ *does not exist.*

It is a subtle point, but, if all path-limits of a particular type exist then it is not necessarily the case the multivariate limit exists. You need all possible path-limits to exist, but, this is nearly impossible to check directly! So, the connection between multivariate and path-limits is primarily a no-go result. We typically use it to prove a particular limit does **not** exist.

3.4 Problems

Problem 70 Calculate the limit below if it exists. If it does not exist then show it fails to exist.

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

Problem 71 Calculate the limit below if it exists. If it does not exist then show it fails to exist.

$$\lim_{(x,y)\to(B,B)} \frac{x^4-y^4}{x^2-y^2}$$

Problem 72 Determine if the limit below exists. If it does exist, calculate it's value. If it does not exist then give an explicit argument which shows it cannot converge to a real value.

$$\lim_{(x,y)\to(0,0)} \left[\frac{y^2x}{y^4+x^2}\right].$$

Problem 73 Calculate the limit below by switching to polar coordinates.

$$\lim_{(x,y)\to(0,0)}\left[\frac{\sin(x^2+y^2)}{2x^2+2y^2}\right].$$

Problem 74 Show that the limit

$$\lim_{(x,y)\to(0,0)}\left[\frac{x-y^2}{x^2+y^2}\right]$$

does not converge to a real number.

Problem 75 Given that f is continuous at the origin, what must the value of A be in the definition below?

$$f(x,y)=\begin{cases}\frac{x^3+y^3}{x^2+y^2} & (x,y)\neq 0\\ A & (x,y)=(0,0)\end{cases}$$

Chapter 4

differentiation

In single variable calculus we learn from the outset that the derivative of a function describes the slope of the function at a point. On the other hand, we also learned that the derivative at a point can be used to construct the best linear approximation to the function. In particular, the derivative at a point shows how the change in the independent variable Δx gives an approximate change $\Delta y = f'(a)\Delta x$. This characterization of the derivative is the one which most readily generalizes to many dimensions. In particular, we generalize Δy and Δx to become vectors and $f'(a)$ is a matrix when $f : \mathbb{R}^n \to \mathbb{R}^m$. I'll explain how the derivative matrix[1] $f'(a)$ is the natural extension of our single-variable calculus to the general case.

The nuts and bolts of this derivative matrix are made from what we call *partial derivatives* of the *component functions*. The partial derivative in turn is naturally defined in the context of directional derivatives. The directional derivative takes the multivariate function and restricts it to a particular line in the domain. By making this restriction we find a way to *do* single variable-type calculations on a multivariate function. Much of the calculation presented in this chapter is little more than single-variable calculus with a few simple rules adjoined. However, connecting the partial derivatives to the general derivative involves multivariate limits and some analysis that is beyond the required content of this course. That said, I include some of those arguments in these notes in the interest of logical completeness.

Most modern treatments ignore the need to discuss the general concept of differentiation and instead just show students an assortment of various partial derivative calculations. I've found students who are thinking are usually unsatisfied with the popular approach because there is no big picture behind the partial differentiation. It's just a seemingly random collection of adhoc rules. This need not be. If we submit ourselves to a little linear algebaic terminology there is a beautiful and quite general context in which all the partial derivatives find a natural purpose.

[1] often called the Jacobian matrix

4.1 directional derivatives

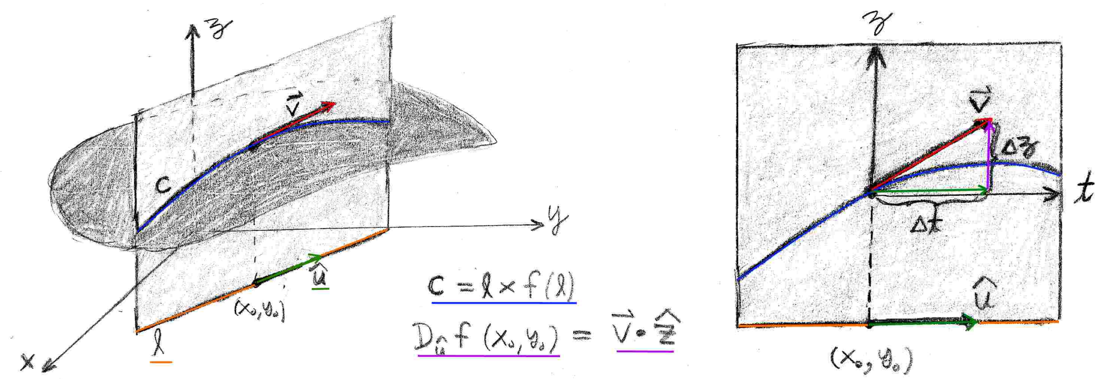
We begin our discussion with a function $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$. Consider a fixed point $(x_o, y_o) \in dom(f)$. Furthermore, picture $\widehat{u} = \langle a, b\rangle$ as a **unit**-vector (it's green) in the domain of f attached to (x_o, y_o) and construct the path in $dom(f)$ with direction $\langle a, b\rangle$ and base-point (x_o, y_o):

$$\vec{r}(t) = \langle x_o + ta,\ y_o + tb\rangle$$

If we feed this path (the orange line) to the function f then we can construct a curve in $\mathbb{R}^3$ which lies on the graph $z = f(x, y)$ and passes through the point $(x_o, y_o, f(x_o, y_o))$. In particular,

$$\vec{\gamma}(t) = \Big(x_o + ta,\ y_o + tb,\ f(x_o + ta,\ y_o + tb)\Big)$$

parametrizes the curve (in blue) formed by the intersection of the graph $z = f(x, y)$ and the vertical plane which contains $\widehat{z}$ and $a\,\widehat{x} + b\,\widehat{y}$.



In the picture above you can see that we identify the xy-plane embedded in $\mathbb{R}^3$ with the plane $\mathbb{R}^2$ which contains $dom(f)$. A natural choice of coordinates on vertical slice containing $\langle a, b, 0\rangle$ and $\langle 0, 0, 1\rangle$ is given by t, z. For the sake of discussion let $g(t) = f(x_o + ta,\ y_o + tb)$ and consider the graph $z = g(t)$. This is a context to which ordinary single-variate calculus applies. The derivative $g'(0)$ describes the slope of the tangent line in the tz-plane at $(0, g(0))$. Of course, from the three-dimensional perspective, $g'(0)$ gives the z-component of the velocity-vector (the red arrow) to the path $t \mapsto \vec{\gamma}(t)$. So what? Well, what is that quantity's meaning for $z = f(x, y)$? It's simply the following:

> **The value of** $\frac{d}{dt}\big[\ f(x_o + ta,\ y_o + tb)\big]\ \big|_{t=0}$ **describes the rate of change in** $f(x, y)$ **in the direction** $\langle a, b\rangle$ **at the point** (x_o, y_o)**.**

This is why we are interested in this calculation. The directional derivative of f in the $\langle a, b\rangle$ direction at (x_o, y_o) is precisely the slope described above.

Definition 4.1.1.

Let $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function with $\vec{p}_o = (x_o, y_o) \in dom(f)$ and suppose $\widehat{u} = \langle a, b\rangle$ is a unit-vector. If the limit below exists, then we define the **directional derivative** of f at $\vec{p}_o$ in the $\widehat{u}$-direction by

$$D_{\widehat{u}} f(\vec{p}_o) = \lim_{t\to 0}\left[\frac{f(\vec{p}_o + t\widehat{u}) - f(\vec{p}_o)}{t}\right] = \lim_{t\to 0}\left[\frac{f(x_o + ta,\ y_o + tb) - f(x_o, y_o)}{t}\right].$$

The definition above can also be written in terms of a derivative followed by an evaluation:

$$D_{\widehat{u}} f(\vec{p}_o) = \frac{d}{dt}\bigg[\ f(x_o + ta,\ y_o + tb)\ \bigg]\bigg|_{t=0}.$$

We pause to look at a few examples.

Example 4.1.2. Problem: *Suppose $f(x,y) = 25xy$ and calculate the rate of change in f at $(1,2)$ in the direction of the $\langle 3,4\rangle$-vector.*

Solution: *we identify this is a directional derivative problem. We need a point and a unit vector. The point is $p_o = (1,2)$. However, $||\langle 3,4\rangle|| = \sqrt{9+16} = 5$ hence we need to rescale the given vector before we calculate. Just divide by 5 to obtain $\widehat{u} = \langle 3/5, 4/5\rangle$. Calculate,*

$$f(\,\vec{p}_o + t\widehat{u}\,) = f(1+3t/5, 2+4t/5) = 25(1+3t/5)(2+4t/5) = (5+3t)(10+4t)$$

Differentiate, and then evaluate,

$$D_{\widehat{u}} f(\vec{p}_o) = \frac{d}{dt}\bigg[\ (5+3t)(10+4t)\ \bigg]\bigg|_{t=0} = \bigg[\ 3(10+4t) + 4(5+3t)\ \bigg]\bigg|_{t=0} = 30 + 20 = 50.$$

Naturally if you would rather calculate the difference quotient and take the limit you are free to do that. I choose to use the tools we've already developed, no sense in reinventing the wheel here. Incidentally, we will find a better way to package this calculation so you should look at this example as a means to better understand the definition. It is not computationally ideal. Neither is what follows, but these help bring understanding to later calculations so here we go.

Example 4.1.3. Problem: *Suppose $f(x,y) = 25xy$ and calculate the rate of change in f at $(1,2)$ in the direction of the* **(a.)** *$\langle 1,0\rangle$-vector,* **(b.)** *$\langle 0,1\rangle$-vector.*

Solution of (a.): *I'll get straight to it here, identify $\widehat{u} = \langle 1,0\rangle$ and $\vec{p}_o = (1,2)$ and calculate*

$$f(\vec{p}_o + t\,\widehat{x}) = f(1+t, 2) = 25(1+t)(2) = 50 + 50t$$

Therefore,

$$D_{\widehat{x}} f(\vec{p}_o) = \frac{d}{dt}\bigg[\ 50 + 50t\ \bigg]\bigg|_{t=0} = 50.$$

Solution of (b.): *Identify $\widehat{u} = \langle 0,1\rangle$ and $\vec{p}_o = (1,2)$ and calculate*

$$f(\vec{p}_o + t\,\widehat{y}) = f(1, 2+t) = 25(1)(2+t) = 50 + 25t$$

Therefore,

$$D_{\widehat{y}} f(\vec{p}_o) = \frac{d}{dt}\bigg[\ 50 + 25t\ \bigg]\bigg|_{t=0} = 25.$$

Notice $50 = \frac{3}{5}(50) + \frac{4}{5}(25)$ hence the previous examples are related in a curious manner:

$$D_{\widehat{u}} f(\vec{p}_o) = \frac{3}{5} D_{\widehat{x}} f(\vec{p}_o) + \frac{4}{5} D_{\widehat{y}} f(\vec{p}_o).$$

In other words, the pattern we see is:

$$D_{\langle a,b\rangle} f(\vec{p}_o) = \langle a,b\rangle \bullet \langle D_{\widehat{x}} f(\vec{p}_o), D_{\widehat{y}} f(\vec{p}_o)\rangle.$$

The directional derivatives in the coordinate directions are apparently important. We may be able to build the directional derivative in other directions[2]. This leads us to the topic of the next section. However, for the sake of logical completeness I define directional derivatives for functions of more than two variables. The visualization of the slopes implicit in the definition below are beyond most of our visual acumen.

Definition 4.1.4.

Let $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ be a function with $\vec{p}_o \in dom(f)$ and suppose $\widehat{u} \in \mathbb{R}^n$ is a unit-vector. If the limit below exists, then we define the **directional derivative** of f at $\vec{p}_o$ in the $\widehat{u}$-direction by

$$D_{\widehat{u}} f(\vec{p}_o) = \lim_{t \to 0} \left[\frac{f(\,\vec{p}_o + t\widehat{u}\,) - f(\,\vec{p}_o\,)}{t} \right] = \frac{d}{dt} \left[\ f(\,\vec{p}_o + t\widehat{u}\,) \ \right] \Big|_{t=0}.$$

We will calculate a few such directional derivatives in the section after the next once we understand the two-dimensional case in some depth.

4.2 partial differentiation in $\mathbb{R}^2$

We continue the discussion of the last section concerning the change in functions of two variables. The formulas and concepts readily generalize to $n \geq 3$ however we postpone such discussion until we have settled the $n = 2$ theory.

Definition 4.2.1.

Let $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function with $(x_o, y_o) \in dom(f)$. If the directional derivative below exists, then we define the **partial derivative** of f at (x_o, y_o) with respect to x by

$$\frac{\partial f}{\partial x}(x_o, y_o) = (D_{\widehat{x}} f)(x_o, y_o).$$

Likewise, we define the **partial derivative** of f at (x_o, y_o) with respect to y by

$$\frac{\partial f}{\partial y}(x_o, y_o) = (D_{\widehat{y}} f)(x_o, y_o)$$

provided the directional derivative $(D_{\widehat{y}} f)(x_o, y_o)$ exists.

Notice that $(x_o, y_o) \mapsto \frac{\partial f}{\partial x}(x_o, y_o)$ and $(x_o, y_o) \mapsto \frac{\partial f}{\partial y}(x_o, y_o)$ define new multivariate functions provided the given function f possesses the necessary directional derivatives. We define higher derivatives by successive partial differentiation in the natural way: $\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left[\frac{\partial f}{\partial x}\right]$. Derivatives such as $\frac{\partial^2 f}{\partial y^2}$ and $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ are similarly defined. A brief notation for partial derivatives is as follows:

$$f_x = \frac{\partial f}{\partial x}, \quad f_y = \frac{\partial f}{\partial y}, \quad f_{xx} = \frac{\partial^2 f}{\partial x^2}, \quad f_{xy} = (f_x)_y = \frac{\partial^2 f}{\partial y \partial x}, \ etc \ldots$$

It is usually the case that $f_{xy} = f_{yx}$ but the proof of that statement is nontrivial and can be found in most advanced calculus texts. Given the connection of the partial derivative and the directional derivative we have the following conceptual guidelines:

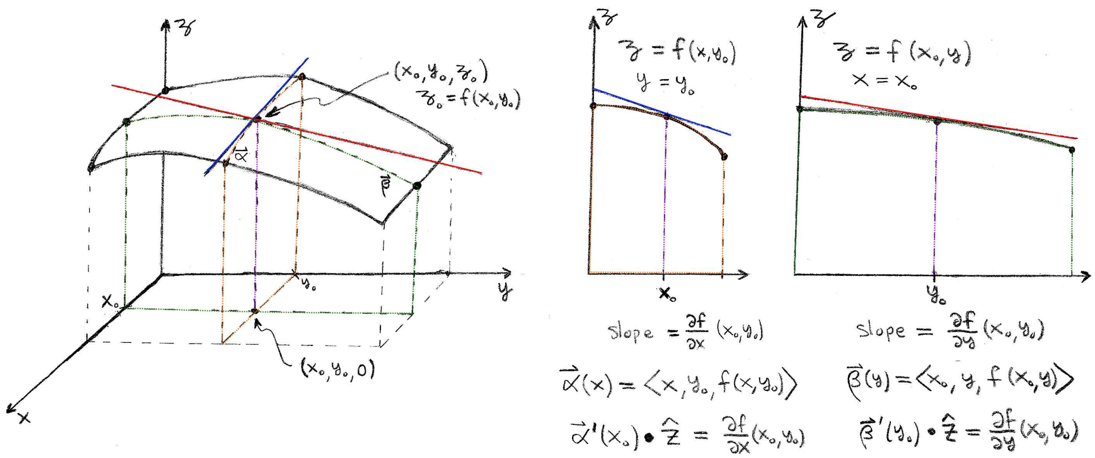
[2]it turns out this is not generally true, but the exceptions are rare in applications

f_x gives the rate of change in f in the x-direction.
f_y gives the rate of change in f in the y-direction.

It is also useful to rewrite the definition of the partial derivatives explicitly in terms of derivatives.

$$\boxed{\frac{\partial f}{\partial x}(x_o, y_o) = \frac{d}{dt}\bigg[\ f(x_o + t, y_o)\ \bigg]\bigg|_{t=0} \qquad \frac{\partial f}{\partial y}(x_o, y_o) = \frac{d}{dt}\bigg[\ f(x_o, y_o + t)\ \bigg]\bigg|_{t=0}.}$$

The geometry is revealed in the diagram below:



Well, how do these really work? The proposition below explains the working calculus of partial derivatives. It is really very simple.

Proposition 4.2.2.

Assume f, g are functions from $\mathbb{R}^2$ to $\mathbb{R}$ whose partial derivatives exist. Then for $c \in \mathbb{R}$,

1. $(f + g)_x = f_x + g_x$ and $(f + g)_y = f_y + g_y$.
2. $(cf)_x = cf_x$ and $(cf)_y = cf_y$.
3. $(fg)_x = f_x g + f g_x$ and $(fg)_y = f_y g + f g_y$.

Moreover, if $h : dom(h) \subseteq \mathbb{R} \to \mathbb{R}$ is a differentiable function then (4.)

$$\frac{\partial}{\partial x}\Big[h(f(x, y))\Big] = \frac{dh}{dt}\bigg|_{f(x,y)} \frac{\partial f}{\partial x} = \frac{dh}{df}\frac{\partial f}{\partial x} \quad \text{and} \quad \frac{\partial}{\partial y}\Big[h(f(x, y))\Big] = \frac{dh}{dt}\bigg|_{f(x,y)} \frac{\partial f}{\partial y} = \frac{dh}{df}\frac{\partial f}{\partial y}.$$

5. $\dfrac{\partial x}{\partial x} = 1, \quad \dfrac{\partial x}{\partial y} = 0, \quad \dfrac{\partial y}{\partial x} = 0, \quad \dfrac{\partial y}{\partial y} = 1.$

Proof: the proofs of 1,2,3 follow immediately from the corresponding properties of single-variable differentiation. Let's work on the x-part of (4.)

$$\begin{aligned}
\frac{\partial}{\partial x}\Big[h(f(x,y))\Big] &= \frac{d}{dt}\Big[\ h(f(x_o+t,y_o))\ \Big]\Big|_{t=0} \\
&= \left(\frac{dh}{dt}\Big|_{f(x_o+t,y_o)}\frac{d}{dt}\Big[\ f(x_o+t,y_o)\ \Big]\right)\Big|_{t=0} \\
&= \frac{dh}{dt}\Big|_{f(x,y)}\frac{\partial f}{\partial x}.
\end{aligned}$$

We find that (4.) follows from the chain-rule of single-variable calculus. The proof in the y-variable is nearly the same. The proof of (5.) requires understanding of the definition. Let $F(x,y) = x$ and calculate

$$\frac{\partial F}{\partial x} = \frac{d}{dt}\Big[\ F(x+t,y)\ \Big]\Big)\Big|_{t=0} = \frac{d}{dt}\Big[\ x+t\ \Big]\Big)\Big|_{t=0} = 1.$$

$$\frac{\partial F}{\partial y} = \frac{d}{dt}\Big[\ F(x,y+t)\ \Big]\Big)\Big|_{t=0} = \frac{d}{dt}\Big[\ x\ \Big]\Big)\Big|_{t=0} = 0.$$

Likewise, let $G(x,y) = y$ and calculate,

$$\frac{\partial G}{\partial x} = \frac{d}{dt}\Big[\ G(x+t,y)\ \Big]\Big)\Big|_{t=0} = \frac{d}{dt}\Big[\ y\ \Big]\Big)\Big|_{t=0} = 0.$$

$$\frac{\partial G}{\partial y} = \frac{d}{dt}\Big[\ G(x,y+t)\ \Big]\Big)\Big|_{t=0} = \frac{d}{dt}\Big[\ y+t\ \Big]\Big)\Big|_{t=0} = 1.$$

Which concludes the proof of (5.) $\square$

Example 4.2.3. *Can you identify which property of the proposition I use in each line below?*

$$\begin{aligned}
\frac{\partial}{\partial x}\Big[\ 2^{x^2+y^2}\ \Big] &= \ln(2)2^{x^2+y^2}\frac{\partial}{\partial x}(x^2+y^2) \\
&= \ln(2)2^{x^2+y^2}\left[\frac{\partial}{\partial x}(x^2)+\frac{\partial}{\partial x}(y^2)\right] \\
&= \ln(2)2^{x^2+y^2}\left[2x\frac{\partial x}{\partial x}+2y\frac{\partial y}{\partial x}\right] \\
&= 2\ln(2)x2^{x^2+y^2}.
\end{aligned}$$

Example 4.2.4. *Can you identify which property of the proposition I use in each line below?*

$$\begin{aligned}
\frac{\partial}{\partial x}[\ \sin(x^2y)\ ] &= \cos(x^2y)\frac{\partial}{\partial x}[\ x^2y\ ] \\
&= \cos(x^2y)\left[\frac{\partial}{\partial x}[x^2]y + x^2\frac{\partial y}{\partial x}\right] \\
&= \cos(x^2y)\left(2x\frac{\partial x}{\partial x}y\right) \\
&= 2xy\cos(x^2y).
\end{aligned}$$

Similarly, you can calculate:

$$\frac{\partial}{\partial y}[\ \sin(x^2y)\ ] = x^2\cos(x^2y).$$

In practice I rarely write as many steps as I just offered in the examples above.

Example 4.2.5. *Power functions and exponential functions are different.*

$$\frac{\partial}{\partial x}\left[\ x^y\ \right] = yx^{y-1} \quad \textit{whereas} \quad \frac{\partial}{\partial y}\left[\ x^y\ \right] = \ln(x)x^y$$

$$\frac{\partial}{\partial y}\left[\ y^x\ \right] = xy^{x-1} \quad \textit{whereas} \quad \frac{\partial}{\partial x}\left[\ y^x\ \right] = \ln(y)y^x.$$

Example 4.2.6.

$$\frac{\partial}{\partial x}\left[\ x^{y^2}\ \right] = y^2x^{y^2-1} \quad \textit{whereas} \quad \frac{\partial}{\partial y}\left[\ x^{y^2}\ \right] = \ln(x)x^{y^2}\frac{\partial y^2}{\partial y} = 2y\ln(x)x^{y^2}.$$

Example 4.2.7.

$$\begin{aligned}\frac{\partial}{\partial x}\left[\ \sin\big(x^2y\cosh(x)\big)\ \right] &= \cos\big(x^2y\cosh(x)\big)\frac{\partial}{\partial x}\big(x^2y\cosh(x)\big). \\ &= \big(2xy\cosh(x) + x^2y\sinh(x)\big)\cos\big(x^2y\cosh(x)\big).\end{aligned}$$

Can I skip the middle step in the example above? Some days yes. Should you? Probably not.

Example 4.2.8.

$$\begin{aligned}\frac{\partial}{\partial y}\left[\ \sin\big(\cos(\sqrt{xy})\big)\ \right] &= \cos\big(\cos(\sqrt{xy})\big)\frac{\partial}{\partial y}\big(\cos(\sqrt{xy})\big). \\ &= \cos\big(\cos(\sqrt{xy})\big)\big(-\sin(\sqrt{xy})\big)\frac{\partial}{\partial y}\sqrt{xy}. \\ &= \cos\big(\cos(\sqrt{xy})\big)\big(-\sin(\sqrt{xy})\big)\frac{1}{2\sqrt{xy}}\frac{\partial}{\partial y}[xy]. \\ &= -\frac{1}{2}\sqrt{\frac{x}{y}}\sin\big(\ \sqrt{xy}\ \big)\cos\big(\ \cos(\sqrt{xy})\ \big).\end{aligned}$$

Example 4.2.9. *Let* $F(x,y) = x^2 + y^2$.

$$\begin{aligned}\frac{\partial F}{\partial x} &= \frac{\partial}{\partial x}[x^2+y^2] = 2x, \\ \frac{\partial F}{\partial y} &= \frac{\partial}{\partial y}[x^2+y^2] = 2y.\end{aligned}$$

Example 4.2.10. *Let* $F(x,y) = xe^{xy}$.

$$\begin{aligned}\frac{\partial F}{\partial x} &= \frac{\partial}{\partial x}[xe^{xy}] = e^{xy} + xe^{xy}\frac{\partial}{\partial x}[xy] = e^{xy}(1+xy), \\ \frac{\partial F}{\partial y} &= \frac{\partial}{\partial y}[xe^{xy}] = x\frac{\partial}{\partial y}[e^{xy}] = xe^{xy}\frac{\partial}{\partial y}[xy] = x^2e^{xy}.\end{aligned}$$

I wrote steps to explicitly indicate the role of the chain-rule. As we gain skill the need to write down such steps diminishes. Unless, of course, one is asked to show such steps.

A common notation for $\frac{\partial F}{\partial x}$ is simply F_x. The seoond derivatives are likewise denoted: $F_{xx} = \frac{\partial}{\partial x}(\frac{\partial}{\partial x}F)$ and $F_{yy} = \frac{\partial}{\partial y}(\frac{\partial}{\partial y}F)$ whereas[3] $F_{xy} = (F_x)_y = \frac{\partial}{\partial y}(\frac{\partial}{\partial x}F)$.

[3]fortunately, for twice continuously differentiable functions $F_{xy} = F_{yx}$ so failure to note the order of operations implicit with the subscript notation is not too dangerous an oversight

Example 4.2.11. *Below we calculate all second derivatives of* $f(x,y) = xy^2$

$$\begin{aligned}
f_{xx} &= \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}xy^2\right) = \frac{\partial}{\partial x}(y^2) = 0. \\
f_{xy} &= \frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}xy^2\right) = \frac{\partial}{\partial y}(y^2) = 2y. \\
f_{yx} &= \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}xy^2\right) = \frac{\partial}{\partial x}(2xy) = 2y. \\
f_{yy} &= \frac{\partial}{\partial y}\left(\frac{\partial}{\partial y}xy^2\right) = \frac{\partial}{\partial y}(2xy) = 2x.
\end{aligned}$$

Example 4.2.12. *Let* $f(x,t) = x^2e^{-ct}$ *where* c *is a constant. Third derivatives are calculated in the natural manner:*

$$\begin{aligned}
f_{ttt} &= \frac{\partial}{\partial t}\frac{\partial}{\partial t}\frac{\partial}{\partial t}\left[x^2e^{-ct}\right] \\
&= x^2\frac{\partial}{\partial t}\frac{\partial}{\partial t}\frac{\partial}{\partial t}\left[e^{-ct}\right] \\
&= x^2\frac{\partial}{\partial t}\frac{\partial}{\partial t}\left[-ce^{-ct}\right] \\
&= x^2\frac{\partial}{\partial t}\left[c^2e^{-ct}\right] \\
&= -c^3x^2e^{-ct}.
\end{aligned}$$

Or,

$$\begin{aligned}
f_{txx} &= \frac{\partial}{\partial x}\frac{\partial}{\partial x}\frac{\partial}{\partial t}\left[x^2e^{-ct}\right] \\
&= \frac{\partial}{\partial x}\frac{\partial}{\partial x}\left[-cx^2e^{-ct}\right] \\
&= \frac{\partial}{\partial x}\left[-2cxe^{-ct}\right] \\
&= -2ce^{-ct}.
\end{aligned}$$

Example 4.2.13. *Let* $f(x,y) = \ln(x + \sqrt{x^2+y^2})$. *Find* $f_x(3,4)$.

Solution: *begin by calculating* f_x *at an arbitrary point:*

$$f_x(x,y) = \frac{1}{x+\sqrt{x^2+y^2}}\frac{\partial}{\partial x}\left[x+\sqrt{x^2+y^2}\right] = \frac{1}{x+\sqrt{x^2+y^2}}\left[1+\frac{x}{\sqrt{x^2+y^2}}\right]$$

We set $x = 3$ *and* $y = 4$ *to obtain* $f_x(3,4) = \frac{1}{5}$.

Example 4.2.14. *Let* $f(x,y) = x^3y^5 + 2x^4y$. *We calculate,*

$$f_{xx} = 6xy^5 + 24x^2y, \qquad f_{yy} = 20x^3y^3, \qquad f_{xy} = 15x^2y^4 + 8x^3.$$

Example 4.2.15. . *Let* $G(s,t) = \frac{st^2}{s^2+t^2}$. *Use the quotient rule for* $\frac{\partial}{\partial s}$ *or* $\frac{\partial}{\partial t}$ *to see:*

$$\begin{aligned}
\frac{\partial G}{\partial s} &= \frac{\partial}{\partial s}\left[\frac{st^2}{s^2+t^2}\right] = \frac{t^2(s^2+t^2)-2s(st^2)}{(s^2+t^2)^2} = \frac{t^2(t^2-s^2)}{(s^2+t^2)^2} \\
\frac{\partial G}{\partial t} &= \frac{\partial}{\partial t}\left[\frac{st^2}{s^2+t^2}\right] = \frac{2st(s^2+t^2)-2t(st^2)}{(s^2+t^2)^2} = \frac{2ts^3}{(s^2+t^2)^2}.
\end{aligned}$$

Example 4.2.16. *The Fundamental Theorem of Calculus Part I (FTC I) is still with us:*

$$\frac{\partial}{\partial x}\int_y^x \cos(t^2)\,dt = \cos(x^2),$$

$$\frac{\partial}{\partial y}\int_y^x \cos(t^2)\,dt = -\frac{\partial}{\partial y}\int_x^y \cos(t^2)\,dt = -\cos(y^2).$$

Example 4.2.17. *We can also use FTC I in combination with the chain rule:*

$$\frac{\partial}{\partial x}\int_{x^2}^{\sin x} \cos(t^2)\,dt = \cos x\cos(\sin^2(x)) - 2x\cos(x^4),$$

I prefer to adopt some notation to unravel this idea: let F be an antiderivative of $\cos(t^2)$ then $\int_a^b \cos(t^2)\,dt = F(b) - F(a)$. We simply allow a, b to be functions of x, y in our current context:

$$\begin{aligned}\frac{\partial}{\partial x}\int_a^b \cos(t^2)\,dt &= \frac{\partial}{\partial x}\big[F(b) - F(a)\big]\\ &= F'(b)\frac{\partial b}{\partial x} - F'(a)\frac{\partial a}{\partial x}\\ &= \cos(b^2)\frac{\partial b}{\partial x} - \cos(a^2)\frac{\partial a}{\partial x}.\end{aligned}$$

This formula explains the initial calculation in this example and a host of other connected problems.

Example 4.2.18. *Let $w = \sin\alpha\cos\beta$.*

$$\frac{\partial w}{\partial \alpha} = \frac{\partial}{\partial \alpha}\big[\sin\alpha\cos\beta\big] = \cos\beta\frac{\partial}{\partial \alpha}\big[\sin\alpha\big] = \cos\beta\cos\alpha,$$

$$\frac{\partial w}{\partial \beta} = \frac{\partial}{\partial \beta}\big[\sin\alpha\cos\beta\big] = \sin\alpha\frac{\partial}{\partial \beta}\big[\cos\beta\big] = -\sin\alpha\sin\beta.$$

Example 4.2.19. *Let $z = y\tan(2x)$. Calculate all the derivatives of z up to second order.*

$$\begin{aligned}\frac{\partial z}{\partial x} &= y\frac{\partial}{\partial x}\big[\tan(2x)\big] = 2y\sec^2(2x)\\ \frac{\partial z}{\partial y} &= \tan(2x)\frac{\partial}{\partial y}\big[y\big] = \tan(2x)\\ \frac{\partial^2 z}{\partial x^2} &= \frac{\partial}{\partial x}\left[\frac{\partial z}{\partial x}\right] = \frac{\partial}{\partial x}\big[2y\sec^2(2x)\big] = 8y\sec^2(2x)\tan(2x).\\ \frac{\partial^2 z}{\partial y^2} &= \frac{\partial}{\partial y}\left[\frac{\partial z}{\partial y}\right] = \frac{\partial}{\partial y}\big[\tan(2x)\big] = 0\\ \frac{\partial^2 z}{\partial x\partial y} &= \frac{\partial}{\partial x}\left[\frac{\partial z}{\partial y}\right] = \frac{\partial}{\partial x}\big[\tan(2x)\big] = 2\sec^2(2x).\end{aligned}$$

It is at times convenient to use the notations $z_x = \frac{\partial z}{\partial x}, z_y = \frac{\partial z}{\partial y}, z_{yx} = \frac{\partial^2 z}{\partial x\partial y}$ etc. However, it is wise to be conversant in both the ∂ and subsript notations for partial differentiation as A_x might denote the x-component function of a vector field $\vec{A}$ in a later Chapter of these notes. It is also wise to adopt the abbreviated notation ∂_x for $\frac{\partial}{\partial x}$ as this gives us a nice language to avoid the ambiguity without an excess of writing: for example, $\frac{\partial A_x}{\partial y} = \partial_y A_x$.

4.2.1 directional derivatives and the gradient in $\mathbb{R}^2$

Now that we have a little experience in partial differentiation let's return to the problem of the directional derivative. We saw that

$$D_{\langle a,b\rangle}f(x_o,y_o) = \langle\ f_x(x_o,y_o),\ f_y(x_o,y_o)\ \rangle \bullet \langle a,b\rangle$$

for the particular example we considered. Is this always true? Is it generally the case that we can build the directional derivative in the $\langle a,b\rangle$-direction from the partial derivatives? If you just try most functions that come to the nonpathological mind then you'd be tempted to agree with this claim. However, many counter-examples exist. We only need one to debunk the claim.

Example 4.2.20. *Suppose that*

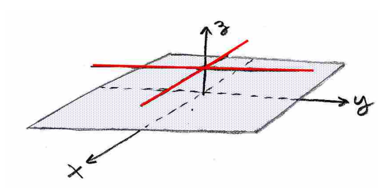
$$f(x,y) = \begin{cases} x+1 & y=0 \\ y+1 & x=0 \\ 0 & xy\neq 0 \end{cases}$$

Clearly $f_x(0,0)=1$ and $f_y(0,0)=1$ however the directional derivative is given by

$$D_{\langle a,b\rangle}f(0,0) = \lim_{t\to 0}\frac{f(ta,tb)-f(0,0)}{t} = \lim_{t\to 0}\frac{-1}{t}$$

which diverges. The directional derivative in any non-coordinate direction does not exist since the function jumps from 0 to 1 at the origin along any line except the axes.

Example 4.2.21. *This example is even easier: let $f(x,y) = \begin{cases} 1 & y=0 \\ 1 & x=0 \\ 0 & xy\neq 0 \end{cases}$. In this case I can graph the function and it is obvious that $f_x(0,0)=0$ and $f_y(0,0)=0$ yet all the directional derivatives in non-coordinate directions fail to exist.*



We can easily see the discontinuity of the function above is the source of the trouble. It is sometimes true that a function is discontinuous and the formula holds. However, the case which we really want to consider, the type of functions for which the derivatives considered are most meaningful, are called **continuously differentiable**. You might recall from single-variable calculus that when a function is differentiable at a point but the derivative function is discontinuous it led to bizzare features for the linearization. That continues to be true in the multivariate case.

Definition 4.2.22.

A function $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ is said to be **continuously differentiable** at (x_o,y_o) iff the partial derivative functions $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ are continuous at (x_o,y_o). We say $f \in C^1(x_o,y_o)$. If all the second-order partial derivatives of f are continuous at (x_o,y_o) then we say $f \in C^2(x_o,y_o)$. If continuous partial derivatives of arbitrary order exist at (x_o,y_o) then we say f is **smooth** and write $f \in C^\infty(x_o,y_o)$.

The continuity of the partial derivative functions implicitly involves multivariate limits and this is what ultimately makes this criteria quite strong.

Proposition 4.2.23.

Suppose f is continuously differentiable at (x_o, y_o) then the directional derivative at (x_o, y_o) in the direction of the unit vector $\langle a, b\rangle$ is given by:

$$D_{\langle a,b\rangle} f(x_o, y_o) = \langle\ f_x(x_o, y_o),\ \ f_y(x_o, y_o)\ \rangle \bullet \langle a, b\rangle$$

Proof: delayed until the next section. $\square$

At this point it is useful to introduce a convenient notation which groups all the partial derivatives together in a particular vector of functions.

Definition 4.2.24.

If the partial derivatives of f exist then we define

$$\nabla f = \langle f_x,\ f_y\rangle = \widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y}.$$

we also use the notation $grad(f)$ and call this the **gradient** of f.

The upside-down triangle ∇ is also known as *nabla*. Identify that $\nabla = \widehat{x}\partial_x + \widehat{y}\partial_y$ is a vector of operators, it takes a function f and produces a vector field ∇f. This is called the **gradient vector field** of f. We'll think more about that after the examples. For a continuously differentiable function we have the following beautiful formula for the directional derivative:

$$\boxed{D_{\langle a,b\rangle} f(x_o, y_o) = (\nabla f)(x_o, y_o) \bullet \langle a, b\rangle.}$$

This is the formula I advocate for calculation of directional derivatives. This formula most elegantly summarizes how the directional derivative works. I'd make it the definition, but the discontinuous[4] counter-Example 4.2.20 already spoiled our fun.

Example 4.2.25. *Suppose* $f(x, y) = x^2 + y^2$*. Then*

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

Calculate the directional derivative of f *at* (x_o, y_o) *in the* $\langle a, b\rangle$*-direction:*

$$D_{\langle a,b\rangle} f(x_o, y_o) = \langle 2x_o, 2y_o\rangle \bullet \langle a, b\rangle = 2x_o a + 2y_o b.$$

It is often useful to write $D_{\langle a,b\rangle} f(x_o, y_o) = (\nabla f)(x_o, y_o) \bullet \langle a, b\rangle$ in terms of the angle θ between the $\nabla f(x_o, y_o)$ and $\langle a, b\rangle$:

$$\boxed{D_{\langle a,b\rangle} f(x_o, y_o) = ||(\nabla f)(x_o, y_o)|| \cos\theta.}$$

With this formula the Proposition below is obvious:

[4] I don't mean to say there are no continuous counter examples,I'd wager there are examples of continuous functions whose partial derivatives exist but are discontinuous. Then the formula fails because some non-coordinate directions fail to possess a directional derivative.

Proposition 4.2.26.

1. ($\theta = 0$) when $\langle a, b\rangle$ is parallel to $(\nabla f)(x_o, y_o)$ the direction $\langle a, b\rangle$ points towards **maximum increase** in f
2. ($\theta = \pi$) when $\langle a, b\rangle$ is antiparallel to $(\nabla f)(x_o, y_o)$ the direction $\langle a, b\rangle$ points towards **maximum decrease** in f
3. ($\theta = \pi/2$) when $\langle a, b\rangle$ is perpendicular to $(\nabla f)(x_o, y_o)$ the direction $\langle a, b\rangle$ points towards where f remains **constant**.

We use this proposition in many of the examples which follow.

Example 4.2.27. Problem: *if $f(x, y) = x^2 + y^2$. Then in what direction(s) is(are) f* **(a.)** *increasing the most at $(2, 3)$,* **(b.)** *decreasing the most at $(2, 3)$,* **(c.)** *not increasing at $(2, 3)$?*

Solution of (a.): *f increases most in the $(\nabla f)(2, 3)$-direction. In particular, $(\nabla f)(2, 3) = \langle 4, 6\rangle$. If you prefer a unit-vector then rescale $\langle 4, 6\rangle\rangle$ to $\widehat{u} = \frac{1}{\sqrt{13}}\langle 2, 3\rangle$. The magnitude $||(\nabla f)(2, 3)|| = \sqrt{13}$ is the rate of increase in the $\widehat{u} = \frac{1}{\sqrt{13}}\langle 2, 3\rangle$-direction.*

Solution of (b.): *f decreases most in the $-(\nabla f)(2, 3)$-direction. In particular, $-(\nabla f)(2, 3) = \langle -4, -6\rangle$. If you prefer a unit-vector then rescale $\langle -4, -6\rangle\rangle$ to $\widehat{u} = \frac{1}{\sqrt{13}}\langle -2, -3\rangle$. The rate of decrease is also $\sqrt{13}$ in magnitude.*

Solution of (c.): *f is constant in directions which are perpendicular to $(\nabla f)(2, 3)$. A unit-vector which is perpendicular to $(\nabla f)(2, 3) = \langle 4, 6\rangle$ satisfied two conditions:*

$$(\nabla f)(2, 3) \bullet \langle a, b\rangle = 4a + 6b = 0 \qquad \textit{and} \qquad a^2 + b^2 = 1$$

These are easily solved by solving the orthogonality condition for $b = -\frac{2}{3}a$ and substituting it into the unit-length condition:

$$1 = a^2 + b^2 = a^2 + \frac{4}{9}a^2 = \frac{13}{9}a^2 \quad \Rightarrow \quad a^2 = \frac{9}{13} \quad \Rightarrow \quad a = \pm\frac{3}{\sqrt{13}} \quad \Rightarrow b = \mp\frac{2}{\sqrt{13}}.$$

Therefore, we find f is constant in either the $\langle 3/\sqrt{13}, -2/\sqrt{13}\rangle$ or the $\langle -3/\sqrt{13}, 2/\sqrt{13}\rangle$ direction.

Example 4.2.28. Problem: *find a point (x_o, y_o) at which the function $f(x, y) = x^2 + y^2$ is constant in all directions.*

Solution: *We need to find a point (x_o, y_o) at which $(\nabla f)(x_o, y_o)$ is perpendicular to all unit-vectors. The only vector which is perpendicular to all other vectors is the zero vector. We seek solutions to $(\nabla f)(x_o, y_o) = \langle 2x_o, 2y_o\rangle = \langle 0, 0\rangle$. The only solution is $x_o = 0$ and $y_o = 0$. Apparently the graph $z = f(x, y)$ levels out at the origin since $f(x, y)$ stays constant in all directions near $(0, 0)$.*

Definition 4.2.29.

We say (x_o, y_o) is a **critical point** of f if $(\nabla f)(x_o, y_o)$ does not exist or $(\nabla f)(x_o, y_o) = \langle 0, 0\rangle$.

The term critical point is appropriate here since these are points where the function f *may* have a local maximum or minimum. Other possibilities exist and we'll spend a few lectures this semester developing tools to carefully discern what the geometry is near a given critical point.

Example 4.2.30. *Let $f(x,y) = \frac{y^2}{x}$ and $P = (1,2)$. Calculate the rate of change in f at P in the $\langle 2, \sqrt{5}\rangle$-direction.*

Solution: *we need to calculate the directional derivative of f at p in the $\langle 2, \sqrt{5}\rangle$-direction. Observe the length of $\langle 2, \sqrt{5}\rangle$ is 3 hence $\widehat{u} = \langle 2/3, \sqrt{5}/3\rangle$ is the unit-vector in the desired direction. Also, note $\nabla f = \langle f_x, f_y\rangle = \langle -y^2/x^2, 2y/x\rangle$. Consider,*

$$\begin{aligned}(D_{\widehat{u}}f)(P) &= (\nabla f(P)) \bullet \widehat{u} \\ &= \langle -4, 4\rangle \bullet \langle 2/3, \sqrt{5}/3\rangle \\ &= \boxed{\frac{4\sqrt{5}-8}{3}.}\end{aligned}$$

Example 4.2.31. *Let $f(x,y) = \sqrt{5x-4y}$ and $P = (4,1)$. What is the rate of change in f at P in the $\theta = -\pi/6$ direction?*

Solution: *The rate of change is given by the directional derivative at P in the direction corresponding to the given angle. Observe $\theta = -\pi/6$ gives the $\widehat{u} = \langle \cos(-\frac{\pi}{6}), \sin(-\frac{\pi}{6})\rangle = \langle \frac{\sqrt{3}}{2}, \frac{-1}{2}\rangle$ direction. Furthermore,*

$$\nabla f(x,y) = \left\langle \frac{5}{2\sqrt{5x-4y}}, \frac{-2}{\sqrt{5x-4y}}, \right\rangle \quad \Rightarrow \quad \nabla f(4,1) = \langle \tfrac{5}{8}, -\tfrac{1}{2}\rangle.$$

Therefore, the directional derivative of f at $(4,1)$ in the $\theta = -\frac{\pi}{6}$-direction is:

$$(D_{\widehat{u}}f)(P) = \left\langle \frac{5}{8}, -\frac{1}{2}\right\rangle \bullet \left\langle \frac{\sqrt{3}}{2}, \frac{-1}{2}\right\rangle = \boxed{\frac{5\sqrt{3}}{16} + \frac{1}{4}.}$$

Example 4.2.32. *Let $f(x,y) = 5xy^2 - 4x^3y$. Find the rate of change in f at $(1,2)$ in the $\langle a, b\rangle$ direction.*

Solution: *observe $\widehat{u} = \frac{1}{\sqrt{a^2+b^2}}\langle a, b\rangle$ is a unit-vector in the $\langle a, b\rangle$ direction. We calculate $(D_{\widehat{u}}f)(1,2)$ for $\widehat{u} = \frac{1}{\sqrt{a^2+b^2}}\langle a, b\rangle$. Notice $\nabla f(x,y) = \langle 5y^2 - 12x^2y, 10xy - 4x^3\rangle$ thus $\nabla f(1,2) = \langle -4, 16\rangle$. Consequently,*

$$(D_{\widehat{u}}f)(1,2) = \frac{1}{\sqrt{a^2+b^2}}\langle a, b\rangle \bullet \langle -4, 16\rangle = \boxed{\frac{-4a+16b}{\sqrt{a^2+b^2}}.}$$

The example above is interesting as it allows us to answer a host of interesting questions about the change in f at $(1,2)$. For example, f remains constant along the solution of $-4a + 16b = 0$ which has two solutions for which $a^2 + b^2 = 1$. Geometrically, that pair of solutions are the unit-vectors which are orthogonal to ∇f at $(1,2)$. Of course, all of these comments merely replicate the general results of Proposition 4.2.26.

Example 4.2.33. *Suppose $f(x,y) = \ln(x^2+y^2)$. Find the rate of change in f at (x,y) in the $\widehat{u} = \langle \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\rangle$ direction.*

Solution: *calculate the gradient at* (x, y) *as follows:*

$$\nabla f(x,y) = \left\langle \frac{2x}{x^2+y^2}, \frac{2y}{x^2+y^2}, \right\rangle$$

Thus,

$$(D_{\widehat{u}} f)(x,y) = \left\langle \frac{2x}{x^2+y^2}, \frac{2y}{x^2+y^2}, \right\rangle \bullet \left\langle \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right\rangle = \frac{(x-y)\sqrt{2}}{x^2+y^2}.$$

For a given circle $x^2 + y^2 = R^2$ *centered about the origin, the point in quadrant IV on the line* $y = -x$ *produces the largest change in the* $\langle \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \rangle$ *direction.*

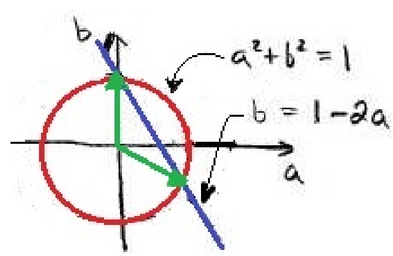
Can[5] you think of an applied problem which requires the analysis of the previous problem?

Example 4.2.34. *Find the direction(s), if any, in which* $f(x, y) = x^2 + \sin(xy)$ *has a rate of change of* 1 *at* $(1, 0)$*.*

Solution: *let* $a, b \in \mathbb{R}$ *such that* $a^2 + b^2 = 1$*. Let* $\widehat{u} = \langle a, b \rangle$*. Furthermore, calculate* $\nabla f(x, y) = \langle 2x + y\cos(xy), x\cos(xy)\rangle$ *hence* $\nabla f(1, 0) = \langle 2, 1\rangle$*. Our goal is to find all solutions for* $(D_{\widehat{u}} f)(1, 0) = 1$*. Thus, study:*

$$\langle 2, 1 \rangle \bullet \langle a, b \rangle = 1 \quad \Rightarrow \quad 2a + b = 1.$$

This gives $b = 1 - 2a$ *and as we imposed* $a^2 + b^2 = 1$ *we find* $a^2 + (1 - 2a)^2 = 1$ *which simplifies to* $a(5a - 4) = 0$ *thus* $a = 0$ *and* $a = 4/5$ *are solutions. But,* $b = 1 - 2a$ *hence* $a = 0$ *gives* $b = 1$ *whereas* $a = 4/5$ *gives* $b = -3/5$*. In summary, we find* $\boxed{\widehat{u} = \langle 0, 1\rangle}$ *and* $\boxed{\widehat{u} = \langle 4/5, -3/5\rangle}$*. Graphically, the algebra respects the geometric solution indicated by the intersection of the unit-circle and* $b = 1 - 2a$ *line in the ab-plane:*



4.2.2 gradient vector fields

We've seen that the value of ∇f at a particular point reveals both the magnitude and the direction of the change in the function f. The gradient vector field is simply the vector field which a differentiable function f generates through the gradient operation.

Definition 4.2.35.

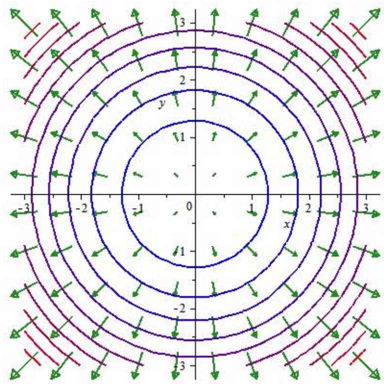
If f is differentiable on $U \subseteq \mathbb{R}^2$ then ∇f defines the gradient vector field on U. We assign to each point $\vec{p} \in U$ the vector $\nabla f(\vec{p})$.

[5] this is not one of those questions where the answer is just obvious and I'm hoping to poke you in the right direction. This question is very open ended.

Example 4.2.36. *Let $f(x,y) = x^2 + y^2$. We calculate,*

$$\nabla f(x,y) = \langle \partial_x(x^2+y^2), \partial_y(x^2+y^2)\rangle = \langle 2x, 2y\rangle$$

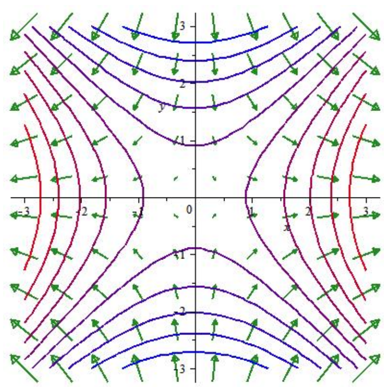
This gradient vector field is easily described; at each point $\vec{p}$ we attach the vector $2\vec{p}$.



Example 4.2.37. *Let $f(x,y) = x^2 - y^2$. We calculate,*

$$\nabla f(x,y) = \langle \partial_x(x^2-y^2), \partial_y(x^2-y^2)\rangle = \langle 2x, -2y\rangle$$

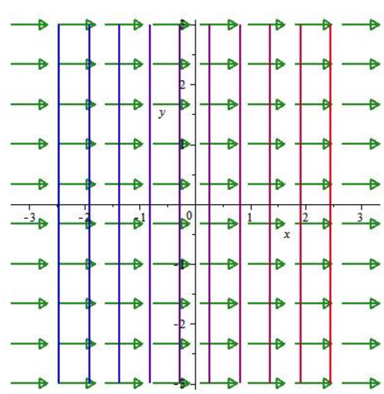
This gradient vector field is not so easily described, however, most CAS will provide nice plots if you are willing to invest a little time.



Example 4.2.38. *Let $f(x,y) = x$. We calculate,*

$$\nabla f(x,y) = \langle \partial_x(x), \partial_y(x)\rangle = \langle 1, 0\rangle = \widehat{x}$$

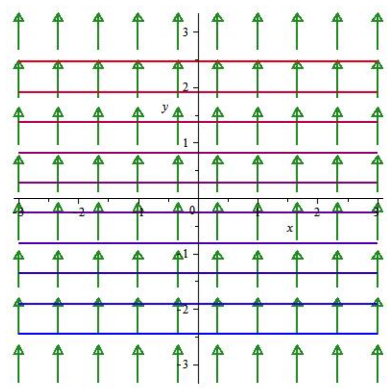
Therefore, $\nabla x = \widehat{x}$. Interesting. The gradient operation reproduces the unit-vector in the direction of increasing x.



Example 4.2.39. *Let $f(x,y) = y$. We calculate,*

$$\nabla f(x,y) = \langle \partial_x(y), \partial_y(y) \rangle = \langle 0, 1 \rangle = \widehat{y}$$

Therefore, $\nabla y = \widehat{y}$. Interesting. The gradient operation reproduces the unit-vector in the direction of increasing y.



Naturally, we are tempted to derive other unit-vector-fields by this method. In the examples above we were a bit lucky, generally when you take the gradient of a coordinate function you'll need to normalize it. But, this is a very nice **algebraic** method to derive the frame of a non-cartesian coordinate system. In particular, if y_1, y_2 are coordinates then there exist differentiable functions f_1, f_2 such that $y_1 = f_1(x,y)$ and $y_2 = f_2(x,y)$ we can calculate the unit-vectors

$$\widehat{y_1} = \frac{\nabla f_1}{||\nabla f_1||} \qquad \text{and} \qquad \widehat{y_2} = \frac{\nabla f_2}{||\nabla f_2||}.$$

Let's see how this method produces the frame for polar coordinates. I initially claimed it could be derived from geometry alone. That is true, but this is also nice:

Example 4.2.40. *Consider polar coordinates r, θ, these were defined by $r = \sqrt{x^2+y^2}$ and $\theta = \tan^{-1}\big[y/x\big]$ for $x > 0$. Calculate,*

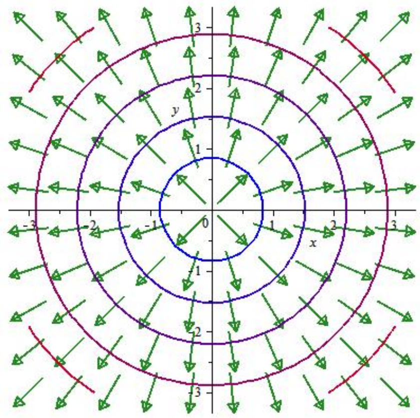
$$\nabla r = \left\langle \frac{\partial}{\partial x}\sqrt{x^2+y^2},\ \frac{\partial}{\partial y}\sqrt{x^2+y^2} \right\rangle = \left\langle \frac{x}{\sqrt{x^2+y^2}},\ \frac{y}{\sqrt{x^2+y^2}} \right\rangle = \left\langle \frac{x}{r},\ \frac{y}{r} \right\rangle$$

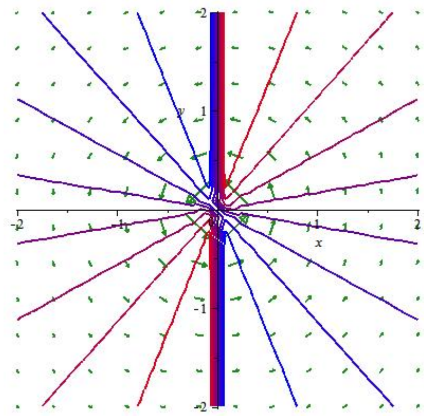
But, $x = r\cos\theta$ and $y = r\sin\theta$ thus we derive $\nabla r = \langle \cos\theta, \sin\theta \rangle$. Since $||\nabla r|| = 1$ we find $\boxed{\widehat{r} = \langle \cos\theta, \sin\theta \rangle.}$ The unit-vector in the direction of increasing θ is likewise calculated,

$$\nabla \theta = \left\langle \frac{\partial}{\partial x}\tan^{-1}\big[y/x\big],\ \frac{\partial}{\partial y}\tan^{-1}\big[y/x\big] \right\rangle = \left\langle \frac{-y}{x^2+y^2},\ \frac{x}{x^2+y^2} \right\rangle = \left\langle \frac{-y}{r^2},\ \frac{x}{r^2} \right\rangle.$$

In this case we find $\nabla\theta = \frac{1}{r}\langle -\sin\theta, \cos\theta \rangle$. Gradients and level curves of r and θ are plotted below[6]:

[6] notice how the software chokes on $x = 0$





The gradient of θ is not a unit-vector so we have to normalize. Since $||\nabla\theta|| = \frac{1}{r}$ we derive $\boxed{\widehat{\theta} = \langle -\sin\theta, \cos\theta\rangle.}$

This is a very nice calculation for coordinates which are not easy to visualize.

Another nice application of the gradient involves level curves. Consider this: a level curve is the set of points which solves $f(x, y) = k$ for some value k. If we consider a point (x_o, y_o) on the level curve $f(x, y) = k$ then the gradient vector $(\nabla f)(x_o, y_o)$ will be perpendicular to the tangent line of the level curve. Remember that when $\theta = \pi/2$ we find a direction in which $f(x, y)$ stays constant near (x_o, y_o). What does this mean? Let's summarize it:

The gradient vector field ∇f is perpendicular to the level curve $f(x, y) = k$.

If you are less than satisfied with my geometric justification for this claim then you'll be happy to hear we can prove it with a simple calculation. However, we need a chain-rule which we have yet to justify. Therefore, further justification is postponed until a later section. That said, let's look at a few examples to appreciate the power of this statement:

Example 4.2.41. *Suppose $V(x, y) = \frac{1}{\sqrt{x^2+y^2}}$ represents the voltage due to a point-charge at the origin. Electrostatics states that the electric field $\vec{E} = -\nabla V$. Geometrically this has a simple meaning; the electric field points along the normal direction to the level-curves of the voltage function. In other words, the electric field is normal to the equipotential lines. What is an "equipotential line", it's a line on which the voltage assumes a constant value. This is nothing more than a level-curve of the voltage function. For the given potential function, using $r = \sqrt{x^2 + y^2}$,*

$$\nabla V = \langle \partial_x(1/r), \partial_y(1/r)\rangle = \langle (-1/r^2)\partial_x r,\ (-1/r^2)\partial_y r\rangle = \frac{-1}{r^2}\langle \partial_x r,\ \partial_y r\rangle = -\frac{1}{r^2}\widehat{r}.$$

Equipotentials $V = V_o = 1/r$ are simply circles $r = 1/V_o$ and the electric field is a purely radial field $\vec{E} = \frac{1}{r^2}\widehat{r}$.

Example 4.2.42. *Consider the ellipse $f(x, y) = x^2/a^2 + y^2/b^2 = k$. At any point on the ellipse the vector field*

$$\nabla f = \frac{2x}{a^2}\,\widehat{x} + \frac{2y}{b^2}\,\widehat{y}$$

points in the normal direction to the ellipse.

Example 4.2.43. *Consider the hyperbolas $g(x, y) = x^2y^2 = k$. At any point on the hyperbolas the vector field*

$$\nabla g = 2xy^2\,\widehat{x} + 2x^2y\,\widehat{y}$$

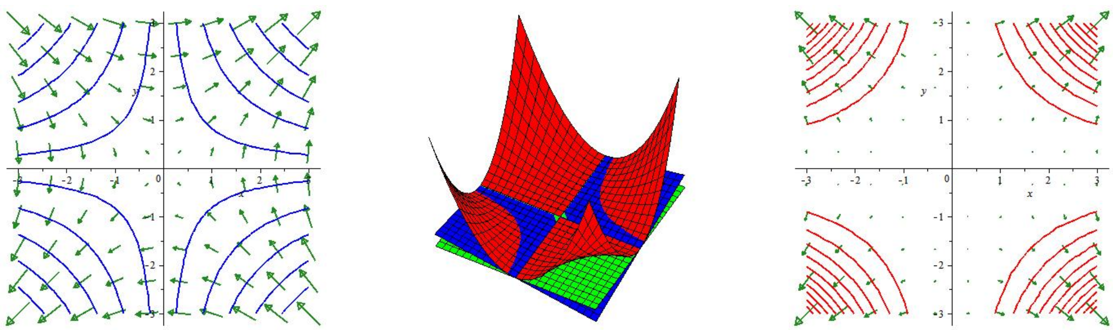
points in the normal direction to the hyperbola. Notice that for $k > 0$ we have $y^2 = k/x^2$ hence $y = \pm\sqrt{k}/x$. When $k = 0$ we find solutions $x = 0$ and $y = 0$. The gradient vector field is identically zero on the coordinate axes in this case. I plot it after the next example for the sake of side-by-side comparison

Example 4.2.44. *Suppose we have a level curve $f(x, y) = xy = k$. This either gives a hyperbola ($k \neq 0$) or the coordinate axes ($k = 0$). The gradient vector field is a bit more descriptive in this case:*

$$\nabla f = y\,\widehat{x} + x\,\widehat{y}.$$

In this case the exceptional solution $x = 0$ has $\nabla f\big|_{x=0} = y\,\widehat{x}$ and $y = 0$ has $\nabla f\big|_{y=0} = x\,\widehat{y}$. The origin $(0, 0)$ is the only critical point for f in this example.

I plot ∇f on the left and ∇g on the right together with a few level curves. The picture in the middle has $z = x^2y^2$ in red and $z = xy$ in blue with $z = 0$ in green for reference.



The last pair of examples goes to show that a given set of points can be described by many different level-functions. In particular notice that $xy = 1$ is covered by $x^2y^2 = 1$ but the level functions $f(x, y) = xy$ and $g(x, y) = x^2y^2$ change to other levels in rather distinct fashions. Just compare the gradient vector fields. Or, use a CAS[7] to graph $z = f(x, y)$ and $z = g(x, y)$. Those graphs will intersect along the curve $(x, 1/x, 1)$ for $x > 0$. Do they intersect anywhere else?

4.2.3 contour plots

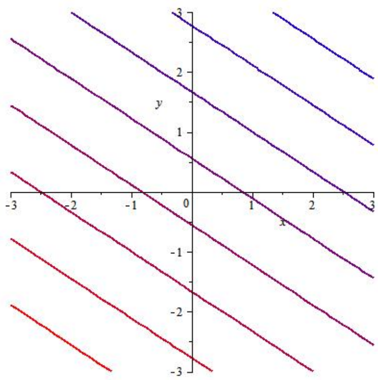
Perhaps you've studied a *topographical map* before. The topographical map uses a two-dimensional chart to plot a three-dimensional landscape. We can make a similar diagram for graphs of the form $z = f(x, y)$. To form such a plot we simply imagine projecting the graph at a few representative z-values down or up to the xy-plane. This is an invaluable tool since we have much better two-dimensional visualization than we do three. Few people can draw excellent three dimensional perspective, but the contour plot requires no understanding of perspective. We just slice and project. Moreover, we can use the gradient vector field as a sort of *compass*[8]. The gradient vector field in the domain of $f(x, y)$ points toward higher contours. I use the term *higher* with the idea of traveling from $f(x, y) = k_1$ to $f(x, y) = k_2$ where $k_1 < k_2$. If $f(x, y)$ was actually the altitude

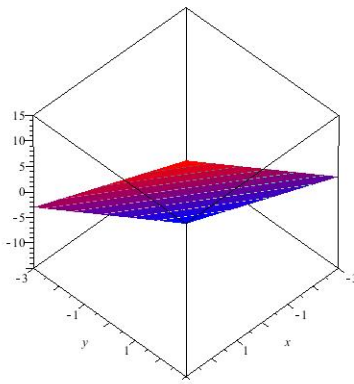
[7] I used Maple to create these graphs, of course you could use Mathematica or any other plotting tool, I have links to free ones on my website... I do expect you use something to aid your visualization.

[8] thanks to Dr. Monty Kester for this particular slogan

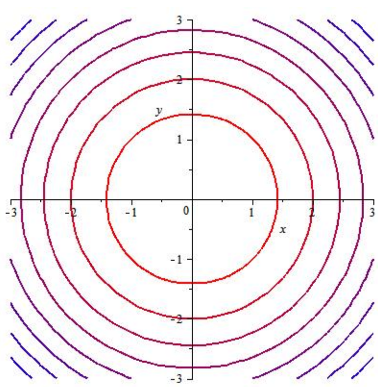
function then the term upward would be literally accurate. Usually the term has nothing to do with an actual height, that's just a mental picture for us to help think through the math.

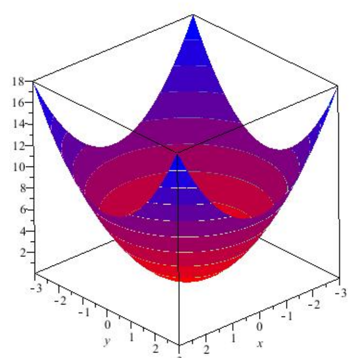
Example 4.2.45. *Suppose $f(x, y) = 2x + 3y$. The graph $z = f(x, y)$ is the plane $z = 2x + 3y$. Contours are level curves of the form $2x + 3y = k$. These contours are simply lines with x-intercept $k/2$ and y-intercept $k/3$. See the plot and graph below to appreciate how the contour plot and graph complement one another. Also, note there is no critical point in this example and the gradient vector field $\nabla f = \langle 2, 3\rangle$ is constant in the domain of f.*



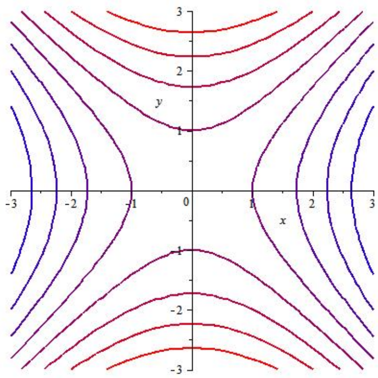
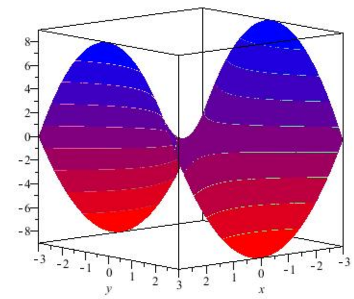


Example 4.2.46. *Suppose $f(x, y) = x^2 + y^2$. The graph $z = f(x, y)$ is the quadratic surface known as a paraboloid. Contours are level curves of the form $x^2 + y^2 = k$. These solutions of $x^2 + y^2 = k$ form circles of radius $\sqrt{k}$ for $k > 0$ and a solitary point $(0, 0)$ for $k = 0$. There are no contours with $k < 0$. Once more see how the graph a contour plot complement one another. Furthermore, observe that $\nabla f = \langle 2x, 2y\rangle$ is zero at the origin which is the only critical point. It's clear from the contours or the graph that $f(0, 0)$ is a local minimum for f. In fact, it's clear it is the global minimum for the function.*





Example 4.2.47. *Suppose $f(x, y) = x^2 - y^2$. The graph $z = f(x, y)$ is the quadratic surface known as a hyperboloid. Contours are level curves of the form $x^2 - y^2 = k$. These solutions of $x^2 - y^2 = k$ form hyperbolas which open up/down for $k < 0$ and open left/right for $k > 0$. If $k = 0$ the $x^2 - y^2 = 0$ yields the special case $y = \pm x$, these are asymptotes for all the hyperbolas from $k \neq 0$. Once more see how the graph and contour plot complement one another. Furthermore, observe that $\nabla f = \langle 2x, -2y\rangle$ is zero at the origin which is the only critical point. It's clear from the contours or the graph that $f(0, 0)$ is a not a local minimum or maximum for f. This sort of critical point is called a* **saddle point**.

Example 4.2.48. *Suppose $f(x,y) = \cos(x)$. The graph $z = f(x,y)$ is sort-of a wavy plane. Contours are solutions of the level curve equation $\cos(x) = k$. In this case y is free, however we only find non-empty solution sets for $-1 \leq k \leq 1$. For a particular $k \in [-1,1]$ we have the level-curve $\{(x,y) \mid \cos(x) = k\}$. Note that the cosine curve will reach k twice for each 2π interval in x. Let me pick on a few special values,*

$$k = 0, \quad \textit{solve} \quad \cos(x) = 0, \quad \textit{to obtain} \;\; x = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2}, \dots$$

The $k = 0$ contours are of the form $x = \frac{\pi}{2}(2n-1)$ for $n \in \mathbb{Z}$, there are infinitely many such contours and they are disconnected from one another. Another case which is easy to think through without a calculator,

$$k = 1/2, \quad \textit{solve} \quad \cos(x) = 1/2, \quad \textit{to obtain} \;\; x = -\frac{\pi}{3} + 2\pi n, \quad \textit{or} \;\; x = \frac{\pi}{3} + 2\pi n$$

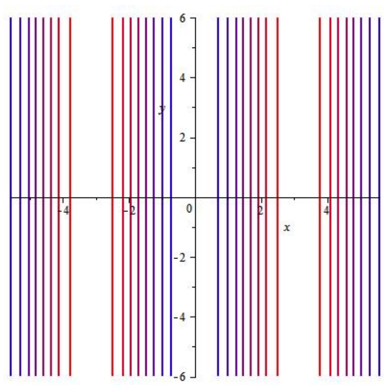
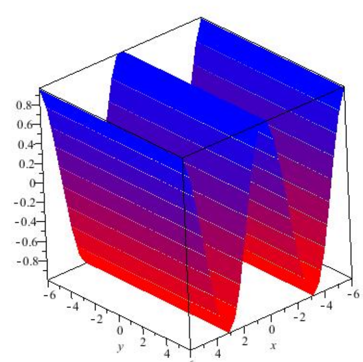
for $n \in \mathbb{Z}$. Once more the level-curves are vertical lines. Continuing, study $k = 1$,

$$k = 1, \quad \textit{solve} \quad \cos(x) = 1, \quad \textit{to obtain} \;\; x = 2\pi n, \textit{for} \;\; n \in \mathbb{Z}.$$

Likewise:

$$k = -1, \quad \textit{solve} \quad \cos(x) = -1, \quad \textit{to obtain} \;\; x = (2n-1)\pi, \textit{for} \;\; n \in \mathbb{Z}.$$

Observe the gradient $\nabla f = \langle -\sin(x), 0 \rangle$ is zero along the $k = \pm 1$ contours. The points on $k = 1$ give a local maximum whereas the points on $k = -1$ give local minima for f. This is a special sort of critical point since they are not isolated, no matter how close we zoom in there are always infinitely many critical points in a neighborhood of a given critical point.

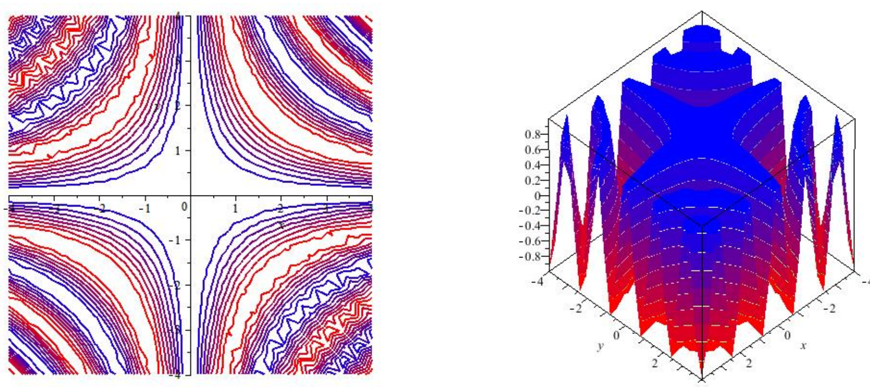
Example 4.2.49. *Suppose $f(x,y) = \cos(xy)$. Calculate $\nabla f = \langle -y\sin(xy), -x\sin(xy)\rangle$ it follows that solutions of $xy = n\pi$ for $n \in \mathbb{Z}$ give critical points of f. Contours are given by the level-curves $\cos(xy) = k$ which have nonempty solutions for $k \in [-1,1]$. For example, note that $\cos(xy) = 1$ has solution $xy = 2j\pi$ for some $j \in \mathbb{Z}$. In particular,*

$$xy = 0,\ xy = \pm 2\pi,\ xy = \pm 4\pi,\ \dots \quad \Rightarrow \quad y = 0, \quad x = 0,\ y = \pm\frac{2\pi}{x}, \quad y = \pm\frac{4\pi}{x},\ \dots$$

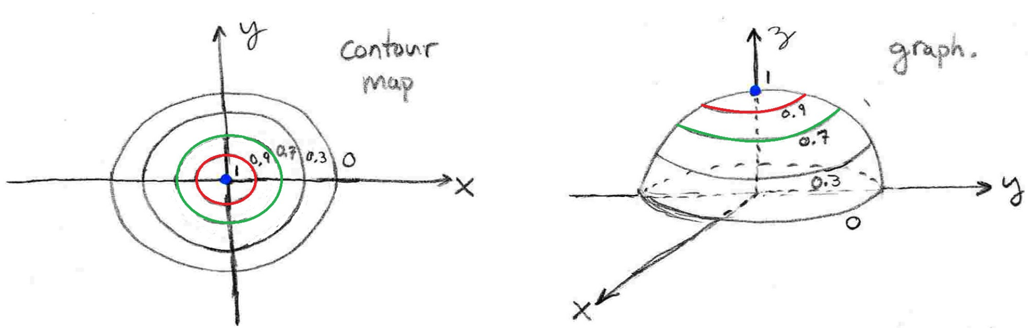
On the other hand, $\cos(xy) = -1$ has solution $xy = (2m-1)\pi$ for some $m \in \mathbb{Z}$. In particular,

$$xy = \pm\pi, \quad xy = \pm 3\pi,, \quad xy = \pm 5\pi \quad \dots \quad \Rightarrow \quad y = \pm\frac{\pi}{x}, \quad y = \pm\frac{3\pi}{x}, \quad y = \pm\frac{5\pi}{x},\ \dots$$

The contours are simply a family of hyperbolas which take the coordinate axes as asymptotes. This is a great example to see both why contour plots help us visualize the graph which we'd rather not illustrate three-dimensionally. Of course we can use a CAS to directly picture $z = f(x,y)$, but such pictures rarely yield the same sort of detailed information as a well-drawn contour plot.



Example 4.2.50. *Nice CAS (in this section I used Maple, but all mature CAS's have built-in contour tools) plots are a luxury we don't always have. Notice we can do much just with hand-drawn sketches. The sketch below is to help envision $z = \sqrt{1 - x^2 - y^2}$:*



4.3 partial differentiation in $\mathbb{R}^3$ and $\mathbb{R}^n$

Definition 4.3.1.

Let $f : dom(f) \subseteq \mathbb{R}^3 \to \mathbb{R}$ be a function with $(x_o, y_o, z_o) \in dom(f)$. If the directional derivative below exists, then we define the **partial derivative** of f at $\vec{p}_o = (x_o, y_o, z_o)$ with respect to x, y, z by

$$\frac{\partial f}{\partial x}(\vec{p}_o) = (D_{\widehat{x}} f)(\vec{p}_o), \qquad \frac{\partial f}{\partial y}(\vec{p}_o) = (D_{\widehat{y}} f)(\vec{p}_o), \qquad \frac{\partial f}{\partial z}(\vec{p}_o) = (D_{\widehat{z}} f)(\vec{p}_o)$$

respective. We also use the notations $\frac{\partial f}{\partial x} = \partial_x f = f_x$, $\frac{\partial f}{\partial y} = \partial_y f = f_y$ and $\frac{\partial f}{\partial z} = \partial_z f = f_z$. Generally, if $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ is a function with $\vec{p}_o \in dom(f)$ and the limit below exists, then we define the **partial derivative** of f at $\vec{p}_o$ with respect to x_j by

$$\frac{\partial f}{\partial x_j}(\vec{p}_o) = (D_{\widehat{x}_j} f)(\vec{p}_o).$$

The notation $\frac{\partial f}{\partial x_j} = \partial_j f$ is at times useful.

Once more we have natural interpretations for these partial derivatives:

f_x **gives the rate of change in** f **in the** x**-direction.**
f_y **gives the rate of change in** f **in the** y**-direction.**
f_z **gives the rate of change in** f **in the** z**-direction.**

It is useful to rewrite the definition of the partial derivatives explicitly in terms of derivatives.

$$\frac{\partial f}{\partial x}(x_o, y_o, z_o) = \frac{d}{dt}\Big[\ f(x_o + t, y_o, z_o)\ \Big]\Big|_{t=0}$$

$$\frac{\partial f}{\partial y}(x_o, y_o, z_o) = \frac{d}{dt}\Big[\ f(x_o, y_o + t, z_o)\ \Big]\Big|_{t=0}$$

$$\frac{\partial f}{\partial z}(x_o, y_o, z_o) = \frac{d}{dt}\Big[\ f(x_o, y_o, z_o + t)\ \Big]\Big|_{t=0}.$$

Partial differentiation is just differentiation where we hold all but one of the **independent variables** constant. Notice that z in the context above is an independent variable. In contrast, when we studied $z = f(x, y)$ the variable z was a **dependent variable**. The symbols x, y, z are not reserved. They have multiple meanings in multiple contexts and you must have the correct conceptual framework if you are to make the correct computations. When z, x are independent we have $\frac{\partial z}{\partial x} = 0$. If z, x are dependent then it is generally some function. In any event, the following proposition should be entirely unsurprising at this point:

Proposition 4.3.2.

Assume f, g are functions from $\mathbb{R}^3$ to $\mathbb{R}$ whose partial derivatives exist. Then for $c \in \mathbb{R}$,

1. $(f+g)_x = f_x + g_x$ and $(f+g)_y = f_y + g_y$ and $(f+g)_z = f_z + g_z$.
2. $(cf)_x = cf_x$ and $(cf)_y = cf_y$ and $(cf)_z = cf_z$.
3. $(fg)_x = f_x g + f g_x$ and $(fg)_y = f_y g + f g_y$ and $(fg)_z = f_z g + f g_z$.

Moreover, if $h : dom(h) \subseteq \mathbb{R} \to \mathbb{R}$ is a differentiable and $x_1 = x,\ x_2 = y,\ x_3 = z$,

4. $\dfrac{\partial}{\partial x_j}\big[h(f(x_1,x_2,x_3))\big] = \dfrac{dh}{dt}\Big|_{f(x_1,x_2,x_3)} \dfrac{\partial f}{\partial x_j} = \dfrac{dh}{df}\dfrac{\partial f}{\partial x_j}$
5. $\dfrac{\partial x_i}{\partial x_j} = \delta_{ij}$ where $x_1 = x, x_2 = y, x_3 = z$.

Proof: The proofs are nearly identical to those given in the $n = 2$ case. However, I will offer a proof of (5.) for arbitrary n. Suppose $f(x_1, x_2, \dots, x_n) = x_i = \vec{x} \bullet \widehat{x}_i$ and calculate

$$\frac{\partial f}{\partial x_j} = \lim_{t\to 0}\left[\frac{f(\vec{x}) - f(\vec{x} + t\,\widehat{x}_j)}{t}\right] = \lim_{t\to 0}\left[\frac{x_i - (\vec{x} + t\,\widehat{x}_j)\bullet \widehat{x}_i}{t}\right] = \lim_{t\to 0}\left[\frac{x_i - x_i + t\delta_{ij}}{t}\right] = \delta_{ij}.$$

Therefore, $\partial_j x_i = \delta_{ij}$ for all $i, j \in \mathbb{N}_n$. In particular, this result applies to the case $n = 3$ hence the proof of (5.) is complete. Naturally this proposition equally well applies to $f : \mathbb{R}^n \to \mathbb{R}$. The proofs are nearly identical to the $n = 2$ case, we just have a few sums to sort through. I leave those to the reader. $\square$

Example 4.3.3. *Let* $g(x, y, z) = xy^2z^3 + \sin(xyz)$

$$\begin{aligned} g_x &= y^2z^3 + yz\cos(xyz), \\ g_y &= 2xyz^3 + xz\cos(xyz), \\ g_z &= 3xy^2z^2 + xy\cos(xyz). \end{aligned}$$

Example 4.3.4. *Let* $f(x, y, z) = \frac{x}{y+z}$. *We calculate* $\frac{\partial f}{\partial z}(3, 2, 1)$*:*

$$\frac{\partial f}{\partial z} = \frac{\partial}{\partial z}\left(\frac{x}{y+z}\right) = \frac{-x}{(y+z)^2}\frac{\partial}{\partial z}(y+z) = \frac{-x}{(y+z)^2} \quad \Rightarrow \quad f_z(3,2,1) = \frac{-1}{3}.$$

Example 4.3.5. *If* $r = \sqrt{x^2 + y^2 + z^2}$ *then*

$$\frac{\partial r}{\partial x} = \frac{1}{2\sqrt{x^2+y^2+z^2}}\frac{\partial}{\partial x}\left[x^2+y^2+z^2\right] = \frac{x}{\sqrt{x^2+y^2+z^2}} = \frac{x}{r} \quad \Rightarrow \quad \frac{\partial r}{\partial x} = \frac{x}{r}.$$

Likewise, by symmetry, $\frac{\partial r}{\partial y} = \frac{y}{r}$ *and* $\frac{\partial r}{\partial z} = \frac{z}{r}$.

The example above extends to n-dimensions without much more work:

Example 4.3.6. *Suppoose* $u = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ *then for* $j \in \mathbb{N}$ *with* $1 \leq j \leq n$

$$\frac{\partial u}{\partial x_j} = \frac{\partial}{\partial x_j}\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \frac{1}{2\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}}\frac{\partial}{\partial x_j}\left[x_1^2 + x_2^2 + \cdots + x_n^2\right] = \frac{x_j}{u}.$$

In other words, we find: $\dfrac{\partial r}{\partial x_j} = \dfrac{x_j}{r}$ *for* $j = 1, 2, \ldots, n$.

Example 4.3.7. *Verify that* $u = \frac{1}{\sqrt{x^2+y^2+z^2}}$ *is a solution of Laplace's equation in three dimensions. That is, show* $u_{xx} + u_{yy} + u_{zz} = 0$.

Solution: *Let* $r = \sqrt{x^2 + y^2 + z^2}$ *and note* $\frac{\partial r}{\partial x} = \frac{x}{r}$, $\frac{\partial r}{\partial y} = \frac{y}{r}$ *and* $\frac{\partial r}{\partial z} = \frac{z}{r}$. *These three relations are nicely summarized as* $\frac{\partial r}{\partial x_j} = \frac{x_j}{r}$ *for* $j = 1, 2, 3$. *Let* $u = 1/r$ *and calculate*

$$\frac{\partial u}{\partial x_j} = \frac{\partial}{\partial x_j}\left[\frac{1}{r}\right] = \frac{-1}{r^2}\frac{\partial r}{\partial x_j} = \frac{-1}{r^2}\cdot\frac{x_j}{r} = \frac{-x_j}{r^3}.$$

Differentiate once more,

$$\frac{\partial^2 u}{\partial x_j^2} = \frac{\partial}{\partial x_j}\left[\frac{-x_j}{r^3}\right] = \frac{-1}{r^3} + \frac{3x_j}{r^4}\frac{\partial r}{\partial x_j} = \frac{-1}{r^3} + \frac{3x_j^2}{r^5}$$

In view of the calculation above, collecting like terms, we find:

$$u_{xx} + u_{yy} + u_{zz} = \frac{-3}{r^3} + \frac{3(x^2 + y^2 + z^2)}{r^5} = \frac{-3}{r^3} + \frac{3r^2}{r^5} = 0.$$

4.3.1 directional derivatives and the gradient in $\mathbb{R}^3$ and $\mathbb{R}^n$

The idea of Example 4.2.20 equally well transfer to functions of three or more variables. We usually require the functions we analyze to be continuously differentiable since that avoids certain pathological examples:

Definition 4.3.8.

A function $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be **continuously differentiable** at $\vec{p}_o \in dom(f)$ iff the partial derivative functions $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots \frac{\partial f}{\partial x_n}$ are continuous at $\vec{p}_o$. We say $f \in C^1(\vec{p}_o)$. If all the second-order partial derivatives of f are continuous at $\vec{p}_o$ then we say $f \in C^2(\vec{p}_o)$. If continuous partial derivatives of arbitrary order exist at $\vec{p}_o$ then we say f is **smooth** and write $f \in C^\infty(\vec{p}_o)$.

We'll see an example in the next section where the formula below holds for a multivariate functions which is not even continuously differentiable, however the geometric analysis which flows from this formula is most meaningful for continuously differentiable functions.

Proposition 4.3.9.

Suppose $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable at $\vec{p}_o \in \mathbb{R}^n$ then the directional derivative at $\vec{p}_o$ in the direction of the unit vector $\widehat{u}$ is given by:

$$D_{\widehat{u}} f(\vec{p}_o) = \left\langle\ \partial_1 f(\vec{p}_o),\ \partial_2 f(\vec{p}_o), \ldots, \partial_n f(\vec{p}_o)\ \right\rangle \bullet \widehat{u}.$$

Proof: delayed until the next section. $\square$

At this point it is useful to introduce a convenient notation which groups all the partial derivatives together in a particular vector of functions. Notice that the length of the gradient vector depends on the context in which it is used.

Definition 4.3.10.

If the partial derivatives of f exist then we define

$$\nabla f = \langle \partial_1 f,\ \partial_1 f,\ \dots,\ \partial_n f\rangle = \widehat{x}_1 \frac{\partial f}{\partial x_1} + \widehat{x}_2 \frac{\partial f}{\partial x_2} + \cdots + \widehat{x}_n \frac{\partial f}{\partial x_n}.$$

we also use the notation $grad(f)$ and call this the **gradient** of f.

The upside-down triangle ∇ is also known as *nabla*. Identify that for $\mathbb{R}^3$ $\nabla = \widehat{x}\partial_x + \widehat{y}\partial_y + \widehat{z}\partial_z$. The operator ∇ takes a function f and produces a vector field ∇f. This is called the **gradient vector field** of f. For a continuously differentiable function we have the following beautiful formula for the directional derivative:

$$\boxed{D_{\widehat{u}} f(\vec{p}_o) = (\nabla f)(\vec{p}_o) \bullet \widehat{u}.}$$

Technically this isn't the definition, but pragmatically this is almost always what we use to work out problems. We can also write the dot-product in terms of lengths and the angle between the gradient vector $(\nabla f)(\vec{p}_o)$ and the unit-direction vector $\widehat{u}$:

$$\boxed{D_{\widehat{u}} f(\vec{p}_o) = ||(\nabla f)(\vec{p}_o)|| \cos\theta.}$$

Just like the $n = 2$ case we can use the gradient vector field to point us in the directions in which f either increases, decreases or simply stays constant.

Example 4.3.11. Problem: *Suppose $f(x, y, z) = x^2 + y^2 + z^2$. Does f increase at a rate of 10 in any direction at the point $(1, 2, 3)$?*

Solution: *Note $\nabla f(x, y, z) = \langle 2x, 2y, 2z\rangle$ thus $\nabla f(1, 2, 3) = \langle 2, 4, 6\rangle$. The magnitude of $\nabla f(1, 2, 3)$ is $||\nabla f(1, 2, 3)|| = \sqrt{4 + 16 + 36} = \sqrt{56}$ and that is the maximum rate possible. Therefore, the answer is no. This function increases at a rate of $\sqrt{56}$ in the direction $\frac{1}{\sqrt{14}}\langle 1, 2, 3\rangle$.*

Example 4.3.12. Problem: *Suppose $f(x, y, z) = 2x + y + 2z$. Does f increase at a rate of 2 in any direction at the point $(1, 1, 1)$?*

Solution: *Note $\nabla f(x, y, z) = \langle 2, 1, 2\rangle$ thus $\nabla f(1, 1, 1) = \langle 2, 1, 2\rangle$. The magnitude $||\nabla f(1, 1, 1)|| = \sqrt{9} = 3$ and that is the maximum rate possible. Therefore, the answer is yes. Now let's find the direction(s) in which this occurs. Solve:*

$$D_{\langle a,b,c\rangle} f(1, 1, 1) = \langle 2, 1, 2\rangle \bullet \langle a, b, c\rangle = 2a + b + 2c = 2$$

subject the unit-vector condition $a^2 + b^2 + c^2 = 1$. I'll eliminate c by solving the linear equation for $c = \frac{1}{2}(2 - 2a - b)$ and substituting:

$$a^2 + b^2 + \frac{1}{4}(2 - 2a - b)^2 = 1.$$

This give an ellipse in a,b-space. Apparently there is not just one direction where f increases at a rate of 2. There are infinitely many. For example, we can easily solve the ellipse equation for its b-intercepts by putting $a = 0$,

$$b^2 + \frac{1}{4}(2-b)^2 = 1 \quad \Rightarrow \quad 4b^2 + 4 - 4b + b^2 = 4 \quad \Rightarrow \quad 5b^2 - 4b = 0 \quad \Rightarrow \quad b(5b-4) = 0.$$

We find the points $(0,0)$ and $(0,4/5)$ are on the ellipse. Returning to the plane equation we find the c-value for these points by substituting them into the equation $c = \frac{1}{2}(2-2a-b)$:

$$(0,0): \quad c = \frac{1}{2}(2-2a-b) = 1 \qquad \& \qquad (0,4/5): \quad c = \frac{1}{2}(2-4/5) = \frac{1}{2}\cdot\frac{6}{5} = \frac{3}{5}.$$

Thus, we find the direction vectors $\langle 0,0,1\rangle$ and $\langle 0,\frac{4}{5},\frac{3}{5}\rangle$ point where f increases at a rate of 2. You can probably see a few more possibilities by just thinking about $(\nabla f)(1,1,1) = \langle 2,1,2\rangle$ directly. For example, I see $\langle 1,0,0\rangle$ also works.

The two-dimensional analogue of this problem is much easier since we have to solve the intersection of a line and the unit-circle. In that case there are either $0,1$ or 2 solutions. The three dimensional case is much more interesting. If f models the temperature at the point (x,y,z) then this calculation shows there are many directions in which the temperature increases at a rate of 2.

Example 4.3.13. *Let $f(x,y,z) = xe^y + ye^z + ze^x$. Find the rate of change of f at $(0,0,0)$ in the $\vec{v} = \langle 5,1,-2\rangle$ direction.*

Solution: *Since $\vec{v}$ has length $v = \sqrt{25+1+4} = \sqrt{30}$ hence $\widehat{v} = \frac{1}{\sqrt{30}}\langle 5,1,-2\rangle$. Calculate,*

$$\nabla f(x,y,z) = \langle e^y + ze^x,\ xe^y + e^z,\ ye^z + e^x\rangle \quad \Rightarrow \quad \nabla f(0,0,0) = \langle 1,1,1\rangle.$$

Therefore, $D_{\widehat{v}}f(0,0,0) = \frac{1}{\sqrt{30}}\langle 5,1,-2\rangle \bullet \langle 1,1,1\rangle = \frac{4}{\sqrt{30}}$.

Example 4.3.14. *Calculate $D_{\widehat{v}}f(4,1,1)$ given that $\vec{v} = \langle 1,2,3\rangle$ and $f(x,y,z) = \frac{x}{y+z}$.*

Solution: *note $v = \sqrt{1+4+9} = \sqrt{14}$ thus $\hat{v} = \frac{1}{\sqrt{14}}\langle 1,2,3\rangle$. We calculate the gradient,*

$$\nabla f(x,y,z) = \left\langle \frac{1}{y+z}, \frac{-x}{(y+z)^2}, \frac{-x}{(y+z)^2}\right\rangle \quad \Rightarrow \quad \nabla f(4,1,1) = \left\langle \frac{1}{2}, -1, -1\right\rangle$$

Therefore, $D_{\hat{v}}f(4,1,1) = \frac{1}{\sqrt{14}}\big\langle 1,2,3\big\rangle \bullet \big\langle \frac{1}{2},-1,-1\big\rangle = \frac{-9}{2\sqrt{14}}$.

Example 4.3.15. *Let $f(x,y,z) = 5\sqrt{x^2+y^2+z^2}$ find the maximum rate of change at $(3,6,-2)$ and determine the direction in which it occurs.*

Solution: *let $r = \sqrt{x^2+y^2+z^2}$. I invite the reader to verify that*

$$\nabla f(x,y,z) = \tfrac{5}{r}\langle x,y,z\rangle.$$

Furthermore, $\nabla f(3,6,-2) = \frac{5}{7}\langle 3,6,-2\rangle$. Notice,

$$||\nabla f(3,6,-2)|| = \tfrac{5}{7}\sqrt{9+36+4} = 5$$

thus, if θ is the angle between $\nabla f(3,6,-2)$ and $\widehat{u}$ then

$$D_{\widehat{u}}f(3,6,-2) = \nabla f(3,6,-2) \bullet \widehat{u} = 5\cos\theta$$

The maximum rate of change is 5 which is attained when $\widehat{u} = \frac{1}{7}\langle 3,6,-2\rangle$.

Example 4.3.16. *Let $f(x, y, z) = x + y^2 + z^3$. Find the directional derivative of f at $(-3, 2, -6)$ in the direction of the origin.*

Solution: *the origin is in the direction of $-P$ at P. In particular, $(0, 0, 0) - (-3, 2, -6) = \langle 3, -2, 6\rangle = \vec{u}$ points from $(-3, 2, -6)$ to $(0, 0, 0)$. Moreover, as $u = 7$ we find $\widehat{u} = \frac{1}{7}\langle 3, -2, 6\rangle$. Note,*

$$\nabla f(x, y, z) = \langle 1, 2y, 3z^2\rangle \quad \Rightarrow \quad \nabla f(-3, 2, -6) = \langle 1, 4, 108\rangle.$$

Therefore,

$$D_{\widehat{u}} f(-3, 2, -6) = \frac{1}{7}\langle 3, -2, 6\rangle \bullet \langle 1, 4, 108\rangle = \frac{1}{7}\big[3 - 8 + 648\big] = \frac{643}{7}.$$

We extend the definition of critical point to the general case in the obvious way:

Definition 4.3.17.

We say $\vec{p}_o$ is a **critical point** of f if $(\nabla f)(\vec{p}_o)$ does not exist or $(\nabla f)(\vec{p}_o) = \vec{0}$.

For the function in Example 4.3.11 the origin $(0, 0, 0)$ is the only critical point. On the other hand, the function in Example 4.3.12 has no critical point.

4.3.2 gradient vector fields in $\mathbb{R}^3$ and $\mathbb{R}^n$

We can calculate the gradient vector field for functions on $\mathbb{R}^n$ with $n \geq 1$ but, visualization is beyond most of us if $n > 3$. I mainly focus on the $n = 3$ case here and we see how the gradient aids our understanding of non-cartesian coordinate systems. Then we examine how the gradient vector field naturally provides a normal vector field to a level surface.

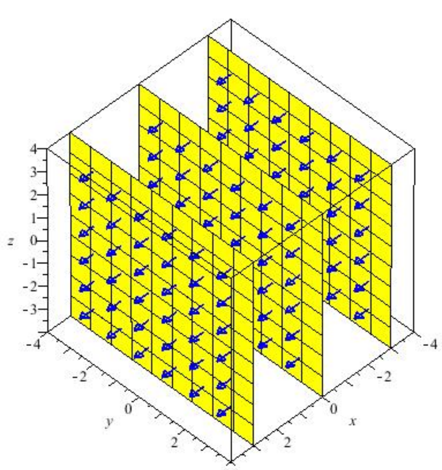
Definition 4.3.18.

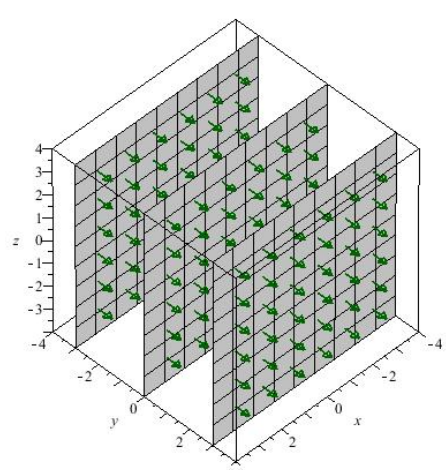
If f is differentiable on $U \subseteq \mathbb{R}^n$ then ∇f defines the gradient vector field on U. We assign to each point $\vec{p} \in U$ the vector $\nabla f(\vec{p})$.

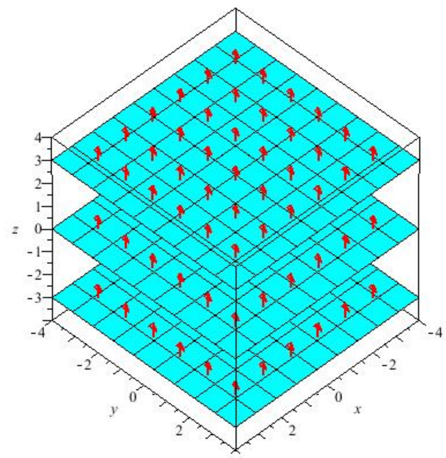
Example 4.3.19. *If x, y, z denote the coordinate functions on $\mathbb{R}^3$ then we find*

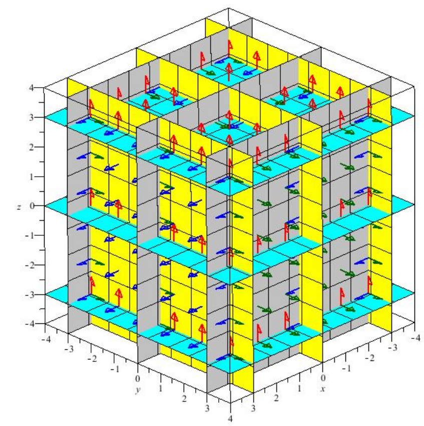
$$\nabla x = \langle 1, 0, 0\rangle = \widehat{x},$$
$$\nabla y = \langle 0, 1, 0\rangle = \widehat{y},$$
$$\nabla z = \langle 0, 0, 1\rangle = \widehat{z}.$$

These define constant vector fields on $\mathbb{R}^3$.









Generally, the gradient vector fields of the coordinate functions of a non-cartesian coordinate system provide a vector fields which point in the direction of increasing coordinates. To obtain unit-vectors we simply normalize the gradient vector fields. In particular, if $y_1, y_2, \dots y_n$ are coordinates on $\mathbb{R}^n$ then there exist differentiable functions $f_1, f_2, \dots, f_n$ such that $y_j = f_j(x_1, x_2, \dots, x_n)$ for $j = 1, 2, \dots, n$. We can define:

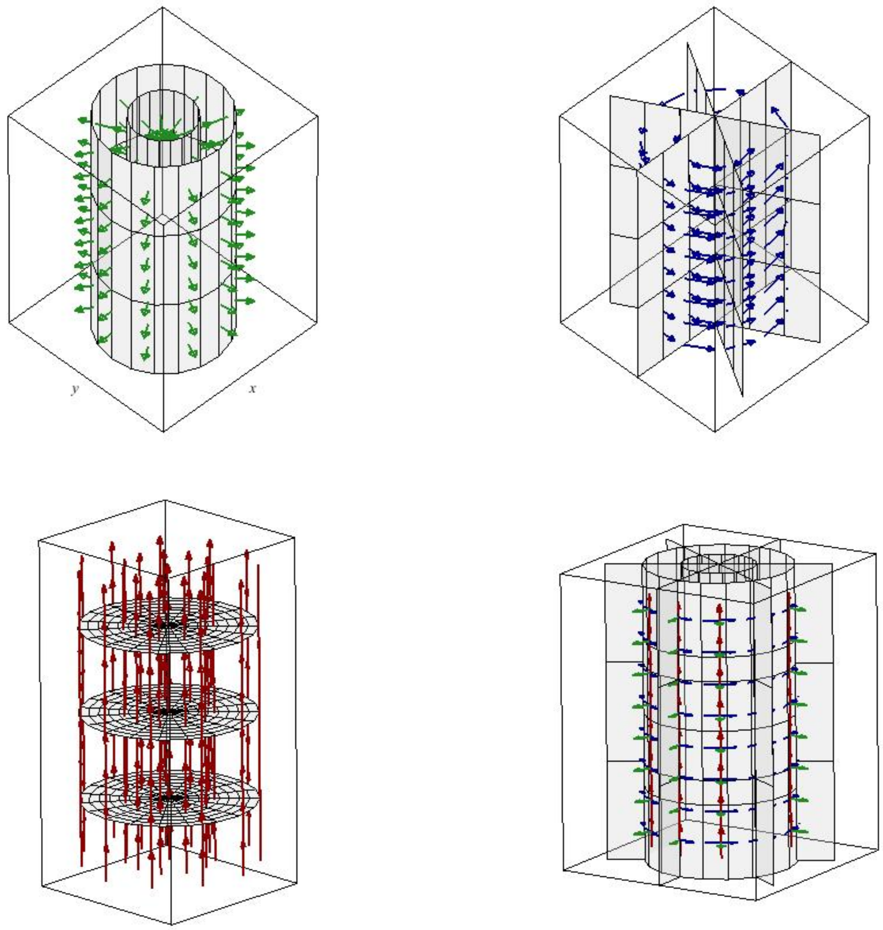
$$\widehat{y}_1 = \frac{\nabla f_1}{||\nabla f_1||} \quad \text{and} \quad \widehat{y}_2 = \frac{\nabla f_2}{||\nabla f_2||}, \ \dots, \quad \widehat{y}_n = \frac{\nabla f_n}{||\nabla f_n||}.$$

I mention this general idea for the interested reader. We are primarily interested in the cylindrical and spherical three dimensional coordinate systems. That's just a custom, we could easily extend these techniques to orthonormal coordinates based on ellipses or hyperbolas. If we are willing to give up on nice distance formulas we could even use coordinates based on tilted lines which meet at angles other than 90 degrees.

Example 4.3.20. *For cylindrical coordinates r, θ, z we can easily derive (following the same calculational steps as the polar two-dimensional case)*

$$\begin{aligned} \widehat{r} &= \tfrac{1}{||\nabla r||}\nabla r = \langle \cos(\theta), \sin(\theta), 0\rangle \\ \widehat{\theta} &= \tfrac{1}{||\nabla \theta||}\nabla \theta = \langle -\sin(\theta), \cos(\theta), 0\rangle \\ \widehat{z} &= \tfrac{1}{||\nabla z||}\nabla z = \langle 0, 0, 1\rangle \end{aligned}$$

The difference between the calculations above and the polar coordinate case is that cylindrical coordinates are three dimensional and that means the gradient vector fields of the coordinate functions are three dimensional vector fields. I advocated a geometric derivation of these cylindrical unit vectors earlier in this course, but we now have computational method which requires almost no geometric intuition.



Example 4.3.21. *Suppose ρ, ϕ, θ denote spherical coordinates. Recall*[9]

$$\rho = \sqrt{x^2+y^2+z^2}, \quad \phi = \cos^{-1}\left(\frac{z}{\sqrt{x^2+y^2+z^2}} \right), \quad \theta = \tan^{-1}\left(\frac{y}{x} \right)$$

You can calculate that

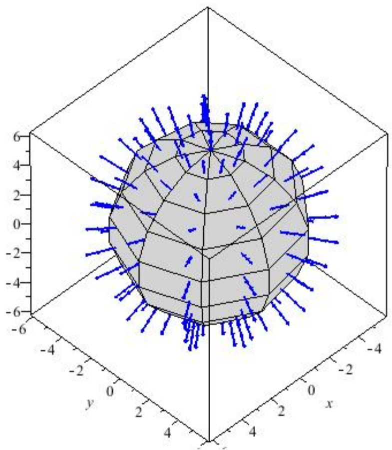
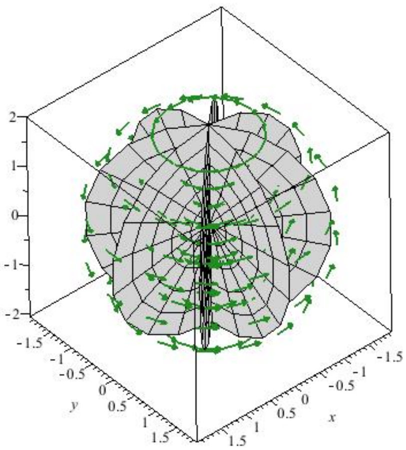
$$\begin{aligned} \widehat{\rho} &= \tfrac{1}{||\nabla\rho||}\nabla\rho = \sin(\phi)\cos(\theta)\,\widehat{x} + \sin(\phi)\sin(\theta)\,\widehat{y} + \cos(\phi)\,\widehat{z} \\ \widehat{\phi} &= \tfrac{1}{||\nabla\phi||}\nabla\phi = \cos(\phi)\cos(\theta)\,\widehat{x} + \cos(\phi)\sin(\theta)\,\widehat{y} - \sin(\phi)\,\widehat{z} \\ \widehat{\theta} &= \tfrac{1}{||\nabla\theta||}\nabla\theta = -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}. \end{aligned}$$

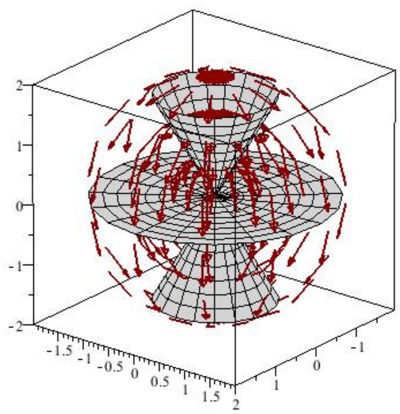
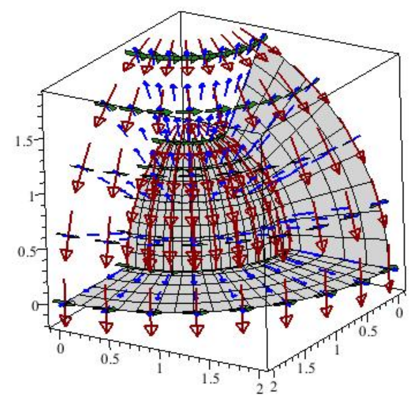
I'll walk you through the ρ calculation. To begin you can show that $\nabla\rho = \langle x/\rho,\ y/\rho,\ z/\rho\rangle$. But, we also know $x = \rho\cos\theta\sin\phi$, $y = \rho\sin\theta\sin\phi$ and $z = \rho\cos\phi$. Therefore,

$$\nabla\rho = \langle \cos\theta\sin\phi,\ \sin\theta\sin\phi,\ \cos\phi\rangle.$$

But, $||\nabla\rho|| = 1$. We derive that $\widehat{\rho} = \langle \cos\theta\sin\phi,\ \sin\theta\sin\phi,\ \cos\phi\rangle$. Perhaps I asked you to verify the formulas for $\widehat{\phi}, \widehat{\theta}$ in your homework. Making nice pictures of the spherical frame is an art I have yet to master... here's my best for now:

[9] these formulas only apply for certain octants, however, the ambiguity for the remaining octants only involves shifting the angular formulas by a constant. As you continue to read you'll notice that differentiation ultimately will kill any such constant so these formulas suffice.

Another nice application of the gradient involves level surfaces. Consider this: a level surface is the set of points which solves $f(x, y, z) = k$ for some value k. If we consider a point (x_o, y_o, z_o) on the level surface $f(x, y, z) = k$ then the gradient vector $(\nabla f)(x_o, y_o, z_o)$ will be perpendicular to the tangent plane of the level surface. Remember that when $\theta = \pi/2$ we find a direction in which $f(x, y, z)$ stays constant near (x_o, y_o, z_o). What does this mean? Let's summarize it:

The gradient vector field ∇f is normal to the level surface $f(x, y, z) = k$.

I use geometric intuition to make this claim here. We will offer a better proof later in this chapter. For now, let's try to appreciate the geometry.

Example 4.3.22. *Suppose $V(x, y, z) = \frac{1}{\sqrt{x^2+y^2+z^2}}$ represents the voltage due to a point-charge at the origin. Electrostatics states that the electric field $\vec{E} = -\nabla V$. Geometrically this has a simple meaning; the electric field points along the normal direction to the level-surfaces of the voltage function*[10]*. In other words, the electric field vectors are normal to the equipotential surfaces where they are attached. What is an "equipotential surface", it's a surface on which the voltage assumes a constant value. This is nothing more than a level-surface of the voltage function. For the given*

[10]The voltage function is the electric potential or simply the potential function in this context

potential function, using $\rho = \sqrt{x^2+y^2+z^2}$,

$$\begin{aligned}\nabla V &= \langle \partial_x(1/\rho), \partial_y(1/\rho), \partial_z(1/\rho)\rangle \\ &= \langle (-1/\rho^2)\partial_x\rho,\ (-1/\rho^2)\partial_y\rho,\ (-1/\rho^2)\partial_z\rho\rangle \\ &= \frac{-1}{\rho^2}\langle \partial_x\rho,\ \partial_y\rho,\ \partial_z\rho\rangle \\ &= -\frac{1}{\rho^2}\widehat{\rho}.\end{aligned}$$

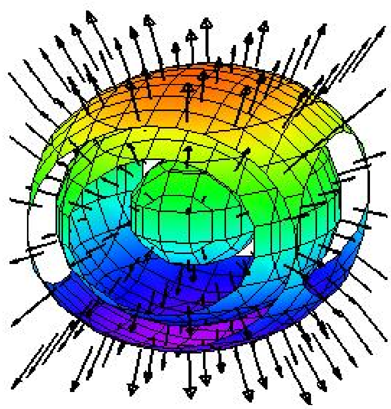
Equipotentials $V = V_o = 1/\rho$ *are simply spheres* $\rho = 1/V_o$ *and the electric field is a purely radial field* $\vec{E} = \frac{1}{r^2}\widehat{\rho}$.

Example 4.3.23. *Consider the ellipsoid* $f(x,y,z) = x^2/a^2 + y^2/b^2 + z^2/c^2 = k$. *At any point on the ellipse the vector field*

$$\nabla f = \frac{2x}{a^2}\,\widehat{x} + \frac{2y}{b^2}\,\widehat{y} + \frac{2z}{c^2}\,\widehat{z}$$

points in the normal direction to the ellipsoid.

It amazes me how easy it is to find a formula to assign a normal-vector to an arbitrary point on an ellipsoid. Imagine solving that problem without calculus.



4.4 the general derivative

Thus far we have primarily discussed partial derivatives in their connection to the rate of change of a given function in a particular direction. However, we would like to characterize the change in the function as a whole. Moreover, even in the one-dimensional case the derivative was closely tied to the best linear approximation to the function. In the single variable case it is as simple as this: the best linear approximation to a differentiable function at a point is the linearization of the function at that point whose graph is the tangent line. The slope of the tangent line is the value of the derivative function at the point. How do these ideas generalize? I take an n-dimensional approach in the beginning of this section because little is gained by talking in lower dimensions for the basic definitions.

Definition 4.4.1.

Suppose that U is open and $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a mapping the we say that $\vec{F}$ is **differentiable** at $\vec{a} \in U$ iff there exists a linear mapping $\vec{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{h} \to 0} \frac{\vec{F}(\vec{a}+\vec{h}) - \vec{F}(\vec{a}) - \vec{L}(\vec{h})}{||\vec{h}||} = 0.$$

In such a case we call the linear mapping $\vec{L}$ the **differential at** $\vec{a}$ and we denote $\vec{L} = d\vec{F}_{\vec{a}}$. The matrix of the differential is called the **derivative of** $\vec{F}$ **at** $\vec{a}$ and we denote $[d\vec{F}_{\vec{a}}] = \vec{F}'(\vec{a}) \in \mathbb{R}^{\, m\times n}$ which means that $d\vec{F}_{\vec{a}}(\vec{v}) = \vec{F}'(\vec{a})\vec{v}$ for all $\vec{v} \in \mathbb{R}^n$.

4.4.1 matrix of the derivative

If we know a function is differentiable at a point then we can calculate the formula for $\vec{L}$ in terms of partial derivatives. In particular, if $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{a} \in U$ then the differential $d\vec{F}_{\vec{a}}$ has the derivative matrix $\vec{F}'(\vec{a})$ which has components expressed in terms of partial derivatives of the component functions:

$$\left[\, d\vec{F}_{\vec{a}} \,\right]_{ij} = \partial_j F_i = \frac{\partial F_i}{\partial x_j}(\, \vec{a}\,)$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. This result is proved in advanced calculus. Let me expand this claim in detail for a few common cases: in each case we note $\vec{L}(\vec{a}+\vec{h}) \approx \vec{F}(a) + \vec{F}'(\vec{a})\vec{h}$

1. **function on** $\mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$, $\boxed{L(a+h) \approx f(a) + f'(a)h}$ the derivative matrix is just the derivative $f'(a)$ at the point.

2. **path into** $\mathbb{R}^n$, $\vec{r} : \mathbb{R} \to \mathbb{R}^n$, $\boxed{\vec{r}(a+h) \approx \vec{r}(a) + \vec{r}\,'(a)h}$. The derivative matrix is just the velocity vector $\vec{r}\,'(a)$ viewed as an $n \times 1$ matrix (it's a column vector).

3. **multivariate real-valued function**, $f : \mathbb{R}^n \to \mathbb{R}$, $\boxed{f(\vec{a}+\vec{h}) \approx f(\vec{a}) + (\nabla f)(\vec{a})\vec{h}}$. The derivative matrix is just the gradient vector $(\nabla f)(\vec{a})$ viewed as an $1 \times n$ matrix (it's a row vector).

4. **coordinate change mapping**, $\vec{T} : \mathbb{R}^{\,3} \to \mathbb{R}^{\,3}$, $\boxed{\vec{T}(\vec{a}+\vec{h}) \approx \vec{T}(\vec{a}) + \vec{T}'(\vec{a})\vec{h}}$. The derivative matrix is a 3×3 matrix. In particular, if we denote $\vec{T} = \langle x, y, z\rangle$ and use u, v, w for cartesian coordinates in the domain of $\vec{T}$

$$\vec{T}\,'(\vec{a}) = [\partial_u \vec{T} | \partial_v \vec{T} | \partial_w \vec{T}] = \left[\begin{array}{ccc} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{array}\right]$$

For two-dimensional coordinate change, $\vec{T} : \mathbb{R}^{\,2} \to \mathbb{R}^{\,2}$ we again write $\boxed{\vec{T}(\vec{a}+\vec{h}) \approx \vec{T}(\vec{a}) + \vec{T}'(\vec{a})\vec{h}}$ but the matrix $\vec{T}'(\vec{a})$ is just a 2×2 matrix

$$\vec{T}'(\vec{a}) = [\partial_u \vec{T} | \partial_v \vec{T}] = \left[\begin{array}{cc} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{array}\right]$$

Example 4.4.2. *Let $f(x) = \sqrt{x}$. The linearization at $x = 4$ is given by $L(x) = 2 + \frac{1}{4}(x-4)$ since $f'(4) = \frac{1}{2\sqrt{4}} = \frac{1}{4}$. We could also express L by $L(4+h) = 2 + h/4$. As an application, note the approximation $\sqrt{5} \approx 2 + 1/4 = 2.25$.*

Example 4.4.3. *Let $\vec{r}(t) = \langle t, t^2, \sin(10t)\rangle$ for $t \in [0,2]$. The linearization of $\vec{r}$ at $t = 1$ is given by $\vec{L}(1+h) = \vec{r}(1) + h\vec{r}\,'(1)$. In particular,*

$$\vec{L}(1+h) = \langle 1+h, 1+2h, \sin(10) + 10h\cos(10)\rangle.$$

Example 4.4.4. *The linearization of $f(x,y) = x/y$ at the point $(6,3)$ is constructed as follows: note $f_x(x,y) = 1/y$ and $f_y(x,y) = -x/y^2$ hence*

$$\begin{aligned} L(x,y) &= f(6,3) + f_x(6,3)(x-6) + f_y(6,3)(y-3) \\ &= 2 + \tfrac{1}{3}(x-6) - \tfrac{2}{3}(y-3). \end{aligned}$$

Example 4.4.5. *The linearization of $f(x,y) = x^2 + y^2$ at the point $(1,2)$ is constructed as follows: note $f_x(x,y) = 2x$ and $f_y(x,y) = 2y$ hence*

$$\begin{aligned} L(x,y) &= f(1,2) + f_x(1,2)(x-1) + f_y(1,2)(y-2) \\ &= 5 + 2(x-1) + 4(y-2). \end{aligned}$$

Example 4.4.6. *The linearization of $f(x,y,z) = x^4 + y^3 + z^2$ at $(1,2,3)$ is given by*

$$\begin{aligned} L(x,y,z) &= f(1,2,3) + f_x(1,2,3)(x-1) + f_y(1,2,3)(y-2) + f_z(1,2,3)(z-3) \\ &= (1 + 2^3 + 3^2) + 4(1)(x-1) + 3(2)^2(y-2) + 2(3)(z-3) \\ &= 18 + 4(x-1) + 12(y-2) + 6(z-3). \end{aligned}$$

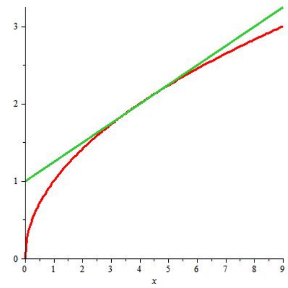
4.4.2 tangent space as graph of linearization

In the section after this I wrestle with why these are good definitions. For now I'll state them without justification.

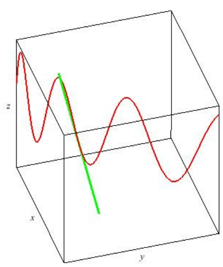
1. $f : \mathbb{R} \to \mathbb{R}$ has tangent line at $(a, f(a))$ with equation $y = f(a) + f'(a)(x-a)$.
2. $\vec{r} : \mathbb{R} \to \mathbb{R}^n$ has tangent line at $\vec{r}(a)$ with natural parametrization $\vec{l}(h) = \vec{r}(a) + \vec{r}\,'(a)h$.
3. $f : \mathbb{R}^2 \to \mathbb{R}$ has tangent plane at $(a, b, f(a,b)$ with equation
   $z = f(a,b) + (\nabla f)(a,b) \bullet \langle x-a, y-b\rangle$.

These are the cases of interest, in case 2 we usually deal with $n = 2$ or $n = 3$ in this course. The following triple of examples mirror those given in the last section. The overall theme is simple: the tangent space to a graph of a function is the graph of the linearization of that function. There are several other viewpoints on the tangent space of a surface and we devote an entire section to that a little later in this chapter. Here I just want you to get what we mean when we say a derivative *gives the best linear approximation to a function.*

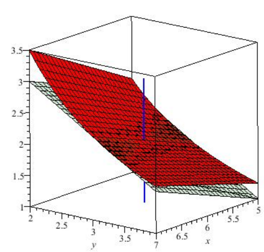
Example 4.4.7. *We continue Example 4.4.2, $f(x) = \sqrt{x}$ and the linearization at $x = 4$ is given by $L(x) = 2 + \frac{1}{4}(x-4)$. The tangent line is the graph $y = L(x)$ which is in green, whereas the $y = f(x)$ is in red.*

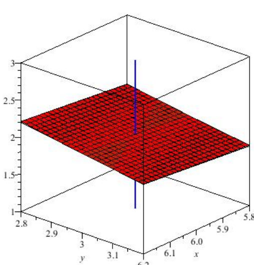


Example 4.4.8. *We continue Example 4.4.3, $\vec{r}(t) = \langle t, t^2, \sin(10t)\rangle$ for $t \in [0,2]$ and the linearization of $\vec{r}$ at $t = 1$ is given by $\vec{L}(1+h) = \langle 1+h, 1+2h, \sin(10) + 10h\cos(10)\rangle$. Once more we plot the curve in red and the tangent line parametrized by $\vec{L}$ in green:*

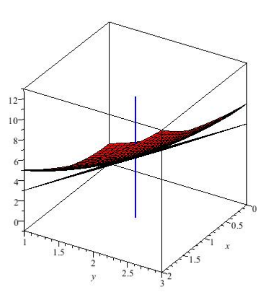


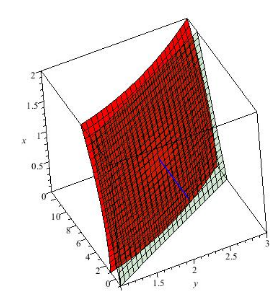
Example 4.4.9. *Continue Example 4.4.4, $f(x,y) = x/y$ and the tangent plane to $z = x/y$ at $(6,3)$ is the solution set of $z = x/3 - 2y/3 + 2$. Below I illustrate the tangent plane, the blue line goes through the point of tangency. See how the surface is locally flat, note the right picture is zoomed further in towards the point of tangency.*





Example 4.4.10. *Continue Example 4.4.5, $f(x,y) = x^2 + y^2$ and the tangent plane to $z = x^2 + y^2$ at $(1,2)$ is the solution set of $z = 5 + (x-1) + 4(y-2)$. Below I illustrate the tangent plane, the blue line goes through the point of tangency. See how the surface is locally flat, these are just two views of the same scale, I put a rotating animation of this on the webpage, take a look.*





There is geometric significance for linearizations such as seen in Example 4.4.6. However, direct visualization would take 4 dimensions. Geometrically, we are mostly interested in $n = 2$ and $n = 3$ for this course[11].

4.4.3 existence and connections to directional differentiation

Existence is usually more troublesome than calculation. But, that is no reason to ignore it. In this subsection I attempt to give you a better sense of what it means for a function to be differentiable at a point. Geometrically we eventually come to the simple realization that a function is differentiable iff it is well-approximated by its linearization. This in turn is tied to the proper definition of the tangent plane. We already gave formulas for important cases in the last subsection, my goal here is to explain why we use those definitions and not something else. Before we get to those more subtle topics, I begin by demonstrating the general derivative recovers single-variable differentiation:

Example 4.4.11. *Suppose $f : dom(f) \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable at x. It follows that there exists a linear function $df_x : \mathbb{R} \to \mathbb{R}$ such that*[12]

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - df_x(h)}{|h|} = 0.$$

Since $df_x : \mathbb{R} \to \mathbb{R}$ is linear there exists a constant matrix m such that $df_x(h) = mh$. In this silly case the matrix m is a 1×1 matrix which otherwise known as a real number. Note that

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - df_x(h)}{|h|} = 0 \quad \Leftrightarrow \quad \lim_{h\to 0^{\pm}} \frac{f(x+h) - f(x) - df_x(h)}{|h|} = 0.$$

In the left limit $h \to 0^-$ we have $h < 0$ hence $|h| = -h$. On the other hand, in the right limit $h \to 0^+$ we have $h > 0$ hence $|h| = h$. Thus, differentiability suggests that $\lim_{h\to 0^{\pm}} \frac{f(x+h)-f(x)-df_x(h)}{\pm h} = 0$. But we can pull the minus out of the left limit to obtain $\lim_{h\to 0^-} \frac{f(x+h)-f(x)-df_x(h)}{h} = 0$. Therefore,

$$\lim_{h\to 0} \frac{f(x+h) - f(x) - df_x(h)}{h} = 0.$$

We seek to show that $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = m$.

$$m = \lim_{h\to 0} \frac{mh}{h} = \lim_{h\to 0} \frac{df_x(h)}{h}$$

A theorem from calculus I states that if $\lim(f - g) = 0$ and $\lim(g)$ exists then so must $\lim(f)$ and $\lim(f) = \lim(g)$. Apply that theorem to the fact we know $\lim_{h\to 0} \frac{df_x(h)}{h}$ exists and

$$\lim_{h\to 0} \left[\frac{f(x+h) - f(x)}{h} - \frac{df_x(h)}{h} \right] = 0.$$

It follows that

$$\lim_{h\to 0} \frac{df_x(h)}{h} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

Consequently,

$$df_x(h) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \quad \text{defined } f'(x) \text{ in calc. I.}$$

[11] if you'd like to talk about higher dimensional geometry in office hours sometime, feel free to stop by

[12] unless we state otherwise, $\mathbb{R}^n$ is assumed to have the euclidean norm, in this case $||x||_{\mathbb{R}} = \sqrt{x^2} = |x|$.

Therefore, $\boxed{df_x(h) = f'(x)h}$. *In other words, if a function is differentiable in the sense we defined at the beginning of this section then it is differentiable in the terminology we used in calculus I. Moreover, the derivative at x is precisely the matrix of the differential. If we use the notation $y = f(x)$ and $h = dx$ then we recover formula for the differential often taught in first semester calculus:*

$$dy_x(dx) = \frac{dy}{dx}(x)dx$$

Or, more compactly, $dy = \frac{dy}{dx}dx$ where dy is the change in y corresponding to the change dx in x. These seemingly heuristic statements take a rigorous meaning in the boxed equation above.

Of course, what really makes the general derivative interesting is its ability to tackle problems such as given below:

Example 4.4.12. *Suppose $F : \mathbb{R}^2 \to \mathbb{R}^3$ is defined by $F(x, y) = (xy,\ x^2,\ x+3y)$ for all $(x, y) \in \mathbb{R}^2$. Consider the difference function $\triangle F$ at (x, y):*

$$\triangle F = F((x, y) + (h, k)) - F(x, y) = F(x + h, y + k) - F(x, y)$$

Calculate,

$$\triangle F = \big((x + h)(y + k),\ (x + h)^2,\ x + h + 3(y + k)\big) - \big(xy,\ x^2,\ x + 3y\big)$$

Simplify by cancelling terms which cancel with $F(x, y)$:

$$\triangle F = \big(xk + hy + hk,\ 2xh + h^2,\ h + 3k)\big)$$

Identify the linear part of $\triangle F$ as a good candidate for the differential. I claim that:

$$L(h, k) = \big(xk + hy,\ 2xh,\ h + 3k\big).$$

is the differential for f at (x,y). Observe first that we can write

$$L(h, k) = \begin{bmatrix} y & x \\ 2x & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}.$$

therefore $L : \mathbb{R}^2 \to \mathbb{R}^3$ is manifestly linear. Use the algebra above to simplify the difference quotient:

$$\lim_{(h,k)\to(0,0)} \left[\frac{\triangle F - L(h, k)}{||(h, k)||} \right] = \lim_{(h,k)\to(0,0)} \left[\frac{(hk, h^2, 0)}{||(h, k)||} \right] = 0.$$

To see why the limit claimed above is true consider the following arguments: since $||(h, k)|| = \sqrt{h^2 + k^2}$ we just need to show

$$(1.)\quad \frac{hk}{\sqrt{h^2 + k^2}} \to 0 \qquad \& \qquad (2.)\quad \frac{h^2}{\sqrt{h^2 + k^2}} \to 0$$

as $(h, k) \to (0, 0)$. Observe that

$$0 \leq \frac{h^2}{\sqrt{h^2 + k^2}} \leq \frac{h^2}{\sqrt{h^2}} = |h| \qquad \& \qquad 0 \leq \frac{hk}{\sqrt{h^2 + k^2}} \leq \frac{|h||k|}{\sqrt{h^2}} = |k|$$

Thus, by the squeeze theorem for multivariate limits, we obtain (2.) and (1.). Therefore,

$$dF_{(x,y)}(h, k) = \begin{bmatrix} y & x \\ 2x & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}.$$

Fortunately we can usually avoid explicit limit calculations due to the nice proposition below.

Example 4.4.13. *Again consider $F(x,y) = (xy,\ x^2,\ x+3y)$. Identify $F_1(x,y) = xy$, $F_2(x,y) = x^2$ and $F_3(x,y) = x+3y$. Calculate,*

$$[F'(x,y)] = \begin{bmatrix} \partial_x F_1 & \partial_y F_1 \\ \partial_x F_2 & \partial_y F_2 \\ \partial_x F_3 & \partial_y F_3 \end{bmatrix} = \begin{bmatrix} y & x \\ 2x & 0 \\ 1 & 3 \end{bmatrix}$$

In single-variable calculus we learn that differentiability implies continuity. However, continuity does not imply differentiability at a given point. The same is true for multivariate functions.

Proposition 4.4.14.

If $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{a} \in U$ then $\vec{F}$ is continuous at $\vec{a}$.

The proof is given in advanced calculus. It's not too difficult. $\square$

The general derivative also reproduces all the directional derivatives we previously discussed.

Proposition 4.4.15.

If $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{a} \in U$ then the directional derivative $D_{\vec{v}}\vec{F}(\vec{a})$ exists for each $\vec{v} \in \mathbb{R}^n$ and $D_{\vec{v}}\vec{F}(\vec{a}) = d\vec{F}_{\vec{a}}(\vec{v})$.

The proof is given in advanced calculus. It's not terribly difficult. $\square$

We should consider the example below. It may challenge some of your misconceptions. It shows that directional differentiation at a point does not give us enough to build the derivative. In fact, the example below has **all** directional derivatives and yet the function is not even continuous.

Example 4.4.16. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = \frac{x^2 y}{x^4+y^2}$ for $(x,y) \neq (0,0)$ and $f(0,0) = 0$. We proved in Example 3.3.4 that this function is not continuous at $(0,0)$. Given the proposition above we also may infer the function is not differentiable at $(0,0)$. You might expect this indicates at least some directional derivative fails to exist. Let's investigate. We turn to the problem of calculating the directional derivative of this function in the unit-vector $\langle a,b\rangle$ direction, suppose $b \neq 0$ to begin,*

$$D_{\langle a,b\rangle} f(0,0) = \frac{d}{dt}\bigg[f(at,bt)\bigg]\bigg|_{t=0} = \frac{d}{dt}\bigg[\frac{a^2bt^3}{a^4t^4+b^2t^2}\bigg]\bigg|_{t=0} = \bigg[\frac{a^2b(a^4t^2+b^2) - a^2bt(2ta^4)}{(a^4t^2+b^2)^2}\bigg]\bigg|_{t=0} = \frac{a^2}{b}.$$

On the other hand, if $b = 0$ then we know $a \neq 0$ since $\langle a,b\rangle$ is a unit-vector[13] *hence $f(at,bt) = \frac{a^2bt^3}{a^4t^4+b^2t^2} = 0$ and it follows $D_{\langle a,0\rangle} f(0,0) = 0$. We find the directional derivatives of f exist in all directions.*

Notice that the directional derivatives do jump from one value to another as we travel around the unit-circle. In particular, as we traverse the arc of the circle through the point $\langle 1,0\rangle$ we have $\langle a,b\rangle$ go from vectors with $b > 0$ which have $\frac{a^2}{b} \to \infty$ to vectors with $b < 0$ which have $\frac{a^2}{b} \to -\infty$.

[13] if $a = 0$ and $b = 0$ then $||\langle a,b\rangle|| = 0 \neq 1$

In the middle, we hit $\langle 1, 0\rangle$ where $D_{\langle a,0\rangle} f(0, 0) = 0$. These directional derivatives may exist but they certainly do not continuously paste together. It turns out that continuity of the directional derivatives in the coordinate directions is a sufficient condition to eliminate the trouble of the previous example.

Definition 4.4.17.

A mapping $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **continuously differentiable** at $a \in U$ iff all the partial derivative mappings $\partial F_i/\partial x_j$ exist on an open set containing a and are continuous at a.

Continuous differentiability is typically easier than differentiability to check. The reason is that partial derivatives are straightforward to calculate. On the other hand, it is sometimes challenging to find the linearization and actually check the appropriate limit vanishes. It follows that the proposition below is welcome news:

Proposition 4.4.18.

If F is continuously differentiable at a then F is differentiable at a

The proof is somewhat involved. The main construction involves breaking a vector into a sum of vector components. Then continuity of the partial derivatives paired with a mean value theorem argument goes to prove the differentiability of the mapping. Again, details are given in my advanced calculus notes (or any good text on the subject). $\square$

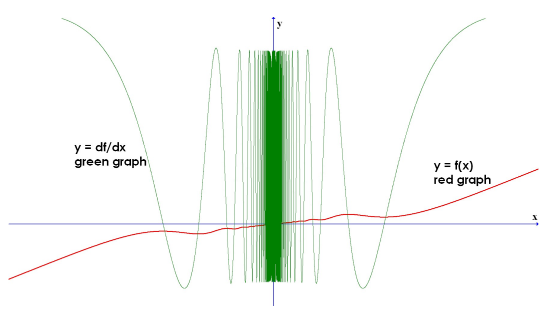
There do exist functions which are differentiable at a point and yet fail to be continuously differentiable at that point. In single variable calculus I usually present the example: Let $f(0) = 0$ and

$$f(x) = \frac{x}{2} + x^2 \sin\frac{1}{x}$$

for all $x \neq 0$. I can be shown that the derivative $f'(0) = 1/2$. Moreover, we can show that $f'(x)$ exists for all $x \neq 0$, we can calculate:

$$f'(x) = \frac{1}{2} + 2x \sin\frac{1}{x} - \cos\frac{1}{x}$$

Notice that $dom(f') = \mathbb{R}$. Note then that the tangent line at $(0, 0)$ is $y = x/2$.



The lack of continuity for the derivative means that the tangent line at the origin does not well-approximate the graph near the point of tangency. In other words, the linearization is not a good approximation near the point of tangency. This is not just a single-variable phenonmenon. Pathological multivariate examples exist. For example,

Example 4.4.19. *Let* $f(0,y)=0$ *and*

$$f(x,y)=x^2\sin\frac{1}{x}$$

for all $(x,y)\in\mathbb{R}^2$ *such that* $x\neq 0$. *You can show that* $D_{\widehat{u}}f(0,0)=0$ *for all unit vectors u. This means that the tangent vectors to any path* $t\to(at,bt,f(at,bt))$ *reside in the xy-plane. It appears the set of all tangent vectors fill out the xy-plane. However, I'm not sure what happens with non-linear paths in the domain. I suspect the curves on the graph* $z=f(x,y)$ *built from composing a smooth, but non-linear, path* $\gamma:\mathbb{R}\to\mathbb{R}^2$ *with* f *might result in a path* $f\circ\gamma$ *which is not even differentiable at the origin.*

Let's investigate the differentiability of f *at* $(0,0)$. *Given the triviality of all the directional derivatives we suspect* $L(h,k)=0$. *Consider,*

$$\frac{|f(h,k)-f(0,0)-L(h,k)|}{||(h,k)||}=\frac{|h^2\sin(1/h)|}{\sqrt{h^2+k^2}}=\frac{|h\sin(1/h)|}{\sqrt{1+k^2/h^2}}\leq|h\sin(1/h)|\leq|h|.$$

It follows that f *is differentiable at* $(0,0)$ *since we have* $|h|\to 0$ *as* $(h,k)\to(0,0)$ *along any path. Therefore, my suspicion was incorrect. Even nonlinear paths composed with* f *yield a differentiable path. However, this does give us another example of a function which is differentiable at* $(0,0)$ *but is not continuously differentiable. If you're wondering it is clear that* f_x *is not continuous along the entire y-axis. Given our experience in the single variable case we suspect the linearization does not approximate the function in a natural way as we leave the point of tangency. We need the continuity of the partial derivatives to insure the function does not wildly misbehave in the locality of the tangent point.*

I haven't proved it yet but I suspect the function below is not differentiable. It gives and example of a function which is continuous but is not differentiable at zero. However, both partial derivatives exist at $(0,0)$, they're just not continuous.

Example 4.4.20. *Let us define* $f(0,0)=0$ *and*

$$f(x,y)=\frac{x^2y}{x^2+y^2}$$

for all $(x,y)\neq(0,0)$ *in* $\mathbb{R}^2$. *It can be shown*[14] *that* f *is continuous at* $(0,0)$. *Moreover, since* $f(x,0)=f(0,y)=0$ *for all* x *and all* y *it follows that* f *vanishes identically along the coordinate axis. Thus the rate of change in the* $\widehat{x}$ *or* $\widehat{y}$ *directions is zero. We can calculate that*

$$\frac{\partial f}{\partial x}=\frac{2xy^3}{(x^2+y^2)^2}\qquad \textit{and}\qquad \frac{\partial f}{\partial y}=\frac{x^4-x^2y^2}{(x^2+y^2)^2}$$

Consider the path to the origin $t\mapsto(t,t)$ *gives* $f_x(t,t)=2t^4/(t^2+t^2)^2=1/2$ *hence* $f_x(x,y)\to 1/2$ *along the path* $t\mapsto(t,t)$, *but* $f_x(0,0)=0$ *hence the partial derivative* f_x *is not continuous at* $(0,0)$. *Therefore, this function has discontinuous partial derivatives. It is not continuously differentiable.*

[14] you did this one in homework... or at least you were supposed to...

Let's return to the question of directional derivatives and differentiability. It is tempting to think that the reason the function in Example 4.4.16 failed to be differentiable is that the tangent vectors to the curves $t \mapsto (at, bt, f(at, bt))$ failed to fill out a plane. This suspicion is further encouraged by Example 4.4.19 where we see the function is differentiable and the tangent vectors to the curves $t \mapsto (at, bt, f(at, bt))$ do fill out the xy-plane. However, this suspicion is false. Think back to our experience with multivariate limits in Example 3.3.2. Differentiability also concerns a multivariate limit so intuitively we may expect something could be hidden if we only think about straight-line approaches to the limit point.

Why all this fuss? Let me try to clarify the confusion which pushed me to this discussion:

1. some authors define the tangent plane to be the union of all tangent vectors at a point.

2. other authors say the tangent plane is a plane which well-approximates the graph of the function near the point of tangency.

Item (2.) begs some questions, what exactly do we mean by "well-approximates". Is the nearness to the graph the concept captured by mere differentiability or is it the stronger version captured by continuous differentiability? Item (1.) is dangerous since it would *seem* that looking at all possible directional derivatives should give a complete picture of the tangent vectors at a point. We just argued this is not the case[15]. It is possible for all tangents to curves built from linear paths to exist whereas the tangent vectors to a path built from a nonlinear path may not even exist. If we are to use item (1.) as a definition we must clarify it a bit:

> **The tangent plane to the graph $z = f(x, y)$ is formed by the union of all possible tangent vectors of curves $f \circ \vec{\gamma}$ where $\vec{\gamma}$ is a smooth curve in $dom(f)$ which pass at $t = 0$ through the xy-coordinates of the point of tangency.Moreover, if these vectors do not form a two-dimensional plane then the tangent plane fails to exist.**

This is just my comment here, I haven't seen this elsewhere. Most authors don't bother with these details or deliberations. In fact, many authors assume continuous differentiability in their definitions. In any event, it seems clear to me that we should prefer a slightly more careful version of (2.) since it has far less technical trouble. With all of this in mind we define (I expand on the most important case to this course after this general definition),

Definition 4.4.21. *general tangent space to a graph.*

Suppose that U is open and $\vec{F} : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a mapping which is differentiable at $\vec{a} \in U$ then the linear mapping $\vec{L} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\vec{h} \to 0} \frac{\vec{F}(\vec{a} + \vec{h}) - \vec{F}(\vec{a}) - \vec{L}(\vec{h})}{||\vec{h}||} = 0.$$

defines the **tangent space** at $(\vec{a}, \vec{F}(\vec{a}))$ to $graph(\vec{F}) = \{(\vec{x}, \vec{F}(\vec{x})) \mid dom(\vec{F})\}$ with equations $\vec{z} = \vec{F}(\vec{a}) + \vec{L}(\vec{x} - \vec{a})$ in $\mathbb{R}^n \times \mathbb{R}^m$. We use the notation $\vec{z} \in \mathbb{R}^m$ whereas $\vec{x}, \vec{a} \in \mathbb{R}^n$ in the equation above.

[15] I have an example if you ask

In particular, for $f : \mathbb{R}^2 \to \mathbb{R}$ we have $L(x-x_o, y-y_o) = (\nabla f)(x_o, y_o) \bullet \langle x-x_o, y-y_o\rangle$ and the tangent plane has equation:

$$\boxed{z = f(x_o, y_o) + (x-x_o)f_x(x_o, y_o) + (y-y_o)f_y(x_o, y_o).}$$

The assumption of differentiability of f at (x_o, y_o) insures that the tangent plane $z = f(x_o, y_o) + L(x, y) \approx f(x, y)$ for points near (x_o, y_o). In other words, the graph $z = f(x, y)$ looks like a plane if we zoom in close to the point $(x_o, y_o, f(x_o, y_o))$. In fact, many authors simply define differentiability in view of this concept:

A function is differentiable at $\vec{p}$ iff it has a tangent plane at $\vec{p}$.

This is less than satisfactory if the text you're reading nowhere defines the tangent plane. I won't name names. The boxed statement is true, but it is not a definition. Not here at least.

In the case a function is differentiable but not continuously differentiable we have the situation that there is a tangent plane, but it fails to well-approximate the graph near the point of tangency.

Continuous differentiability is needed for many of the calculations we perform in the remainder of this course. I conclude this section with an example of how it may happen that $f_{xy} \neq f_{yx}$ at a point which is merely differentiable. On the other hand, Clairaut's Theorem states that $f_{xy} = f_{yx}$ for continuously differentiable functions.

Example 4.4.22. *This example studies a curious function for which $f_{xy}(0,0) \neq f_{yx}(0,0)$. We define, $f : \mathbb{R}^2 \to \mathbb{R}$ as follows: for $(x, y) \neq (0,0)$*

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

and $f(0,0) = 0$. To calculate the partial derivatives at $(0,0)$ we need to examine explicit limits:

$$f_x(0,0) = \lim_{h\to 0}\left(\frac{f(h,0) - f(0,0)}{h}\right) = \lim_{h\to 0}\left(\frac{0}{h}\right) = 0$$
$$f_y(0,0) = \lim_{h\to 0}\left(\frac{f(0,h) - f(0,0)}{h}\right) = \lim_{h\to 0}\left(\frac{0}{h}\right) = 0$$

Next, to calculate $f_{xy}(0,0$ and $f_{yx}(0,0)$ we need f_x and f_y along the coordinate axes. Thus, calculate, for $(x, y) \neq (0,0)$ the quotient rule and a bit of algebra shows:

$$f_x(x, y) = \frac{x^4y + 4x^2y^3 - y^5}{(x^2 + y^2)^2} \qquad \& \qquad f_y(x, y) = \frac{x^5 - 4y^2x^3 - y^4x}{(x^2 + y^2)^2}$$

Consider,

$$f_{xy}(0,0) = \frac{\partial f_x}{\partial y}(0,0) = \lim_{h\to 0}\left(\frac{f_x(0,h) - f_x(0,0)}{h}\right) = \lim_{h\to 0}\left(\frac{-h^5/h^4}{h}\right) = \lim_{h\to 0}(-1) = -1.$$
$$f_{yx}(0,0) = \frac{\partial f_y}{\partial x}(0,0) = \lim_{h\to 0}\left(\frac{f_x(h,0) - f_x(0,0)}{h}\right) = \lim_{h\to 0}\left(\frac{h^5/h^4}{h}\right) = \lim_{h\to 0}(1) = 1.$$

Therefore, $f_{xy}(0,0) \neq f_{yx}(0,0)$.

Fortunately, it usually takes considerable effort to achieve the result of the Example above. However, such examples exist and we would like to have a criteria with which we can reliably assume $f_{xy} = f_{yx}$. The theorem that follows give us some guidance:

Theorem 4.4.23. *Clairaut's Theorem:*

If $f : dom(f) \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ is a function where $dom(f)$ contains an open disk D centered at (a, b) and the function f_{xy} and f_{yx} are both continuous on D then

$$\boxed{f_{xy}(a, b) = f_{yx}(a, b).}$$

Notice that the existence of f_{xy} and f_{yx} does not suffice, we need **continuity** of the derivatives. Once again, continuous differentiability makes calculation with partial derivatives natural. The proof is found in most advanced calculus texts. Finally, I should mention that the concerns and examples of this section readily generalize to functions from $\mathbb{R}^m$ to $\mathbb{R}^n$.

4.4.4 properties of the derivative

Suppose $\vec{F}_1 : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $\vec{F}_2 : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ are differentiable at $\vec{a} \in U$ then $\vec{F}_1 + \vec{F}_2$ is differentiable at $\vec{a}$ and $d(\vec{F}_1 + \vec{F}_2)_a = (d\vec{F}_1)_a + (d\vec{F}_2)_a$ which means for the Jacobian matrices we also have $(\vec{F}_1 + \vec{F}_2)'(\vec{a}) = \vec{F}_1'(\vec{a}) + \vec{F}_2'(\vec{a})$. Likewise, if $c \in \mathbb{R}$ then $d(c\vec{F}_1)_a = c(d\vec{F}_1)_a$ hence for the Jacobian matrices we have $(c\vec{F}_1)'(\vec{a}) = c(\vec{F}_1'(\vec{a})$. Nothing terribly surprising here. What is much more fascinating is the following general version of the chain rule:

Proposition 4.4.24.

If $\vec{F} : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^p$ is differentiable at $\vec{a}$ and $\vec{G} : V \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^m$ is differentiable at $\vec{F}(\vec{a}) \in V$ then $\vec{G} \circ \vec{F}$ is differentiable at $\vec{a}$ and $d(\vec{G} \circ \vec{F})_{\vec{a}} = (d\vec{G})_{\vec{F}(\vec{a})} \circ d\vec{F}_{\vec{a}}$. Moreover, in Jacobian matrix notation,

$$(\vec{G} \circ \vec{F})'(\vec{a}) = \vec{G}\,'(\vec{F}(\vec{a}))\vec{F}\,'(\vec{a}).$$

In words, the Jacobian matrix of the composite of $\vec{G}$ with $\vec{F}$ is simply the matrix product of the Jacobian matrices of $\vec{G}$ with the Jacobian matrix of $\vec{F}$. Unfortunately, not all students really learned matrix algebra in highschool so this statement lacks the power it should have in your mind. This proposition builds the foundation for the multivariate version of u-substution. All the chain rules in the next section are derivable from this general proposition. For this reason I offer no proofs in the next section. The calculations in the next section all follow from the calculation below[16]:

Proof: $\approx$ Suppose $\vec{F} : dom(\vec{F}) \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^p$ and $\vec{G} : dom(\vec{G}) \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^m$. Let $\vec{x}_o \in \mathbb{R}^n$ for which $\vec{F}(\vec{x}_o) = \vec{y}_o \in dom(\vec{G})$ and suppose that $\vec{F}$ is differentiable at $\vec{x}_o$ and $\vec{G}$ is differentiable at $\vec{y}_o$. We seek to show that $\vec{G} \circ \vec{F}$ is differentiable at $\vec{x}_o$ with Jacobian matrix $\vec{G}\,'(\vec{y}_o)\vec{F}\,'(\vec{x}_o)$. Observe that the existence of $\vec{G}\,'(\vec{y}_o) \in \mathbb{R}^{\,m \times p}$ and $\vec{F}\,'(\vec{x}_o) \in \mathbb{R}^{\,p \times n}$ follow from the differentiability of $\vec{G}$ at $\vec{y}_o$ and $\vec{F}$ at $\vec{x}_o$. In particular, if $||\vec{k}|| \approx 0$ then

$$\vec{G}(\vec{y}_o + \vec{k}) \approx \vec{G}(\vec{y}_o) + \vec{G}\,'(\vec{y}_o)\vec{k}.$$

[16] this is a plausibility argument, not a formal proof, all the $\approx$ symbols are shorthands for a more detailed estimation which is not given in these notes, however, you guessed it, can be found in a good advanced calculus text.

Likewise, if $||\vec{h}|| \approx 0$ then
$$\vec{F}(\vec{x}_o + \vec{h}) \approx \vec{F}(\vec{x}_o) + \vec{F}'(\vec{x}_o)\vec{h}.$$
Suppose $\vec{h}$ is given such that $||\vec{h}|| \approx 0$. It follows that $\vec{F}'(\vec{x}_o)\vec{h} \approx 0$. Let $\vec{k} = \vec{F}'(\vec{x}_o)\vec{h}$ and note that
$$\underbrace{\vec{G}(\vec{F}(\vec{x}_o + \vec{h})) \approx \vec{G}(\vec{F}(\vec{x}_o) + \vec{F}'(\vec{x}_o)\vec{h})}_{continuity\ of\ G\ at\ y_o} = \vec{G}(\vec{y}_o + \vec{k}) \approx \vec{G}(\vec{y}_o) + \vec{G}'(\vec{y}_o)\vec{k}$$
Therefore, for $||\vec{h}|| \approx 0$,
$$\vec{G}(\vec{F}(\vec{x}_o + \vec{h})) \approx \vec{G}(\vec{F}(\vec{x}_o)) + \vec{G}'(\vec{F}(\vec{x}_o))\vec{F}'(\vec{x}_o)\vec{h}.$$
Thus $\vec{G}(\vec{F}(\vec{x}_o + \vec{h})) - \vec{G}(\vec{F}(\vec{x}_o)) - \vec{G}'(\vec{F}(\vec{x}_o))\vec{F}'(\vec{x}_o)\vec{h} \approx 0$. In fact, if we worked out the careful details we could show that
$$\lim_{\vec{h}\to 0} \frac{\vec{G}(\vec{F}(\vec{x}_o + \vec{h})) - \vec{G}(\vec{F}(\vec{x}_o)) - \vec{G}'(\vec{F}(\vec{x}_o))\vec{F}'(\vec{x}_o)\vec{h}}{||\vec{h}||} = 0$$
and it follows that $(\vec{G} \circ \vec{F})'(\vec{x}_o) = \vec{G}'(\vec{F}(\vec{x}_o))\vec{F}'(\vec{x}_o)$. Technically, this is not a proof, but perhaps it makes the rule a bit more plausible. The chain rule is primarily a consequence of matrix multiplication when we look at it the right way. $\square$.

Example 4.4.25. *Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable. Define polar coordinate change map $X : \mathbb{R}^2 \to \mathbb{R}^2$ by $X(r,\theta) = (r\cos\theta, r\sin\theta)$. Let $g = f \circ X$ then the general chain rule says $dg = df \circ dX$. In particular, this means $[dg] = [df](X)[dX]$ where $[df](X)$ denotes the Jacobian matrix of df evaluated at the value of X. In more detail:*
$$[df] = \left[\frac{\partial f}{\partial x}\Big|\frac{\partial f}{\partial y}\right], \qquad [dX] = \left[\frac{\partial X}{\partial r}\Big|\frac{\partial X}{\partial \theta}\right] = \left[\begin{array}{cc} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{array}\right], \quad [dg] = \left[\frac{\partial g}{\partial r}\Big|\frac{\partial g}{\partial \theta}\right]$$
Thus, $g(r,\theta) = (f \circ X)(r,\theta)$ has chain rules given by the matrix equation below:
$$\left[\frac{\partial g}{\partial r}\Big|\frac{\partial g}{\partial \theta}\right] = \left[\frac{\partial f}{\partial x}\Big|\frac{\partial f}{\partial y}\right]\left[\begin{array}{cc} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{array}\right]$$
Let us abbreviate partial notation and calculate the matrix product:
$$[\partial_r g, \partial_\theta g] = [\partial_x f, \partial_y f]\left[\begin{array}{cc} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{array}\right] = [\cos\theta\partial_x f + \sin\theta\partial_y f,\ -r\sin\theta\partial_x f + r\cos\theta\partial_y f]$$
Thus, reverting once more to the full partial notation,
$$\frac{\partial g}{\partial r} = \cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y}, \qquad \frac{\partial g}{\partial \theta} = -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}$$
Step away from the details of the polar coordinate map to better appreciate the pattern above. If we denote $X = (x,y)$ then the expression above can be written as:
$$\frac{\partial g}{\partial r} = \frac{\partial x}{\partial r}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial f}{\partial y}, \qquad \frac{\partial g}{\partial \theta} = \frac{\partial x}{\partial \theta}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial \theta}\frac{\partial f}{\partial y}$$
Perhaps you can see how the pattern above is easy to remember. The point of this example is simply to show you that these expressions are not disjoint thoughts, rather, they taken together form the Jacobian matrix of g from the Jacobians of f and X. Calculationally, we use the pattern (or diagrams if you like) as a guide since the matrix multiplication is overkill given the simplicity of these patterns. In the next section, we study how the pattern seen here is replicated and generalized to other contexts.

4.5 chain rules

In this section we explain how the chain rule generalizes to functions of several variables.

4.5.1 one independent variable

You might have noticed that we already learned one new chain-rule for space curves:

$$\boxed{\frac{d}{dt}\big[\ \vec{r}(u(t))\ \big] = \frac{d\vec{r}}{du}\frac{du}{dt}.}$$

For example, if $\vec{r}(t) = \langle t, t^2, t^3\rangle$ and $u(t) = \sin t$ then

$$\frac{d}{dt}\big[\ \vec{r}(u(t))\ \big] = \langle 1, 2u, 3u^2\rangle \cos t = \langle \cos t, 2\sin t\cos t, 3\sin^2 t\cos t\rangle.$$

This chain rule was important to understand how the Frenet Serret formulas are reformulated for non-unit-speed curves. It was the source of the speed factors ds/dt in those equations.

Next consider the composite of $f : \mathbb{R}^2 \to \mathbb{R}$ and $\vec{r} : \mathbb{R} \to \mathbb{R}^2$ where $\vec{r} = \langle x, y\rangle$. Here's the rule:

$$\boxed{\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \nabla f \bullet \frac{d\vec{r}}{dt}}$$

In this case the **independent variable** is t and the **intermediate variables** are x, y. All of the expressions above are understood to be functions of t. A more pedantic statement of the same rule is as follows:

$$\boxed{\frac{d}{dt}f(\vec{r}(t)) = \frac{\partial f}{\partial x}(\vec{r}(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(\vec{r}(t))\frac{dy}{dt}.}$$

Example 4.5.1. *Suppose $f(x, y) = x^2 - xy$ and $x = e^t, y = t^2$ then*

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = (2x - y)e^t - x(2t) = (2e^t - t^2)e^t - 2te^t.$$

Now, some of you will doubtless note that we could just as well subsitute $x = e^t$ and $y = t^2$ at the outset and just do ordinary differentiation on $g(t) = (e^t)^2 - t^2e^t$. Will you obtain the same answer? Yes. Is that the right method to count on generally? No.

Example 4.5.2. *Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable and $t \mapsto \vec{r}(t) = \langle x(t), y(t)\rangle$ defines a smooth path. What is the geometric relation between the tangent vector to the path and the gradient vector field of f? Use the chain rule,*

$$\frac{d}{dt}\big[f(\vec{r}(t))\big] = \frac{\partial f}{\partial x}(\vec{r}(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(\vec{r}(t))\frac{dy}{dt} = \nabla f(\vec{r}(t)) \bullet \left\langle \frac{dx}{dt}, \frac{dy}{dt}\right\rangle = \nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt}.$$

This doesn't really tell us much of anything for an arbitrary function and path. However, if we suppose the path parametrizes a level curve $f(x, y) = k$ then we find something nice. To say $\vec{r}$ parametrizes $f(x, y) = k$ is to insist $f(\vec{r}(t)) = k$ for all t. Differentiate this equation and we again use the chain rule on the l.h.s. whereas $\frac{d}{dt}(k) = 0$. Thus,

$$\nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt} = 0.$$

We find the gradient vector field is normal to the tangent vector field of the level curve. The chain rule has given us the calculational tool to verify what we argued geometrically earlier in this chapter.

Notice we can't just substitute in the formulas for $x(t)$ and $y(t)$ in the example above. Why? Because we are not given them. The chain rule allows us to discover general relationships which may not be obvious if we always just work at the level of the independent variable.

Example 4.5.3. *Let $z = x\ln(x+2y)$ and $x = \sin t$ and $y = \cos t$ then*

$$\begin{aligned}
\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \\
&= \left[\ln(x+2y) + \frac{x}{x+2y}\right]\cos t + \left[\frac{2x}{x+2y}\right](-\sin t) \qquad (\star) \\
&= \boxed{\left[\ln(\sin t + 2\cos t) + \frac{\sin t}{\sin t + 2\cos t}\right]\cos t - \frac{2\sin^2 t}{\sin t + 2\cos t}}
\end{aligned}$$

Note the boxed answer has all the x, y expressions explicitly written in terms of t. In contrast, if we were content with an answer which was written in terms of t as well as x, y which implicitly contain t-dependence then $\star$ would have consitituted a valid answer.

As a point of brevity on some tests, some instructors allow a $\star$-style answer. The complete thought is that $\star$ is paired with $x = \sin t$ and $y = \cos t$. Since the term $\star$-style is a bit adhoc, let us instead agree that the process of leaving an answer in the form such as $\star$ is an answer which is written **implicitly**. If I say to give the answer **explicitly** in terms of t then $\star$ is not a correct answer.

Example 4.5.4. *Let $z = xy$ and suppose $x = e^t$ and $y = \sin t$.*

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = ye^t + x\cos t = e^t(\sin t + \cos t).$$

In the example above t is the *independent* variable, z is the *dependent* variable and x, y are *intermediate* variables. Notice, the context matters. It is completely possible that x is independent, dependent or intermediate for a given problem. You must judge by context.

The chain rule below is a natural generalization of what we just discussed: if $\vec{r} = \vec{r}(t)$ where $\vec{r} = \langle x, y, z\rangle$ and[17] $f = f(x, y, z)$ then

$$\boxed{\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = \nabla f \bullet \frac{d\vec{r}}{dt}.}$$

In this case the **independent variable** is t and the **intermediate variables** are x, y, z. All of the expressions above are understood to be functions of t. A more pedantic statement of the same rule is as follows:

$$\frac{d}{dt}f(\vec{r}(t)) = \frac{\partial f}{\partial x}(\vec{r}(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(\vec{r}(t))\frac{dy}{dt} + \frac{\partial f}{\partial z}(\vec{r}(t))\frac{dz}{dt}.$$

[17] this notation means $f : \mathbb{R}^3 \to \mathbb{R}$ and $\vec{r} : \mathbb{R} \to \mathbb{R}^3$, it is a bit sloppy, but it is popular and I suppose I should expose you to it.

Example 4.5.5. *Suppose $\vec{r}(t) = \langle \cos t \sin t, \sin t \sin t, \cos t\rangle$ and $f(x,y,z) = x^2+y^2+z^2$. This means $x = \cos t \sin t$, $y = \sin^2 t$ and $z = \cos t$. Calculate,*

$$\begin{aligned} \frac{d}{dt} f(\vec{r}(t)) &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} \\ &= 2x(\cos^2 t - \sin^2 t) + 2y(2\sin t \cos t) + 2z(-\sin t) \\ &= 2\cos t \sin t(\cos^2 t - \sin^2 t) + 2\sin^2 t(2\sin t\cos t) + 2\cos t(-\sin t) \\ &= 2\cos t \sin t(1 - 2\sin^2 t) + 2\sin^2 t(2\sin t\cos t) + 2\cos t(-\sin t) \\ &= 0. \end{aligned}$$

Why is this? Simple. The path given by $x = \cos t \sin t$, $y = \sin^2 t$ and $z = \cos t$ parametrizes a curve which lies on the sphere $x^2+y^2+z^2 = 1$. It follows that $f(\cos t \sin t, \sin^2 t, \cos t) = 1$ hence differentiation by t yields zero. Geometrically we find the gradient vector field $\nabla f = \langle 2x, 2y, 2z\rangle$ is normal to the tangent vector field of the curve wherever they intersect.

Of course there are many other curves which reside in the level surface $f(x,y,z) = 1$. I just picked one to illustrate that the gradient vectors are normal to the curves on the surface. We can argue this in general.

Example 4.5.6. *Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is differentiable and $t \mapsto \vec{r}(t) = \langle x(t), y(t), z(t)\rangle$ defines a smooth path. Use the chain rule, omitting explicit point dependence on the partials,*

$$\frac{d}{dt}\big[f(\vec{r}(t))\big] = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = \nabla f(\vec{r}(t)) \bullet \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right\rangle = \nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt}.$$

If we suppose the path $t \mapsto \vec{r}(t)$ parametrizes a curve which is on the level surface $f(x,y,z) = k$ then $f(\vec{r}(t)) = k$ for all t. Differentiate this equation and we again use the chain rule on the l.h.s. whereas $\frac{d}{dt}(k) = 0$. Thus,

$$\nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt} = 0.$$

We find the gradient vector field is normal to the tangent vector field of an arbitrary curve on the surface. If the function f is continuously differentiable then it follows that the union of all such tangent vectors forms the tangent space to the level surface. The gradient vector at the point of tangency gives the normal to the tangent plane.

For example, a sphere of radius R centered at (a,b,c) has equation $(x-a)^2+(y-b)^2+(z-c)^2 = R^2$. This sphere is naturally viewed as a level surface of $F(x,y,z) = (x-a)^2+(y-b)^2+(z-c)^2$. We calculate,

$$\nabla F(x,y,z) = \langle 2(x-a), 2(y-b), 2(z-c)\rangle$$

The equation of the tangent plane at (x_o, y_o, z_o) on this sphere is

$$2(x_o - a)(x - x_o) + 2(y_o - b)(y - y_o) + 2(z_o - c)(z - z_o) = 0.$$

In particular, if $a = b = c = 0$ then we have a tangent plane

$$2x_o(x - x_o) + 2y_o(y - y_o) + 2z_o(z - z_o) = 0.$$

For this case the vector pointing to (x_o, y_o, z_o) and the normal vector $\langle 2x_o, 2y_o, 2z_o\rangle$ point along the same line.

Example 4.5.7. *Let $w = xyz$ and suppose $x = t, y = t^2$ and $z = t^3$.*

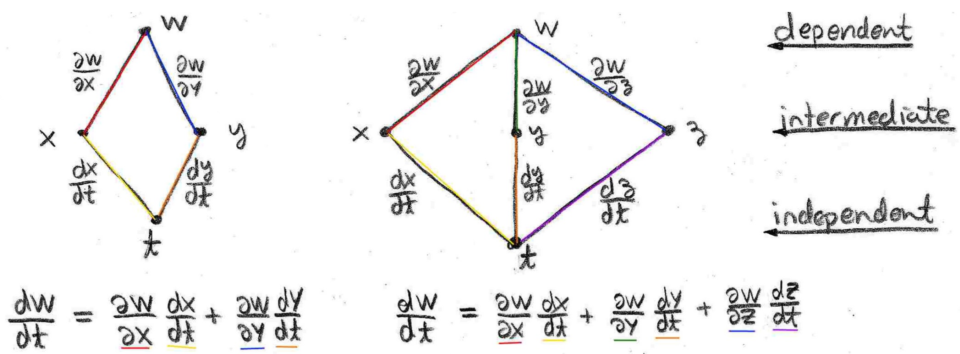
$$\begin{aligned}\frac{dw}{dt} &= \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt} \\ &= yz(1) + xz(2t) + xy(3t^2) \\ &= t^5 + 2t^5 + 3t^5 \\ &= 6t^5.\end{aligned}$$

Notice, if we plug-in the expressions $x = t, y = t^2$ and $z = t^3$ then we obtain $w = t^6$ hence $\frac{dw}{dt} = 6t^5$ is also obtained by direct computation with no use of the chain-rule.

Moving on to our next case, if $\vec{r} = \langle x_1, x_2, \dots, x_n\rangle$ and[18] $f = f(x_1, x_2, \dots, x_n)$ then

$$\boxed{\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt} = \nabla f \bullet \frac{d\vec{r}}{dt}.}$$

In this case the **independent variable** is t and the **intermediate variables** are $x_1, x_2, \dots, x_n$. All of the expressions above are understood to be functions of t. I'm not a big fan, but, another trick to remember the chain-rules above is given by the mneumonic device below:



The way the picture works: multiply down and add together all possible products.

Example 4.5.8. *Suppose $f(\vec{x}) = \vec{x} \bullet \vec{x}$ where $\vec{x} \in \mathbb{R}^n$. Moreover, suppose $\vec{r} : \mathbb{R} \to \mathbb{R}^n$ is a path which parametrizes the level set $f(\vec{x}) = R^2$ (this is a higher-dimensional sphere). We have $f(\vec{r}(t)) = R^2$ for all t. Differentiate to find*

$$\nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt} = 0.$$

Once more we find the tangents to curves on the level set are orthogonal to the gradient vector field. Don't ask me to draw the picture here. The tangent space is an $(n-1)$-dimensional hyperplane embedded in $\mathbb{R}^n$, the normal vector field ∇f always points in the one remaining dimension if there are no critical points for f.

Another case[19] is $\vec{F} = \vec{F}(x_1, x_2, \dots, x_n)$ composed with a path. In particular, if $\vec{F} = \langle F_1, F_2, \dots, F_m\rangle : \mathbb{R}^n \to \mathbb{R}^m$ is composed with $\vec{r} = \langle x_1, x_2, \dots, x_n\rangle : \mathbb{R} \to \mathbb{R}^n$ then we have the chain rule

$$\boxed{\frac{d\vec{F}}{dt} = \frac{\partial \vec{F}}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial \vec{F}}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial \vec{F}}{\partial x_n}\frac{dx_n}{dt} = \left\langle \nabla F_1 \bullet \frac{d\vec{r}}{dt}, \nabla F_2 \bullet \frac{d\vec{r}}{dt}, \dots, \nabla F_m \bullet \frac{d\vec{r}}{dt} \right\rangle.}$$

[18] this notation means that $f : \mathbb{R}^n \to \mathbb{R}$.

[19] geometrically, this derivative studies the change in a vector field along a path

Another nice way to think of this rule is as follows:

$$\begin{aligned}\frac{d}{dt}\langle F_1, F_2, \dots, F_m\rangle &= \left\langle \frac{dF_1}{dt}, \frac{dF_2}{dt}, \dots, \frac{dF_m}{dt}\right\rangle \\ &= \left\langle \frac{\partial F_1}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial F_1}{\partial x_n}\frac{dx_n}{dt},\ \dots,\ \frac{\partial F_m}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial F_m}{\partial x_n}\frac{dx_n}{dt}\right\rangle\end{aligned}$$

Example 4.5.9. *Suppose $\vec{F}(x, y, z) = \langle xy,\ y + z^2\rangle$ and suppose $x = t, y = t^2, z = t^3$. Calculate,*

$$\begin{aligned}\frac{d\vec{F}}{dt} &= \left\langle \frac{\partial F_1}{\partial x}\frac{dx}{dt} + \frac{\partial F_1}{\partial y}\frac{dy}{dt} + \frac{\partial F_1}{\partial z}\frac{dz}{dt},\ \frac{\partial F_2}{\partial x}\frac{dx}{dt} + \frac{\partial F_2}{\partial y}\frac{dy}{dt} + \frac{\partial F_2}{\partial z}\frac{dz}{dt}\right\rangle \\ &= \left\langle y\frac{dx}{dt} + x\frac{dy}{dt} + 0\frac{dz}{dt},\ 0\frac{dx}{dt} + \frac{dy}{dt} + 2z\frac{dz}{dt}\right\rangle \\ &= \big\langle y + x(2t),\ 2t + 2z(3t^2)\big\rangle \\ &= \big\langle 3t^2,\ 2t + 6t^5\big\rangle\end{aligned}$$

All of the examples up to this point have considered chain rules for functions of just one independent variable which we have denoted by t for the sake of conceptual uniformity. We now consider differentiation of composite functions of two or more independent variables.

4.5.2 two independent variables

Suppose $f = f(x, y)$ and $x = x(u, v)$ and $y = y(u, v)$. This means $f : \mathbb{R}^2 \to \mathbb{R}$ and $x : \mathbb{R}^2 \to \mathbb{R}$ and $y : \mathbb{R}^2 \to \mathbb{R}$. We have two interesting partial derivatives to compute:

$$\boxed{\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} \qquad \& \qquad \frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}.}$$

In this case the **independent variables** are u, v and the **intermediate variables** are x, y. All of the expressions above are understood to be functions of u, v. To be a bit more pedantic we can use $\vec{r}(u, v) = \langle x(u, v), y(u, v)\rangle$ and write

$$\boxed{\frac{\partial}{\partial u}\big[f(\vec{r}(u, v))\big] = \nabla f(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial u} \qquad \& \qquad \frac{\partial}{\partial v}\big[f(\vec{r}(u, v))\big] = \nabla f(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial v}.}$$

Notation aside, these rules are very natural extensions of what we have already seen. Note, in many examples which follow the notation $z = f(x, y)$ is utilized so you see z in the place of f.

Example 4.5.10. *Suppose $z = e^{xy}$ and $x = u^2 + v^2$ and $y = uv$. Calculate,*

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u} = ye^{xy}(2u) + xe^{xy}(v) = \big[3u^2v + v^3\big]e^{u^3v + uv^3}.$$

$$\frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v} = ye^{xy}(2v) + xe^{xy}(u) = \big[3uv^2 + u^3\big]e^{u^3v + uv^3}.$$

Once more, if we ignored the chain rule and instead directly substituted the expressions for u, v at the outset then we will still obtain the same result. However, if we are faced with extremely ugly formulas for $x(u, v)$ or $y(u, v)$ then this is a useful organizaing principle. Or we could encounter the situtation that formulas for $x(u, v)$ and $y(u, v)$ are not given and the chain rule still helps uncover general patterns.

Example 4.5.11. *Suppose $z = f(x, y)$ has continuous partial derivatives of at least order two. Furthermore, suppose $x = r^2 + s^2$ and $y = 2rs$. Note,*

$$\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r} = 2r\frac{\partial z}{\partial x} + 2s\frac{\partial z}{\partial y}.$$

If we differentiate again we must be careful to use the product rule and account for any implicit r-dependence:

$$\begin{aligned}
\frac{\partial^2 z}{\partial r^2} &= \frac{\partial}{\partial r}\left[2r\frac{\partial z}{\partial x} + 2s\frac{\partial z}{\partial y}\right] \\
&= 2\frac{\partial z}{\partial x} + 2r\frac{\partial}{\partial r}\left[\frac{\partial z}{\partial x}\right] + 2s\frac{\partial}{\partial r}\left[\frac{\partial z}{\partial y}\right] \\
&= 2\frac{\partial z}{\partial x} + 2r\left[\frac{\partial^2 z}{\partial x^2}\frac{\partial x}{\partial r} + \frac{\partial^2 z}{\partial y \partial x}\frac{\partial y}{\partial r}\right] + 2s\left[\frac{\partial^2 z}{\partial x \partial y}\frac{\partial x}{\partial r} + \frac{\partial^2 z}{\partial y^2}\frac{\partial y}{\partial r}\right] \\
&= 2\frac{\partial z}{\partial x} + 2r\left[2r\frac{\partial^2 z}{\partial x^2} + 2s\frac{\partial^2 z}{\partial y \partial x}\right] + 2s\left[2r\frac{\partial^2 z}{\partial x \partial y} + 2s\frac{\partial^2 z}{\partial y^2}\right] \\
&= 2\frac{\partial z}{\partial x} + 4r^2\frac{\partial^2 z}{\partial x^2} + 8sr\frac{\partial^2 z}{\partial x \partial y} + 4s^2\frac{\partial^2 z}{\partial y^2}
\end{aligned}$$

We can express the result above as: $f_{rr} = 2f_x + 4r^2 f_{xy} + 8sr f_{xy} + 4s^2 f_{yy}$. Notice we made use of Clairaut's Theorem ($f_{xy} = f_{yx}$) in several calculations.

Example 4.5.12. *Suppose $w = e^x \sin(y)$ and $y = st^2$ and $x = \ln(s - t)$. Then*

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial s} = e^x \sin(y) \cdot \frac{1}{s-t} + e^x \cos(y)(t^2) = \sin(st^2) + t^2(s-t)\cos(st^2).$$

The algebra of the last step is mostly due the identity $e^x = e^{\ln(s-t)} = s - t$. Likewise:

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} = e^x \sin(y) \cdot \frac{-1}{s-t} + e^x \cos(y)(2st) = -\sin(st^2) + 2st(s-t)\cos(st^2).$$

Example 4.5.13. *Suppose $z = e^r \cos\theta$ where $r = st$ and $\theta = \sqrt{s^2 + t^2}$.*

$$\begin{aligned}
\frac{\partial z}{\partial s} &= \frac{\partial z}{\partial r}\frac{\partial r}{\partial s} + \frac{\partial z}{\partial \theta}\frac{\partial \theta}{\partial s} \\
&= (e^r \cos\theta)(t) - (e^r \sin\theta)\left(\frac{s}{\sqrt{s^2+t^2}}\right) \\
&= e^{st}\left(t\cos\sqrt{s^2+t^2} - \frac{s}{\sqrt{s^2+t^2}}\sin\sqrt{s^2+t^2}\right).
\end{aligned}$$

Likewise,

$$\begin{aligned}
\frac{\partial z}{\partial t} &= \frac{\partial z}{\partial r}\frac{\partial r}{\partial t} + \frac{\partial z}{\partial \theta}\frac{\partial \theta}{\partial t} \\
&= (e^r \cos\theta)(s) - (e^r \sin\theta)\left(\frac{t}{\sqrt{s^2+t^2}}\right) \\
&= e^{st}\left(s\cos\sqrt{s^2+t^2} - \frac{t}{\sqrt{s^2+t^2}}\sin\sqrt{s^2+t^2}\right).
\end{aligned}$$

For a change of pace, the next example only uses the subscript notation:

Example 4.5.14. *Let $z = x^2y^3$ where $x = s\cos t$ and $y = s\sin t$. Observe $z_x = 2xy^3$ and $z_y = 3x^2y^2$. On the other hand, $x_s = \cos t$, $x_t = -s\sin t$, $y_s = \sin t$ and $y_t = s\cos t$. Therefore,*

$$\begin{aligned} z_t &= z_x x_t + z_y y_t \\ &= (2xy^3)(-s\sin t) + (3x^2y^2)(s\cos t) \\ &= s^5\big[-2\cos t\sin^4 t + 3\sin^2 t\cos^3 t\big]. \end{aligned}$$

Likewise, $z_s = z_x x_s + z_y y_s = (2xy^3)(\cos t) + (3x^2y^2)(\sin t) = 5s^4\cos^2 t\sin^3 t$.

Example 4.5.15. *Suppose $f : \mathbb{R} \to \mathbb{R}$ is a smooth function. Consider $z = f(xy)$. In such a case:*

$$\frac{\partial z}{\partial x} = \frac{df}{dt}(xy)\frac{\partial}{\partial x}\big[xy\big] = y\frac{df}{dt}(xy).$$

$$\frac{\partial z}{\partial y} = \frac{df}{dt}(xy)\frac{\partial}{\partial y}\big[xy\big] = x\frac{df}{dt}(xy).$$

Similarly, if $z = f(x/y)$

$$\frac{\partial z}{\partial x} = \frac{df}{dt}(x/y)\frac{\partial}{\partial x}\left[\frac{x}{y}\right] = \frac{1}{y}\frac{df}{dt}(x/y).$$

$$\frac{\partial z}{\partial y} = \frac{df}{dt}(x/y)\frac{\partial}{\partial y}\left[\frac{x}{y}\right] = \frac{-x}{y^2}\frac{df}{dt}(x/y).$$

The notations $\frac{df}{dt}(xy)$ and $\frac{df}{dt}(x/y)$ mean to differentiate the function f at t then set $t = xy$ or $t = x/y$ respectively. In the prime notation,

$$\frac{df}{dt}(xy) = f'(xy) \qquad \& \qquad \frac{df}{dt}(x/y) = f'(x/y).$$

Example 4.5.16. *Another application of the chain rule is coordinate change for the differentiation operators. For example, suppose $x = r\cos\theta, y = r\sin\theta$. How do we convert a partial derivative with respect to x for an equivalent differentiation in terms of the polar coordinates? Suppose f is an arbitrary function on $\mathbb{R}^2$, notice by the chain rule,*

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = \cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y}$$

$$\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}.$$

But, these relations hold for any function f hence we find the following operator equations:

$$\boxed{\frac{\partial}{\partial r} = \frac{x}{\sqrt{x^2+y^2}}\frac{\partial}{\partial x} + \frac{y}{\sqrt{x^2+y^2}}\frac{\partial}{\partial y}} \qquad \boxed{\frac{\partial}{\partial \theta} = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}}$$

Algebra challenge: solve the operator equations above for $\partial/\partial x$ and $\partial/\partial y$. Then compare your answers to what we obtain from the chain rules below:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial x} \qquad \& \qquad \frac{\partial f}{\partial y} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial y}$$

We need the formulas $r = \sqrt{x^2+y^2}$ *and* $\theta = \tan^{-1}(y/x) + c$ *where* c *is a constant that is either zero for* $x > 0$ *or* π *for* $x < 0$. *(ok, maybe constant is the wrong word, but it certainly differentiates to zero at most points). Calculate that* $\frac{\partial r}{\partial x} = \frac{x}{r} = \cos\theta$ *and* $\frac{\partial r}{\partial y} = \frac{y}{r} = \sin\theta$. *Also,*

$$\frac{\partial \theta}{\partial x} = \frac{\partial}{\partial x}\big[\tan^{-1}(y/x) + c\big] = \frac{1}{1+y^2/x^2} \cdot \frac{-y}{x^2} = \frac{-y}{x^2+y^2} = -\frac{\sin\theta}{r}$$

$$\frac{\partial \theta}{\partial y} = \frac{\partial}{\partial y}\big[\tan^{-1}(y/x) + c\big] = \frac{1}{1+y^2/x^2} \cdot \frac{1}{x} = \frac{x}{x^2+y^2} = \frac{\cos\theta}{r}.$$

Now substitute these back into the chain rules,

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial r}\cos\theta - \frac{\partial f}{\partial \theta}\frac{\sin\theta}{r} \qquad \& \qquad \frac{\partial f}{\partial y} = \frac{\partial f}{\partial r}\sin\theta + \frac{\partial f}{\partial \theta}\frac{\cos\theta}{r}$$

We obtain,

$$\boxed{\frac{\partial}{\partial x} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}} \qquad \& \qquad \boxed{\frac{\partial}{\partial y} = \sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}.}$$

We may also be faced with the problem of changing coordinates for higher derivatives. The differentiatial equation $\nabla^2\Phi = 0$ is called Laplace's equation. It is important to the theory of electrostatics as well as fluid flow. In cartesian coordinates $\nabla = \widehat{x}\partial_x + \widehat{y}\partial_y$ and it follows that the Laplacian operator $\nabla \bullet \nabla = \partial_x^2 + \partial_y^2$. (we'll explore this sort of differentiation more at the end of this course). The example below builds off the results of the previous example, keep that in mind.

Example 4.5.17. *Laplace's equation in Cartesian coordinates is* $\Phi_{xx} + \Phi_{yy} = 0$. *If a problem happens to have circular boundary conditions it may be better to solve Laplace's equation in polar coordinate form. What follows is a partial sketch on how this is accomplished, there are a few lines of algebra missing between the third and fourth equalities.*

$$\begin{aligned}
\nabla^2\Phi &= \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} \\
&= \frac{\partial}{\partial x}\left[\frac{\partial\Phi}{\partial x}\right] + \frac{\partial}{\partial y}\left[\frac{\partial\Phi}{\partial y}\right] \\
&= \left[\cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta}\right]\left[\cos\theta\frac{\partial\Phi}{\partial r} - \frac{\sin\theta}{r}\frac{\partial\Phi}{\partial\theta}\right] \\
&\quad + \left[\sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial\theta}\right]\left[\sin\theta\frac{\partial\Phi}{\partial r} + \frac{\cos\theta}{r}\frac{\partial\Phi}{\partial\theta}\right] \\
&= \boxed{\frac{\partial^2\Phi}{\partial r^2} + \frac{1}{r}\frac{\partial\Phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\theta^2}}
\end{aligned}$$

I invite the reader to fill in the details missing in the last step.

Example 4.5.18. *Let* $w(s,t) = F(u(s,t), v(s,t))$ *where* F, u, v *are differentiable functions and we're given that:*

$$\begin{array}{lll}
u(1,0) = 2 & v(1,0) = 3 & F_u(2,3) = -1 \\
u_s(1,0) = -2 & v_s(1,0) = 5 & F_v(2,3) = 10 \\
u_t(1,0) = 6 & v_t(1,0) = 4 &
\end{array}$$

We can calculate $w_s(1,0)$ *and* $w_t(1,0)$ *given the data above. The chain-rules below give us a guide:*

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial u}\frac{\partial u}{\partial s} + \frac{\partial w}{\partial v}\frac{\partial v}{\partial s} \quad \& \quad \frac{\partial w}{\partial t} = \frac{\partial w}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial w}{\partial v}\frac{\partial v}{\partial t}.$$

We apply the chain-rule above in what follows, note $w = F$,

$$\begin{aligned} w_s(1,0) &= F_u(u(1,0), v(1,0))u_s(1,0) + F_v(u(1,0), v(1,0))v_s(1,0) \\ &= -2F_u(2,3) + 5F_v(2,3) \\ &= 52. \end{aligned}$$

Furthermore,

$$\begin{aligned} w_t(1,0) &= F_u(u(1,0), v(1,0))u_t(1,0) + F_v(u(1,0), v(1,0))v_t(1,0) \\ &= 6F_u(2,3) + 4F_v(2,3) \\ &= 34. \end{aligned}$$

You may recall implicit differentiation from first semester calculus. What we do in the following example is similar. The full justification for such calculus is given in advanced calculus and perhaps this example makes more sense in light of the later section on *constrained differentiation*. That said, I leave it here as the calculation is fairly natural:

Example 4.5.19. *Suppose* $x^2+y^2+z^2 = 3xyz$*. Assume this equation implicitly defines* $z = z(x,y)$ *and calculate* $\frac{\partial z}{\partial x}$ *and* $\frac{\partial z}{\partial y}$*. We proceed just as we did in first semester calculus: differentiate the constraining equation. However, now we have a choice* $\frac{\partial}{\partial x}$ *or* $\frac{\partial}{\partial y}$*. We begin with* $\frac{\partial}{\partial x}$*: not the use of the product rule on the r.h.s.*

$$2x + 2y\frac{\partial y}{\partial x} + 2z\frac{\partial z}{\partial x} = 3yz + 3xy\frac{\partial z}{\partial x}$$

However, $\frac{\partial y}{\partial x} = 0$ *as these are independent variables. Observe*

$$2x - 3yz = (3xy - 2z)\frac{\partial z}{\partial x} \qquad \Rightarrow \qquad \boxed{\frac{\partial z}{\partial x} = \frac{3yz - 2x}{2z - 3xy}.}$$

Next, we calculate $\frac{\partial}{\partial y}$ *on the equation:*

$$2x\frac{\partial x}{\partial y} + 2y + 2z\frac{\partial z}{\partial y} = 3xz + 3xy\frac{\partial z}{\partial y}$$

By independence of x, y, $\frac{\partial x}{\partial y} = 0$*. Observe,*

$$2y - 3xz = (3xy - 2z)\frac{\partial z}{\partial y} \qquad \Rightarrow \qquad \boxed{\frac{\partial z}{\partial y} = \frac{3xz - 2y}{2z - 3xy}.}$$

Example 4.5.20. *Let* $z = z(x,y)$ *be implicitly defined by* $z^2 = \sin(xy) + x + \ln(y)$*. We find* $\frac{\partial z}{\partial x}$ *by implicit differentiation of the given equation:*

$$2z\frac{\partial z}{\partial x} = y\cos(xy) + 1 \quad \Rightarrow \quad \boxed{\frac{\partial z}{\partial x} = \frac{y\cos(xy)+1}{2z}.}$$

Likewise, take $\frac{\partial}{\partial y}$ *of the given equation to derive:*

$$2z\frac{\partial z}{\partial y} = x\cos(xy) + \frac{1}{y} \quad \Rightarrow \quad \boxed{\frac{\partial z}{\partial y} = \frac{xy\cos(xy)+1}{2zy}.}$$

The answers above are **implicitly** in terms of x, y. However, they are not **explicit** in x, y as the variable z appears in both solutions.

Example 4.5.21. *Let $x - z = \tan^{-1}(yz)$ implicitly define $z = z(x, y)$ and take $\frac{\partial}{\partial x}$ of the equation to find:*

$$1 - \frac{\partial z}{\partial x} = \left(\frac{1}{1+y^2z^2}\right) y\frac{\partial z}{\partial x} \quad \Rightarrow \quad \boxed{\frac{\partial z}{\partial x} = \frac{1+y^2z^2}{1+y^2z^2+y}.}$$

We have omit some algebra in the implication above. Next, calculate $\frac{\partial}{\partial y}$ of the given equation:

$$-\frac{\partial z}{\partial y} = \left(\frac{1}{1+y^2z^2}\right)\left(z + y\frac{\partial z}{\partial y}\right) \quad \Rightarrow \quad \boxed{\frac{\partial z}{\partial y} = \frac{-z}{1+y^2z^2+y}.}$$

Likewise,

Next, consider $F = F(x, y, z)$ and $\vec{r} = \vec{r}(u, v)$. In particular, we wish to differentiate the composite of $F : \mathbb{R}^3 \to \mathbb{R}$ with $\vec{r} : \mathbb{R}^2 \to \mathbb{R}^3$. The chain rules in this case are as follows:

$$\boxed{\frac{\partial F}{\partial u} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial u} \qquad \& \qquad \frac{\partial F}{\partial v} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial v}.}$$

You can write these rules in terms of gradients and partial derivatives of vectors,

$$\frac{\partial F}{\partial u} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial u} \qquad \frac{\partial F}{\partial v} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial v}.$$

I explained in the preceding section that we can derive chain rules from the general derivative.

Let us return once more to the conclusion of the preceding section to appreciate how these rules may be derived from matrix multiplication of the Jacobian matrices(I call the Jacobian matrix **the derivative** of the map). For example, suppose $F = F(x, y, z)$ which implicits $F : \mathbb{R}^3 \to \mathbb{R}$. It follows that the derivative matrix of F is a 1×3 matrix:

$$F' = \left[\frac{\partial F}{\partial x}, \ \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}\right]$$

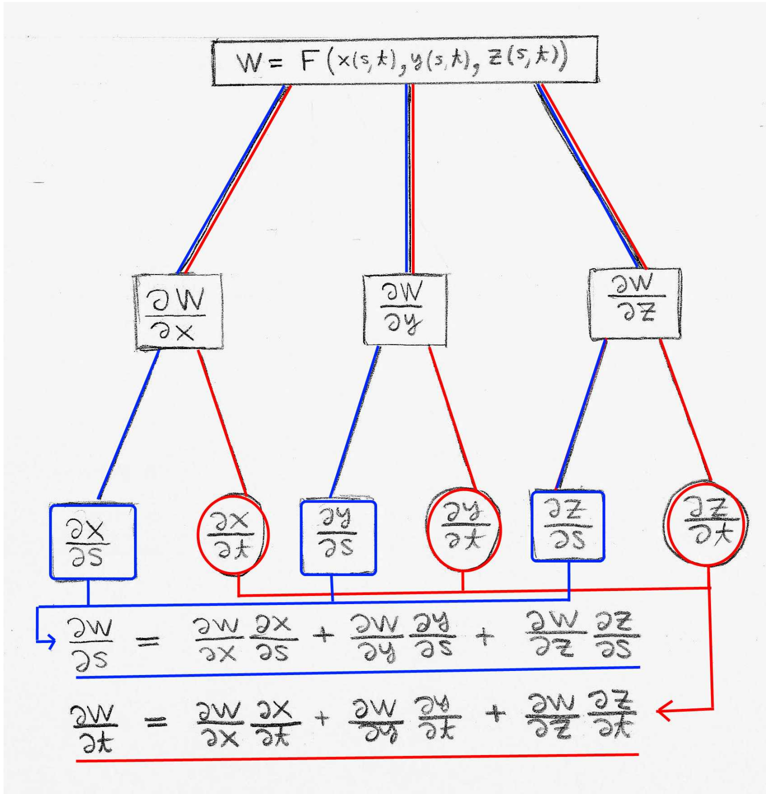
Also, consider $\vec{X} = \vec{X}(s, t) = \langle x(s, t), y(s, t), z(s, t)\rangle$ which implies $\vec{X} : \mathbb{R}^2 \to \mathbb{R}^3$ and hence the derivative matrix of $\vec{X}$ is the 3×2 matrix:

$$\vec{X}' = \left[\frac{\partial \vec{X}}{\partial s}, \frac{\partial \vec{X}}{\partial t}\right] = \begin{bmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial t} \end{bmatrix}$$

Very well, set $w = F(\vec{X}(s, t))$ and we may extract the chain-rules for w from the general chain-rule as follows: $w = F \circ \vec{X}$ hence

$$\begin{aligned} w' = (F \circ \vec{X})' &= F'\vec{X}' \\ &= \left[\frac{\partial F}{\partial x}, \ \frac{\partial F}{\partial y}, \frac{\partial F}{\partial z}\right] \begin{bmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial t} \end{bmatrix} \\ &= \Bigg[\underbrace{\frac{\partial F}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial s}}_{\frac{\partial w}{\partial s}}, \ \underbrace{\frac{\partial F}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial t}}_{\frac{\partial w}{\partial s}}\Bigg] \end{aligned}$$

If you would rather divorce the concept of the chain rule from linear algebra[20] and deeper math then feel free to use these silly tree diagrams to remember the chain rule:



Personally, I much prefer to calculate with understanding as opposed to inventing new and unecessary mnemonics. To each his own, you just need to find the way that works for you. Context is everything with chain rules.

Finally, to finish this section we consider problems with two independent variables involving vectors of functions. I begin with the easy cases:

Example 4.5.22. *Consider $\vec{r} = \langle x, y, z\rangle$ where $x, y, z : \mathbb{R}^2 \to \mathbb{R}$ meaning $x = x(u, v)$ and $y = y(u, v)$ and $z = z(u, v)$. In such a case,*

$$\frac{\partial \vec{r}}{\partial u} = \left\langle \frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u} \right\rangle \qquad \& \qquad \frac{\partial \vec{r}}{\partial v} = \left\langle \frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v} \right\rangle$$

Partial differentiation of a vector is completed just the same way as ordinary differentiation of a vector: componentwise.

Example 4.5.23. *Consider $\vec{r}(u, v) = \langle u, v, f(u, v)\rangle$ then*

$$\partial_u \vec{r} = \langle 1, 0, \partial_u f\rangle \qquad \& \qquad \partial_v \vec{r} = \langle 0, 1, \partial_v f\rangle$$

[20] I realize most of you have not seen linear algebra in its manifest beauty, however, sometime soon you will and I write these notes not just for the now, but, also for you later.

Example 4.5.24. *Let $\vec{A}(x,y,z) = \langle y^2, x, \sin(z)\rangle$ and $\vec{r}(u,v) = \langle u, v, f(u,v)\rangle$. Find partial derivatives of $\vec{A}\circ\vec{r}$ with respect to u and v.*

Solution: *we use the same chain rule as in previous work, however, now we have to calculate one such rule for each component of $\vec{A}$:*

$$\begin{aligned}\frac{\partial}{\partial u}(\vec{A}\circ\vec{r}) &= \frac{\partial}{\partial u}\langle y^2, x, \sin(z)\rangle \\ &= \left\langle\ 2y\frac{\partial y}{\partial u},\ \frac{\partial x}{\partial u},\ \cos(z)\frac{\partial z}{\partial u}\ \right\rangle \\ &= \left\langle\ 0,\ 1,\ \cos(f(u,v))\frac{\partial f}{\partial u}\ \right\rangle.\end{aligned}$$

Likewise,

$$\begin{aligned}\frac{\partial}{\partial v}(\vec{A}\circ\vec{r}) &= \frac{\partial}{\partial v}\langle y^2, x, \sin(z)\rangle \\ &= \left\langle\ 2y\frac{\partial y}{\partial v},\ \frac{\partial x}{\partial v},\ \cos(z)\frac{\partial z}{\partial v}\ \right\rangle \\ &= \left\langle\ 2v,\ 0,\ \cos(f(u,v))\frac{\partial f}{\partial v}\ \right\rangle.\end{aligned}$$

It might be more clear to consider a somewhat abstract example.

Example 4.5.25. *Let $\vec{F} = \langle A, B, C\rangle$ where $A, B, C : \mathbb{R}^3 \to \mathbb{R}$. Furthermore, suppose $\vec{r} = \langle x, y, z\rangle$ where $x, y, z : \mathbb{R}^2 \to \mathbb{R}$. In other words, $x = x(u,v), y = y(u,v)$ and $z = z(u,v)$. Let us examine the chain rule in this context:*

$$\frac{\partial}{\partial u}(\vec{F}\circ\vec{r}) = \frac{\partial}{\partial u}\langle A, B, C\rangle = \left\langle \frac{\partial A}{\partial u},\ \frac{\partial B}{\partial u},\ \frac{\partial C}{\partial u}\right\rangle = \left\langle\ \nabla A \bullet \frac{\partial \vec{r}}{\partial u},\ \nabla B \bullet \frac{\partial \vec{r}}{\partial u},\ \nabla C \bullet \frac{\partial \vec{r}}{\partial u}\ \right\rangle$$

Likewise,

$$\frac{\partial}{\partial v}(\vec{F}\circ\vec{r}) = \left\langle \frac{\partial A}{\partial v},\ \frac{\partial B}{\partial v},\ \frac{\partial C}{\partial v}\right\rangle = \left\langle\ \nabla A \bullet \frac{\partial \vec{r}}{\partial v},\ \nabla B \bullet \frac{\partial \vec{r}}{\partial v},\ \nabla C \bullet \frac{\partial \vec{r}}{\partial v}\ \right\rangle$$

Just to expand on the notation for two terms:

$$\frac{\partial A}{\partial u} = \underbrace{\frac{\partial A}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial A}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial A}{\partial z}\frac{\partial z}{\partial u}}_{\nabla A \bullet \partial_u \vec{r}} \quad or \quad \frac{\partial C}{\partial v} = \underbrace{\frac{\partial C}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial C}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial C}{\partial z}\frac{\partial z}{\partial v}}_{\nabla C \bullet \partial_v \vec{r}}.$$

We will return to this particular type of chain rule when we study the question of coordinate dependence of the surface integral. I include these examples to make that discussion more accessible.

4.5.3 several independent variables

We proceed by extending the results of this section thus far to three and more independent variables. To begin, if $f : \mathbb{R} \to \mathbb{R}$ is a function and $g : \mathbb{R}^n \to \mathbb{R}$ is a function of $n$variables $u_1, u_2, \dots, u_n$ then

$$\frac{\partial}{\partial u_j}\big[f(g(u_1,\dots,u_n))\big] = f'(g(u_1,\dots,u_n))\frac{\partial g}{\partial u_j}$$

Example 4.5.26. *Suppose* $f(x) = \sin^2(x)$ *and* $g(u_1, \dots, u_n) = u_1 + 2u_2 + \cdots + nu_n$ *then*

$$\frac{\partial}{\partial u_j}\left[\sin^2(g)\right] = 2\sin(g)\cos(g)\frac{\partial g}{\partial u_j} = j\sin(2[u_1 + 2u_2 + \cdots + nu_n]).$$

Example 4.5.27. *Let* $r = \sqrt{x^2 + y^2 + z^2}$ *and consider* $f(x, y, z) = \sinh(\sqrt{x^2 + y^2 + z^2}) = \sinh(r)$ *observe,*

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\sinh(r) = \cosh(r)\frac{\partial r}{\partial x} = \frac{x\cosh(r)}{r}$$
$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\sinh(r) = \cosh(r)\frac{\partial r}{\partial y} = \frac{y\cosh(r)}{r}$$
$$\frac{\partial f}{\partial z} = \frac{\partial}{\partial z}\sinh(r) = \cosh(r)\frac{\partial r}{\partial z} = \frac{z\cosh(r)}{r}$$

In fact, we already made calculations like the one above in our study of basic partial differentiation. Next, consider $f : \mathbb{R}^m \to \mathbb{R}$ and $x, y : \mathbb{R}^n \to \mathbb{R}$. For $m = 2$, we suppose $x = x(u_1, u_2, \dots, u_n)$ and $y = y(u_1, u_2, \dots, u_n)$ and we seek to analyze $\frac{\partial f}{\partial u_j}$ for some $j \in \mathbb{N}_n$. The answer is by now familar,

$$\frac{\partial f}{\partial u_j} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u_j} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u_j}.$$

Moreover,$m = 3$ means $f = f(x, y, z)$ and

$$\frac{\partial f}{\partial u_j} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u_j} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u_j} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial u_j}.$$

Then, the glorious general case: $f = f(x_1, x_2, \dots, x_m)$

$$\frac{\partial f}{\partial u_j} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial u_j} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial u_j} + \cdots + \frac{\partial f}{\partial x_m}\frac{\partial x_m}{\partial u_j}.$$

All of these can be written as dot products of gradients. I leave that to the interested reader and conclude our study with a few examples: often $u_1 = u$ and $u_2 = v$ and $u_3 = w$ for examples. Sometimes, $u_1 = r$ and $u_2 = s$ and $u_3 = t$.

Example 4.5.28. *Let* $f(x, y) = x^2 + xy$ *and* $x = u^2 + v^2 + w^2$ *and* $y = uvw$ *then*

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} = (2x + y)\frac{\partial x}{\partial u} + x\frac{\partial y}{\partial u} = (2x + y)(2u) + x(vw).$$

$$\frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} = (2x + y)\frac{\partial x}{\partial v} + x\frac{\partial y}{\partial v} = (2x + y)(2v) + x(uw).$$

$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial w} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial w} = (2x + y)\frac{\partial x}{\partial w} + x\frac{\partial y}{\partial w} = (2x + y)(2w) + x(uv).$$

The answers above could be made explicit by setting $x = u^2 + v^2 + w^2$ *and* $y = uvw$.

Example 4.5.29. *Let* $f(x, y, z) = x + y^3 + e^z$ *and* $x = r^2$ *and* $y = r + s$ *and* $z = \sin(s^2 + t^2)$

$$\begin{aligned}\frac{\partial f}{\partial r} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial r} \\ &= 1(2r) + 3y^2(1) + e^z(0) \\ &= 2r + 3(r + s)^2.\end{aligned}$$

Likewise,

$$\begin{aligned}\frac{\partial f}{\partial t} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial t} \\ &= 1(0) + 3y^2(0) + e^z(2t\cos(s^2+t^2)) \\ &= 2t\cos(s^2+t^2)e^{\sin(s^2+t^2)}.\end{aligned}$$

I leave $\frac{\partial f}{\partial s}$ to the reader.

Example 4.5.30. *Suppose $z = x^2 + xy^3$ and $x = uv^2 + w^3$ and $y = u + ve^w$. Calculate $\frac{\partial z}{\partial u}$ when $u = 2, v = 1$ and $w = 0$.*

Solution: *By the chain rule for $z = z(x, y)$,*

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u} = (2x + y^3)(v^2) + (3xy^2)(1).$$

Observe $u = 2, v = 1$ and $w = 0$ implies $x(2, 1, 0) = 2(1)^2 + 0^3 = 2$ and $y(2, 1, 0) = 2 + 1 \cdot e^0 = 3$. Therefore,

$$\frac{\partial z}{\partial u}\bigg|_{(2,1,0)} = (2(2) + 3^3)(1^2) + (3 \cdot 2 \cdot 3^2)(1) = 4 + 27 + 54 = 85.$$

It is probably good for you to see an example where the outside function has variables not labeled as Cartesians. For example:

Example 4.5.31. *Suppose $f(r, \theta, \phi) = r^3 + \tan^{-1}(\theta)$ and $r = \sqrt{x^2 + y^2 + z^2}$, $\tan\theta = y/x$ and $z = r\cos\phi$ then we calculate,*

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial x} + \frac{\partial f}{\partial \phi}\frac{\partial \phi}{\partial x} = 3r^2\frac{\partial r}{\partial x} + \frac{1}{1+\theta^2}\frac{\partial \theta}{\partial x}$$

We calculate $\frac{\partial r}{\partial x} = \frac{x}{r}$ whereas $\sec^2\theta\frac{\partial \theta}{\partial x} = \frac{-y}{x^2}$ thus after a small bit of algebra:

$$\frac{\partial f}{\partial x} = 3xr + \frac{1}{1+\theta^2}\left(-\cos^2\theta\frac{y}{x^2}\right)$$

Next,

$$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial z} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial z} + \frac{\partial f}{\partial \phi}\frac{\partial \phi}{\partial z} = 3r^2\frac{\partial r}{\partial z} + \frac{1}{1+\theta^2}\frac{\partial \theta}{\partial z} = 3rz.$$

as we note $\frac{\partial \theta}{\partial z} = 0$ and $\frac{\partial r}{\partial z} = z/r$.

Example 4.5.32. *Supopse $f(r, \theta, z) = r^2 + z\theta^2$ and $r = \sqrt{x^2 + y^2}$ and $x = r\cos\theta$ whereas $y = r\sin\theta$ hence $\tan\theta = y/x$ once more. Observe:*

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial f}{\partial \theta}\frac{\partial \theta}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 2x + 2z\theta\left(-\cos^2\theta\frac{y}{x^2}\right).$$

On the other hand, $\frac{\partial f}{\partial z} = \theta^2$.

These sort of calculations arise when we ask questions about coordinate change in three variables.

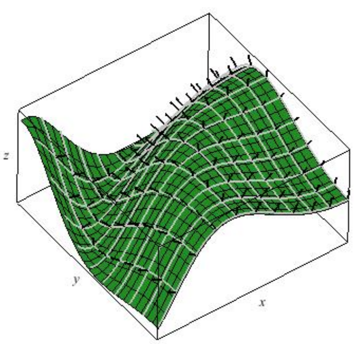
4.6 tangent spaces and the normal vector field

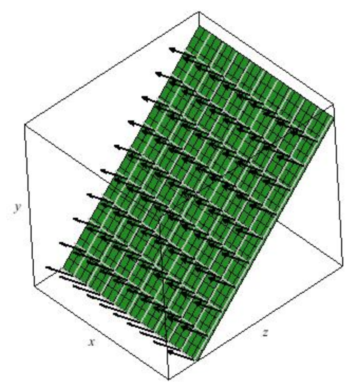
In this section we wish to analyze the tangent space for a smooth surface. We assume the surface in consideration is smooth so that the calculations are not complicated by exceptional cases. In particular, we wish to analyze a surface S in three particular views:

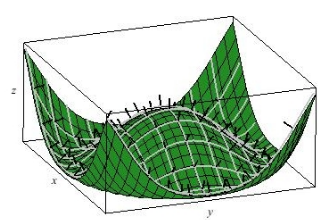
1. as a **level surface** the set S is the solution set of $F(x, y, z) = 0$
2. as a **parametrized surface** we see S as the image of $\vec{r} : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$
3. as a **graph** we see S as the solution set of $z = f(x, y)$

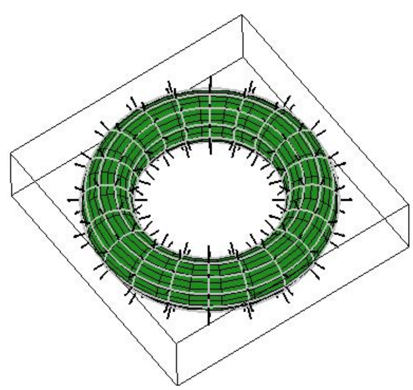
As we have discussed previously it is only sometimes possible to cover all of S as a graph. Moreover, each view has its advantages. My goal in this section is to explain how to find the tangent space and normal vector field for S in each of these views. We've already done a lot of calculation towards these questions in the last section.

For your viewing enjoyment I have included a few figures of surfaces which have coordinate curves in gray and little normal vectors in black. I have animations of these on the webpage, perhaps it helps bring to life the fact the normals pick out a side of orientable surfaces.









Our goal in this section is to find formulas for the little black arrows.

4.6.1 level surfaces and tangent space

In Example 4.5.6 we proved that curves in the solution set of $F(x, y, z) = k$ have tangent vectors which are perpendicular to ∇F. It follows that the normal vector for the tangent plane at $(x_o, y_o, z_o) \in S$ is simply $\nabla F(x_o, y_o, z_o)$. The tangent plane has equation:

$$\boxed{\nabla F(x_o, y_o, z_o) \bullet \langle x - x_o, y - y_o, z - z_o\rangle = 0.}$$

The **normal vector field on** S is given by the assignment

$$(x, y, z) \to \nabla F(x, y, z)$$

for each $(x, y, z) \in S$. A **normal line** to a point $p \in S$ is simply a line through p in the direction of the normal to S at p.

Remark 4.6.1.

The choice of level function matters. If we multiply the equation by a negative quantity the direction of the gradient flips over and hence the normal vector field flips to the other side of the surface. As an example, $F(x, y, z) = x^2 + y^2 + z^2 = 1$ has $\nabla F = \langle 2x, 2y, 2z\rangle$ whereas $G(x, y, z) = -x^2 - y^2 - z^2 = -1$ has $\nabla G = \langle -2x, -2y, -2z\rangle$. We say $F = 1$ is the sphere **oriented outwards** whereas $G = -1$ is the sphere **oriented inwards**.

Example 4.6.2. *Find the equation of hte tangent plane and normal line to the surface S:* $F(x, y, z) = x^2 - 2y^2 + z^2 + yz = 2$ *at* $(2, 1, -1)$.

Solution: *the normal vector field is given by* ∇F. *In particular,*

$$\nabla F(x, y, z) = \langle 2x, -4y + z, 2z + y\rangle$$

Therefore, the normal vector to S *at* $(2, 1, -1)$ *is:*

$$\nabla F(2, 1, -1) = \langle 4, -5, -1\rangle.$$

The tangent plane has normal $\langle 4, -5, -1\rangle$ *and we can use* $(2, 1, -1)$ *as the basepoint for the equation of the tangent plane;* $4(x - 2) - 5(y - 1) - (z + 1) = 0$. *Likewise, the parametric equation for the normal line is:* $\vec{r}(t) = (2, 1, -1) + t\langle 4, -5, -1\rangle$.

4.6.2 parametrized surfaces and tangent space

Suppose a surface S can either be viewed as a level surface $F(x, y, z) = k$ or as a parametrized surface by the mapping $\vec{r} : D \subseteq \mathbb{R}^2 \to \mathbb{R}^3$. In particular, if we denote the parameters by u, v and write $\vec{r} = \langle x, y, z\rangle$ then these viewpoints are connected by the equation $F(\vec{r}(u, v)) = k$ for all $u, v \in D$. The chain rules in this case are as follows:

$$\frac{\partial F}{\partial u} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial u} \qquad \& \qquad \frac{\partial F}{\partial v} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial v}.$$

You can write these rules in terms of gradients and partial derivatives of vectors,

$$\frac{\partial F}{\partial u} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial u} \qquad \frac{\partial F}{\partial v} = \nabla F(\vec{r}(u, v)) \bullet \frac{\partial \vec{r}}{\partial v}.$$

Differentiate $F(\vec{r}(u,v)) = k$ with respect to u or v to obtain,

$$\frac{\partial F}{\partial u} = \nabla F(\vec{r}(u,v)) \bullet \frac{\partial \vec{r}}{\partial u} = 0 \qquad \frac{\partial F}{\partial v} = \nabla F(\vec{r}(u,v)) \bullet \frac{\partial \vec{r}}{\partial v} = 0.$$

The vectors $\frac{\partial \vec{r}}{\partial u}$ and $\frac{\partial \vec{r}}{\partial v}$ are perpendicular to $\nabla F(\vec{r}(u,v))$. We envision all three of these vectors attached at the point $\vec{r}(u,v)$ of S. The curves

$$\vec{\alpha}(u) = \vec{r}(u, v_o) \qquad \vec{\beta}(v) = \vec{r}(u_o, v)$$

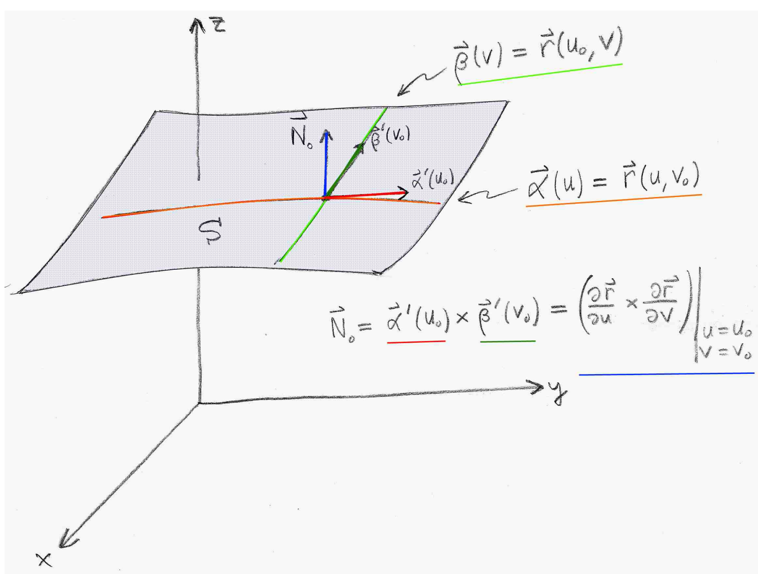
are the coordinate curves through $\vec{r}(u_o, v_o)$. The tangent vectors at $\vec{r}(u_o, v_o)$ to these curves are given by

$$\vec{\alpha}\,'(u_o) = \frac{d}{du}\big[\vec{r}(u,v_o)\big]_{u_o} = \frac{\partial \vec{r}}{\partial u}(u_o) \;\;\&\;\; \vec{\beta}\,'(v_o) = \frac{d}{dv}\big[\vec{r}(u_o,v)\big]_{v_o} = \frac{\partial \vec{r}}{\partial v}(v_o)$$

Therefore, the tangent vectors to the coordinate curves are perpendicular to the gradient vector of the corresponding level curve. In three dimensional space it follows that the cross product of $\vec{\alpha}\,'(u_o)$ with $\vec{\beta}\,'(v_o)$ must be colinear to $\nabla F(\vec{r}(u_o, v_o))$. Therefore, we define

$$\boxed{\vec{N}(u,v) = \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v}.}$$

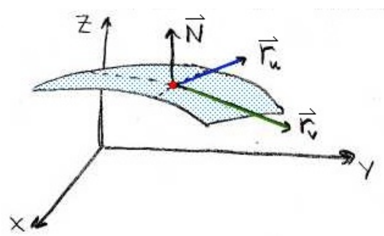
The vector field $\vec{r}(u,v) \to \vec{N}(u,v)$ defines the **normal vector field** of $\vec{r}$. If a surface S has a non-vanishing normal vector field then it is said to be **oriented**. Clearly it is easier to calculate the normal in the level surface formulation since gradients are way easier than cross products. However, we will find that the parametric viewpoint is an essential part of the definition of the surface integral for a vector field. The diagram below indicates how a particular vector in the normal vector field is calculated in the parametric setting:



Remark 4.6.3.

The ordering of the parameters matters. If we swap the order of the parameters it flips the normal vector field. Suppose S_1 is oriented by $\vec{r}_1(u,v) = \langle f(u,v), g(u,v), h(u,v)\rangle$ and $\vec{r}_2(v,u) = \langle f(u,v), g(u,v), h(u,v)\rangle$. The normal vector field induced from $\vec{r}_1$ by our conventions is $\vec{N}_1 = \partial_u \vec{r}_1 \times \partial_v \vec{r}_1$ whereas the normal vector field induced from $\vec{r}_2$ is $\vec{N}_2 = \partial_v \vec{r}_2 \times \partial_u \vec{r}_2$. Since $\vec{r}_1(u,v) = \vec{r}_2(v,u)$ it follows that $\vec{N}_1 = -\vec{N}_2$. My point? Beware the order.

Example 4.6.4. *Suppose S is parametrized by $\vec{r}(u,v) = \langle u^2, 2u\sin v, u\cos v\rangle$. To find the tangent plane at $\vec{r}(1,0) = \langle 1,0,1\rangle$ we need to find the normal vector to S at the point. A natural approach here is to find two tangent vectors from which we can produce a perpendicular vector by the cross product. Partial differentiation with respect to u and v gives us tangent vectors as pictured below:*



Notice, at (u,v),

$$\frac{\partial \vec{r}}{\partial u} = \langle 2u, 2\sin v, \cos v\rangle \qquad \& \qquad \frac{\partial \vec{r}}{\partial v} = \langle 0, 2u\cos v, -u\sin v\rangle$$

Evaluate at $(1,0)$ yields

$$\frac{\partial \vec{r}}{\partial u}(1,0) = \langle 2,0,1\rangle \qquad \& \qquad \frac{\partial \vec{r}}{\partial v}(1,0) = \langle 0,2,0\rangle$$

The cross-product of the tangent vectors above gives us the normal,

$$\frac{\partial \vec{r}}{\partial u}(1,0) \times \frac{\partial \vec{r}}{\partial v}(1,0) = det\begin{bmatrix} \widehat{x} & \widehat{y} & \widehat{z} \\ 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix} = \langle -2, 0, 4\rangle.$$

Remember $(1,0,1)$ is the point of tangencyy thus we find the equation of the tangent plane is $-2(x-1)+4(z-1)=0$.

Example 4.6.5. *Suppose the plane $F(x,y,z) = a(x-x_o)+b(y-y_o)+c(z-z_o) = 0$ contains non-colinear vectors $\vec{A}$ and $\vec{B}$. Note $\nabla F = \langle a,b,c\rangle$, the normal derived from the level function matches the natural normal suggested by the equation for the plane. Next, consider the parametrization naturally induced from $\vec{A}, \vec{B}$ and the base-point (x_o, y_o, z_o),*

$$\vec{r}(u,v) = \langle x_o, y_o, z_o\rangle + u\vec{A} + v\vec{B}.$$

In this case calculation of the tangent vectors to the coordinate curves is easy:

$$\frac{\partial \vec{r}}{\partial u} = \vec{A} \qquad \frac{\partial \vec{r}}{\partial v} = \vec{B}$$

Thus $\vec{N}(u,v) = \vec{A} \times \vec{B}$. The normal vector field to a plane is a constant vector field. Geometry indicates that $\vec{A} \times \vec{B} = \lambda\langle a,b,c\rangle$ for some nonzero constant λ.

Example 4.6.6. *Find the tangent plane at $(1,0,1)$ for the surface S parametrized by $\vec{r}(s,t) = \big\langle s, \ln(st), t\big\rangle$.*

Solution: *partial differentiation of the parametrization yields:*

$$\frac{\partial \vec{r}}{\partial s} = \left\langle 1, \frac{1}{s}, 0\right\rangle \quad \& \quad \frac{\partial \vec{r}}{\partial t} = \left\langle 0, \frac{1}{t}, 1\right\rangle$$

these vectors are tangent to S at $\vec{r}(s,t)$. Therefore, in particular, if $s = 1$ and $t = 1$ we obtain tangent vectors to S at $(1,0,1)$ of:

$$\frac{\partial \vec{r}}{\partial s}(1,1) = \big\langle 1,1,0 \big\rangle \quad \& \quad \frac{\partial \vec{r}}{\partial t}(1,1) = \big\langle 0,1,1 \big\rangle$$

The cross-product of these tangent vectors provides a normal to the tangent plane. Consider,

$$\frac{\partial \vec{r}}{\partial s}(1,1) \times \frac{\partial \vec{r}}{\partial t}(1,1) = det \begin{bmatrix} \widehat{x} & \widehat{y} & \widehat{z} \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \langle 1,-1,1 \rangle.$$

Therefore, the tangent plane to S at $(1,0,1)$ has equation $(1)(x-1)+(-1)(y-0)+(1)(z-1) = 0$ which simplifies to $x - y + z = 2$.

4.6.3 tangent plane to a graph

The graphical viewpoint is connected to the level-surface view and the parametric view by the following: given that S is the solution set of $z = f(x,y)$ we can

1. write S as the level surface $F(x,y,z) = 0$ for $F(x,y,z) = z - f(x,y)$.

2. write S as a parametrized surface with parameters x, y and $\vec{r}(x,y) = \langle x, y, f(x,y) \rangle$.

Notice there is some ambiguity in the normal vectors which are induced. If we chose $-F = 0$ then that flips over the normal and if we swapped the order of the parameters x, y then that would also flip the normal vector $\vec{N}(x,y)$. These ambiguities must be dealt with as we do calculations on surfaces. Picking an orientation specifies a side to the surface. Equivalently, an **oriented surface is a set of points paired with a normal vector field on the surface.**

Remark 4.6.7.

Question: if S is oriented and we describe S_1 by $F(x,y,z) = 0$ for $F(x,y,z) = z - f(x,y)$ then is the same oriented surface as the parametrized surface S_2 with parameters x, y and $\vec{r}(x,y) = \langle x, y, f(x,y) \rangle$?

Solution: to begin note that as point-sets it is clear that $S_1 = S_2$ so the question reduces to the problem of ascertaining if the normal vector fields match-up. Calculate, from the level-surface viewpoint the normal vector field at (x,y,z) on S_1 is

$$\vec{N}(x,y,z) = \nabla F = \langle -f_x, -f_y, 1 \rangle$$

On the other hand, from the parametric viewpoint we calculate for $(x,y) \in dom(f)$,

$$\frac{\partial \vec{r}}{\partial x} = \langle 1, 0, f_x \rangle \qquad \& \qquad \frac{\partial \vec{r}}{\partial y} = \langle 0, 1, f_y \rangle$$

and the cross-product

$$\begin{aligned} \vec{N}(x,y) = \frac{\partial \vec{r}}{\partial x} \times \frac{\partial \vec{r}}{\partial y} &= (\,\widehat{x} + f_x\,\widehat{z}) \times (\,\widehat{y} + f_y\,\widehat{z}) \\ &= \widehat{x} \times \widehat{y} + f_y\,\widehat{x} \times \widehat{z} + f_x\,\widehat{z} \times \widehat{y} \\ &= \widehat{z} - f_y\,\widehat{y} - f_x\,\widehat{x} \\ &= \langle -f_x, -f_y, 1 \rangle. \end{aligned}$$

Therefore, if we change viewpoints as advocated at the beginning of the subsection we will maintain the natural orientation. This is the reason I wrote $F(x,y,z) = z - f(x,y)$ as opposed to $G(x,y,z) = f(x,y) - z$.

Continuing the thought of the remark above, it is useful for us to write the standard equation for the tangent plane to a graph: if we consider the point of tangency $(a, b, f(a,b))$ we have normal vector $\langle -f_x(a,b), -f_y(a,b), 1\rangle$ hence the equation is $-(x-a)f_x(a,b) - (y-b)f_y(a,b) + z = 0$. It is convenient to write this as:

$$\boxed{z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)} \qquad \text{tangent plane as a graph}$$

You should recall that $y = f(a) + f'(a)(x-a)$ was the equation of the tangent line to $y = f(x)$ at $(a, f(a))$. The formula above is the natural generalization.

Example 4.6.8. *Let $f(x,y) = x^2 + y^2$. The tangent plane at $(1,2,5)$ is:*

$$z = f(1,2) + f_x(1,2)(x-1) + f_y(1,2)(y-2) \quad \Rightarrow \quad \boxed{z = 5 + 2(x-1) + 4(y-2).}$$

We should remember that the formula of the previous example is not magic, it is the natural outgrowth of our geometric discussion either from the perspective of level surfaces or from the parametric formulation.

Example 4.6.9. *Let $f(x,y) = \exp(x^2 - y^2)$. Let us find the equation of the tangent plane to $z = f(x,y)$ when $x = 1$ and $y = -1$. Note:*

$$f_x(x,y) = 2x\exp(x^2 - y^2) \qquad \& \qquad f_y(x,y) = -2y\exp(x^2 - y^2).$$

We use these formulas to simplify the equation of the tangent plane below:

$$\begin{aligned} z &= f(1,-1) + f_x(1,-1)(x-1) + f_y(1,-1)(y+1) \\ &= \exp(0) + 2(1)\exp(0)(x-1) - 2(-1)\exp(0)(y+1) \\ &= 1 + 2(x-1) + 2(y+1) \end{aligned}$$

Thus, the tangent plane of $z = \exp(x^2 - y^2)$ to $(1,-1,1)$ is $\boxed{z = 1 + 2(x-1) + 2(y+1)}$.

Example 4.6.10. *Let $f(x,y) = 4x^2 - y^2 + 2y$. Find the equation of the tangent plane at $x = -1$ and $y = 2$. Note, as this is a graph, this unambiguously indicates that $z = f(-1,2) = 4$. We could proceed as in the previous pair of examples and use $z = f(-1,2) + f_x(-1,2)(x+1) + f_y(-1,2)(y-2)$, however, I add zero instead:*

$$\begin{aligned} z &= 4(x+1-1)^2 - (y-2+2)^2 + 2(y-2+2) \\ &= 4(x+1)^2 - 8(x+1) + 4 - (y-2)^2 - 4(y-2) - 4 + 2(y-2) + 4 \\ &= \underbrace{4 - 8(x+1) - 2(y-2)}_{\text{first order part}} + 4(x+1)^2 - (y-2)^2 \end{aligned}$$

The tangent plane at $(-1,2,4)$ to $z = 4x^2 - y^2 + 2y$ is given by $z = 4 - 8(x+1) - 2(y-2)$. Notice, close to the point of tangency, the quadratic terms in the expansion above are very small. In comparison, the linear part dominates.

The example above gives us a window into the theory of multivariate power series. It turns out that we can expand a function of several variables in terms of a multivariate Taylor expansion. The initial terms in such a series correspond to the tangent plane for a function of two variables, or the tangent hyperplane if we have additional variables. This is the natural generalization of the one-variable result; an analytic function $f(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2 + \cdots$ has a Taylor expansion centered at $x = a$ which begins with the terms which form the tangent line to $(a, f(a))$. In any event, we have more to say about multivariate power series as we consider the problem of optimization in several variables in a later chapter.

4.7 partial differentiation with side conditions

Every chain rule in the preceding section follows as a subcase of the chain rule for the general derivative. In this section the rigorous justification is given by the implicit or inverse function theorems. I will not even state those here[21]. I discuss them in advanced calculus and those notes are available if you'd like to read about the theoretical underpinning for the calculations in this section. I will show how to formally calculate in this section. In other words, I will teach you symbol pushing techniques. To begin, we define the total differential.

Definition 4.7.1.

If $f = f(x_1, x_2, \dots, x_n)$ then $df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \cdots + \frac{\partial f}{\partial x_n}dx_n$.

Example 4.7.2. *Suppose $E = pv + t^2$ then $dE = vdp + pdv + 2tdt$. In this example the dependent variable is E whereas the independent variables are p, v and t.*

Example 4.7.3. *What are $\partial F/\partial x$ and $\partial F/\partial y$ if we know that $F = F(x,y)$ and $dF = (x^2+y)dx - \cos(xy)dy$.*

Solution: *if $F = F(x,y)$ then the total differential has the form $dF = F_x dx + F_y dy$. We simply compare the general form to the given $dF = (x^2+y)dx - \cos(xy)dy$ to obtain:*

$$\frac{\partial F}{\partial x} = x^2 + y, \qquad \frac{\partial F}{\partial y} = -\cos(xy).$$

Example 4.7.4. *Suppose $u = e^{-t}\sin(s+2t)$. The total differential is:*

$$\begin{aligned} du &= \frac{\partial}{\partial s}\big[e^{-t}\sin(s+2t)\big]ds + \frac{\partial}{\partial t}\big[e^{-t}\sin(s+2t)\big]dt \\ &= e^{-t}\cos(s+2t)ds + e^{-t}\big[2\cos(s+2t) - \sin(s+2t)\big]dt. \end{aligned}$$

Example 4.7.5. *Let $w = xy\exp(xz)$. The total differential is:*

$$\begin{aligned} dw &= \frac{\partial}{\partial x}\big[xye^{xz}\big]dx + \frac{\partial}{\partial y}\big[xye^{xz}\big]dy + \frac{\partial}{\partial z}\big[xye^{xz}\big]dz \\ &= e^{xz}\big(y + xyz\big)dx + xe^{xz}dy + x^2ye^{xz}dz \end{aligned}$$

[21] I do give the easiest version of the implicit function theorem later in this section, but it does not really play a computational role beyond the question of existence

One popular application of differentials is to estimate error in a dependent variable due to known errors in the indendent variable. Roughly speaking, we replace the infinitesimal changes with finite changes to obtain an estimate. To be more precise, the calculation which follows actually amounts to replacing the function (which might be nonlinear in general) with its linearization. Then, the error in the linearization gives a good approximation to the actual error in the function.

Example 4.7.6. *Suppose a spherical shell of inner radius A and outer radius B has $A = 100 \pm 1$ and $B = 50 \pm 2$ where A, B are both given in terms of inches. In this case, $\triangle A = 1$ and $\triangle B = 2$ are the given uncertainties in the inner and outer radius. We can approximate the uncertainty in the volume of the shell as follows: to begin, note*

$$V = \frac{4}{3}\pi(A^3 - B^3)$$

The total differential is easily calculated,

$$dV = 4\pi(A^2 dA - B^2 dB).$$

Thus, approximately,

$$\triangle V = 4\pi(A^2\triangle A - B^2\triangle B) = 4\pi\big[100^2(1) - 50^2(2)\big] = 62831.9$$

In other words, the uncertainty in the volume is $\triangle V = \pm 6 \times 10^4\ in^3$.

I hope the method is clear from the example above. It should be emphasized this is merely an approximation technique and the real use of differentials is far more interesting.

Example 4.7.7. *Suppose $w = xyz$ then $dw = yzdx + xzdy + xydz$. On the other hand, we can solve for $z = z(x, y, w)$*

$$z = \frac{w}{xy} \qquad \Rightarrow \qquad dz = -\frac{w}{x^2y}dx - \frac{w}{xy^2}dy + \frac{1}{xy}dw. \ \star$$

If we solve $dw = yzdx + xzdy + xydz$ directly for dz we obtain:

$$dz = -\frac{z}{x}dx - \frac{z}{y}dy + \frac{1}{xy}dw \quad \star\star.$$

Are $\star$ and $\star\star$ consistent? Well, yes. Note $\frac{w}{x^2y} = \frac{xyz}{x^2y} = \frac{z}{x}$ and $\frac{w}{xy^2} = \frac{xyz}{xy^2} = \frac{z}{y}$.

Which variables are independent/dependent in the example above? It depends. In this initial portion of the example we treated x, y, z as independent whereas w was dependent. But, in the last half we treated x, y, w as independent and z was the dependent variable. Consider this, if I ask you what the value of $\frac{\partial z}{\partial x}$ is in the example above then this question is ambiguous!

$$\underbrace{\frac{\partial z}{\partial x} = 0}_{z \text{ independent of } x} \qquad \text{versus} \qquad \underbrace{\frac{\partial z}{\partial x} = \frac{-z}{x}}_{z \text{ depends on } x}$$

Obviously this sort of ambiguity is rather unpleasant. A natural solution to this trouble is simply to write a bit more when variables are used in multiple contexts. In particular,

$$\underbrace{\frac{\partial z}{\partial x}\bigg|_{y,z} = 0}_{\text{means } x,y,z \text{ independent}} \qquad \text{is different than} \qquad \underbrace{\frac{\partial z}{\partial x}\bigg|_{y,w} = \frac{-z}{x}}_{\text{means } x,y,w \text{ independent}}.$$

The key concept is that all the other independent variables are held fixed as an independent variable is partial differentiated. Holding y, z fixed as x varies means z does not change hence $\frac{\partial z}{\partial x}\big|_{y,z} = 0$. On the other hand, if we hold y, w fixed as x varies then the change in z need not be trivial; $\frac{\partial z}{\partial x}\big|_{y,w} = \frac{-z}{x}$. Let me expand on how this notation interfaces with the total differential.

Definition 4.7.8.

If w, x, y, z are variables then

$$dw = \frac{\partial w}{\partial x}\bigg|_{y,z} dx + \frac{\partial w}{\partial y}\bigg|_{x,z} dy + \frac{\partial w}{\partial z}\bigg|_{x,y} dz.$$

Alternatively,

$$dx = \frac{\partial x}{\partial w}\bigg|_{y,z} dw + \frac{\partial x}{\partial y}\bigg|_{w,z} dy + \frac{\partial x}{\partial z}\bigg|_{w,y} dz.$$

The larger idea here is that we can identify partial derivatives from the coefficients in equations of differentials. I'd say a differential equation but you might get the wrong idea... Incidentally, there is a whole theory of solving differential equations by clever use of differentials, it's called the *method of characteristics*. I have books if you are interested.

Example 4.7.9. *Suppose $w = x + y + z$ and $x + y = wz$ then calculate $\frac{\partial w}{\partial x}\big|_y$ and $\frac{\partial w}{\partial x}\big|_z$. Notice we must choose dependent and independent variables to make sense of partial derivatives in question.*

1. *suppose w, z both depend on x, y. Calculate,*

$$\frac{\partial w}{\partial x}\bigg|_y = \frac{\partial}{\partial x}\bigg|_y (x+y+z) = \frac{\partial x}{\partial x}\bigg|_y + \frac{\partial y}{\partial x}\bigg|_y + \frac{\partial z}{\partial x}\bigg|_y = 1 + 0 + \frac{\partial z}{\partial x}\bigg|_y \quad \star$$

To calculate further we need to eliminate w by substituting $w = x + y + z$ into $x + y = wz$; thus $x + y = (x + y + z)z$ hence $dx + dy = (dx + dy + dz)z + (x + y + z)dz$

$$(2z + x + y)dz = (1 - z)dx + (1 - z)dy \quad \star\star$$

Therefore,

$$dz = \frac{1-z}{2z+x+y}dx + \frac{1-z}{2z+x+y}dy = \frac{\partial z}{\partial x}\bigg|_y dx + \frac{\partial z}{\partial y}\bigg|_x dy \quad \Rightarrow \quad \frac{\partial z}{\partial x}\bigg|_y = \frac{1-z}{2z+x+y}.$$

Returning to $\star$ we derive

$$\boxed{\frac{\partial w}{\partial x}\bigg|_y = 1 + \frac{1-z}{2z+x+y}.}$$

2. *suppose w, y both depend on x, z. Calculate,*

$$\frac{\partial w}{\partial x}\bigg|_z = \frac{\partial}{\partial x}\bigg|_z (x+y+z) = \frac{\partial x}{\partial x}\bigg|_z + \frac{\partial y}{\partial x}\bigg|_z + \frac{\partial z}{\partial x}\bigg|_z = 1 + \frac{\partial y}{\partial x}\bigg|_z + 0$$

To complete this calculation we need to eliminate w as before, using $\star\star$,

$$(1 - z)dy = (2z + x + y)dz - (1 - z)dx \quad \Rightarrow \quad \frac{\partial y}{\partial x}\bigg|_z = -1.$$

Therefore,

$$\boxed{\frac{\partial w}{\partial x}\bigg|_z = 0.}$$

I hope you can begin to see how the game is played. Basically the example above generalizes the idea of implicit differentiation to several equations of many variables. This is actually a pretty important type of calculation for engineering. The study of thermodynamics is full of variables which are intermittently used as either dependent or independent variables. The so-called equation of state can be given in terms of about a dozen distinct sets of state variables.

Example 4.7.10. *The ideal gas law states that for a fixed number of particles n the pressure P, volume V and temperature T are related by $PV = nRT$ where R is a constant. Calculate,*

$$\frac{\partial P}{\partial V}\bigg|_T = \frac{\partial}{\partial V}\left[\frac{nRT}{V}\right]\bigg|_T = -\frac{nRT}{V^2},$$

$$\frac{\partial V}{\partial T}\bigg|_P = \frac{\partial}{\partial T}\left[\frac{nRT}{P}\right]\bigg|_T = \frac{nR}{P},$$

$$\frac{\partial T}{\partial P}\bigg|_V = \frac{\partial}{\partial P}\left[\frac{PV}{nR}\right]\bigg|_T = \frac{V}{nR}.$$

You might expect that $\frac{\partial P}{\partial V}\big|_T \frac{\partial V}{\partial T}\big|_P \frac{\partial T}{\partial P}\big|_V = 1$. Is it true?

$$\frac{\partial P}{\partial V}\bigg|_T \frac{\partial V}{\partial T}\bigg|_P \frac{\partial T}{\partial P}\bigg|_V = -\frac{nRT}{V^2}\cdot\frac{nR}{P}\cdot\frac{V}{nR} = \frac{-nRT}{PV} = -1.$$

This is an example where naive cancellation of partials fails.

The example above is merely a special case of a general result shown below.

Example 4.7.11. *You can repeat the example above for x, y, z constrained by $F(x, y, z) = 0$. The differential of F is*

$$dF = F_x dx + F_y dy + F_z dz$$

Solve for dx, dy or dz to derivatives of $x = x(y, z)$, $y = y(x, z)$ or $z = z(x, y)$,

$$dx = -\frac{F_y}{F_x}dy - \frac{F_z}{F_x}dz \qquad \Rightarrow \qquad \frac{\partial x}{\partial y}\bigg|_z = -\frac{F_y}{F_x} \quad \& \quad \frac{\partial x}{\partial z}\bigg|_y = -\frac{F_z}{F_x}$$

$$dy = -\frac{F_x}{F_y}dx - \frac{F_z}{F_y}dz \qquad \Rightarrow \qquad \frac{\partial y}{\partial x}\bigg|_z = -\frac{F_x}{F_y} \quad \& \quad \frac{\partial y}{\partial z}\bigg|_x = -\frac{F_z}{F_y}$$

$$dz = -\frac{F_x}{F_z}dx - \frac{F_y}{F_z}dy \qquad \Rightarrow \qquad \frac{\partial z}{\partial x}\bigg|_y = -\frac{F_x}{F_z} \quad \& \quad \frac{\partial z}{\partial y}\bigg|_x = -\frac{F_y}{F_z}$$

Notice that the factors will cancel if we choose the right triple from the list above:

$$\frac{\partial x}{\partial y}\bigg|_z \frac{\partial y}{\partial z}\bigg|_x \frac{\partial z}{\partial x}\bigg|_y = -\frac{F_y}{F_x}\cdot\frac{F_z}{F_y}\cdot\frac{F_x}{F_z} = -1.$$

The identity above reliably holds if all the partial derivatives of F are nonzero. We need $F_x \neq 0$,$F_y \neq 0$ and $F_z \neq 0$. Incidentally, and not coincidentally, the implicit function theorem[22] *needs precisely these three conditions to solve for $x = x(y, z)$, $y = y(x, z)$ and $z = z(x, y)$ respective.*

[22] covered in advanced calculus

Example 4.7.12. *Here's a different take on the example as above. We differentiate $F(x,y,z)=0$ directly and use the chain-rule: assuming $z=z(x,y)$ hence $\frac{\partial y}{\partial x}=0$ and*

$$0 = F_x + F_y \frac{\partial y}{\partial x}\bigg|_y + F_z \frac{\partial z}{\partial x}\bigg|_y \quad \Rightarrow \quad \frac{\partial z}{\partial x}\bigg|_y = -\frac{F_x}{F_z}.$$

Also, again assuming $z=z(x,y)$ once more:

$$0 = F_x \frac{\partial x}{\partial y}\bigg|_x + F_y + F_z \frac{\partial z}{\partial y}\bigg|_x \quad \Rightarrow \quad \frac{\partial z}{\partial y}\bigg|_x = -\frac{F_y}{F_z}.$$

Next, consider $y=y(x,z)$ (have $\frac{\partial z}{\partial x}|_z=0$) and differentiate $F(x,y(x,z),z)=0$ with respect to x:

$$0 = F_x + F_y \frac{\partial y}{\partial x}\bigg|_z + F_z \frac{\partial z}{\partial x}\bigg|_z \quad \Rightarrow \quad \frac{\partial y}{\partial x}\bigg|_z = -\frac{F_x}{F_y}.$$

Also, again assuming $y=y(x,z)$ once again: differentiate w.r.t. z holding x fixed,

$$0 = F_x \frac{\partial x}{\partial z}\bigg|_x + F_y \frac{\partial y}{\partial z}\bigg|_x + F_z \quad \Rightarrow \quad \frac{\partial y}{\partial z}\bigg|_x = -\frac{F_z}{F_y}.$$

We could reproduce the other two results in the previous example by differentiating $F(x(y,z),y,z)=0$ with respect to y and z.

I tend to prefer the method of differentials as given in Example 4.7.11 over the method of Example 4.7.12. In short, the method of differentials allows us to calculate without making any initial decision as to the dependence or independence of variables. Our choice of dependent or independent is made when we solve a given *equation with differentials* for a particular differential. That solution process and conclusions we draw from it are based on the demarcation of dependent and independent variables. However, the initial total differentiation calculation makes no such demarcation!

Example 4.7.13. *Consider $F(x,y,z)=x^2+y^2+z^2-1=0$. Then,*

$$\frac{\partial z}{\partial x}\bigg|_y = -\frac{F_x}{F_z} = -\frac{2x}{2z} = -\frac{x}{z}.$$

On the other hand, we could just calculate

$$dF = 2xdx + 2ydy + 2zdz = 0 \quad \Rightarrow \quad dz = -\tfrac{x}{z}dx - \tfrac{y}{z}dy \quad \Rightarrow \quad \frac{\partial z}{\partial x}\bigg|_y = -\frac{x}{z} \;\; \& \;\; \frac{\partial z}{\partial y}\bigg|_x = -\frac{y}{z}$$

You be the judge. Which method is better?

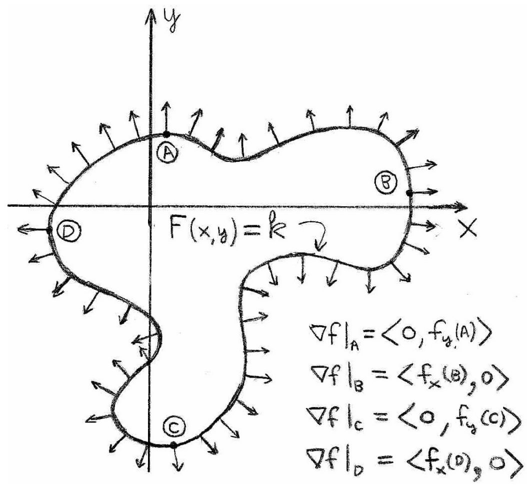
You might be curious about level curves or volumes given the interesting results above for the level surface $F(x,y,z)=0$. Consider the curve case first. Here we recover the implicit differentiation of single-variable calculus.

Example 4.7.14. *Suppose $F(x,y)=0$ then $dF=F_xdx+F_ydy=0$ and it follows that $dx=-\frac{F_y}{F_x}dy$ or $dy=-\frac{F_x}{F_y}dx$. Hence, $\frac{\partial x}{\partial y}=-\frac{F_y}{F_x}$ and $\frac{\partial y}{\partial x}=-\frac{F_x}{F_y}$. Therefore,*

$$\frac{\partial x}{\partial y}\frac{\partial y}{\partial x} = \frac{F_y}{F_x}\cdot\frac{F_x}{F_y} = 1$$

for (x,y) such that $F_x\neq 0$ and $F_y\neq 0$. The condition $F_x\neq 0$ suggests we can solve for $y=y(x)$ whereas the condition $F_y\neq 0$ suggests we can solve for $x=x(y)$.

If you pause to think about the geometry of $F(x,y)=0$ as it relates to $\nabla F = \langle F_x, F_y \rangle$ you can see why the conditions $F_x \neq 0$ and $F_y \neq 0$ are necessary.



There is no way to find y as single-valued function of x on a open set about the point where $F_y = 0$. Likewise, when $F_x = 0$ this means that there may be no way to find x as a single-valued function of y for a neighborhood centered at the point in question. If the point where $F_x = 0$ or $F_y = 0$ is on the edge of an interval then there is still hope, but the implicit function theorem does not apply. For example, $y - x^2 = 0$ for $x \geq 0$ can be solved for x as a function of y by $x = \sqrt{y}$. On the other hand, we cannot solve $y - x^2 = 0$ for x as a function of y in an open set centered about $x = 0$, each y value must return two x-values and that is not a function. Ok, enough about this.

Example 4.7.15. *Suppose* $F(x,y) = \sin x \cos y - \sin x - \cos y = 0$.

$$dF = (\cos x \cos y - \cos x)dx + (\sin y - \sin x \sin y)dy = 0$$

Thus,

$$\frac{dy}{dx} = \frac{\cos x \cos y - \cos x}{\sin x \sin y - \sin y}.$$

You might recall, we solved problems such as the one above in first semester calculus by the method of *implicit differentiation.*

Example 4.7.16. *Let C be the curve defined by* $\sin x \cos y + y^2 = x^3$. *Let* $F(x,y) = \sin x \cos y + y^2 - x^3$. *Observe that C is the solution set of* $F(x,y) = 0$. *Observe, if we follow the method of Example 4.7.14 we have*

$$\frac{dy}{dx} = -\frac{F_x}{F_y} = -\frac{\cos x \cos y - 3x^2}{-\sin x \sin y + 2y} = \frac{3x^2 - \cos x \cos y}{-\sin x \sin y + 2y}.$$

The solution set of $F(x,y,z,w) = 0$ gives a volume embedded in four-dimensional space. I invite the reader to demonstrate

$$\frac{\partial x}{\partial y}\bigg|_{z,w} \frac{\partial y}{\partial z}\bigg|_{x,w} \frac{\partial z}{\partial w}\bigg|_{x,y} \frac{\partial w}{\partial x}\bigg|_{y,z} = 1.$$

Again, this formula is only valid if all the partial derivatives of F are nontrivial at the point in question. In the next example we see why this identity holds in thermodynamics:

Example 4.7.17. *Suppose $PV = NRT$ where P is pressure, V is volume, N is the number of gas particles, R is a constant and T is the temperature. This is known as the* **ideal gas law**. *We differentiate to find:*

$$VdP + PdV = RTdN + NRdT \quad (\star)$$

Notice, if we consider $P = P(T, V, N)$ then T, V, N are independent variables and we can read from $\star$ that:

$$\frac{\partial P}{\partial T}\bigg|_{V,N} = \frac{NR}{V}$$

intuitively, we can think that fixing V, N makes $dV = dN = 0$ in $\star$. Likewise, to calculate $\frac{\partial T}{\partial V}\big|_{P,N} = \frac{P}{NR}$ we set $dP = dN = 0$ in $\star$. If this method seems suspect, you can also proceed by more traditional arguments such as:

$$\frac{\partial V}{\partial N}\bigg|_{P,T} = \frac{\partial}{\partial N}\bigg|_{P,T}\left[\frac{NRT}{P}\right] = \frac{RT}{P} \qquad \text{and} \qquad \frac{\partial N}{\partial P}\bigg|_{T,V} = \frac{\partial}{\partial P}\bigg|_{T,V}\left[\frac{PV}{RT}\right] = \frac{V}{RT}.$$

You may ask, why did I calculate these particular derivatives? Well, it was merely to point out the following:

$$\frac{\partial P}{\partial T}\bigg|_{V,N}\frac{\partial T}{\partial V}\bigg|_{P,N}\frac{\partial V}{\partial N}\bigg|_{P,T}\frac{\partial N}{\partial P}\bigg|_{T,V} = 1.$$

Example 4.7.18. *Suppose the internal energy U is a function of the pressure P, volume V and temperature T subject the ideal gas law; that is, $U = f(P, V, T)$ subject $PV = nRT$. Here n, R are constants. Consider,*

$$\frac{\partial U}{\partial P}\bigg|_V = \frac{\partial}{\partial P}\bigg|_V [f(P, V, T)] = \frac{\partial f}{\partial P} + \frac{\partial f}{\partial T}\frac{\partial T}{\partial P}\bigg|_V = \frac{\partial f}{\partial P} + \frac{\partial f}{\partial T}\frac{V}{nR}.$$

$$\frac{\partial U}{\partial T}\bigg|_V = \frac{\partial}{\partial T}\bigg|_V [f(P, V, T)] = \frac{\partial f}{\partial P}\frac{\partial P}{\partial T}\bigg|_V + \frac{\partial f}{\partial T} = \frac{\partial f}{\partial P}\frac{nR}{V} + \frac{\partial f}{\partial T}.$$

In the following pair of examples the additional notation declaring the independent variables is arguably not necessary.

Example 4.7.19. *The kinetic energy of a mass m with speed v is given by $K = \frac{1}{2}mv^2$. Note,*

$$\left(\frac{\partial K}{\partial m}\right)_v = \frac{\partial}{\partial m}\bigg|_v\left[\frac{1}{2}mv^2\right] = \frac{1}{2}v^2.$$

$$\left(\frac{\partial K}{\partial v}\right)_m = \frac{\partial}{\partial v}\bigg|_m\left[\frac{1}{2}mv^2\right] = mv \quad \Rightarrow \quad \left(\frac{\partial^2 K}{\partial v^2}\right)_m = \frac{\partial}{\partial v}\bigg|_m\Big[mv\Big] = m.$$

Therefore, we find the cute formula: $K = \left(\frac{\partial K}{\partial m}\right)_v\left(\frac{\partial^2 K}{\partial v^2}\right)_m$.

Example 4.7.20. *Suppose $x^2 + y^2 = r^2$ and $x = r\cos\theta$ and $y = r\sin\theta$.*

$$\left(\frac{\partial x}{\partial r}\right)_\theta = \frac{\partial}{\partial r}\Big[r\cos\theta\Big]\bigg|_{\theta \text{ fixed}} = \cos\theta.$$

$$\left(\frac{\partial r}{\partial x}\right)_y = \frac{\partial}{\partial x}\left[\sqrt{x^2+y^2}\right]\bigg|_{y \text{ fixed}} = \frac{x}{\sqrt{x^2+y^2}} = \frac{r\cos\theta}{r} = \cos\theta.$$

Notice, the results above are **not** *reciprocal. Curious.*

Finally, for the unsatisfied reader I remind you once more that these calculations are justified by the implicit function theorem of advanced calculus. Here is a brief discussion of the simplest version of the theorem:

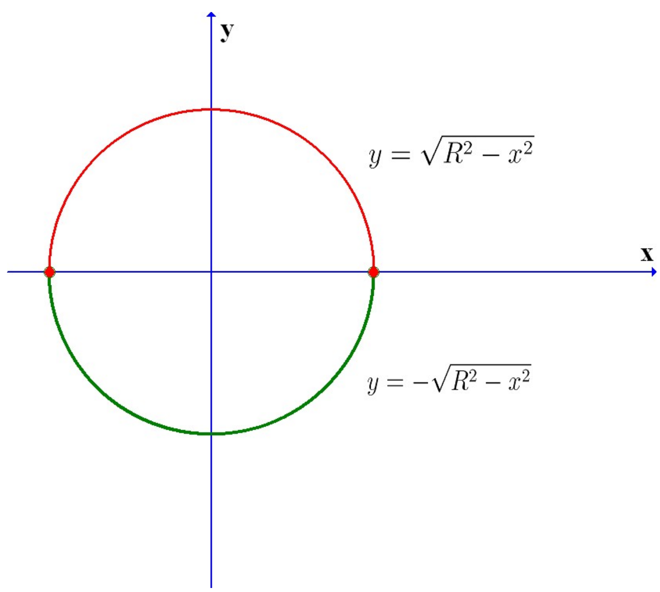
Theorem 4.7.21. *sometimes a level curve can be locally represented as the graph of a function.*

Suppose (x_o, y_o) is a point on the level curve $F(x,y) = k$ hence $F(x_o, y_o) = k$. We say the level curve $F(x,y) = k$ is **locally represented by a function** $y = f(x)$ at (x_o, y_o) iff $F(x, f(x)) = k$ for all $x \in B_\delta(x_o)$ for some $\delta > 0$. Claim: if

$$\frac{\partial F}{\partial y}(x_o, y_o) = \left(\frac{d}{dy}F(x_o, y)\right)\bigg|_{y=y_o} \neq 0$$

and the $\frac{\partial F}{\partial y}$ is continuous near (x_o, y_o) then $F(x,y) = k$ is locally represented by some function near (x_o, y_o).

The theorem above is called the **implicit function theorem** and its proof is nontrivial. Its proper statement is given in Advanced Calculus (Math 332). I'll just illustrate with the circle: $F(x,y) = x^2 + y^2 = R^2$ has $\frac{\partial F}{\partial y} = 2y$ which is continuous everywhere, however at $y = 0$ we have $\frac{\partial F}{\partial y} = 0$ which means the implicit function theorem might fail. On the circle, $y = 0$ when $x = \pm R$ which are precisely the points where we cannot write $y = f(x)$ for just one function. For any other point we may write either $y = \sqrt{R^2 - x^2}$ or $y = -\sqrt{R^2 - x^2}$ as a local solution of the level curve.



4.8 gradients in curvelinear coordinates

In this section we derive formulas for the gradient in polar, cylindrical and spherical coordinates. These formulas are important since many problems are more naturally phrased in polar, cylindrical or spherical coordinates.

4.8.1 polar coordinates

Our goal is to convert $\nabla f = \widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y}$ to polar coordinates. The unit-vectors for polar coordinates are given by

$$\begin{aligned} \widehat{r} &= \cos\theta\,\widehat{x} + \sin\theta\,\widehat{y} \\ \widehat{\theta} &= -\sin\theta\,\widehat{x} + \cos\theta\,\widehat{y}. \end{aligned} \tag{4.1}$$

We need to solve the equations above for $\widehat{x}$, $\widehat{y}$. I'll use multiplication by inverse:

$$\begin{bmatrix} \widehat{r} \\ \widehat{\theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \widehat{x} \\ \widehat{y} \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \widehat{x} \\ \widehat{y} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \widehat{r} \\ \widehat{\theta} \end{bmatrix} = \begin{bmatrix} \cos\theta\,\widehat{r} - \sin\theta\,\widehat{\theta} \\ \sin\theta\,\widehat{r} + \cos\theta\,\widehat{\theta} \end{bmatrix}$$

Therefore, $\widehat{x} = \cos\theta\,\widehat{r} - \sin\theta\,\widehat{\theta}$ and $\widehat{y} = \sin\theta\,\widehat{r} + \cos\theta\,\widehat{\theta}$. Recall that we worked out in the chain rule section that $\frac{\partial}{\partial x} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta}$ and $\frac{\partial}{\partial y} = \sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial\theta}$. Let's put these together,

$$\begin{aligned} \nabla f &= \widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} \\ &= \big(\cos\theta\,\widehat{r} - \sin\theta\,\widehat{\theta}\big)\left[\cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta}\right] f + \big(\sin\theta\,\widehat{r} + \cos\theta\,\widehat{\theta}\big)\left[\sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial\theta}\right] f \\ &= \big(\cos^2\theta + \sin^2\theta\big)\widehat{r}\frac{\partial f}{\partial r} + \widehat{\theta}\,\frac{1}{r}\big(\cos^2\theta + \sin^2\theta\big)\frac{\partial f}{\partial\theta} \\ &= \widehat{r}\frac{\partial f}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial f}{\partial\theta} \end{aligned}$$

Therefore, $\boxed{\nabla = \widehat{r}\frac{\partial}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial}{\partial\theta}}$ and $\boxed{\nabla f = \widehat{r}\frac{\partial f}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial f}{\partial\theta}.}$

Example 4.8.1. *Suppose* $f(r,\theta) = r^3$ *then* $\nabla r^3 = \widehat{r}\frac{\partial r^3}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial r^3}{\partial\theta} = 3\widehat{r}\,r^2$.

The geometry of the function above is fairly clear in polar coordinates. If we did the same calculation in cartesians then you'd face the trouble of sorting through the derivatives of $f(x,y) = (x^2+y^2)^{3/2}$ paired with sorting out the radial pattern hidden in the $\widehat{x}$, $\widehat{y}$ notation.

Example 4.8.2. *Suppose* $f(r,\theta) = \theta$ *then* $\nabla\theta = \widehat{r}\frac{\partial\theta}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial\theta}{\partial\theta} = \frac{1}{r}\widehat{\theta}$.

4.8.2 cylindrical coordinates

There is not much to do here. We follow the same calculations as in the polar case with the slight modication of adjoining a z coordinate. It's not hard to see that we'll find $\nabla f = \widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} + \widehat{z}\frac{\partial f}{\partial z}$ converts to

$$\boxed{\nabla f = \widehat{r}\frac{\partial f}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial f}{\partial\theta} + \widehat{z}\frac{\partial f}{\partial z}.}$$

Example 4.8.3. *Suppose* $f(r,\theta,z) = rz\theta$ *then we calculate,*

$$\nabla f = \widehat{r}\frac{\partial(rz\theta)}{\partial r} + \widehat{\theta}\,\frac{1}{r}\frac{\partial(rz\theta)}{\partial\theta} + \widehat{z}\frac{\partial(rz\theta)}{\partial z} = z\theta\,\widehat{r} + z\widehat{\theta} + r\theta\,\widehat{z}.$$

4.8.3 spherical coordinates

We could derive the formula for ∇f in spherical coordinates in the same way as we did for polar and cylindrical coordinates. However, I take a different approach to illustrate a few calculational techniques. The basic observation is this: ∇f is a vector field and we can write it as a sum of the spherical unit-vector fields at each point in space;

$$\nabla f = (\nabla f \bullet \widehat{\rho})\,\widehat{\rho} + (\nabla f \bullet \widehat{\phi})\,\widehat{\phi} + (\nabla f \bullet \widehat{\theta})\,\widehat{\theta}$$

Hence the problem reduces to converting $\nabla f \bullet \widehat{\rho}$, $\nabla f \bullet \widehat{\phi}$ and $\nabla f \bullet \widehat{\theta}$ to spherical coordinates. Recall that unit vectors in the direction of increasing $\rho,\ \phi,\ \theta$ by $\widehat{\rho},\ \widehat{\phi},\ \widehat{\theta}$ are given by:

$$\begin{aligned}
\widehat{\rho} &= \sin(\phi)\cos(\theta)\,\widehat{x} + \sin(\phi)\sin(\theta)\,\widehat{y} + \cos(\phi)\,\widehat{z} \\
\widehat{\phi} &= \cos(\phi)\cos(\theta)\,\widehat{x} + \cos(\phi)\sin(\theta)\,\widehat{y} - \sin(\phi)\,\widehat{z} \\
\widehat{\theta} &= -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}.
\end{aligned} \tag{4.2}$$

We calculate: (remember $x = \rho\cos\theta\sin\phi$, $y = \rho\sin\theta\sin\phi$ and $z = \rho\cos\phi$ in order to understand the chain rule calculation below)

$$\begin{aligned}
\nabla f \bullet \widehat{\rho} &= \left(\widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} + \widehat{z}\frac{\partial f}{\partial z}\right) \bullet \left(\sin(\phi)\cos(\theta)\,\widehat{x} + \sin(\phi)\sin(\theta)\,\widehat{y} + \cos(\phi)\,\widehat{z}\right) \\
&= \sin(\phi)\cos(\theta)\frac{\partial f}{\partial x} + \sin(\phi)\sin(\theta)\frac{\partial f}{\partial y} + \cos(\phi)\frac{\partial f}{\partial z} \\
&= \frac{\partial x}{\partial \rho}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial \rho}\frac{\partial f}{\partial y} + \frac{\partial z}{\partial \rho}\frac{\partial f}{\partial z} \\
&= \frac{\partial f}{\partial \rho}
\end{aligned}$$

Continuing, calculate the ϕ-component of ∇f

$$\begin{aligned}
\nabla f \bullet \widehat{\phi} &= \left(\widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} + \widehat{z}\frac{\partial f}{\partial z}\right) \bullet \left(\cos(\phi)\cos(\theta)\,\widehat{x} + \cos(\phi)\sin(\theta)\,\widehat{y} - \sin(\phi)\,\widehat{z}\right) \\
&= \cos(\phi)\cos(\theta)\frac{\partial f}{\partial x} + \cos(\phi)\sin(\theta)\frac{\partial f}{\partial y} - \sin(\phi)\frac{\partial f}{\partial z} \\
&= \frac{1}{\rho}\frac{\partial x}{\partial \phi}\frac{\partial f}{\partial x} + \frac{1}{\rho}\frac{\partial y}{\partial \phi}\frac{\partial f}{\partial y} + \frac{1}{\rho}\frac{\partial z}{\partial \phi}\frac{\partial f}{\partial z} \\
&= \frac{1}{\rho}\frac{\partial f}{\partial \phi}
\end{aligned}$$

One more component to go:

$$\begin{aligned}
\nabla f \bullet \widehat{\theta} &= \left(\widehat{x}\frac{\partial f}{\partial x} + \widehat{y}\frac{\partial f}{\partial y} + \widehat{z}\frac{\partial f}{\partial z}\right) \bullet \left(-\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}\right) \\
&= -\sin(\theta)\frac{\partial f}{\partial x} + \cos(\theta)\frac{\partial f}{\partial y} \\
&= \frac{-\rho\sin(\phi)\sin(\theta)}{\rho\sin(\phi)}\frac{\partial f}{\partial x} + \frac{\rho\sin(\phi)\cos(\theta)}{\rho\sin(\phi)}\frac{\partial f}{\partial y} \\
&= \frac{1}{\rho\sin(\phi)}\frac{\partial x}{\partial \theta}\frac{\partial f}{\partial x} + \frac{1}{\rho\sin(\phi)}\frac{\partial y}{\partial \theta}\frac{\partial f}{\partial y} + \frac{1}{\rho\sin(\phi)}\frac{\partial z}{\partial \theta}\frac{\partial f}{\partial z} \\
&= \frac{1}{\rho\sin(\phi)}\frac{\partial f}{\partial \theta}.
\end{aligned}$$

Therefore, we find

$$\boxed{\nabla f \;=\; \widehat{\rho}\,\frac{\partial f}{\partial \rho} + \widehat{\phi}\,\frac{1}{\rho}\frac{\partial f}{\partial \phi} + \widehat{\theta}\,\frac{1}{\rho\sin(\phi)}\frac{\partial f}{\partial \theta}.}$$

Spherical coordinate formulas are important for studying applications with spherical symmetry.

Example 4.8.4. *In spherical coordinates the potential due to a point charge is simply $V(\rho,\phi,\theta) = \frac{1}{\rho}$. The theory of electrostatics says this generates an electric field $\vec{E} = -\nabla V$. We find the field easily using our formula for the gradient in sphericals,*

$$\vec{E} = \nabla V = \widehat{\rho}\,\frac{\partial V}{\partial \rho} + \widehat{\phi}\,\frac{1}{\rho}\underbrace{\frac{\partial V}{\partial \phi}}_{zero} + \widehat{\theta}\,\frac{1}{\rho\sin(\phi)}\underbrace{\frac{\partial V}{\partial \theta}}_{zero} = -\frac{1}{\rho^2}\widehat{\rho}.$$

More generally, the spherical gradient formula allows us to evaluate how a given function changes in spherical coordinates.

Example 4.8.5. *Suppose $f(x,y,z) = y/x$. To find how f changes in spherical coordinates we convert to sphericals[23]; $f(\rho,\phi,\theta) = \tan\theta$. It is clear that f is constant in ρ and ϕ. In particular,*

$$\nabla f = \widehat{\rho}\,\frac{\partial}{\partial \rho}\big[\tan\theta\big] + \widehat{\phi}\,\frac{1}{\rho}\frac{\partial}{\partial \phi}\big[\tan\theta\big] + \widehat{\theta}\,\frac{1}{\rho\sin(\phi)}\frac{\partial}{\partial \theta}\big[\tan\theta\big] = \frac{\sec^2(\theta)}{\rho\sin(\phi)}\,\widehat{\theta}.$$

This means that $\tan\theta$ increases at the $\frac{\sec^2(\theta)}{\rho\sin(\phi)}$ rate in the $\widehat{\theta}$-direction.

Remark 4.8.6.

> There are other slicker methods to derive the formulas in this section. My goal here is not to be particularly clever. I merely wish to obtain these formulas for our future use and hopefully to illustrate once more the structure of vector algebra and the chain rules of multivariate calculus. If you'd like to know about alternate ways to derive these formulas I have a source or two for further reading.

4.9 Problems

Problem 76 Suppose $f(x,y) = x\cosh(x+y^2)$. Calculate f_x and f_y

Problem 77 Calculate ∇f for each of the functions below:

1. $f(x,y) = 2x + 3y$
2. $f(x,y) = exp(-x^2 + 2x - y^2)$
3. $f(x,y) = \sin(x+y)$

Problem 78 What is the rate of change in the functions given in Problem 77 at the point $(1,3)$ in the direction of the vector $\langle 1,-1\rangle$.

[23]this is a slight abuse of notation, the function is not really f with this modication. Instead, we should perhaps denote it $\tilde{f}$ where to be technical $\tilde{f}(\rho,\phi,\theta) = f(\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$. The underlying motivation for this abuse is the idea that f is really an object which exists w/o regard to the particulars of the coordinate system we use, so it's appropriate to use the same letter for both the cartesian and spherical. Well, perhaps, but they are not the same actual function. This is similar to the problem of the sine function. $\sin(90)$ and $\sin(\pi/2)$ are usually both taken to be 1 but this is an overloading of the symbol sin. The degree-based sine function and the radian-based sine function are in fact different functions on $\mathbb{R}$.

Problem 79 Again, concerning the functions given in Problem 77, in what directions are the functions locally constant at the point $(1,3)$? (give answers in terms of unit-direction vectors)?

Problem 80 Find parametrizations for the normal lines through $(1,3)$ for the functions given in Problem 77. These lines will be perpendicular to the level curve of f through $(1,3)$.

Problem 81 find all critical points of (use integer notation for (b.) since there are many answers!)

1. $f(x,y) = exp(-x^2 + 2x - y^2)$
2. $f(x,y) = \sin(x+y)$

Problem 82 Suppose the temperature T is a function of the coordinates x, y in a large plane of battle. Furthermore, suppose the enemy ninja is carefully building a large attack by molding chakra over some time. During the preparation of the attack the enemy is vulnerable to your attack. Knowing this he has obscured the field of vision with multiple smoke bombs. However, the mass of energy building actually heats the ground. Fortunately one of your ninja skills is temperature sensitivity. You extrapolate from the temperature of the ground near your location that the temperature function has the form $T(x,y) = 50 + x - y$. In what direction should you attack?

Problem 83 [use of technology to solve algebraic and/or transcendental equation that the problem suggests] The temperature in an air conditioned room is set at 65. A ninja with expert ocular jutsu disguises himself in plain sight by bending light near him with his art. However, his art does not extend to the infared spectrum and his body heat leaves a signature variation in the otherwise constant room temperature. In particular,

$$T(x,y,z) = 33exp\big[-(x-3)^2 - (y-4)^2 - (z-1)^2\big] + 65.$$

Shino searches for the cloaked ninja by sending insect scouts which are capable of sensing a change in temperature as minute as 0.1 degree per meter. How close do the scout insects have to get before they sense the hidden ninja? (also, where is the hidden ninja and what is his body temperature on the basis of the given T which is in meters and degrees Farenheight)

Problem 84 Suppose A, B, C are constants. Calculate all nonzero partial derivatives for $z = Ax^2 + Bxy + Cy^2$.

Problem 85 Assume g, h are differentiable functions on $\mathbb{R}$. Calculate f_x and f_{xy} for

$$f(x,y) = xg(x^2 + y^2) + h(x)$$

Problem 86 The ideal gas law states that $P = kT/V$ for a volume V of gas at temperature T and pressure P. Show that

$$V\frac{\partial P}{\partial V} = -P \qquad \text{and} \qquad V\frac{\partial P}{\partial V} + T\frac{\partial P}{\partial T} = 0$$

Problem 87 The operation of $\nabla = \sum_{j=1}^{n} \widehat{x}_j \frac{\partial}{\partial x_j}$ takes in a function f with domain in $\mathbb{R}^n$ and creates a vector field ∇f which assigns an n-vector at each point in $\mathbb{R}^n$. This operation has several nice properties to prove here: for differentiable real-valued functions f, g and constant c,

(a.) $\nabla(f+g) = \nabla f + \nabla g$

(b.) $\nabla(cf) = c\nabla f$

(c.) $\nabla(fg) = g\nabla f + f\nabla g$

Problem 88 Set-aside the polar coordinate notation. Define in $\mathbb{R}^n$ the spherical radius by

$$r = \sqrt{x_1^2 + x_2^2 + \cdots x_n^2} = ||\vec{r}|| = \sqrt{\vec{r} \bullet \vec{r}}$$

Show that:

(a.) $\nabla r = \frac{1}{r}\vec{r}$.

(b.) $\nabla\left(\frac{1}{r}\right) = -\frac{1}{r^3}\vec{r}$.

Problem 89 A "power" rule? Show $\nabla r^n = nr^{n-1}\widehat{r}$.

Problem 90 Find the gradient of:

1. $f(x,y,z,w) = x + y^2 + z^3 + w^4$
2. $f(x,y,z) = xyz\ln(x+y+z)$

Problem 91 Suppose Paccun speeds towards the base of a valley with paraboloid shape given by the equation $z = x^2 + 3y^2$. What is the direction of steepest **descent** at the point $(1,1,4)$?

Problem 92 Let $f(x,y) = x^3 - xy$. Let $A = (0,1)$ and $B = (1,3)$. Find a point C on the line-segment $\overline{AB}$ such that $f(B) - f(A) = \nabla f(C) \bullet (B - A)$. (*this illustrates a mean-value theorem which is known for real-valued functions of several variables*)

Problem 93 Suppose $\vec{F}(x,y,z) = \langle 2xy^2, 2x^2y, 3\rangle$. What scalar function f yields $\vec{F}$ as a gradient vector field? Find f such that $\nabla f = \vec{F}$.
(here we have to work backwards, write down what you want and guess, by the way, the function $-f$ is the pontial energy function for the force field $\vec{F}$.)

Problem 94 A basic wave equation is

$$\frac{\partial^2 y}{\partial t^2} = v^2\frac{\partial^2 y}{\partial x^2}.$$

The wave can be viewed as a graph in the xy-plane which animates with time t. Calculate appropriate partial derivatives to test if the functions give a solution to the given wave equation. ($c_1, c_2, v \neq 0$ are constants)

(a.) $y(x,t) = x - vt$

(b.) $y(x,t) = c_1\sin(x - vt) + c_2\cos(x - vt)$

(c.) $y(x,t) = \sin(vt)\sin(x)$

Problem 95 Show that $u = e^x\cos(y)$ and $v = e^x\sin(y)$ solves Laplaces equation $\Phi_{xx} + \Phi_{yy} = 0$.

Problem 96 Find the best linear approximation of each object at the given point. Also, write either and equation or a parametrization of the tangent space in each case ("space" could mean line, surface, space curve or other things. . .)

(a.) $f(x) = x^2$ at $a = 2$

(b.) $f(x,y) = x^2 - 2xy$ at $(3,4)$

(c.) $\vec{r}(t) = \langle t, 3, t^2 2^t \rangle$ at $t = 0$

Problem 97 Given that $x = u^2 + v^2$ and $y = 3uv$ and $z = 3\sin(uv)$ and $w = ze^{xy}$ calculate w_u and w_v and finally w_z.

Problem 98 Calculate the Jacobian matrix of $\vec{r}(u, v) = \langle u^2 + v^2,\ 3uv,\ 3\sin(uv)\rangle$ and that of $f(x, y, z) = ze^{xy}$. Multiply these matrices and identify how this relates to the previous problem.

Problem 99 Suppose $z = xy$ and $x = \sinh[g(t)]$ and $y = h(t^2)$ for some differentiable functions g, h. Calculate dz/dt by the chain rule(s).

Problem 100 Suppose a car speeds over a hill with equation $x^2 + 4y^2 + z^2 = 1$. If at the point with $x = 1m$ and $y = 0.5m$ the car has an x-velocity of $10m/s$ and a y-velocity of $20m/s$ then what z-velocity does the car have? (assume the car stays on hill)

Problem 101 Suppose S is the surface defined by $F(x, y, z) = xyz = 1$. Find the equation of the tangent plane and the parametrization of the normal line through $(1, 1, 1)$.

Problem 102 Find a parametrization $\vec{X}$ of S from the previous problem which provides a patch in the locality of $(1, 1, 1)$. Use α, β for your parameters and find the normal-vector field $\vec{N}(\alpha, \beta)$ by computing $\vec{N}(\alpha, \beta) = \frac{\partial \vec{X}}{\partial \alpha} \times \frac{\partial \vec{X}}{\partial \beta}$. Do you obtain the same normal vector at $(1, 1, 1)$ with this patch?

Problem 103 Label the solution set of $x^2 = y - z^2$ as M.

(a.) present M as a level-surface for some function F. Explicitly state the formula for F. Find the normal vector field on M.

(b.) parametrize M and once more find the normal vector field. This time find $\vec{N}$ explicitly in terms of your chosen parameters.

Problem 104 Suppose a level surface $G(x, y, z) = 2$ has $\nabla G(x, y, z) = \langle x, y, z\rangle$ and another level surface $F(x, y, z) = 42$ with $\nabla F(x, y, z) = \langle 1, x, 3\rangle$. Suppose these surfaces intersect along some curve and at the point (a, b, c) the curve of intersection has tangent line with direction vector colinear to $\langle 0, 0, 10\rangle$. Find (a, b, c).

Problem 105 Consider the line-segment $\mathcal{L} = \overline{PQ}$ where $P = (1, 2, 3)$ to $Q = (5, 0, -1)$.

(a.) describe $\mathcal{L}$ parametrically as a path from $t = 0$ at P to $t = 1$ at Q.

(b.) describe $\mathcal{L}$ parametrically as a path from $s = 0$ at Q to $s = 6$ at P.

(c.) describe $\mathcal{L}$ as a graph. In particular, find $h(x)$ and $g(x)$ such that $f(x) = \langle g(x), h(x)\rangle$ and $graph(f) = \{(x, f(x)) \mid x \in dom(f)\ \}$.

(d.) describe $\mathcal{L}$ as a level-curve. In particular, find F such that $\mathbb{R}^3 \xrightarrow{F} \mathbb{R}^2$ and $\mathcal{L} = F^{-1}\{(0, 0)\}$.

Remark: personally, I view the last two parts of the previous problem as less natural than the parametric presentation. It is in fact possible to present lines, surfaces, volumes etc... as either graphs, level-sets or as parametrized objects. Which to use depends on the context.

Problem 106 Ohms' Law says that $V = IR$ where V is the voltage of a battery which delivers a current I to a resistor R. As the current flows the battery will wear down and the voltage will drop. On the other hand, as the resitor heats-up the resistance will increase. Given that $R = 600$ ohms and $I = 0.04$ amp, if the resistance is increasing at a rate of 0.5 ohm/sec and the voltage is dropping at 0.01 volt/sec then what is the rate of change in the current I at this time.

Problem 107 Suppose that the temperature T in the xy-plane changes according to

$$\frac{\partial T}{\partial x} = 8x - 4y \qquad \& \qquad \frac{\partial T}{\partial y} = 8y - 4x.$$

Find the maximum and minimum temperatures of T on the unit circle $x^2+y^2 = 1$. To achieve this goal you should parametrize the the circle by $x = \cos t$ and $y = \sin t$ and calculate dT/dt and d^2T/dt^2 by the chain-rule. (you have no other option since the formula for T is not given!)

Problem 108 Suppose $f(u, v, w)$ is the formula for a differentiable f and $u = x - y$, $v = y - z$ and $w = z - x$. Show that $f_x + f_y + f_z = 0$.

Problem 109 Suppose $w = xy^2 + z^3$ and $x = f(u, v)$, $y = g(u, v)$ and $z = h(u, v)$ where f, g, h are differentiable functions. If $f(1, 2) = 2$ and $g(1, 2) = 3$ and $h(1, 2) = 4$ and $g_v(1, 2) = 0$, $h_v(1, 2) = 7$ and $f_v(1, 2) = 42$ calculate $\frac{\partial w}{\partial v}(1, 2)$.

Problem 110 Suppose $f(x, y) = x^2 - 3xy + 5$. A theorem states that for twice continuously differentiable f the error $E(x, y) = f(x, y) - L(x, y)$ in the linearization for each (x, y) in some rectangle R centered at (x_o, y_o) is bounded by

$$|E(x, y) \leq \frac{1}{2}M\bigg(|x - x_o + |y - y_o|\bigg)^2$$

where M bounds $|f_{xx}|$, $|f_{yy}|$ and $|f_{xy}|$ on R. In other words, if you can find such an M to bound the second order partial derivatives then the error is given by the inequality above.

(a.) find the linearization of f at $(2, 1)$.

(b.) bound the error $E(x, y)$ for the square $[1.9, 2.1] \times [0.9, 1.1]$

Remark: not that I plan to derive it this semester, but this is a consequence of the error estimate for the single-variable Taylor series as it applies to the construction of the multivariate Taylor expansion. The multivariate Taylor expansion derives from the chain-rule and Taylor's theorem from calculus II.

Problem 111 The area of a triangle is given by $A = \frac{1}{2}ab\sin\gamma$ where a, b are the lengths of two sides which have angle γ between them. Suppose that $\gamma = \pi/3 \pm 0.01$ and $a = 150 \pm 1$ ft and $b = 200 \pm 1$ ft. Find the corresponding uncertainty in the area. (leave answer as $A \pm \delta$ where the δ is calculated from the differential of A)

Problem 112 Suppose you know $x = 3 \pm 0.01$ and $y = 4 \pm 0.01$ what are the corresponding polar coordinate ranges.

Problem 113 The kinetic energy in 2D-problem with cartesian coordinates is given by $K = \frac{1}{2}mv^2$ or explicitly in terms of the x, y velocities $\dot{x}, \dot{y}$ we have $K = \frac{1}{2}m(\dot{x}^2+\dot{y}^2)$. Calculate the formula for K in terms of polar coordinates r, θ and their velocities $\dot{r}, \dot{\theta}$.

Problem 114 Suppose $w = x^2 + y - z + \sin(t)$ and $x + y = t$. Calculate the following constrained partial derivatives:

$$(\mathbf{a}.) \left(\frac{\partial w}{\partial y}\right)_{x,z} \qquad (\mathbf{b}.) \left(\frac{\partial w}{\partial y}\right)_{z,t} \qquad (\mathbf{c}.) \left(\frac{\partial w}{\partial z}\right)_{x,y} \qquad (\mathbf{d}.) \left(\frac{\partial w}{\partial z}\right)_{y,t}$$

Problem 115 Suppose $\vec{F} = \rho^2\,\widehat{\rho} + \frac{1}{\rho}\theta^3 \sin(\phi)\,\widehat{\phi}$ find f such that $\nabla f = \vec{F}$.

Problem 116 Suppose $f(x,y,z) = (x^2+y^2+z^2)\tan^{-1}(y/x) + \cos^{-1}(z/\sqrt{x^2+y^2+z^2})$. Calculate ∇f.

Problem 117 Suppose a formula for $f(x,y)$ is given. Furthermore, suppose you are asked to calculate $\frac{\partial f}{\partial r}$ where $r = \sqrt{x^2+y^2}$. Techincally, this question is ambiguous. Why? Because you need to know what other variable besides r is to be used in concert with r. If we use the usual polar coordinates then $\tan(\theta) = \frac{y}{x}$ and all is well. We adopt the following (standard) interpretation:

$$f_r = \frac{\partial f}{\partial r} = \frac{\partial}{\partial r}\Big[f(r\cos\theta, r\sin\theta)\Big] = \frac{\partial f}{\partial x}\bigg|_{(r\cos\theta, r\sin\theta)} \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\bigg|_{(r\cos\theta, r\sin\theta)} \frac{\partial y}{\partial r}$$

In other words, we define the derivative of f with respect to some curvelinear coordinate by the derivative of $f \circ \vec{T}$ where $\vec{T} : \mathbb{R}^2 \to \mathbb{R}^2$ is the coordinate transformation to which the curvelinear coordinate belongs. Denoting $\vec{T}(r,\theta) = (r\cos\theta, r\sin\theta)$ we define,

$$f_\theta = \frac{\partial f}{\partial \theta} = \frac{\partial}{\partial \theta}\Big[(f \circ \vec{T})(r,\theta)\Big] = \frac{\partial f}{\partial x}\bigg|_{(r\cos\theta, r\sin\theta)} \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\bigg|_{(r\cos\theta, r\sin\theta)} \frac{\partial y}{\partial \theta}$$

A short calculation reveals that:

$$f_r = f_x \cos\theta + f_y \sin\theta \qquad \& \qquad f_\theta = -f_x r\sin\theta + f_y r\cos\theta$$

Solve the equations above for f_x and f_y.

Problem 118 Recall that $\nabla \bullet \nabla\Phi = \nabla^2\Phi = 0$ is called Laplace's equation. In cartesian coordinates, in two dimensions, Laplace's equation reads $\Phi_{xx} + \Phi_{yy} = 0$. Show that Laplace's equation in polar coordinates is

$$\frac{\partial^2\Phi}{\partial r^2} + \frac{1}{r}\frac{\partial\Phi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\Phi}{\partial\theta^2} = 0.$$

(yes, most of this is in the notes, but I'd like to see the rest of the details)

Problem 119 Given the potential functions Φ below show they are solutions to Laplace's equations either via computation in cartesian coordinates or polar coordinates.

(a.) $\Phi(x,y) = \sqrt{x^2+y^2}$

(b.) $\Phi(x,y) = \tan^{-1}(y/x)$

(c.) $\Phi(r,\theta) = r^2\cos\theta\sin\theta$

Problem 120 Define hyperbolic coordinates h, ϕ by the following equations

$$x = h\cosh\phi \qquad \& \qquad y = h\sinh(\phi)$$

Let's study these coordinates by answering the following:

(a.) solve the equations above for h and ϕ.

(b.) Find hyperbolic coordinates for $(1,1),(-1,1),(-1,-1)$ and $(1,-1)$. Write a diagram which explains the signs for h and ϕ in each quadrant.

(c.) What do are level curves of h ?

(d.) What are level curves of ϕ?

Problem 121 Continuing the study from the previous problem,

(a.) find functions A, B, C, D of hyperbolic coordinates h, ϕ which give unit-vectors

$$\widehat{h} = A\,\widehat{x} + B\,\widehat{y} \qquad \& \qquad \widehat{\phi} = C\,\widehat{x} + D\,\widehat{y}$$

(b.) derive a formula for ∇f in terms of hyperbolic coordinate derivatives, however express it in terms of $\widehat{x}$ and $\widehat{y}$. That is find $\nabla f = \langle f_x, f_y\rangle$ but express f_x and f_y in terms of the hyperbolic coordinates and derivatives.

(c.) derive ∇f is purely hyperbolic notation: that is find E, F such that

$$\nabla f = E\widehat{h} + F\widehat{\phi}.$$

partial derivative computation is fun... but, what does it mean? We explore this question in the pair of problems below

Problem 122 Let $(a,b) \in \mathbb{R}^2$ be a particular point. Explain geometrically the meaning of the equations given below:

(a.) $\frac{\partial f}{\partial r}(a,b) = -1$

(b.) $\frac{\partial f}{\partial \theta}(a,b) = 1$

(c.) $\frac{\partial f}{\partial \phi}(a,b) = 0$ (same notation as in previous pair of problems)

As an example: $\frac{\partial f}{\partial x}(a,b) = 0$ indicates that the function stays constant along the line passing through (a,b) on which y is held fixed at value b (parametrically f is constant along the path $t \to (a+ta, b)$ near $t=0$).

Problem 123 Joshua asked if $\frac{\partial}{\partial(xy)}$ had meaning. I would say yes. In fact, it has many meanings.

(a.) Define $u = xy$ and $v = y/x$ for $(x,y) \in (0,\infty)^2$. Find inverse transformations. That is, solve for $x = x(u,v)$ and $y = y(u,v)$ in view of the definition just given and comment on the level curves of u, v (if they are a named curve then name them).

(b.) explain what $\frac{\partial f}{\partial u} = 0$ means for a function f at a given point. (use meaning suggested from part (a.))

(c.) Define $u = xy$ and $w = y$ for $(x,y) \in (0,\infty)^2$. Find inverse transformations. That is, solve for $x = x(u,w)$ and $y = y(u,w)$ in view of the definition just given and comment on the level curves of u, w (if they are a named curve then name them).

(d.) explain what $\frac{\partial f}{\partial u} = 0$ means for a function f at a given point. (use meaning suggested from part (c.)) (it is **not** a directional derivative in the traditional sense of the term.

the previous problem is important for applications. Think about this, what variable is most interesting to your model? It is important to be able to write the equations which describe the model in terms of those variables. On the other hand, it may be simple to express the physics of the model in cartesian coordinates. Hopefully these problems give you an idea about how to translate from one formalism to the other and vice-versa.

Chapter 5

optimization

The problem of optimizing a function of several variables is in many ways similar to the problem of optimization in single variable calculus. There is a fermat-type theorem; extrema are found at crtical points if anywhere. Also, there is an analogue of the closed interval method for continuous functions on some closed domain; the absolute extrema either occur at a critical point in the interior or somewhere on the boundary. However, there is no simple analogue of the first derivative test. In higher dimensions we can approach a potential extremum in infinitely many directions, in one-dimension you just have left and right approaches. The second derivative test does have a fairly simple analogue for functions of several variables. To understand the multivariate second derivative test we must first understand multivariate Taylor series. Once those are understood the second derivative test is easy to motivate. Not all instructors emphasize this point, but even in the single variable case the Taylor series expansion is probably the best tool to really understand the second derivative test. As a starting point for this chapter I assume you know what a Taylor series is, have memorized all the standard expansions and tricks, and are ready and willing to think. To the more mathematical reader, I apologize for the lack of rigor. I will not even discuss finer points of convergence or divergence. The theory of multivariate series is found in many good advanced calculus texts. I'll break from my usual format and offer the main terms in this overview:

Definition 5.0.1.

A function $f : \text{dom}(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ has a **local maximum** at (a, b) if $f(x, y) \leq f(a, b)$ for all (x, y) in some disk centered at (a, b). Likewise, $f(a, b)$ is a **local minimum** if there exists some disk D centered on (a, b) for which $(x, y) \in D$ implies $f(x, y) \geq f(a, b)$. If $S \subseteq \text{dom}(f)$ and $f(a, b) \geq f(x, y)$ for all $(x, y) \in S$ then $f(a, b)$ is a maximum of f on S. Similarly, if $S \subseteq \text{dom}(f)$ and $f(a, b) \leq f(x, y)$ for all $(x, y) \in S$ then $f(a, b)$ is a minimum of f on S. If f has a maximum or minimum on $\text{dom}(f)$ then f is said to have a **global** maximum or minimum. Maximum and minimum values are collectively called **extreme values**.

Given the terms above, let me briefly outline the chapter. In the first section we study Lagrange Multipliers which gives us a method to find extrema on constraint curves or surfaces. These constraint curves or surfaces can often be thought of as boundaries of areas or volumes. The problem of finding extrema in the interior of areas or volumes is revealed by the theory of critical points. In short, the structure of the quadratic terms in the multivariate Taylor series expanded about a critical point classify the type as maximum, minimum or saddle. However, degenerate cases such as troughs or constants require a more delicate analysis. In the second section we introduce multivariate power series and in the third section we study the second derivative test for functions of two variables. Our derivation of the second derivative test is partially based in a Lagrange multiplier argument to circles of arbitrary radii about the critical point. See Theorem 5.1.12 and 5.1.13. This allows us to avoid some linear algebra. However, we describe in Subsection 5.1.4 how the theory of quadratic forms in linear algebra allows vast generalization of our two-dimensional result. Finally, in the fourth section we discuss the closed set test which unifies the efforts of the first two sections into a common goal.

5.1 lagrange multipliers

The method of Lagrange Multipliers states the following: for smooth functions f, g with non-vanishing gradients[1] on $g = 0$

If $f(\vec{p})$ is a maximum/minimum of f on the level-set $g = 0$ then for some constant λ

$$\boxed{\nabla f = \lambda \nabla g.}$$

Notice that the method does not provide the existence of maximums or minimums of the **objective function** f on the constraint equation $g = 0$. If no max/min for f exists on $g = 0$ then it may be possible to solve the Lagrange multiplier equation $\nabla f = \lambda \nabla g$ and find points which do not provide extrema for f on $g = 0$. We'll see examples that show that when $g = 0$ is a closed and bounded set then the extrema for f do exist. We return to this subtle points in the examples which follow the proof. Finally, I apply the method to a whole class of functions on $\mathbb{R}^2$. The last subsection is difficult but it lays the foundation for the two-dimensional second derivative test we derive later in this chapter. The logic of the test rests on a combination of the final subsection in this section and the multivariate taylor series discussed in the next section.

5.1.1 proof of the method

Proof: **($n = 2$ case)** Suppose f has a local maximum at (x_o, y_o) on the level curve $g(x, y) = 0$. Let I be an interval containing zero and $\vec{r} : I \subseteq \mathbb{R} \to \mathbb{R}^2$ be a smooth path parametrizing $g(x, y) = 0$ with $\vec{r}(0) = (x_o, y_o)$. This means $g(\vec{r}(t)) = 0$ for all $t \in I$. It is intuitively clear that the function of one-variable $h = f \circ \vec{r}$ has a maximum at $t = 0$. Therefore, by Fermat's theorem from single-variable calculus, $h'(0) = 0$. But, h is a composite function so the multivariate chain rule applies. In particular,

$$\frac{d}{dt}\Big[f(\vec{r}(t))\Big]\Bigg|_{t=0} = \nabla f(\vec{r}(0)) \bullet \frac{d\vec{r}}{dt}(0) = 0.$$

But, we also know $g(\vec{r}(t)) = 0$ for all $t \in I$ hence

$$\frac{d}{dt}\Big[g(\vec{r}(t))\Big] = \nabla g(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt}(t) = 0.$$

for each $t \in I$. In particular, put $t = 0$ in the equation above to find

$$\frac{d}{dt}\Big[g(\vec{r}(t))\Big]\Bigg|_{t=0} = \nabla g(\vec{r}(0)) \bullet \frac{d\vec{r}}{dt}(0) = 0.$$

We find that both $\nabla f(x_o, y_o)$ and $\nabla g(x_o, y_o)$ are orthogonal to the tangent vector $\frac{d\vec{r}}{dt}(0)$. In two dimensions geometry forces us to conclude that $\nabla f(x_o, y_o)$ and $\nabla g(x_o, y_o)$ are colinear[2] thus there

[1] this means there are no critical points for f and g on the region of interest

[2] I assume $\nabla f(x_o, y_o) \neq 0$ and $\nabla g(x_o, y_o) \neq 0$ as mentioned at the outset of this section.

exists some nonzero constant λ such that $\nabla f(x_o, y_o) = \lambda \nabla g(x_o, y_o)$. $\square$

Proof: ($n = 3$ case) Suppose f has a local maximum at (x_o, y_o, z_o) on the level surface $g(x, y, z) = 0$. Let I be an interval containing zero and $\vec{r} : I \subseteq \mathbb{R} \to \mathbb{R}^3$ be a smooth path on the level surface $g(x, y, z) = 0$ with $\vec{r}(0) = (x_o, y_o, z_o)$. This means $g(\vec{r}(t)) = 0$ for all $t \in I$. It is intuitively clear that the function of one-variable $h = f \circ \vec{r}$ has a maximum at $t = 0$. Therefore, by Fermat's theorem from single-variable calculus, $h'(0) = 0$. But, h is a composite function so the multivariate chain rule applies. In particular,

$$\frac{d}{dt}\Big[f(\vec{r}(t))\Big]\Big|_{t=0} = \nabla f(\vec{r}(0)) \bullet \frac{d\vec{r}}{dt}(0) = 0.$$

But, we also know $g(\vec{r}(t)) = 0$ for all $t \in I$ hence

$$\frac{d}{dt}\Big[g(\vec{r}(t))\Big] = \nabla g(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt}(t) = 0.$$

for each $t \in I$. In particular, put $t = 0$ in the equation above to find

$$\frac{d}{dt}\Big[g(\vec{r}(t))\Big]\Big|_{t=0} = \nabla g(\vec{r}(0)) \bullet \frac{d\vec{r}}{dt}(0) = 0.$$

We find that both $\nabla f(x_o, y_o, z_o)$ and $\nabla g(x_o, y_o, z_o)$ are orthogonal to the tangent vector $\frac{d\vec{r}}{dt}(0)$. We derive this result for every smooth curve on $g(x, y, z) = 0$ thus $\nabla f(x_o, y_o, z_o)$ and $\nabla g(x_o, y_o, z_o)$ are normal to the tangent plane to $g(x, y, z) = 0$ at (x_o, y_o, z_o). It follows that $\nabla f(x_o, y_o, z_o)$ and $\nabla g(x_o, y_o, z_o)$ are colinear thus there exists some nonzero constant λ such that $\nabla f(x_o, y_o, z_o) = \lambda \nabla g(x_o, y_o, z_o)$. $\square$

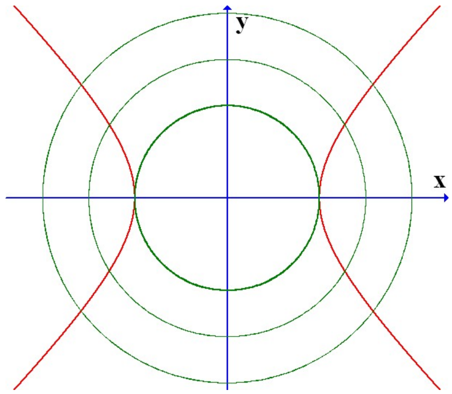
In advanced calculus I discuss an more general version of the Lagrange multiplier method which solves a wider array of problems. I think these two cases suffice for calculus III. If you are curious about the general method then perhaps you should take advanced calculus.

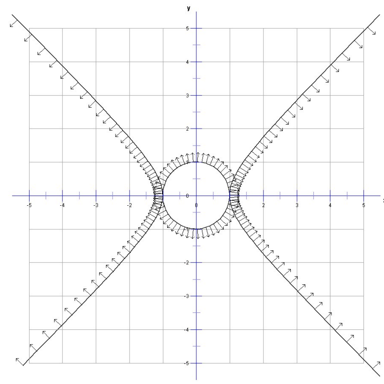
5.1.2 examples of the method

Example 5.1.1. *Suppose we wish to find maximum and minimum distance to the origin for points on the curve $x^2 - y^2 = 1$. In this case we can use the distance-squared function as our objective $f(x, y) = x^2 + y^2$ and the single constraint function is $g(x, y) = x^2 - y^2$. Observe that $\nabla f = \langle 2x, 2y \rangle$ whereas $\nabla g = \langle 2x, -2y \rangle$. We seek solutions of $\nabla f = \lambda \nabla g$ which gives us $\langle 2x, 2y \rangle = \lambda \langle 2x, -2y \rangle$. Hence $2x = 2\lambda x$ and $2y = -2\lambda y$. We must solve these equations subject to the condition $x^2 - y^2 = 1$. Observe that $x = 0$ is not a solution since $0 - y^2 = 1$ has no real solution. On the other hand, $y = 0$ does fit the constraint and $x^2 - 0 = 1$ has solutions $x = \pm 1$. Consider then*

$$2x = 2\lambda x \ \ \text{and} \ \ 2y = -2\lambda y \qquad \Rightarrow \qquad x(1 - \lambda) = 0 \ \ \text{and} \ \ y(1 + \lambda) = 0$$

Since $x \neq 0$ on the constraint curve it follows that $1 - \lambda = 0$ hence $\lambda = 1$ and we learn that $y(1 + 1) = 0$ hence $y = 0$. Consequently, $(1, 0)$ and $(-1, 0)$ are the two point where we expect to find extreme-values of f. In this case, the method of Lagrange multipliers served it's purpose, as you can see in the left graph. Below the green curves are level curves of the objective function whereas the particular red curve is the given constraint curve.





The picture on the right above is a screen-shot of the Java applet created by David Lippman and Konrad Polthier to explore 2D and 3D graphs. Especially nice is the feature of adding vector fields to given objects, many other plotters require much more effort for similar visualization. See more at the website: http://dlippman.imathas.com/g1/GrapherLaunch.html. Note how the gradient vectors to the objective function and constraint function line-up nicely at those points.

In the previous example, we actually got lucky. There are examples of this sort where we could get false maxima due to the nature of the constraint function.

Example 5.1.2. *Suppose we wish to find the points on the unit circle $g(x,y) = x^2 + y^2 = 1$ which give extreme values for the objective function $f(x,y) = x^2 - y^2$. Apply the method of Lagrange multipliers and seek solutions to $\nabla f = \lambda \nabla g$:*

$$\langle 2x, -2y\rangle = \lambda \langle 2x, 2y\rangle$$

We must solve $2x = 2x\lambda$ which is better cast as $(1-\lambda)x = 0$ and $-2y = 2\lambda y$ which is nicely written as $(1+\lambda)y = 0$. On the basis of these equations alone we have several options:

1. *if $\lambda = 1$ then $(1+1)y = 0$ hence $y = 0$*

2. *if $\lambda = -1$ then $(1-(-1))x = 0$ hence $x = 0$*

But, we also must fit the constraint $x^2 + y^2 = 1$ hence we find four solutions:

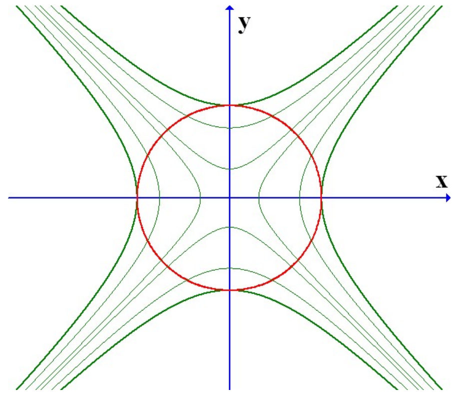
1. *if $\lambda = 1$ then $y = 0$ thus $x^2 = 1 \quad \Rightarrow \quad x = \pm 1 \quad \Rightarrow \quad (\pm 1, 0)$*

2. *if $\lambda = -1$ then $x = 0$ thus $y^2 = 1 \quad \Rightarrow \quad y = \pm 1 \quad \Rightarrow \quad (0, \pm 1)$*

We test the objective function at these points to ascertain which type of extrema we've located:

$$f(0, \pm 1) = 0^2 - (\pm 1)^2 = -1 \qquad \& \qquad f(\pm 1, 0) = (\pm 1)^2 - 0^2 = 1$$

When constrained to the unit circle we find the objective function attains a maximum value of 1 at the points $(1,0)$ and $(-1,0)$ and a minimum value of -1 at $(0,1)$ and $(0,-1)$. Let's illustrate the answers as well as a few non-answers to get perspective. Below the green curves are level curves of the objective function whereas the particular red curve is the given constraint curve.



The success of the last example was no accident. The fact that the constraint curve was a circle which is a closed and bounded subset of $\mathbb{R}^2$ means that is is a **compact** subset of $\mathbb{R}^2$. A well-known theorem of analysis states that any real-valued continuous function on a compact domain attains both maximum and minimum values. The objective function is continuous and the domain is compact hence the theorem applies and the method of Lagrange multipliers succeeds. In contrast, the constraint curve of the preceding example was a hyperbola which is not compact. We have no assurance of the existence of any extrema. Indeed, we only found minima but no maxima in Example 5.1.1.

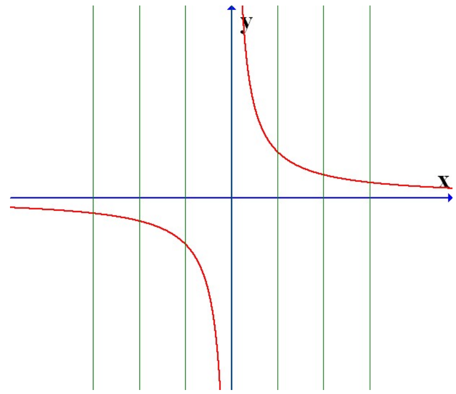
The generality of the method of Lagrange multipliers is naturally limited to smooth constraint curves and smooth objective functions. We must insist the gradient vectors exist at all points of inquiry. Otherwise, the method breaks down. If we had a constraint curve which has sharp corners then the method of Lagrange breaks down at those corners. In addition, if there are points of discontinuity in the constraint then the method need not apply. This is not terribly surprising, even in calculus I the main attack to analyze extrema of function on $\mathbb{R}$ assumed continuity, differentiability and sometimes twice differentiability. Points of discontinuity require special attention in whatever context you meet them.

At this point it is doubtless the case that some of you are, to misquote an ex-student of mine, "not-impressed". Perhaps the following examples better illustrate the dangers of non-compact constraint curves.

Example 5.1.3. *Suppose we wish to find extrema of $f(x,y) = x$ when constrained to $xy = 1$. Identify $g(x,y) = xy = 1$ and apply the method of Lagrange multipliers and seek solutions to $\nabla f = \lambda \nabla g$:*

$$\langle 1, 0\rangle = \lambda\langle y, x\rangle \quad \Rightarrow \quad 1 = \lambda y \ \ \textit{and} \ \ 0 = \lambda x$$

If $\lambda = 0$ then $1 = \lambda y$ is impossible to solve hence $\lambda \neq 0$ and we find $x = 0$. But, if $x = 0$ then $xy = 1$ is not solvable. Therefore, we find no solutions. Well, I suppose we have succeeded here in a way. We just learned there is no extreme value of x on the hyperbola $xy = 1$. Below the green curves are level curves of the objective function whereas the particular red curve is the given constraint curve.



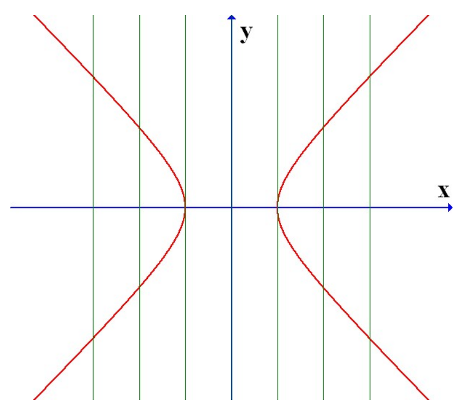
Example 5.1.4. *Suppose we wish to find extrema of $f(x,y) = x$ when constrained to $x^2 - y^2 = 1$. Identify $g(x,y) = x^2 - y^2 = 1$ and apply the method of Lagrange multipliers and seek solutions to $\nabla f = \lambda \nabla g$:*

$$\langle 1, 0\rangle = \lambda\langle 2x, -2y\rangle \quad \Rightarrow \quad 1 = 2\lambda x \quad \textit{and} \quad 0 = -2\lambda y$$

If $\lambda = 0$ then $1 = 2\lambda x$ is impossible to solve hence $\lambda \neq 0$ and we find $y = 0$. If $y = 0$ and $x^2 - y^2 = 1$ then we must solve $x^2 = 1$ whence $x = \pm 1$. We are tempted to conclude that:

1. *the objective function $f(x,y) = x$ attains a maximum on $x^2 - y^2 = 1$ at $(1,0)$ since $f(1,0) = 1$*
2. *the objective function $f(x,y) = x$ attains a minimum on $x^2 - y^2 = 1$ at $(-1,0)$ since $f(-1,0) = -1$*

But, both conclusions are false. Note $\sqrt{2}^2 - 1^2 = 1$ hence $(\pm\sqrt{2}, 1)$ are points on the constraint curve and $f(\sqrt{2}, 1) = \sqrt{2}$ and $f(-\sqrt{2}, 1) = -\sqrt{2}$. The error of the method of Lagrange multipliers in this context is the supposition that there exists extrema to find, in this case there are no such points. It is possible for the gradient vectors to line-up at points where there are no extrema. Below the green curves are level curves of the objective function whereas the particular red curve is the given constraint curve.



Incidentally, if you want additional discussion of Lagrange multipliers for two-dimensional problems one very nice source I certainly profitted from was the YouTube video by Edward Frenkel of Berkley. See his website http://math.berkeley.edu/ frenkel/ for links.

Example 5.1.5. *Suppose we wish to find extrema of $f(x,y) = x^2+3y^2$ on the unit circle $g(x,y) = x^2+y^2 = 1$. Identify that f is the objective function and g is the contraint function for this problem. The method of Lagrange multipliers claims that extrema for f subject to $g = 1$ are found from solutions of $\nabla f = \lambda \nabla g$. In particular we face the algebra problem below:*

$$\langle 2x, 6y\rangle = \lambda\langle 2x, 2y\rangle$$

Therefore, $x = \lambda x$ and $3y = \lambda y$. We must solve simultaneously

$$x(1-\lambda) = 0, \quad y(3-\lambda) = 0, \quad x^2+y^2 = 1$$

If $x = 0$ then $\lambda = 3$ and hence $x^2+y^2 = 1$ implies $y = \pm 1$. On the other hand, if $\lambda = 1$ then $y = 0$ hence $x^2+y^2 = 1$ implies $x = \pm 1$. Thus, we find the four extremal points: $(0,1), (0,-1), (1,0), (-1,0)$ and evaluation will reveal which is max/min

$$f(0,\pm 1) = 3 \qquad f(\pm 1, 0) = 1$$

Therefore, f restricted to the unit circle $x^2+y^2 = 1$ reaches an absolute maxium value of 3 at the points $(0,-1)$ and $(0,1)$ and an absolute minimum of 1 at the points $(1,0)$ and $(-1,0)$.

I know we found the absolute maximum and minimum because the constraint curve is closes and bounded and the objective function is smooth with non-vanishing gradient near the constraint curve. These two criteria imply that extreme values exist and the method of Lagrange can find them.

Example 5.1.6. Problem: *find the closest point on the plane $2x-2y+6z = 12$ to the point $(2,3,4)$.*

Solution: *we wish to minimize the distance between the (x,y,z) on the plane and the point $(2,3,4)$. This suggests our objective function is $f(x,y,z) = (x-2)^2+(y-3)^2+(z-4)^2$. The constraint surface is simply $g(x,y,z) = 2x-2y+6z-12 = 0$. Examine the lagrange multiplier equations:*

$$\nabla f = \lambda \nabla g \qquad \Rightarrow \qquad \Big\langle\ 2(x-2),\ 2(y-3),\ 2(z-4)\ \Big\rangle = \lambda\Big\langle 2,-2,6\Big\rangle$$

Therefore, $x = 2+\lambda$, $y = 3-\lambda$, $z = 4+3\lambda$. Substituting into the plane equation $2x-2y+6z = 12$,

$$2(2+\lambda)-2(3-\lambda)+6(4+3\lambda) = 12 \qquad \Rightarrow \qquad 2+\lambda-3+\lambda+12+9\lambda = 6$$

Hence, $11\lambda = 6-11$ so $\lambda = -5/11$. We deduce that the closest point at

$$x = 2-\frac{5}{11} = \frac{17}{11}, \quad y = 3+\frac{5}{11} = \frac{38}{11}, \quad z = 4-\frac{15}{11} = \frac{29}{11}.$$

The closest point is $\boxed{(17/11,\ 38/11,\ 29/11)}$.

The plane $2x-2y+6z = 12$ is not a bounded subset of $\mathbb{R}^3$ so we shouldn't necessarily expect to find extrema for the objective function in the last example. In fact, we found no maximally distant point. In a case such as the last example we use common sense to supplement the method. Proof of that a closest point exists involves a bit more than common sense. I'll leave it to your imagination, or a future course.

Example 5.1.7. *Let $f(x,y) = e^{xy}$ then find the extrema of f on the curve $x^3 + y^3 = 16$.*

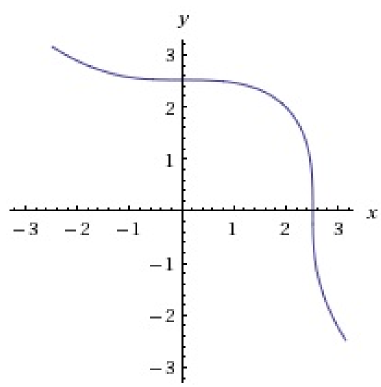
Solution: *identify our constraint as the level curve $g(x,y) = 16$ for $g(x,y) = x^3 + y^3$. The method of Lagrange multipliers suggests we solve simultaneously $g = 16$ and $\nabla f = \lambda \nabla g$. Explicitly, this yields,*

$$\langle ye^{xy},\ xe^{xy} \rangle = \lambda \langle 3x^2,\ 3y^2 \rangle$$

Therefore, if we solve both component equations for λ we obtain

$$\lambda = \frac{ye^{xy}}{3x^2} = \frac{xe^{xy}}{3y^2} \quad \Rightarrow \quad \frac{y}{x^2} = \frac{x}{y^2} \quad \Rightarrow \quad y^3 = x^3.$$

Now, return to $g(x,y) = x^3 + y^3$ to see $2x^3 = 16$ hence $x = \sqrt[3]{8} = 2$. It follows that $y^3 = 8$ hence $y = 2$. Thus $f(2,2) = e^4$. This is the global maximum of $f(x,y) = e^{xy}$ on $x^3 + y^3 = 16$. This claim is seen from examining the exponential function in each quadrant as the point gets far away from the origin on the given constraint curve. The constraint curve plotted with Wolfram Alpha is given below:



Notice in both quadrants II. and IV. we have $xy < 0$ hence $e^{xy} < 1$. It follows the maximum found is indeed the global maximum. Also, asymptotically, the values of f approach 0 as we travel along the constraint curve far from the origin.

Example 5.1.8. *Let $f(x,y,z) = xyz$ find the extreme values of f on the surface $x^2 + 2y^2 + 3z^2 = 6$.*

Solution: *let $g(x,y,z) = x^2 + 2y^2 + 3z^2$ hence the constraint surface is given by the solution set of $g(x,y,z) = 6$. Apply the method of Lagrange multipliers to solve $g = 6$ and $\nabla f = \lambda \nabla g$ simultaneously. Explicitly,*

$$\langle yz, xz, xy \rangle = \lambda \langle 2x, 4y, 6z \rangle$$

Thus,

$$yz = 2\lambda x, \quad xz = 4\lambda y, \quad xy = 6\lambda z.$$

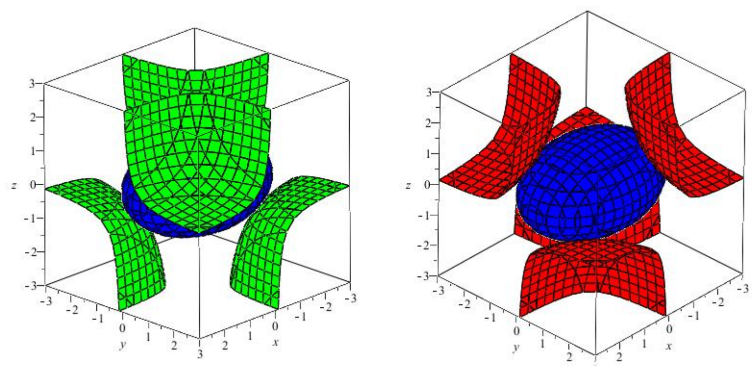
If any of the variables are zero then the equations above force the remaining two variables to be zero as well. Therefore, as $(0,0,0)$ is not a solution of $g(x,y,z) = 6$ we may assume $x, y, z \neq 0$ in the algebra which follows. Multiply the equations by x, y, z respectively to obtain:

$$xyz = 2\lambda x^2, \quad xyz = 4\lambda y^2, \quad xyz = 6\lambda z^2.$$

From which we find $x^2 = 2y^2 = 3z^2$ hence $g(x,y,z) = x^2 + 2y^2 + 3z^2 = 3x^2 = 6$. We find $x^2 = 2$ thus $x = \pm\sqrt{2}$ and it follows $2y^2 = 2$ hence $y = \pm 1$ and $3z^2 = 2$ thus $z = \pm\frac{\sqrt{2}}{\sqrt{3}}$. It follows we have *eight points to consider:*

$$f(\sqrt{2}, 1, \tfrac{\sqrt{2}}{\sqrt{3}}) = \tfrac{2}{\sqrt{3}} \qquad f(\sqrt{2}, 1, -\tfrac{\sqrt{2}}{\sqrt{3}}) = -\tfrac{2}{\sqrt{3}}$$
$$f(\sqrt{2}, -1, \tfrac{\sqrt{2}}{\sqrt{3}}) = -\tfrac{2}{\sqrt{3}} \qquad f(\sqrt{2}, -1, -\tfrac{\sqrt{2}}{\sqrt{3}}) = \tfrac{2}{\sqrt{3}}$$
$$f(-\sqrt{2}, 1, \tfrac{\sqrt{2}}{\sqrt{3}}) = -\tfrac{2}{\sqrt{3}} \qquad f(-\sqrt{2}, 1, -\tfrac{\sqrt{2}}{\sqrt{3}}) = \tfrac{2}{\sqrt{3}}$$
$$f(-\sqrt{2}, -1, \tfrac{\sqrt{2}}{\sqrt{3}}) = \tfrac{2}{\sqrt{3}} \qquad f(-\sqrt{2}, -1, -\tfrac{\sqrt{2}}{\sqrt{3}}) = -\tfrac{2}{\sqrt{3}}.$$

The maximum value $\frac{2}{\sqrt{3}}$ *is attained on the surface at the four points which have either all positive or just two negative coordinates. The minimum value* $-\frac{2}{\sqrt{3}}$ *is attained on the surface at the four points which have an odd number of negative components. Our analysis is illustrated in diagrams below. The blue ellipsoid is the constraint and the green and red illustrate surfaces on which the objective function takes its maximum and minimum values respective:*



Example 5.1.9. *Find the point on the plane* $x - y + z = 8$ *which is closest to* $(1, 2, 3)$.

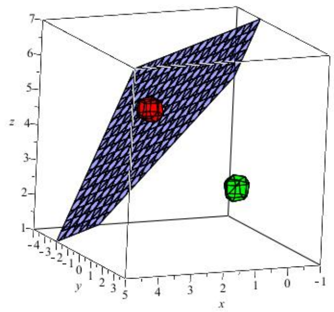
Solution: *we seek to minimize* $f(x, y, z) = (x-1)^2 + (y-2)^2 + (z-3)^2$ *subject* $g(x, y, z) = x - y + z = 8$. *Lagrange's method gives us:*

$$\nabla f = \lambda \nabla g \quad \Rightarrow \quad \begin{array}{c} 2(x-1) = \lambda \\ 2(y-2) = -\lambda \\ 2(z-3) = \lambda \end{array}$$

Thus,

$$\frac{\lambda}{2} = x - 1 = 2 - y = z - 3$$

hence $x = 3 - y$ *and* $z = 5 - y$ *and we can substitute these into the plane equation to find* $3 - y - y + 5 - y = -3y + 8 = 8$ *thus* $y = 0$ *and we find* $x = 3$ *and* $z = 5$. *Therefore, the closest point on the plane is* $(3, 0, 5)$. *A silly picture of this is as follows: the green point is* $(1, 2, 3)$ *and the red is* $(3, 0, 5)$



Example 5.1.10. *A rectangular box without a lid is made from 12 square feet of cardboard. Find the maximum volume of such a box.*

Solution: *let x, y, z be the lengths of the sides of the box then $V = xyz$ is the volume. The contraint is given by $g(x, y, z) = 2xz + 2yz + xy = 12$. The xy is the area of the base and the top is open so this distinguishes the term from the sides of the box. We apply the method of Lagrange: consider $\nabla V = \lambda \nabla g$ yields*

$$\langle yz, xz, xy\rangle = \lambda\langle 2z + y, 2z + x, 2x + 2y\rangle.$$

Thus,

$$\begin{aligned} yz = \lambda(2z + y) \quad &\Rightarrow \quad xyz = \lambda(2zx + xy) \\ xz = \lambda(2z + x) \quad &\Rightarrow \quad xyz = \lambda(2zy + xy) \\ xy = \lambda(2x + 2y) \quad &\Rightarrow \quad xyz = \lambda(2xz + 2yz) \end{aligned}$$

Hence,

$$\lambda(2zx + xy) = \lambda(2zy + xy) = \lambda(2xz + 2yz)$$

From which we find,

$$zx = zy \qquad xy = 2xz$$

Thus $y = x$ and $y = 2z$ hence $x = 2z$ (note $x, z = 0$ are not interesting physically). The constraint equation can be reduced to an equation in z:

$$12 = 2xz + 2yz + xy = 4z^2 + 4z^2 + 4z^2 = 12z^2 \quad \Rightarrow \quad z = 1.$$

Therefore, $x = y = 2$ and we conclude the box should have dimensions $2 \times 2 \times 1$ in feet. Thus $V = 4ft^3$ is the maximum volume.

The answer above comes as no surprise, there is no difference between x, y in the problem thus by symmetry $x = y$.

Example 5.1.11. *Let $f(x,y) = xy$. Find the extrema of f on the ellipse $x^2/8 + y^2/2 = 1$.*

Solution: *identify the constraint function $g(x,y) = x^2/8 + y^2/2 = 1$. Note, $\nabla f = \lambda \nabla g$ yields*

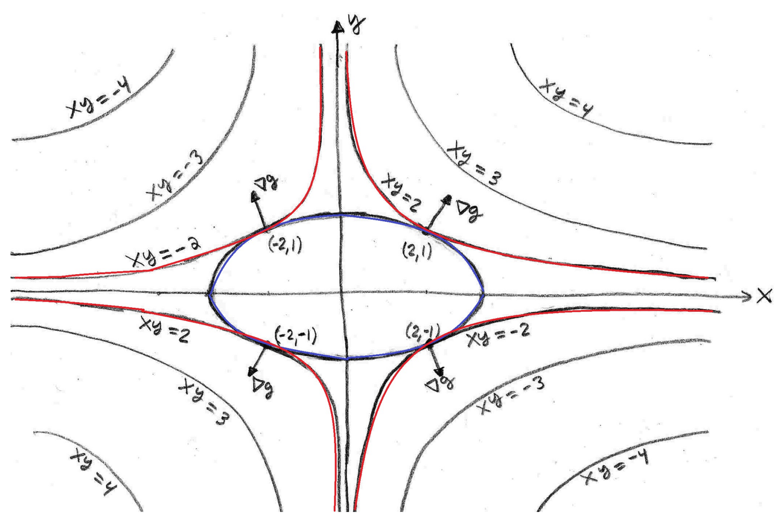
$$\langle y, x\rangle = \lambda\langle x/4, y\rangle \quad \Rightarrow \quad 4y = \lambda x \;\;\&\;\; x = \lambda y \quad \Rightarrow \quad y = \frac{\lambda^2 y}{4}$$

Therefore, $y(1 - \frac{\lambda^2}{4}) = 0$ from which we find solutions $y = 0$ or $\lambda = \pm 2$.

If $y = 0$ then $x = \lambda y = 0$ but $(0,0)$ is not on the ellipse.

If $\lambda = 2$ then $x = 2y$ and $4y = 2x$ and thus $x^2 = 4$ hence $x = \pm 2$ and $y = \pm 1$ hence $(-2,-1)$ and $(2,1)$ are solutions. Note $f(\pm 2, \pm 1) = (\pm 2)(\pm 1) = 2$.

If $\lambda = -2$ then $x = -2y$ and $4y = -2x$ hence $y = -\frac{1}{2}x$ thus $x^2 = 4$ and $x = \pm 2$. However, $y = -\frac{1}{2}(\pm 2) = \mp 1$. The solutions $(-2,1)$ and $(2,-1)$ follow. Note $f(\pm 2, \mp 1) = (\pm 2)(\mp 1) = -2$. In summary, we find the maximum of 2 is attained at $(-2,-1)$ and $(2,1)$ whereas the minimum of -2 is attained at $(2,-1)$ and $(-2,1)$. The picture below illustrates why:



5.1.3 extreme values of a quadratic form on a circle

In the next example we generalize the results of several past examples. In particular we intend to find the max/min for an arbitrary quadratic function in x, y on a circle of radius R. The result of this discussion will be of great use later in this chapter.

Problem: Suppose $Q(x, y) = ax^2 + 2bxy + cy^2$ for some constants a, b, c. Determine general formulas for the extrema of Q on a circle of radius R given by $g(x, y) = x^2 + y^2 = R^2$.

Solution: Apply the method of Lagrange, we seek to solve $\nabla Q = \lambda \nabla g$ subject to $g(x, y) = x^2 + y^2 = R^2$,

$$\langle 2ax + 2by, 2bx + 2cy \rangle = \lambda \langle 2x, 2y \rangle \qquad \Rightarrow \qquad ax + by = \lambda x, \;\; bx + cy = \lambda y$$

We must solve simultaneously the following triple of equations:

$$(a - \lambda)x + by = 0, \quad bx + (c - \lambda)y = 0, \quad x^2 + y^2 = R^2.$$

As a matrix problem, setting aside the circle equation for a moment,

$$\begin{bmatrix} a - \lambda & b \\ b & c - \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

If $(a - \lambda)(c - \lambda) - b^2 \neq 0$ then the inverse matrix of the 2×2 exists and multiplication by of the equation above by the matrix $\frac{1}{(a-\lambda)(c-\lambda)-b^2} \begin{bmatrix} c - \lambda & -b \\ -b & a - \lambda \end{bmatrix}$ yields the solution $x = y = 0$. However, in that case there is only a solution to the circle equation $x^2 + y^2 = R^2$ if it happens that $R = 0$, we are more interested in the case $R \neq 0$ so we must look for solutions elsewhere. In other words, for interesting solutions we must insist that $(a - \lambda)(c - \lambda) - b^2 = 0$. The constants a, b, c are given so we face a quadratic equation in λ:

$$\lambda^2 - (a + c)\lambda + ac - b^2 = 0 \;\; \star$$

Completing the square yields solutions:

$$\lambda = \frac{a + c \pm \sqrt{(a + c)^2 - 4(ac - b^2)}}{2}$$

What type of solutions are possible from this expresssion? Simplify the expression in the radical,

$$\lambda = \frac{a + c \pm \sqrt{(a - c)^2 + 4b^2}}{2}$$

The radicand is manifestly non-negative, there is clearly no way to obtain a complex solution for real values of a, b, c. Values of λ are either zero, positive or negative. This is an important observation once we pair it with the calculation that follows. Casewise logic is needed:

Case I. Suppose $b = 0$ then $(a - \lambda)(c - \lambda) = 0$ and there are three ways to solve this:

1. $a = \lambda \neq c$ hence $y = 0$ and it follows $x = \pm R$. Therefore, extreme values of $Q(\pm R, 0) = aR^2$ are attained at $(\pm R, 0)$.
2. $c = \lambda \neq a$ hence $x = 0$ and it follows $y = \pm R$. Therefore, extreme values of $Q(0, \pm R) = cR^2$ are attained at $(0, \pm R)$.
3. $a = c = \lambda$ hence $(a - \lambda)x + by = 0$ and $bx + (c - \lambda)y = 0$ are solved, this leaves only the circle equation $x^2 + y^2 = R^2$. We find infinitely many solutions. At each point of the unit circle the value $Q(x, y) = ax^2 + cy^2 = a(x^2 + y^2) = aR^2$ is attained.

To summarize the results above, if $b = 0$ and $a \neq c$ then the extreme values of aR^2 and cR^2 are attained at the points $(\pm R, 0)$ and $(0, \pm R)$. However, if $b = 0$ and $a = c$ then Q is constant on the radius R circle with value aR^2.

Case II: Suppose $b \neq 0$. We already worked out that

$$\lambda = \frac{a + c \pm \sqrt{(a - c)^2 + 4b^2}}{2}$$

solves

$$(a - \lambda)x + by = 0, \quad bx + (c - \lambda)y = 0, \quad x^2 + y^2 = R^2.$$

Our current goal is to solve the equations above for x, y. Solve for y, I'll aim for solutions in terms of λ since we have a clear method to calculate it already,

$$y = \frac{x}{b}(\lambda - a)$$

Substitute into $x^2 + y^2 = R^2$ to find

$$x^2 + \frac{(\lambda - a)^2}{b^2}x^2 = R^2 \quad \Rightarrow \quad x_\pm = \pm R\sqrt{\frac{b^2}{b^2 + (\lambda - a)^2}}.$$

Return once more to $y = \frac{x}{b}(\lambda - a)$ to find:

$$y_\pm = \pm R\left[\frac{\lambda - a}{b}\right]\sqrt{\frac{b^2}{b^2 + (\lambda - a)^2}}.$$

Therefore, the points (x_-, y_-) and (x_+, y_+) are solutions to the Lagrange multiplier equations for each λ. Moreover, the extreme values attained via these points are given by:

$$\begin{aligned}
Q(x_\pm, y_\pm) &= ax_\pm^2 + 2bx_\pm y_\pm + cy_\pm^2 \\
&= aR^2\frac{b^2}{b^2 + (\lambda - a)^2} + 2bR^2\left[\frac{\lambda - a}{b}\right]\frac{b^2}{b^2 + (\lambda - a)^2} + cR^2\left[\frac{\lambda - a}{b}\right]^2\frac{b^2}{b^2 + (\lambda - a)^2} \\
&= \frac{R^2 b^2}{b^2 + (\lambda - a)^2}\left[a + 2(\lambda - a) + c\left[\frac{\lambda - a}{b}\right]^2\right] \\
&= \frac{R^2}{b^2 + (\lambda - a)^2}\left[\lambda b^2 + b^2(\lambda - a) + c(\lambda - a)^2\right] \\
&= \frac{R^2}{b^2 + (\lambda - a)^2}\left[\lambda b^2 + (\lambda - a)\left[b^2 - ac + c\lambda\right]\right]
\end{aligned}$$

From $\star$ we know $b^2 - ac = \lambda^2 - (a+c)\lambda = \lambda^2 - a\lambda - c\lambda$ hence,

$$\begin{aligned} Q(x_\pm, y_\pm) &= \frac{R^2}{b^2 + (\lambda - a)^2}\Big[\lambda b^2 + (\lambda - a)\big[\lambda^2 - a\lambda - c\lambda + c\lambda\big]\Big] \\ &= \frac{R^2}{b^2 + (\lambda - a)^2}\Big[\lambda b^2 + \lambda(\lambda - a)^2\Big] \\ &= \lambda R^2. \end{aligned}$$

I invite the reader to prove that $b \neq 0$ implies solutions of the equation $(a - \lambda)(c - \lambda) - b^2 = 0$ are distinct. That is, given $b \neq 0$, the solutions λ_1, λ_2 must have $\lambda_1 \neq \lambda_2$. If we label these with $\lambda_1 < \lambda_2$ then it follows that $\lambda_1 R^2$ is the minimum value whereas $\lambda_2 R^2$ is the maximum value of Q on the circle $x^2 + y^2 = R^2$.

The theorem below summarizes our analysis thus far:

Theorem 5.1.12.

Suppose $Q(x, y) = ax^2 + 2bxy + cy^2$ for some constants $a, b, c \in \mathbb{R}$. Let $R > 0$ be the radius of the circle S_R with equation $x^2 + y^2 = R^2$. The **characteristic equation** $(a-\lambda)(c-\lambda)-b^2 = 0$ has only real solutions. Furthermore, the extreme values of Q on the circle are simply given by λR^2 where λ is a solution of the characteristic equation.

There are several cases implicit within the theorem above: let's denote the solutions to the characteristic equation by λ_1, λ_2,

1. if $\lambda_1 = \lambda_2$ then the Q is constant on S_R.
2. if $\lambda_1 \neq \lambda_2$ and $\lambda_1 \langle \lambda_2$ then $Q_{min} = \lambda_1 R^2$ whereas $Q_{max} = \lambda_2 R^2$.

Case (1.) is when the level curves of Q are circles. The graph $z = Q(x, y) = a(x^2 + y^2)$ either opens up $(a > 0)$ or down $(a < 0)$ from the origin where $Q(0, 0) = 0$ is either the minimum or maxium of Q on any disk of radius R. Think geometrically for the moment, imagine shrinking $R \to 0$ to obtain this result on the disk.

Part of Case (2.) is almost the same as Case (1.) if λ_1, λ_2 share the same sign. For instance, if $0 < \lambda_1 < \lambda_2$ then $z = Q(x, y)$ opens upward with each contour being an ellipse and clearly $Q(0, 0)$ is a minimum. On the other hand if $\lambda_1 < \lambda_2 < 0$ then $z = Q(x, y)$ opens downwards and each contour is an ellipse and $Q(0, 0)$ is a maximum.

However, when Case (2.) has λ_1, λ_2 with different signs we find $\lambda_1 < 0 < \lambda_2$. In this case $z = Q(x, y)$ opens upward in the direction associated with λ_2 and it opens downward in the direction associated to λ_1. It has a saddle shape, and the contours of the graph are hyperbolae.

Finally, in Case (2.) if $\lambda_1 = 0$ and $\lambda_2 > 0$ then $z = Q(x, y)$ is constant along the direction corresponding to λ_1 and it opens upward along the direction corresponding to λ_2. Likwise, if $\lambda_1 = 0$ and $\lambda_2 < 0$ then $z = Q(x, y)$ is constant along the direction corresponding to λ_1 and it opens downward along the direction corresponding to λ_2.

The theorem below summarizes the relation between the characteristic equation for the quadratic form Q and its extrema in the plane $\mathbb{R}^2$. The values λ are usually called **eigenvalues** so this theorem can be essentially summarized as: the eigenvalues determine the nature of the extreme values for a quadratic form:

Theorem 5.1.13.

The graph of $z = Q(x, y) = ax^2+2bxy+cy^2$ for some constants $a, b, c \in \mathbb{R}$ can be categorized by real solutions of the **characteristic equation** $(a - \lambda)(c - \lambda) - b^2 = 0$. In particular,

1. if $\lambda_1, \lambda_2 > 0$ then $Q(0, 0)$ is an minimum value for Q (the graph $z = Q(x, y)$ is a paraboloid which opens up)
2. if $\lambda_1, \lambda_2 < 0$ then $Q(0, 0)$ is a maximum value for Q (the graph $z = Q(x, y)$ is a paraboloid which opens down)
3. if $\lambda_1 < 0 < \lambda_2$ then $Q(0, 0)$ is neither a maximum or minimum for Q (the graph $z = Q(x, y)$ is a hyperbolic paraboloid which opens up and down)
4. if $\lambda_1 = 0$ and $\lambda_2 > 0$ then $Q(0, 0)$ is a minimum value for Q (the graph $z = Q(x, y)$ is a parabolic trough which opens upward)
5. if $\lambda_1 = 0$ and $\lambda_2 < 0$ then $Q(0, 0)$ is a maximum value for Q (the graph $z = Q(x, y)$ is a parabolic trough which opens down)

In cases (4.) and (5.) above the local extrema is not isolated, there is a whole line on which Q is extremal. In contrast, cases (1.) and (2.) have isolated local extrema. As we apply this result later in this section cases (1-3) will play a larger role than cases (4-5).

5.1.4 quadratic forms in n-variables

optional section: I briefly explain how the last section generalizes. a good linear algebra text will provide further detail for the interested student. Notice that the last section did not use linear algebraic technique, we just brute-force solved the $n = 2$ case. To go further it is wise to learn linear algebraic techniques to organize the calculation, otherwise it could get difficult.

There is a better way to derive the results of the last section. In linear algebra we define a quadratic form on $\mathbb{R}^n$ as a function $Q : \mathbb{R}^n \to \mathbb{R}$ defined by $Q(\vec{x}) = \vec{x}^T A\vec{x}$ for a *symmetric* matrix A. It turns out that the values of Q on a sphere of radius R in $\mathbb{R}^n$ are given by the eigenvalues of A. In particular, λ is a solution to $det(A - \lambda I) = 0$ and if $\vec{x} \neq 0$ solves $(A - \lambda)\vec{x} = 0$ and $||\vec{x}|| = R$ then $Q(\vec{x}, \vec{x}) = \lambda R^2$. The value λ is called an **eigenvalue** with **eigenvector** $\vec{x}$. When you work out the details it becomes clear that $det(A-\lambda I) = 0$ is an n-th order polynomial equation in λ and, while it is not entirely trivial to prove, these solutions are all real. The list of all eigenvalues for Q is called the *spectrum*. If we order the spectrum in increasing values $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$ then $\lambda_1 R^2$ is the minimum value whereas $\lambda_n R^2$ is the maxium value of Q on the sphere $x_1^2 + x_2^2 + \cdots + x_n^2 = R^2$ in $\mathbb{R}^n$. If you study the equations of the last section once you've studied eigenvectors and eigenvalues then you'll find that the equations provided by the Lagrange multiplier method are just the characteristic and eigenvector equations for $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. Moreover, you learn that the big idea is that for quadratic forms is that they reduce to $Q(y_1, y_2, \dots, y_n) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots \lambda_n y_n^2$ for appropriate choice of coordinates $(y_1, y_2, \dots, y_n)$. For $Q(x, y) = ax^2 + 2bxy + cy^2$, the middle term $2bxy$ is an artifact of the cartesian coordinates that framed the given Q, a simple rotation will remove the non-diagonal terms in A and leave us with $Q(y_1, y_2) = \lambda_1 y_1^2 + \lambda_2 y_2^2$. Then in the (y_1, y_2) coordinates it becomes manifestly obvious that a quadratic form $Q(x, y) = ax^2 + 2bxy + cy^2$ has contours which are either hyperbolas, lines, parabolas or ellipses.

5.2 multivariate taylor series

We begin this section with a brief overview of single-variate power series. The results presented are important as we often use the single-variable results paired with a substitution to generate interesting multivariate series.

5.2.1 taylor's polynomial for one-variable

If $f : U \subseteq \mathbb{R} \to \mathbb{R}$ is analytic at $x_o \in U$ then we can write

$$f(x) = f(x_o) + f'(x_o)(x - x_o) + \frac{1}{2}f''(x_o)(x - x_o)^2 + \cdots = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_o)}{n!}(x - x_o)^n$$

We could write this in terms of the operator $D = \frac{d}{dt}$ and the evaluation of $t = x_o$

$$f(x) = \left[\sum_{n=0}^{\infty} \frac{1}{n!}(x - t)^n D^n f(t)\right]_{t=x_o}$$

I remind the reader that a function is called **entire** if it is analytic on all of $\mathbb{R}$, for example $e^x, \cos(x)$ and $\sin(x)$ are all entire. In particular, you should know that:

$$e^x = 1 + x + \frac{1}{2}x^2 + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}x^n$$

$$\cos(x) = 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}x^{2n}$$

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}x^{2n+1}$$

Since $e^x = \cosh(x) + \sinh(x)$ it also follows that

$$\cosh(x) = 1 + \frac{1}{2}x^2 + \frac{1}{4!}x^4 \cdots = \sum_{n=0}^{\infty} \frac{1}{(2n)!}x^{2n}$$

$$\sinh(x) = x + \frac{1}{3!}x^3 + \frac{1}{5!}x^5 \cdots = \sum_{n=0}^{\infty} \frac{1}{(2n+1)!}x^{2n+1}$$

The geometric series is often useful, for $a, r \in \mathbb{R}$ with $|r| < 1$ it is known

$$a + ar + ar^2 + \cdots = \sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$$

This generates a whole host of examples, for instance:

$$\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \cdots$$

$$\frac{1}{1-x^3} = 1 + x^3 + x^6 + x^9 + \cdots$$

$$\frac{x^3}{1-2x} = x^3(1+2x+(2x)^2+\cdots) = x^3+2x^4+4x^5+\cdots$$

Moreover, the term-by-term integration and differentiation theorems yield additional results in conjuction with the geometric series:

$$\tan^{-1}(x) = \int \frac{dx}{1+x^2} = \int \sum_{n=0}^{\infty}(-1)^n x^{2n}dx = \sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}x^{2n+1} = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 + \cdots$$

$$\ln(1-x) = \int \frac{d}{dx}\ln(1-x)dx = \int \frac{-1}{1-x}dx = -\int \sum_{n=0}^{\infty} x^n dx = \sum_{n=0}^{\infty}\frac{-1}{n+1}x^{n+1}$$

Of course, these are just the basic building blocks. We also can twist things and make the student use algebra,

$$e^{x+2} = e^x e^2 = e^2(1+x+\frac{1}{2}x^2+\cdots)$$

or trigonmetric identities,

$$\sin(x) = \sin(x-2+2) = \sin(x-2)\cos(2) + \cos(x-2)\sin(2)$$

$$\Rightarrow \quad \sin(x) = \cos(2)\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)!}(x-2)^{2n+1} + \sin(2)\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n)!}(x-2)^{2n}.$$

Feel free to peruse my most recent calculus II materials to see a host of similarly sneaky calculations.

5.2.2 taylor's multinomial for two-variables

Suppose we wish to find the taylor polynomial centered at $(0,0)$ for $f(x,y) = e^x\sin(y)$. It is as simple as this:

$$f(x,y) = \left(1+x+\frac{1}{2}x^2+\cdots\right)\left(y-\frac{1}{6}y^3+\cdots\right) = y + xy + \frac{1}{2}x^2y - \frac{1}{6}y^3 + \cdots$$

the resulting expression is called a multinomial since it is a polynomial in multiple variables. If all functions $f(x,y)$ could be written as $f(x,y) = F(x)G(y)$ then multiplication of series known from calculus II would often suffice. However, many functions do not possess this very special form. For example, how should we expand $f(x,y) = \cos(xy)$ about $(0,0)$?. We need to derive the two-dimensional Taylor's theorem[3].

In previous chapters we have discussed the best linear approximation for a function of several variables. The next step is the best quadratic approximation. In particular, we seek to find formulas to fix the constants $c_o, c_1, c_2, c_{11}, c_{12}, c_{22}$ as given below:

$$\boxed{f(x,y) \approx c_o + c_1(x-x_o) + c_2(y-y_o) + c_{11}(x-x_o)^2 + c_{12}(x-x_o)(y-y_o) + c_{22}(y-y_o)^2.}$$

[3] A more careful proof will be found in most advanced calculus texts, it turns out the multivariate expansion follow from differentiating $g = f\circ\vec{r}$ where $\vec{r}:\mathbb{R}\to\mathbb{R}^2$ has $\vec{r}(t) = \langle x_o+at, y_o+bt\rangle$. The single-variable Taylor theorem applies and we therefore generalize the remainder estimation theorems to higher dimensions. I will not dig deeper into questions about the remainder for multivariate taylor expansions in these notes. Intuitively, we have one assumption: higher order terms are small near the center of the series.

The expression above is a quadratic polynomial in x, y centered at (x_o, y_o). Observe that it is already clear that $f(x_o, y_o) = c_o$. Take partial derivatives in x and y,

$$f_x(x,y) \approx c_1 + 2c_{11}(x - x_o) + c_{12}(y - y_o) \qquad f_y(x,y) \approx c_2 + c_{12}(x - x_o) + 2c_{22}(y - y_o).$$

Therefore, it is clear that: $f_x(x_o, y_o) = c_1$ and $f_y(x_o, y_o) = c_2$. Differentiating once more,

$$f_{xx}(x,y) \approx 2c_{11} \qquad f_{xy}(x,y) \approx c_{12} \qquad f_{yy}(x,y) \approx 2c_{22}.$$

Therefore, $f_{xx}(x_o, y_o) = 2c_{11}$, $f_{xy}(x_o, y_o) = c_{12}$ and $f_{yy}(x_o, y_o) = 2c_{22}$. It follows that we can construct the best quadratic approximation near (x_o, y_o) by the formula below: let $\vec{p}_o = (x_o, y_o)$

$$\boxed{f(x,y) \approx f(x_o, y_o) + L(x - x_o, y - y_o) + Q(x - x_o, y - y_o)}$$

Where, I denoted $L(x - x_o, y - y_o) = f_x(\vec{p}_o)(x - x_o) + f_y(\vec{p}_o)(y - y_o)$ and

$$Q(x - x_o, y - y_o) = \frac{1}{2} f_{xx}(\vec{p}_o)(x - x_o)^2 + f_{xy}(\vec{p}_o)(x - x_o)(y - y_o) + \frac{1}{2} f_{yy}(\vec{p}_o)(y - y_o)^2.$$

Notice that $f(\vec{p}_o) + L(x - x_o, y - y_o)$ gives the first-order approximation of f, it is the linearization of f at $\vec{p}_o$. We can also write the expansion as

$$\boxed{f(x_o + h, y_o + k) \approx f(\vec{p}_o) + f_x(\vec{p}_o)h + f_y(\vec{p}_o)k + \frac{1}{2} f_{xx}(\vec{p}_o)h^2 + f_{xy}(\vec{p}_o)hk + \frac{1}{2} f_{yy}(\vec{p}_o)k^2.}$$

Example 5.2.1. *Suppose* $f(x,y) = \sqrt{1 + x + y}$. *Differentiating yields:*

$$f_x(x,y) = f_y(x,y) = \frac{1}{2}(1 + x + y)^{-1/2}.$$

Differentiate once more,

$$f_{xx}(x,y) = f_{yy}(x,y) = f_{xy}(x,y) = \frac{-1}{4}(1 + x + y)^{-3/2}.$$

Observe that $f(0,0) = 1$, $f_x(0,0) = f_y(0,0) = \frac{1}{2}$ *and* $f_{xx}(0,0) = f_{xy}(0,0) = f_{yy}(0,0) = \frac{-1}{4}$. *Therefore,*

$$\sqrt{1 + x + y} \approx 1 + \frac{1}{2}(x + y) - \frac{1}{8}(x^2 + 2xy + y^2)$$

As an application, let's calculate $\sqrt{1.11}$. *Notice* $1.11 = 1 + 0.1 + 0.01$ *so apply the formula with* $x = 0.1$ *and* $y = 0.01$,

$$\begin{aligned}
\sqrt{1 + 0.1 + 0.01} &\approx 1 + \frac{1}{2}(0.1 + 0.01) - \frac{1}{8}((0.1)^2 + 2(0.1)(0.01) + (0.01)^2) \\
&\approx 1 + (0.5)(0.11) - (0.125)\big(0.01 + 0.002 + 0.0001\big) \\
&\approx 1 + (0.5)(0.11) - (0.125)(0.0121) \\
&\approx 1 + 0.055 - 0.0015125 \\
&\approx 1 + 0.055 - 0.0015125 \\
&\approx 1.0534875.
\end{aligned}$$

In contrast, my Casio fx-115 ES claims $\sqrt{1.11} = 1.053565375$. *If we trust my calculator then we have correctly calculated the four correct digits for* $\sqrt{1.11}$. *Not too shabby for our trouble.*

Computation of third, fourth or higher order terms reveals the multivariate taylor expansion below. We denote $h = h_1$ and $k = h_2$,

$$\boxed{f(x_o + h, y_o + k) = \sum_{n=0}^{\infty} \sum_{i_1=0}^{2} \sum_{i_2=0}^{2} \cdots \sum_{i_n=0}^{2} \frac{1}{n!} \frac{\partial^{(n)} f(x_o, y_o)}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_n}} h_{i_1} h_{i_2} \cdots h_{i_n}}$$

Example 5.2.2. *Expand $f(x,y) = \cos(xy)$ about $(0,0)$. We calculate derivatives,*

$$f_x = -y\sin(xy) \qquad f_y = -x\sin(xy)$$

$$f_{xx} = -y^2\cos(xy) \qquad f_{xy} = -\sin(xy) - xy\cos(xy) \qquad f_{yy} = -x^2\cos(xy)$$

$$f_{xxx} = y^3\sin(xy) \qquad f_{xxy} = -y\cos(xy) - y\cos(xy) + xy^2\sin(xy)$$

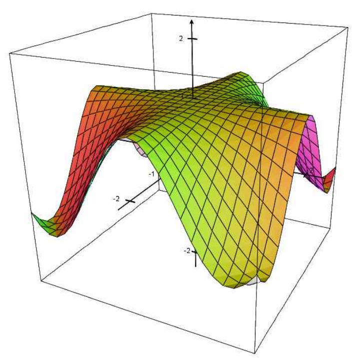
$$f_{xyy} = -x\cos(xy) - x\cos(xy) + x^2y\sin(xy) \qquad f_{yyy} = x^3\sin(xy)$$

Next, evaluate at $x = 0$ and $y = 0$ to find $f(x,y) = 1 + \cdots$ to third order in x, y about $(0,0)$. We can understand why these derivatives are all zero by approaching the expansion a different route: simply expand cosine directly in the variable (xy),

$$f(x,y) = 1 - \frac{1}{2}(xy)^2 + \frac{1}{4!}(xy)^4 + \cdots = 1 - \frac{1}{2}x^2y^2 + \frac{1}{4!}x^4y^4 + \cdots.$$

Apparently the given function only has nontrivial derivatives at $(0,0)$ at orders $0, 4, 8, \ldots$. We can deduce that $f_{xxxxy}(0,0) = 0$ without further calculation.

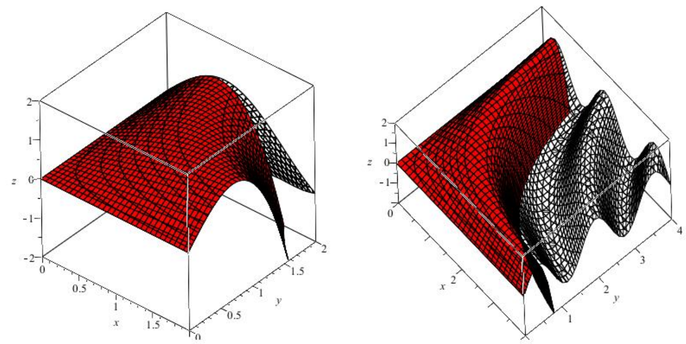
This is actually a very interesting function, I think it defies our analysis in the later portion of this chapter. The second order part of the expansion reveals nothing about the nature of the critical point $(0,0)$. Of course, any student of trigonometry should recognize that $f(0,0) = 1$ is likely a local maximum, it's certainly not a local minimum. The graph reveals that $f(0,0)$ is a local maxium for f restricted to certain rays from the origin whereas it is constant on several special directions (the coordinate axes).



Example 5.2.3. *Suppose $f(x,y) = \sin(xy)$. Once more I'll use the substitution trick. Let $u = xy$ hence $f(x,y) = \sin(u) = u - \frac{1}{6}u^3 + \cdots$ and to quadratic 6-th order we find*

$$f(x,y) = xy - \frac{1}{6}x^3y^3 + \cdots.$$

It is interesting to compare the graph $z = f(x,y)$ and $z = xy - \frac{1}{6}x^3y^3$, note how closely they correspond near the origin: the red graph is the approximating surface $z = xy - \frac{1}{6}x^3y^3$ and the transparent wire-frame is the actual function $z = \sin(xy)$. Roughly, they are within 0.1 units a distance of 1 from the origin. You can see in the right picture as we zoom away they difference between the function and the approximation is appreciable.



Example 5.2.4. *Suppose $f(x,y) = \sin(\sqrt{x^2+y^2})$ then in polar coordinates $f(r,\theta) = \sin(r)$. In this case the natural expansion to use is $f(x,y) = r - \frac{1}{6}r^3 + \frac{1}{120}r^5 + \cdots$ which is not technically a multivariate power series in x, y. In fact, $\sqrt{x^2+y^2}$ is not even differentiable at $(0,0)$.*

Example 5.2.5. *Suppose $f(x,y) = \sin(x^2+y^2)$ then in polar coordinates $f(r,\theta) = \sin(r^2)$. In this case the natural expansion to use is $f(x,y) = r^2 - \frac{1}{6}r^6 + \frac{1}{120}r^{10} + \cdots$ which is easily rewritten as a multivariate power series since $r^2 = x^2+y^2$. You use the series to observe that the first, third, fourth, fifth, seventh, eighth and ninth derivatives of f at $(0,0)$ are zero.*

5.2.3 taylor's multinomial for many-variables

Suppose $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ is a function of n-variables. It turns out that the Taylor series centered at $\vec{a} = (a_1, a_2, \dots, a_n)$ has the form:

$$\boxed{f(\vec{a}+\vec{h}) = \sum_{k=0}^{\infty}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n}\frac{1}{k!}\big(\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k}f\big)(\vec{a})\ h_{i_1}h_{i_2}\cdots h_{i_k}.}$$

Naturally, we sometimes prefer to write the series expansion about $\vec{a}$ as an expresssion in $\vec{x} = \vec{a}+\vec{h}$. With this substitution we have $\vec{h} = \vec{x}-\vec{a}$ and $h_{i_j} = (x-a)_{i_j} = x_{i_j} - a_{i_j}$ thus

$$\boxed{f(x) = \sum_{k=0}^{\infty}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n}\frac{1}{k!}\big(\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k}f\big)(\vec{a})\ (x_{i_1}-a_{i_1})(x_{i_2}-a_{i_2})\cdots(x_{i_k}-a_{i_k}).}$$

Proof of these claims is found in advanced calculus. Let me illustrate how these formulas work for $n = 3$.

Example 5.2.6. *Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ let's unravel the Taylor series centered at $(0,0,0)$ from the general formula boxed above. Utilize the notation $x = x_1, y = x_2$ and $z = x_3$ in this example.*

$$f(\vec{x}) = \sum_{k=0}^{\infty}\sum_{i_1=1}^{3}\sum_{i_2=1}^{3}\cdots\sum_{i_k=1}^{3}\frac{1}{k!}\big(\partial_{i_1}\partial_{i_2}\cdots\partial_{i_k}f\big)(0)\ x_{i_1}x_{i_2}\cdots x_{i_k}.$$

The terms to order 2 are as follows:

$$\begin{aligned} f(\vec{x}) &= f(0) + f_x(0)x + f_y(0)y + f_z(0)z \\ &\quad + \tfrac{1}{2}\Big(f_{xx}(0)x^2 + f_{yy}(0)y^2 + f_{zz}(0)z^2 + \\ &\qquad + f_{xy}(0)xy + f_{xz}(0)xz + f_{yz}(0)yz + f_{yx}(0)yx + f_{zx}(0)zx + f_{zy}(0)zy \Big) + \cdots \end{aligned}$$

Partial derivatives commute for smooth functions hence,

$$\begin{aligned} f(\vec{x}) &= \underbrace{f(0) + f_x(0)x + f_y(0)y + f_z(0)z}_{linearization} + \\ &+ \underbrace{\frac{1}{2}\Big(\quad f_{xx}(0)x^2 + f_{yy}(0)y^2 + f_{zz}(0)z^2 + 2f_{xy}(0)xy + 2f_{xz}(0)xz + 2f_{yz}(0)yz \quad \Big)}_{quadratic\ form\ Q(x,y,z)} + \cdots \end{aligned}$$

Identify that $f(0) + f_x(0)x + f_y(0)y + f_z(0)z$ is the linearization of f at the origin and the quadratic terms are simply the analogue of $Q(x,y) = f_x x(0)x^2 + 2f_{xy}(0)xy + f_{yy}(0)y^2$ for $n = 3$.

In the $n = 2$ case the graph $z = f(x,y)$ is relatively easy to visualize. Intuitively, the lineariization gives a plane which resembles the graph and then the linearization plus the quadratic form give some quadratic surface which better models the graph $z = f(x,y)$ neat the point of the expansion. Something similar is true in $n = 3$ however visualization is hard since the graph $w = f(x,y,z)$ is a four-dimensional picture.

Example 5.2.7. *Suppose $f(x,y,z) = e^{xyz}$. Find a quadratic approximation to f near $(0,1,2)$. Observe:*

$$f_x = yze^{xyz} \qquad f_y = xze^{xyz} \qquad f_z = xye^{xyz}$$

$$f_{xx} = (yz)^2 e^{xyz} \qquad f_{yy} = (xz)^2 e^{xyz} \qquad f_{zz} = (xy)^2 e^{xyz}$$

$$f_{xy} = ze^{xyz} + xyz^2 e^{xyz} \qquad f_{yz} = xe^{xyz} + x^2 yze^{xyz} \qquad f_{xz} = ye^{xyz} + xy^2 ze^{xyz}$$

Evaluating at $x = 0, y = 1$ and $z = 2$,

$$f_x(0,1,2) = 2 \qquad f_y(0,1,2) = 0 \qquad f_z(0,1,2) = 0$$

$$f_{xx}(0,1,2) = 4 \qquad f_{yy}(0,1,2) = 0 \qquad f_{zz}(0,1,2) = 0$$

$$f_{xy}(0,1,2) = 2 \qquad f_{yz}(0,1,2) = 0 \qquad f_{xz}(0,1,2) = 1$$

Hence, as $f(0,1,2) = e^0 = 1$ we find

$$f(x,y,z) = 1 + 2x + 2x^2 + 2x(y-1) + x(z-2) + \cdots$$

Another way to calculate this expansion is to make use of the adding zero trick,

$$f(x,y,z) = e^{x(y-1+1)(z-2+2)} = 1 + x(y-1+1)(z-2+2) + \frac{1}{2}\big[x(y-1+1)(z-2+2)\big]^2 + \cdots$$

Keeping only terms with two or less of x, $(y-1)$ and $(z-2)$ variables,

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

Which simplifies once more to $f(x,y,z) = 1 + 2x + 2x(y-1) + x(z-2) + 2x^2 + \cdots$.

Example 5.2.8. *Suppose $f(x,y,z) = \frac{1}{1-z^2}e^{x^2}\cos(y^3)$. Find the multivariate series expansion to quadratic order about the origin. In this case we can just multiply expansions known from calculus II, no need to do partial derivatives!*

$$\begin{aligned} f(x,y,z) &= \Big(1 + z^2 + z^4 + \ldots\Big)\Big(1 + x^2 + \frac{1}{2}x^4 + \cdots\Big)\Big(1 - \frac{1}{2}y^6 + \cdots\Big) \\ &= 1 + x^2 + z^2 + \cdots. \end{aligned}$$

5.3 critical point analysis

Let's focus on the $n = 2$ case since that is the only case we can work out in general[4]. If $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ and (x_o, y_o) is a critical point then $f_x(x_o, y_o) = f_y(x_o, y_o) = 0$ hence the taylor expansion at (x_o, y_o) provides the following representation of f:

$$f(x_o + h, y_o + k) = f(x_o, y_o) + \underbrace{\frac{1}{2}f_{xx}(x_o, y_o)h^2 + f_{xy}(x_o, y_o)hk + \frac{1}{2}f_{yy}(x_o, y_o)k^2}_{Q(h,k)} + T$$

where T is the tail of the series which factors in higher-derivative corrections. We worked out the general behaviour of a quadratic form in a previous section. Let me quote the result here: The graph of $z = Q(h, k) = ah^2 + 2bhk + ck^2$ for some constants $a, b, c \in \mathbb{R}$ can be categorized by real solutions λ_1, λ_2 of the **characteristic equation** $\lambda^2 - (a + c)\lambda + ac - b^2 = 0$. In particular, if $\lambda_1 \leq \lambda_2$ then $\lambda_1 R^2 \leq Q(h, k) \leq \lambda_2 R^2$ for all (h, k) on the circle $h^2 + k^2 = R^2$. We identify that

$$a = \frac{1}{2}f_{xx}(x_o, y_o), \quad b = \frac{1}{2}f_{xy}(x_o, y_o), \quad c = \frac{1}{2}f_{yy}(x_o, y_o).$$

Let's reason through the cases. If both λ_1, λ_2 share the same sign then we can be sure that $|Q(h, k)| >> |T|$ since T depends on third and higher order powers of the coordinates which are much smaller than quadratic powers near the origin. It follows that $0 < \lambda_1 \leq \lambda_2$ imply $f(x_o, y_o)$ is a local minimum of f. Likewise, $\lambda_1 \leq \lambda_2 < 0$ imply $f(x_o, y_o)$ is a local maximum of f. On the other hand, if $\lambda_1 < 0 < \lambda_2$ then the values of Q are sure to increase and decrease near the point of tangency in such a way that T cannot possibly squelch the behaviour and we find $f(x_o, y_o)$ is not a local extremum. The case $\lambda_1 = 0$ is not as useful since the contributions of T are dominant in the direction associated to $\lambda_1 = 0$, we could find a saddle or a minimum or maximum in such a case, so the final two cases in Theorem 5.1.13 are silenced by the tailed beast.

Put all of this together and we have a generalization of the second derivative test for functions of two variables! We need to work out the formulas for λ_1, λ_2 in our current context to make it useful. Solutions of the quadratic equation $\lambda^2 - (a + c)\lambda + ac - b^2 = 0$ are given by

$$\lambda = \frac{a + c \pm \sqrt{(a + c)^2 - 4(ac - b^2)}}{2}$$

We have $a = \frac{1}{2}f_{xx}(x_o, y_o)$, $b = \frac{1}{2}f_{xy}(x_o, y_o)$, $c = \frac{1}{2}f_{yy}(x_o, y_o)$. To reduce clutter, drop the (x_o, y_o) for the next few computations, the two's in a, b, c nicely cancel with the quadratic formula to yield:

$$\lambda_{\pm} = f_{xx} + f_{yy} \pm \sqrt{(f_{xx} + f_{yy})^2 - 4(f_{xx}f_{yy} - f_{xy}^2)}$$

Let $D = f_{xx}f_{yy} - f_{xy}^2$. We have a few cases to consider:

1. If $D < 0$ then clearly

$$|f_{xx} + f_{yy}| = \sqrt{(f_{xx} + f_{yy})^2} < \sqrt{(f_{xx} + f_{yy})^2 - 4(f_{xx}f_{yy} - f_{xy}^2)}.$$

   This inequality indicates that the radical dominates the sign of the solution; given $D < 0$ we have $\lambda_- < 0$ and $\lambda_+ > 0$. Hence, the condition $D < 0$ signifies a saddle shape for $graph(f)$.

[4] I will work special cases in the $n = 3$ case, but the general problem is too hard w/o the help of linear algebra

2. If $D > 0$ then clearly
$$|f_{xx} + f_{yy}| = \sqrt{(f_{xx} + f_{yy})^2} < \sqrt{(f_{xx} + f_{yy})^2 - 4(f_{xx}f_{yy} - f_{xy}^2)}.$$
This inequality indicates that $f_{xx} + f_{yy}$ dominates the sign of the solution; in particular:
   (a) if $f_{xx} + f_{yy} > 0$ then $\lambda_\pm > 0$ hence f attains a local minimum at (x_o, y_o)
   (b) if $f_{xx} + f_{yy} < 0$ then $\lambda_\pm < 0$ hence f attains a local maximum at (x_o, y_o)
3. If $D = 0$ then either $\lambda_+ = 0$ or $\lambda_- = 0$ hence the quadratic data is inconclusive. The function may attain a maximum, minimum, a saddle or a trough at the critical point.

Notice that in Case (2.) we can simplify the criteria a bit. If $D > 0$ then $f_{xx}f_{yy} - f_{xy}^2 > 0$ thus $0 \leq f_{xy}^2 < f_{xx}f_{yy}$. It follows that either both f_{xx} and f_{yy} are positive or both are negative. Therefore, given $D > 0$, the criteria $f_{xx} + f_{yy} > 0$ can be replaced with criteria $f_{xx} > 0$ or $f_{yy} > 0$. Likewise, given $D > 0$, the criteria $f_{xx} + f_{yy} < 0$ can be replaced with criteria $f_{xx} < 0$ or $f_{yy} < 0$.

Let us collect these thoughts for future use.

Theorem 5.3.1.

Suppose $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ is smooth and (x_o, y_o) is a critical point of f.
If $D = f_{xx}(x_o, y_o)f_{yy}(x_o, y_o) - f_{xy}(x_o, y_o)^2$ then

1. $D < 0$ implies $f(x_o, y_o)$ is not a local extrema,
2. $D > 0$ and $f_{xx}(x_o, y_o) > 0$ (or $f_{yy}(x_o, y_o) > 0$) implies $f(x_o, y_o)$ is a local minimum,
3. $D > 0$ and $f_{xx}(x_o, y_o) < 0$ (or $f_{yy}(x_o, y_o) < 0$) implies $f(x_o, y_o)$ is a local maximum.

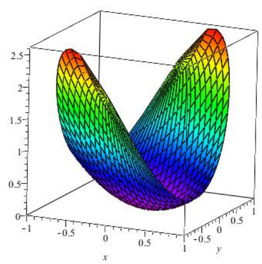
Example 5.3.2. *Suppose $f(x, y) = x^2 + 2xy + 2y^2$ then $\nabla f = \langle 2x + 2y, 2x + 4y\rangle$. The origin $(0, 0)$ is a critical point since $\nabla f(0, 0) = \langle 0, 0\rangle$. Let's use the theorem to test what type of critical point we've found. We should calculate all the second derivatives,*
$$f_{xx} = 2, \quad f_{xy} = 2, \quad f_{yy} = 4.$$
Calculate $D = f_{xx}f_{yy} - f_{xy}^2 = 8 - 4 = 4 > 0$ and note $f_{xx} = 2 > 0$ hence $f(0, 0)$ is a local minimum. The graph $z = f(x, y)$ opens upward at the origin.

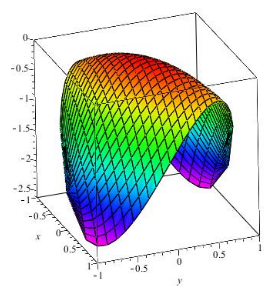
Example 5.3.3. *Suppose $f(x, y) = -x^2 + 2xy - 2y^2$ then $\nabla f = \langle -2x + 2y, 2x - 4y\rangle$. The origin $(0, 0)$ is a critical point since $\nabla f(0, 0) = \langle 0, 0\rangle$. Let's use the theorem to test what type of critical point we've found. We should calculate all the second derivatives,*
$$f_{xx} = -2, \quad f_{xy} = 2, \quad f_{yy} = -4.$$
Calculate $D = f_{xx}f_{yy} - f_{xy}^2 = 8 - 4 = 4 > 0$ and note $f_{xx} = -2 < 0$ hence $f(0, 0)$ is a local maximum. The graph $z = f(x, y)$ opens downward at the origin.

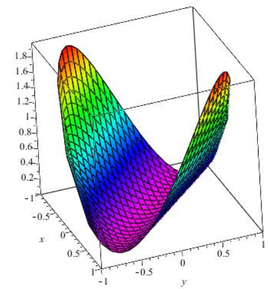
Example 5.3.4. *Suppose $f(x, y) = x^2 + 2xy + y^2$ then $\nabla f = \langle 2x + 2y, 2x + 2y\rangle$. The origin $(0, 0)$ is a critical point since $\nabla f(0, 0) = \langle 0, 0\rangle$. Let's use the theorem to test what type of critical point we've found. We should calculate all the second derivatives,*
$$f_{xx} = 2, \quad f_{xy} = 2, \quad f_{yy} = 2.$$

Calculate $D = f_{xx}f_{yy} - f_{xy}^2 = 4 - 4 = 0$. The multivariate second derivative test fails. We can easily see why in this case. Note that the formula for $f(x,y)$ factors $f(x,y) = (x+y)^2$. The graph $z = (x+y)^2$ is zero all along the line $y = -x, z = 0$ and it opens like a parabola in planes normal to this line. In other words, this is just $z = x^2$ rotated 45 degrees around the z-axis. It's a parabolic trough. Notice there are infinitely many critical points in this example.

Let us constrast the graphs of the past three examples: I plot the graphs over the unit-disk $x^2+y^2 \leq 1$ for Examples 5.3.2, 5.3.3 and 5.3.4 from left to right respective:







Example 5.3.5. *Let $f(x,y) = x^3 - 12xy + 8y^3$. Find and classify any local extrema of f.*

Solution: *begin by locating all critical points:*

$$\nabla f(x,y) = \langle 3x^2 - 12y,\ -12x + 24y^2 \rangle = \langle 0, 0 \rangle$$

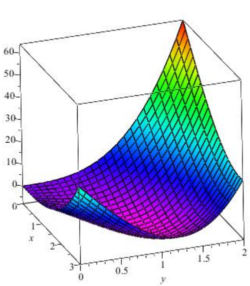
thus, $3x^2 - 12y = 0$ and $-12x + 24y^2 = 0$. Hence, $4y = x^2$ and $x = 2y^2$ from which we obtain $4y = (2y^2)^2 = 4y^4$. Therefore, $0 = y^4 - y = y(y^3 - 1)$ hence $y = 0$ or $y = 1$ and so $x = 2(0)^2 = 0$ and $x = 2(1)^2 = 2$ respectively. We find critical points $(0,0)$ and $(2,1)$. The Hessian at (x,y) is calculated:

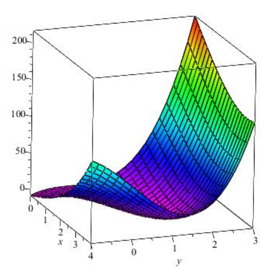
$$f_{xx} = 6x, \quad f_{yy} = 48y, \quad f_{xy} = -12 \quad \Rightarrow \quad D = 288xy - 144 = 144(2xy - 1).$$

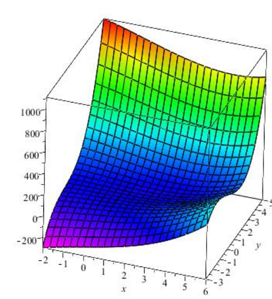
Consider the

critical point	D	f_{xx}	*conclusion*
$(0,0)$	-144	*no need*	*saddle at* $(0,0)$
$(2,1)$	432	12	$f(2,1) = -8$ *is local minimum*

Therefore, we conclude, by the second derivative test, there is only one local extrema. The local minimum value of -8 is attained at $(2,1)$. Graphically, you could easily miss this valley. Consider: the plots below are centered about $(2,1)$ and zoom out as you read from left to right:







It only gets worse as we zoom out further. Thankfully, we need not rely on graphs. I use them to check the answer, not to find it. I think the reader can appreciate why.

Example 5.3.6. *Suppose $f(x,y) = (2x - x^2)(2y - y^2)$. Find and classify any local extrema of f.*

Solution: *we must find the critical points where $\nabla f = 0$. Consider,*

$$\langle f_x, f_y \rangle = \big\langle (2-2x)(2y-y^2), (2x-x^2)(2-2y) \big\rangle$$

Factoring $2, y, x$ reveal:

$$f_x: \quad 2(1-x)y(2-y) = 0 \qquad \& \qquad f_y: \quad 2x(2-x)(1-y)$$

We must simultaneously solve the equations above. To solve $f_x = 0$ we have three cases:

(i.) *If $x = 1$ then we require $y = 1$ to solve $f_y = 0$.*

(ii.) *If $y = 0$ then we either need $x = 0$ or $x = 2$ to solve $f_y = 0$.*

(iii.) *If $y = 2$ then we either need $x = 0$ or $x = 2$ to solve $f_y = 0$.*

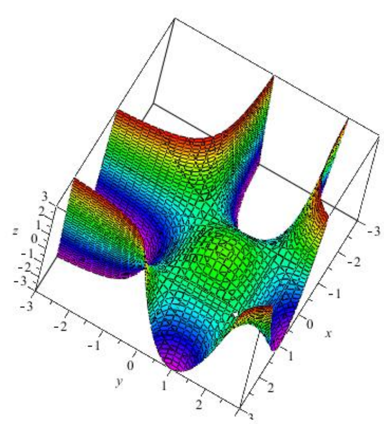
In summary, we find critical points $(1,1), (0,0), (2,0), (0,2), (2,2)$. The Hessian is derived below:

$$\begin{aligned} D(x,y) &= f_{xx}f_{yy} - f_{xy}^2 \\ &= [(-2)(2y-y^2)][(2x-x^2)(-2)] - [(2-2x)(2-2y)]^2 \\ &= 4xy(2-y)(2-x) - 16(1-x)^2(1-y)^2 \end{aligned}$$

Therefore, we find:

critical point	D	f_{xx}	*conclusion*
$(1,1)$	4	-2	*$f(1,1) = 1$ is local maximum*
$(0,0)$	-16	*no need*	*saddle at $(0,0)$*
$(2,0)$	-16	*no need*	*saddle at $(2,0)$*
$(0,2)$	-16	*no need*	*saddle at $(0,2)$*
$(2,2)$	-16	*no need*	*saddle at $(2,2)$*

The plot below uses the "zhue" option to indicate z-values by color. You can clearly see which point is $(1,1)$ and the saddle points are situated symmetrically about the point as our analysis predicted:



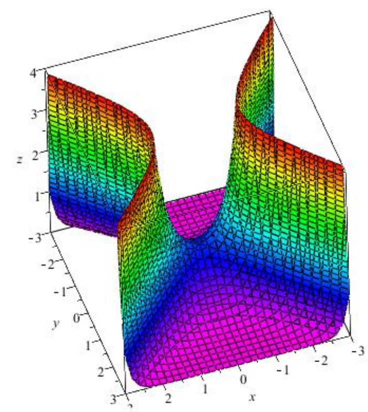
Example 5.3.7. *Suppose $f(x,y) = e^{-x^2+y^2}$ calculate $\nabla f = \langle -2xe^{-x^2+y^2}, 2ye^{-x^2+y^2}\rangle$ and note the origin is the only critical point since exponential functions are strictly positive. Once more we use the multivariate second derivative test at the origin. We need to calculate second derivatives,*

$$f_{xx} = (-2+4x^2)e^{-x^2+y^2}, \quad f_{xy} = -4xye^{-x^2+y^2}, \quad f_{yy} = (2+4y^2)e^{-x^2+y^2}$$

Hence, $f_{xx}(0,0) = -2$, $f_{xy} = 0$ and $f_{yy} = 2$. Note then that

$$D = f_{xx}f_{yy} - f_{xy}^2 = -4 < 0$$

Therefore, $f(0,0)$ is not a local extremum. The graph of $z = f(x,y)$ is saddle shaped over $(0,0)$.



Notice that in the last example it is easy to see why we find the result we did since

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

The fourth order and higher terms are very small compared to the quadratic terms near the origin hence to a good approximation the graph $z = f(x,y)$ looks like $z = 1 + y^2 - x^2$. This is the type of function we can analyze without the help of linear algebra. Let me illustrate by example.

Example 5.3.8. *Suppose $f(x,y,z) = \sin(x^2+y^2+z^2)$ then you can calculate that $\nabla f(0,0,0) = \langle 0,0,0\rangle$ hence the origin is a critical point. Applying the power series expansion for sine,*

$$f(x,y,z) = x^2+y^2+z^2 - \frac{1}{6}(x^2+y^2+z^2)^3 + \cdots$$

clearly $f(0,0,0)$ is a local minimum for f since the values clearly increase. This is clear because the quadratic terms dominate near $(0,0,0)$. On the other hand, if $g(x,y,z) = \sin(x^2+y^2-z^2)$ then

$$g(x,y,z) = x^2+y^2-z^2 - \frac{1}{6}(x^2+y^2-z^2)^3 + \cdots$$

and it is clear that the values of g both increase and decrease near $(0,0,0)$. For example, $g(x,0,0) = x^2 + \cdots$ whereas $g(0,0,z) = -z^2 + \cdots$. It follows that $g(0,0,0)$ is neither a maximum nor a minimum.

The logic used in the example above is not so easy if there are cross terms. For example, $f(x,y,z) = x^2+y^2+z^2+2xy+2xz$ has critical point $(0,0,0)$ but I wouldn't ask you to ascertain the behaviour of f at $(0,0,0)$ because we need linear algebra to understand clearly how f behaves.

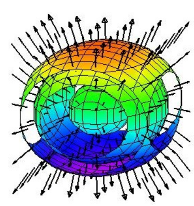
5.3.1 a view towards higher dimensional critical points*

If the function $f : dom(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ is *analytic* at $\vec{p}$ then that means it is well-approximated by its multivariate Taylor series near $\vec{p}$. For such a function f the statement $\vec{p}$ is a critical point is to say $\nabla f(\vec{p}) = 0$. It follows the taylor series at $\vec{p}$ has the form

$$\boxed{f(\vec{p} + \vec{h}) = f(\vec{p}) + Q(\vec{h}) + T}$$

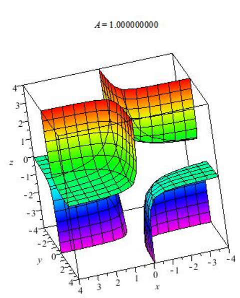
where $|T|$ is usually smaller than $|Q(\vec{h})|$. We call T the **tail** of the expansion. To judge if Q or T dominates the behaviour of f near $\vec{p}$ we must calculate the spectrum of Q. If the spectrum consists of all positive eigenvalues then $f(\vec{p})$ is a local minimum. If the spectrum consists of all negative eigenvalues then $f(\vec{p})$ is a local maximum. If the spectrum consists of both positive and negative eigenvalues then $f(\vec{p})$ is a not a local extrema. If zero is an eigenvalue of Q then further analysis beyond quadratic data may be needed to ascertain the nature of the critical point.

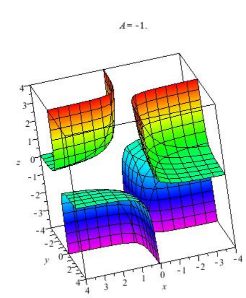
Incidentally, there is a way to visualize maxima for functions of three variables in terms of level surfaces. It's the analogue of using two-dimensional contour plots for finding max/min of a three-dimensional graph. For example, the function $f(x, y, z) = x^2 + 2y^2 + 3z^2$ has level surfaces which are ellipsoids centered at the origin.



You can *see* how the ellipsoids enfold the origin. The larger ellipsoids correspond to higher levels and there does not exist a negative level surface. Intuitively it is clear that $f(0, 0, 0) = 0$ is a local minimum of the function f near the origin. I don't teach this as a method because few of us are capable of mastering such visualization with any reliability. On the other hand, contour plots are extremely useful because our minds are much more adept at handling two-dimensional data.

Consider $f(x, y, z) = xyz$. The origin $(0, 0, 0)$ is a critical point. Plotted below are the level surfaces $xyz = 1$ and $xyz = -1$. In this case $f(0, 0, 0)$ is not a local extreme.





Intuitively, if we have a critical point where f has a trivial quadratic term and a nontrivial cubic term then I expect it is not a local extreme. On the other hand, if the first nontrivial term beyond the constant term is fourth order then max/min or saddle-type points ought to exist. For example,

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

5.4 closed set method

The analog of the extreme value theorem of first semester calculus is given below.

Theorem 5.4.1.

If D is a closed and bounded subset of $\mathbb{R}^2$ then a continuous function $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ attains a global maximum and global minimum somewhere in D.

To say D is closed means that it has edges which are not fuzzy in our usual contexts. There is a better *topological* method to describe such terms[5] but I leave that for another course. To say D is bounded simply means we can find a point $(x_o, y_o) \in D$ and $\epsilon > 0$ such that $D \subset B_\epsilon(x_o, y_o) = \{(x, y) \in \mathbb{R}^2 \mid \sqrt{(x - x_o)^2 + (y - y_o)^2} < \epsilon\}$. In other words, D is bounded if there exists a finite open disk which properly contains D. Or, in plain-English, if you can draw a circle big enough to enclose D. These terms don't usually bother students in practice, if anything, the attempt to define them here is the most troubling part. Common examples of closed and bounded sets are: disks, rectangles, areas bounded by curves which we studied in first semester calculus, polygons regular or otherwise.

The extreme value theorem told us that the maximum and minimum values of a continuous function on a closed interval $[a, b]$ were attained somewhere in $[a, b]$. That data motivated the **closed-interval test** which said, given a continuous function on a closed interval $[a, b]$,

(i.) find any critical numbers for f in the interval

(ii.) evaluate the function at critical numbers and endpoints

(iii.) select the minimum and maximum from the values found in step **(ii.)**

The theorem that follows is the analog of the closed interval test for functions of several variables.

Theorem 5.4.2.

Suppose D is a closed and bounded subset of $\mathbb{R}^2$ and $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ is a continuous function. Extreme values of f on D may be found as follows:

(i.) find the value of f at any critical points in the interior of D

(ii.) find any extreme values for f on the boundary of D

(iii.) select the minimum and maximum on D from the values found in steps **(i.)** and **(ii.)**

I leave the proof of this assertion to another course. That said, it is useful to think about the two cases[6]. As we consider closed and bounded D it follows $D = \text{int}(D) \cup \partial D$. The boundary ∂D is the edge whereas $\text{int}(D)$ is D with the edge removed. The basic idea is that we can apply the theory of local extrema to the interior; that is, use the second derivative test to classify any critical points in the interior. On the other hand, the boundary is a curve or set of curves where we might apply the method of Lagrange multipliers. However, sometimes the boundary admits a better solution in terms of a parametric formulation. We'll see that techinque in the examples to follow below.

[5] indeed, you may learn later that closed and bounded is synonomous with compact

[6] I should admit, I assume f is continuously differentiable in this discussion as to avoid certain pathological cases.

Example 5.4.3. *Let $f(x,y) = x^4 + y^4 - 4xy + 1$. Find the maximum and minimum of $f(x,y)$ on the half-disk $H = \{(x,y) \mid x^2 + y^2 \leq 4,\ y \geq 0\}$.*

Solution: *we begin by searching for local maxima and minima. Consider,*

$$\nabla f(x,y) = \langle 4x^3 - 4y, 4y^3 - 4x \rangle = \langle 0, 0 \rangle \quad \Rightarrow \quad y = x^3 \quad \& \quad x = y^3.$$

It follows $x^9 = x$. This is solved by factoring,

$$0 = x^9 - x = x(x^8 - 1) = x(x^4 + 1)(x^4 - 1) = x(x^4 + 1)(x^2 + 1)(x + 1)(x - 1)$$

Hence $x = 0, 1, -1$ and we find critical points $(0,0), (1,1)$ and $(-1,-1)$. Note, $(-1,-1) \notin H$ hence we ignore it. Notice $f_{xx} = 12x^2$ and $f_{yy} = 12y^2$ and $f_{xy} = -4$ give Hessian $D = 144x^2y^2 - 16$. Hence,

critical point	D	f_{xx}	*conclusion*
$(0,0)$	-16	*no need*	$f(0,0) = 1$ *is a saddle*
$(1,1)$	128	12	$f(1,1) = -1$ *is local minimum*

Logically, we do not need the Hessian or the analysis of the table above (In include it here for curiousity alone). It suffices to calculate $f(0,0), f(1,1)$ and $f(-1,-1)$ for future comparison to extreme values on the boundary ∂H. There are two cases in the boundary:

(i.) *the diameter of the half-circle boundary is given by $y = 0$ and $-2 \leq x \leq 2$. Let $g(x) = f(x,0) = x^4 + 1$. We analyze the behaviour of g on $[-2,2]$ by the closed interval test. Notice $g'(x) = 4x^3$ hence $x = 0$ is the only critical number. Observe,*

$$g(-2) = 17, \quad g(0) = 1, \quad g(2) = 17.$$

Thus, $f(-2,0) = 17$, $f(2,0) = 17$ are two new candidates we should consider as we seek the extreme values of f on H.

(ii.) *curved part of the half-circle has parameterization $x = 2\cos t$ and $y = 2\sin t$ for $0 \leq t \leq \pi$. Let $h(t) = f(2\cos t, 2\sin t)$ which gives $h(t) = 16(\sin^4 t + \cos^4 t - \sin t \cos t) + 1$. We find extrema of h on $[0,\pi]$ by the closed interval test. Consider,*

$$h'(t) = 16(-4\sin t\cos^3 t + 4\cos t\sin^3 t - \cos^2 t + \sin^2 t).$$

thus $h'(t) = 0$ yields:

$$-(\cos(t) - \sin(t))(\sin(t) + \cos(t))(4\sin(t)\cos(t) + 1) = 0$$

or,

$$\tan t = 1, \quad \tan t = -1, \quad \sin(2t) = -1/2.$$

We seek solutions on $[0,\pi]$. Observe, $t = \pi/4$ give $\tan(\pi/4) = 1$. Also, $t = 3\pi/4$ gives $\tan(3\pi/4) = -1$. Solutions of $\sin(2t) = -1/2$ are $2t = 7\pi/6$ and $2t = 11\pi/6$ hence $t = 7\pi/12$ and $t = 11\pi/12$. These four values of t yield points:

$$(\sqrt{2}, \sqrt{2}), \quad (-\sqrt{2}, \sqrt{2}), \quad (2\cos(7\pi/12), 2\sin(7\pi/12)), \quad (2\cos(11\pi/12), 2\sin(11\pi/12)).$$

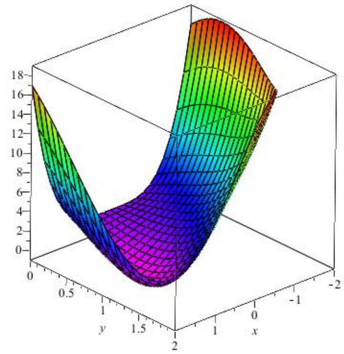
Or, approximately,

$$(1.41, 1.41), \quad (-1.41, 1.41), \quad (-0.52, 1.93), \quad (-1.93, 0.52).$$

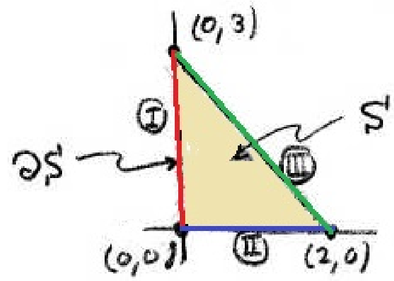
These yield (approximate) values:

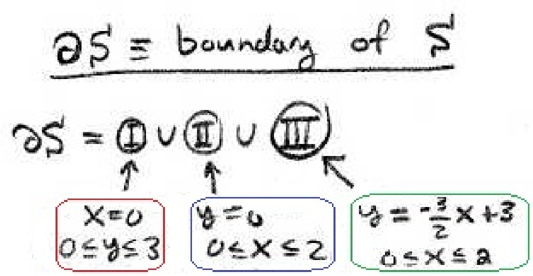
$$f(1.41, 1.41) = 0.95, \ f(-1.41, -1.41) = 16.9, \ f(-0.52, 1.93) = 19.0, \ f(-1.93, 0.52) = 19.0.$$

Of the nine possible extremal points, we observe the minimum value is -1 *which is attained at* $(1,1)$ *and the maximum value is approximately* 19 *which is attained at* $(2\cos(7\pi/12), 2\sin(7\pi/12))$ *and* $(2\cos(11\pi/12), 2\sin(11\pi/12))$*. The graph below illustrates our analysis:*



Example 5.4.4. *Consider* $f(x,y) = 1 + 4x - 5y$*. Find the absolute extrema of* f *on the set* S *pictured below: I have taken the step of labeling the edges for convenience of discussion.*





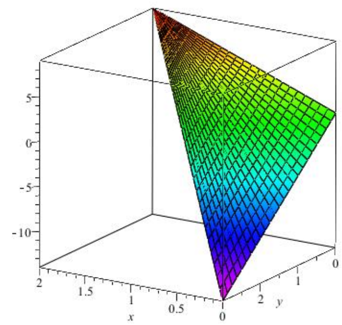
To begin note that $\nabla f(x,y) = \langle 4, -5\rangle \neq 0$ *thus there is no local extrema in the interior of* S*. We need only consider* ∂S

(I.) $x = 0$ *and* $y \in [0,3]$*. Let* $g(y) = f(0,y) = 1 - 5y$*. Note* $g'(y) = -5 \neq 0$ *hence the closed interval test need only consider* $g(0) = 1$ *and* $g(3) = -14$*. For future reference, we should remember to consider* $f(0,0) = 1$ *and* $f(0,3) = -14$ *as possible extrema on* S*.*

(II.) $y = 0$ *and* $x \in [0,2]$*. Let* $h(x) = f(x,0) = 1 + 4x$*. Note* $h'(x) = 4 \neq 0$ *hence the closed interval test faces no critical numbers. We consider the endpoints;* $h(0) = f(0,0) = 1$ *and* $h(2) = f(2,0) = 9$*. This shows* $1 \leq f(x,0) \leq 9$ *for* $0 \leq x \leq 2$*.*

(III.) $y = -\frac{3}{2}x+3$ *for* $x \in [0,2]$*. Let* $l(x) = f(x, -\frac{3}{2}x+3) = -14+\frac{23}{2}x$*. Once again,* $l'(x) = \frac{23}{2} \neq 0$ *hence* $l(0) = -14$ *and* $l(2) = 9$ *are possible extrema for* l *on* $[0,2]$*. Once more, we are prompted to consider* $f(0,3) = -14$ *and* $f(2,0) = 9$ *as possible extreme values for* S*.*

In summary, only the vertices of the triangular region appear as possible extrema and we conclude the **maximum of** f **on** S *is* 9 *which is attained at* $(2,0)$ *and the* **minimum of** f **on** S *is* -14 *which is attained at* $(0,0)$*. Geometrically, our analysis is easy to see: here I plot* $z = 1 + 4x - 5y$ *for* $(x,y) \in S$



The result above generalizes to any closed polygon in the plane. If we find extrema of $f(x,y) = ax+by+c$ for some a,b with $ab \neq 0$ then we need only consider vertices of the polygon. This well-known result is often taught in high-school algebra as **linear programming**. If we consider the higher-dimensional problem with linear constraints in three variables which enclose some polyhedral surface then almost the same analysis shows the vertices must provide extreme values. Graphically, three or more variables is difficult, however if you take a course in **Operations Research** it is likely you will learn the **simplex method** which provides an algebraic method to find the vertices through the introduction of so-called **slack variables**.

Example 5.4.5. *Let $f(x,y) = 2x^3 + y^4$. Find the absolute extrema of f on the unit-disk $D = \{(x,y) \mid x^2+y^2 \leq 1\}$.*

Solution: *note $\nabla f(x,y) = \langle 6x^2, 4y^3 \rangle$ hence $(0,0)$ is the only critical point of f. Note $f(0,0)=0$. Continuing, we analyze the boundary ∂D where $y^2 = 1-x^2$ hence*

$$f\big|_{\partial D}(x,y) = 2x^3 + (1-x^2)^2 = \underbrace{x^4+2x^3-2x^2+1}_{\text{let this be } g(x)}.$$

Note,

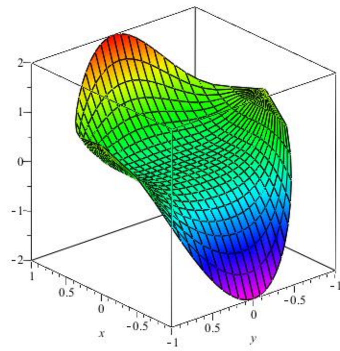
$$g'(x) = 4x^3+6x^2-4x = 2x(2x^2+3x-2) = 2x(2x-1)(x+2).$$

Thus $x = 0, 1/2, -2$ are critical numbers for g. Note $x^2+y^2=1$ yields points $(0,\pm 1)$, $(1/2, \pm\sqrt{3}/2)$ whereas $x=-2$ gives no solutions in the unit-circle. We calculate,

$$f(0,\pm 1) = 2(0)^3 + (\pm 1)^4 = 1 \quad \& \quad f(1/2,\pm\sqrt{3}/2) = 2(1/2)^3 + (\pm\sqrt{3}/2)^4 = 13/16.$$

A subtle point[7] which matters to this problem, y is not a differentiable function of x on an open set centered about $x = \pm 1$. Note the points $(\pm 1, 0)$ are on the unit-circle and we obtain $f(1,0)=2$ and $f(-1,0)=-2$. We find the maximum of f is 2 is attained at $(1,0)$. Whereas the mininum of f is -2 which is attained at $(-1,0)$. Below I plot $z = f(x,y)$ for $(x,y) \in D$:

[7] this is a good example of why you ignore the implicit function theorem to your own peril. Look at my old notes to see I speak from experience here.



To deal with ∂D in the problem above we could have studied $\nabla f = \lambda \nabla g$ for $g(x,y) = x^2 + y^2$ or we could have set $x = \cos t$ and $y = \sin t$ and sought out extreme values of $h(t) = 2\cos^3 t + \sin^4 t$. There are several ways to analyze the boundary in a given problem.

5.5 Problems

Problem 124 Suppose that the temperature T in the xy-plane changes according to

$$\frac{\partial T}{\partial x} = 8x - 4y \qquad \& \qquad \frac{\partial T}{\partial y} = 8y - 4x.$$

Find the maximum and minimum temperatures of T on the unit circle $x^2+y^2 = 1$. This time use the method of Lagrange multipliers. Hopefully we find agreement with Problem 107.

Problem 125 Use the method of Lagrange multipliers to find the point on the plane $x+2y-3z = 10$ which is closest to the point $(8,8,8)$.

Problem 126 Apply the method of Lagrange multipliers to solve the following problem: Let a, b be constants. Maximize xy on the ellipse $b^2x^2 + a^2y^2 = a^2b^2$.

Problem 127 Apply the method of Lagrange multipliers to solve the following problem: Find the distance from $(1,0)$ to the parabola $x^2 = 4y$.

Problem 128 Apply the method of Lagrange multipliers to solve the following problem: Suppose the base of a rectangular box costs twice as much per square foot as the sides and the top of the box. If the volume of the box must be $12ft^3$ then what dimensions should we build the box to minimize the cost? *[Please state the dimensions of the base and altitude clearly. Include a picture in your solution to explain the meaning of any variables you introduce, thanks!]*

Problem 129 Taking a break from the method of Lagrange. Assume a, b, c are constants: Show that the surfaces $xy = az^2$, $x^2 + y^2 + z^2 = b$ and $z^2 + 2x^2 = c(z^2 + 2y^2)$ are mutually perpendicular.

Problem 130 Apply the method of Lagrange multipliers to derive a formula for the distance from the plane $ax + by + cz + d = 0$ to the origin. If necessary, break into cases.

Problem 131 Suppose you want to design a soda can to contain volume V of soda. If the can must be a right circular cylinder then what radius and height should you use to minimize the cost of producing the can? *assume the cost is directly proportional to the surface area of the can*

Problem 132 Find any extreme values of xy^2z on the sphere $x^2 + y^2 + z^2$. *note the sphere is compact and the function* $f(x, y, z) = xy^2z$ *is continuous so this problem will have at least two interesting answers*

Problem 133 Again, breaking from optimization, this problem explores a concept some of you have not yet embraced. Find the point(s) on $x^2 + y^2 + z^2 = 4$ which the curve $\vec{r}(t) = \langle \sin(t), \cos(t), t\rangle$ intersects.

Problem 134 Consider $f(x, y) = x^3 - 3x - y^2$. Find any critical points for f and use the second derivative test for functions of two variables to judge if any of the critical points yield local extrema.

Problem 135 Consider $f(x, y) = x^2 - y^2$. Find any critical points for f and use the second derivative test for functions of two variables to judge if any of the critical points yield local extrema.

Problem 136 Consider $f(x, y) = x^3 + y^3 - 3xy$. Find any critical points for f and use the second derivative test for functions of two variables to judge if any of the critical points yield local extrema.

Problem 136+i An armored government agent decides to investigate a disproportionate use of electricity in a gated estate. Foolishly entering without a warrant he find himself at the mercy of Ron Swanson (at $(1, 0, 0)$), Dwight Schrute (at $(-1, 1, 0)$) and Kakashi (in a tree at $(1, 1, 3)$). Supposing Ron Swanson inflicts damage at a rate of 5 units inversely proportional from the square of his distance to the agent, and Dwight inflicts constant damage at a rate of 3 in a sphere of radius 2. If Kakashi inflicts a damage at a rate of 5 units directly proportional to the square of his distance from his location (because if you flee it only gets worse the further you run as he attacks you retreating) then where should you assume a defensive position as you call for back-up? What location minimizes your damage rate? Assume the ground is level and you have no jet-pack and/or antigravity devices.

Problem 137 Find global extrema for $f(x, y) = exp(x^2 - 2x + y^2 - 6y)$ on the closed region bounded by $x^2/4 + y^2/16 = 1$.

Problem 138 Find the maximum and minimum values for $f(x, y) = x^2 + y^2 - 1$ on the region bounded by the triangle with vertices $(-3, 0), (1, 4)$ and $(0, -3)$.

Problem 139 Find the maximum and minimum values for $f(x, y) = x^4 - 2x^2 + y^2 - 2$ on the closed disk with boundary $x^2 + y^2 = 9$.

Problem 140 Find the multivariate power series expansion for $f(x, y) = ye^x \sin(y)$ centered at $(0, 0)$

Problem 141 Expand $f(x, y, z) = xyz + x^2$ about the center $(1, 0, 3)$.

Problem 142 Given that $f(x, y) = 3 + 2x^2 + 3y^2 - 2xy + \cdots$ determine if $(0, 0)$ is a critical point and is $f(0, 0)$ a local extremum.

Problem 143 Use Clairaut's Theorem to show it is impossible for $\vec{F} = \langle y^3 + x, x^2 + y\rangle = \nabla f$.

Problem 144 Suppose $\vec{F} = \langle P, Q\rangle$ and suppose $P_y = Q_x$ for all points in some subset $U \subseteq \mathbb{R}^2$. Does it follow that $\vec{F} = \nabla f$ on U for some scalar function f? Discuss.

Hint: the polar angle θ has total differential $d\theta = d(\tan^{-1}(y/x)) = \frac{y}{x^2+y^2}dx - \frac{x}{x^2+y^2}dx$, think about the example $\vec{F} = \langle \frac{y}{x^2+y^2}, -\frac{x}{x^2+y^2} \rangle$. This function has domain $U = \mathbb{R}^2 - \{(0,0)\}$, can you find f such that $\vec{F} = \nabla f$ on all of U?

Problem 145 We say $U \subseteq \mathbb{R}^n$ is path-connected iff any pair of points in U can be connected by a polygonal-path (this is a path made from stringing together finitely many line-segments one after the other) . Show that if $\nabla f = 0$ on a path-connected set $U \subseteq \mathbb{R}^n$ then $f(\vec{x}) = c$ for each $\vec{x} \in U$. You may use the theorem from calculus I which states that if $f'(t) = 0$ for all t in a connected domain then $f = c$ on that domain.

Problem 146 Show that if $\nabla f = \nabla g$ on a path-connected set $U \subseteq \mathbb{R}^n$ then $f(\vec{x}) = g(\vec{x}) + c$ for each $\vec{x} \in U$. *Hint: you can use Problem 145.*

Problem 147 Prove the mean-value theorem for functions $\mathbb{R}^n \xrightarrow{f} \mathbb{R}$. In particular, show that if f is differentiable at each point of the line-segment connecting $\vec{P}$ and $\vec{Q}$ then there exists a point $\vec{C}$ on the line-segment $\overline{PQ}$ such that $\nabla f(\vec{C}) \bullet (\vec{Q} - \vec{P}) = f(\vec{Q}) - f(\vec{P})$.
Hint: parametrize the line-segment and construct a function on $\mathbb{R}$ to which you can apply the ordinary mean value theorem, use the multivariate chain-rule and win.

Problem 148 The method of characteristics is one of the many calculational techniques suggested by the total differential. The idea is simply this: given $dx/dt = f(x,y)$ and $dy/dt = g(x,y)$ we can solve both of these for dt to eliminate time. This leaves a differential equation in just the cartesian coordinates x, y and we can usually use a separation of variables argument to solve for the level curves which the solutions to $dx/dt = f(x,y)$ and $dy/dt = g(x,y)$ parametrize. Use the technique just described to solve

$$\frac{dx}{dt} = -y \qquad \& \qquad \frac{dy}{dt} = x.$$

Problem 149 Suppose that the force $\vec{F} = q(\vec{v} \times \vec{B} + \vec{E})$ is the net-force on a mass m. Furthermore, suppose $\vec{B} = B\,\widehat{z}$ and $\vec{E} = E\,\widehat{z}$ where E and B are constants. Find the equations of motion in terms of the initial position $\vec{r}_o = \langle x_o, y_o, z_o \rangle$ and velocity $\vec{v}_o = \langle v_{ox}, v_{oy}, v_{oz} \rangle$ by solving the differential equations given by $\vec{F} = m\frac{d\vec{v}}{dt}$. If $E = 0$ and $v_{oz} = 0$ then find the radius of the circle in which the charge q orbits.
Hint: first solve for the velocity components via the technique from Problem 148 then integrate to get the components of the position vector.

Problem 150 Suppose objective function $f(x,y)$ has an extremum on $g(x,y) = 0$. Show that F defined by $F(x,y,\lambda) = f(x,y) - \lambda g(x,y)$ recovers the extremum as a critical point. From this viewpoint, the adjoining of the multiplier converts the constrained problem in n-dimensions to an unconstrained problem in $(n+1)$-dimensions (you can easily generalize your argument to $n > 2$).

Chapter 6

integration

In this chapter we study integration of functions of two, three or more variables. The integral is a continuous summation. In particular, we can begin with a finite summation which approximates some quantity. However, as the quantity we consider such as area, volume, mass etc. depends on several variables we have to use a sum which covers some area or volume which describes the possible values of the variables. It is convenient to write such a sum in terms of a mesh which labels each approximating region in terms of the variables. For example, if there are two variables the the approximation is naturally written as a double sum $\sum_i \sum_j$ whereas if there are three variables then we face $\sum_i \sum_j \sum_k$. If we allow the number of approximating areas, or volumes, etc. to shrink to zero as we take the number of such approximating objects to infinity then this brings us to the integral. This is in direct analogy with our development of the single-variate integral from the Riemann sum.

Let me briefly describe the structure of this chapter. In the first section I give the definitions in terms of multiple summations and we detail the properties of multiple integrals. Significantly, we also cover Fubini's wonderful theorems which allow us to calculate multiple integrals by simple iteration of ordinary single-variate integrals. The first section concludes by examining a variety of applications to area, volume and net-quantity as seen from integrals over various area or volume densities. In Section 2 we study double integrals over TYPE I and II Cartesian domains. In Section 3 we study triple integrals with Cartesian coordinates. Section 4 is largely qualitative, we seek to describe the motivation for the change of variables theorem for multiple integrals. Then in Sections 5 and 6 we study the analog of u-substitution for double and triple integrals. Coordinate change is an important tool going forward as the choice of the right coordinate system can sometimes reduce the computational complexity of a problem by a great measure. Finally, in Section 8 we introduce a direct geometric method to understand the structure of dA or dV in non-Cartesian coordinates. In addition, a novel construction called the **wedge product** is introduced and we see how it recovers the calculations of determinants in a simple algebraic fashion.

6.1 definition and interpretations of integration

We begin by defining the double and triple integrals over rectangular domains as the natural extension of the single-variate Riemann sum. Recall,

$$\int_a^b f(x)\,dx \equiv \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\Delta x \qquad : \qquad \Delta x = \frac{b-a}{n}$$

In other words, the integral is an $f(x)$-weighted sum. Of course, this represents the signed-area. But, the essence of the formula is that the integral is a continuous summation. In view of this observation, the following definitions are natural:

Definition 6.1.1.

integrals are defined as the limit of a weighted sum of f

$$\iint_R f(x,y)\,dA \equiv \lim_{n\to\infty}\lim_{k\to\infty}\sum_{i=1}^{n}\sum_{j=1}^{k} f(x_i^*,y_j^*)\Delta x\Delta y$$

$$R=[a,b]\times[c,d] \text{ and } \Delta x=\frac{b-a}{n} \text{ while } \Delta y=\frac{d-c}{k}$$

$$\iiint_B f(x,y,z)\,dV \equiv \lim_{m\to\infty}\lim_{n\to\infty}\lim_{k\to\infty}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{k} f(x_i^*,y_j^*,z_l^*)\Delta x\Delta y\Delta z$$

$$B=[a,b]\times[c,d]\times[p,q] \text{ and } \Delta x=\frac{b-a}{m}, \Delta y=\frac{d-c}{n}, \Delta z=\frac{q-p}{k}$$

We note that in Cartesian coordinates $dA = dx\,dy =$ infinitesimal area element in xy-plane. $dV = dx\,dy\,dz =$ infinitesimal volume element. As in calculus I and II, the sample points are chosen randomly and the method in which they are chosen is washed away in the limit. In practice the limit is rarely seen, instead the F.T.C or evaluation rule and here the Fubini theorem will keep us from ever using the definition directly[1]

Let me expand on the specialization of the definition offered above. I have stated the definition in terms of rectangular regions, but in general we might like to calculate integrals over more general regions. For example, we might like to integrate $f(x,y)$ over a disk, or $f(x,y,z)$ over an ellipsoid. I will not attempt to write the multiple Riemann sum for an integral over such a region, however, I will make some unjustified claims which relate the rectangular region integrals to the more general type. The basic idea is this: if $S = S_1 \cup S_2$ then $\int_S f = \int_{S_1} f + \int_{S_2} f$ where I intend this notation to include integrals over areas, volumes and even n-volumes for $n > 3$. In words, the integral over some region is given by adding the integrals over subregions whose union forms the total region. It is geometrically evident that a general region can always be written as a union of rectangular regions. This is not too hard to **see** in $n = 2, 3$, however, it is also clear the union may need to be over an infinity of rectangular regions. If we suppose the integration region is closed and bounded[2] then analysis beyond this course verifies what we have already *seen*. We can arrange weighted sums over non-rectangular regions as sums of rectangular regions. Furthermore, as we refine the partition

[1] THANKFULLY!
[2] this makes the region compact

of the region into finer and finer subregions the approximate sums will coverge[3] to a unique value which we call the **integral**. Fortunately, the subtlety I describe here has little to do with our aims in this chapter. The properties and theorems which we soon discover allow us to trade the direct computation of multiple Riemann sums for simple algebraic maniupulation. Moreover, when those methods fail, in this modern age, we may rely on numerical techniques for problems which defy algebraic methods.

Several properties of the integral follow directly from the properties of the limit itself: let R be a closed and bounded region in what follows below:

Proposition 6.1.2.

$$\iint_R [f(x,y)+g(x,y)]dA = \iint_R f(x,y)dA + \iint_R g(x,y)dA$$
$$\iint_R cf(x,y)dA = c\iint_R f(x,y)dA$$
$$f(x,y) \geq g(x,y)\,\forall (x,y) \in R \quad \Rightarrow \quad \iint_R f(x,y)dA \geq \iint_R g(x,y)dA$$

Likewise for $f(x,y,z)$ and $g(x,y,z)$ over a closed and bounded solid region. We assume that f, g are continuous *almost everywhere*[4]. Meaning we can integrate $f(x,y)$ if it has a finite number of curve discontinuities, or $f(x,y,z)$ if it has a family number of planar discontinuities. We just chop the integral into a finite number of regions on which f is continuous. The theorem below was known to Cauchy for continuous f in early 19th century.

Theorem 6.1.3. *Fubini's Theorem(weak form) :*

Let $R = [a,b] \times [c,d]$ and let f be a mostly continuous function of $f(x,y)$

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

where the expression on the RHS are "iterated integrals" which you work inside out, treating the outside variable as a constant to begin.

Example 6.1.4. *Let* $f(x,y) = \sin(x) + y$ *and* $R = [0,\pi] \times [0,2]$ *which means* $(x,y) \in R$ *iff* $0 \leq x \leq \pi$ *and* $0 \leq y \leq 2$. *Integrate* f *over* R.

[3]here is where the analysis is needed both here and in calculus I where we were also vague on this point if you do some soul searching. Measure theory makes this process careful and general.

[4]one should study measure theory where this is given a precise meaning, we leave that to a later course in analysis

$$\begin{aligned}
\iint_R f(x,y)dA &= \int_0^{\pi}\left(\int_0^2 [\sin(x)+y]dy\right)dx && : (\dots) \textit{ added to emphasize order of } \int \\
&= \int_0^{\pi} [y\sin(x) + \frac{1}{2}y^2]\Big|_0^2 dx && : \textit{ note } \sin(x) \textit{ is a constant w.r.t. the } dy \textit{ integration} \\
&= \int_0^{\pi} [2\sin(x)+2]dx \\
&= -2\cos(x)|_0^{\pi} + 2x|_0^{\pi} \\
&= -2\cos\pi + 2\cos(0) + 2\pi \\
&= \boxed{4+2\pi}
\end{aligned}$$

A good exercise for the reader: calculate $\int_0^2 \int_0^{\pi} (\sin(x)+y)dx\,dy$ and compare with the result above.

Example 6.1.5. *Let $R = \{(x,y) \mid 0 \leq x \leq \frac{\pi}{2},\ 0 \leq y \leq 2\}$. Consider the integral of $y\cos(xy)$ over R.*

$$\begin{aligned}
\iint_R y\cos(xy)dA &= \int_0^2\left(\int_0^{\pi/2} y\cos(xy)dx\right)dy && : \int \cos(ax)dx = \frac{1}{a}\sin(ax)+c \\
&= \int_0^2 \left(\sin(xy))\Big|_0^{\pi/2}\right)dy \\
&= \int_0^2 \left(\sin\left(\frac{\pi y}{2}\right) - \sin(0)\right)dy \\
&= \frac{-2}{\pi}\cos\left(\frac{\pi y}{2}\right)\Big|_0^2 \\
&= \frac{-2}{\pi}\Big(\cos(\pi) - \cos(0)\Big) \\
&= \boxed{\frac{4}{\pi}}
\end{aligned}$$

Remark 6.1.6.

Notice, if we had integrated first over dy then dx in the preceding example then the calculation would have required integration by parts in the dy integration. The point? Swapping the order of the integration can change the difficulty of the calculation.

6.1.1 interpretations of integrals

In this section thus far we have gained some basic experience in **how** to calculate a multivariate integral. In this subsection we turn to the question of **what** the integral represents. The answer for a given integral is far from unique. We must be prepared to think of the integral as a multifaceted tool which solves more than one problem. Geometrically, integration finds signed areas, volumes and more generally n-volumes which defy direct visualization. Physically, integration of density with respect to some quantity over some space yields the total amount of the quantity in the space. I used the term *space* to be deliberately vague, it might be one, two, or three or even n-dimensional.

Let us begin with the basic interpretation:

Geometry: the double integral of f(x,y) over R is the volume of the solid bounded by $z = f(x, y)$ and $z = 0$ for $(x, y) \in R$. The theorem of Fubini can be seen as merely saying you can slice up the volume along x or y crossections. Infinitesimally

$$dV = \underbrace{(Z_{top} - Z_{base})}_{\text{height of box}} \underbrace{dx\, dy}_{\text{area of box}}$$

So if $Z_{base} = 0$ and $Z_{top} \geq 0$ then we get the volume, however as in $\int_a^b f(x)dx$ we count volume below the xy-plane as negative so the integral calculates the "signed" volume.

Example 6.1.7. *Let* $B = \{f(x, y, z) | 0 \leq x \leq a, 0 \leq y \leq b, 0 \leq z \leq c\}$

$$\begin{aligned}
\iiint dV &= \int_0^c \int_0^b \int_0^a dx\, dy\, dz \\
&= \int_0^c \int_0^b x\Big|_0^a dy\, dz \\
&= \int_0^c \left(\int_0^b ady \right) dz \\
&= \int_0^c abdz \\
&= \boxed{abc = V}
\end{aligned}$$

If we integrate 1 over B we find the volume of B. Likewise if we integrate 1 *over a rectangle* $R \subset \mathbb{R}^2$ *we obtain the area.*

Example 6.1.8. *Let* $B = [0, 1] \times [0, 2] \times [0, 3]$, *Let* $\rho = \frac{dm}{dV} = xyz$. *Consider, note, generally the step made in the second equality is only allowed if the integrand f factors into a product of functions of one variable:* $f(x, y, z) = f_1(x)f_2(y)f_3(z)$,

$$\begin{aligned}
\iiint_B xyzdV &= \int_0^3 \int_0^2 \int_0^1 xyz\, dx\, dy\, dz \\
&= \int_0^3 zdz \int_0^2 ydy \int_0^1 xdx \\
&= \left(\frac{1}{2} z^2 \Big|_0^3 \right) \left(\frac{1}{2} y^2 \Big|_0^2 \right) \left(\frac{1}{2} x^2 \Big|_0^1 \right) \\
&= \frac{1}{8} (3)^2 (2)^2 \\
&= \boxed{\frac{9}{2}}
\end{aligned}$$

What is the meaning of such an integration? Well, if ρ *denotes mass density then* $\rho = dm/dV$ *and it follows* $\rho\, dV = dm$. *Therefore, an object occupying the space B with density* $\rho = xyz$ *has mass m as calculated below:*

$$m = \int_B dm = \iiint_B \rho dV = \frac{9}{2}.$$

Or you could interpret it as $\rho = dq/dV$ where dq is the tiny bit of charge in the tiny volume dV hence the total charge in B is $q = \int \rho dV = 9/2$. I'm sure you could imagine other densities.

Example 6.1.9. *Another interpretataion of $\iint_R f(x,y)dA$is that $f(x,y)$ represents an area density. So say $f(x,y) = \sigma(x,y)$*

$$\sigma(x,y) = \frac{dq}{dA} \quad \Rightarrow \quad q = \iint_R \sigma(x,y)dA = \text{ charge on the planar region } R.$$
$$\sigma(x,y) = \frac{dm}{dA} \quad \Rightarrow \quad m = \iint_R \sigma(x,y)dA = \text{ mass of the rectangle } R.$$

Students often insist that $\int_a^b f(x)\,dx$ represents a signed-area. This is true, but, it is not the only interpretation. Consider:

Example 6.1.10. *Another interpretation of $\int_a^b f(x)dx$ is that $f(x)$ reoresents a linear density. So say $f(x) = \lambda(x)$ and,*

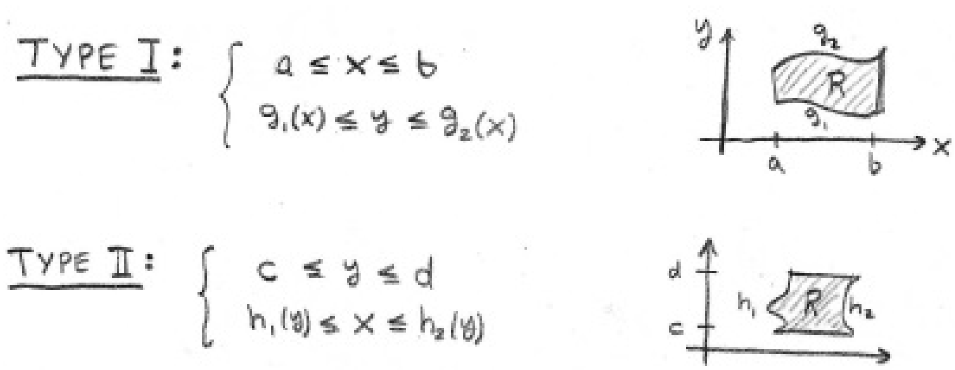
$$\lambda(x) = \frac{dq}{dx} \quad \Rightarrow \quad q = \int_a^b \lambda(x)dx \qquad \& \qquad \lambda = \frac{dm}{dx} \Rightarrow m = \int_a^b \lambda(x)dx.$$

Remark 6.1.11.

Linear density is more exciting once we know about line-integrals. At the moment, we just have the technology to calculate the total amount of some substance whose density is given along a line-segment. The integral with respect to arclength we discuss later will allow us to generalize such calculation to curves.

6.2 Double Integration over General Regions

Given an arbitrary connected region in the xy-plane there are two primary descriptions of the region, say R (not necessarily a rectangle anymore). We **define** TYPE I and II as follows:



Of course, you can imagine region which don't conveniently fit either TYPE. On the other hand a rectange is both TYPE I and II at once. Observe, for a rectangle $R = [a,b] \times [c,d]$ we have $g_1(x) = c, g_2(x) = d$ to show R is TYPE I and we set $h_1(y) = a, h_2(y) = b$ to show R is TYPE II.

Theorem 6.2.1. *(Fubini, Strong version): Suppose f is mostly continuous.*

Given $R_I = \{(x, y) \mid a \leq x \leq b, g_1(x) \leq y \leq g_2(x)\}$a TYPE I region,

$$\iint_{R_I} f(x, y)dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x, y)dydx$$

Given $R_{II} = \{(x, y) \mid c \leq y \leq d, h_1(y) \leq x \leq h_2(y)\}$a TYPE II region,

$$\iint_{R_{II}} f(x, y)dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x, y)dxdy$$

Geometric Proof: A justification of the Theorem above is given by the following geometric argument. Suppose R is TYPE I with $a \leq x \leq b$ and $g_1(x) \leq y \leq g_2(x)$ for each $x \in [a, b]$. Furthermore, suppose $f(x, y) \geq 0$ for $(x, y) \in R$. In such a case, $\iint_R f(x, y)dA$ represents the volume bounded by $z = 0$, $z = f(x, y)$ and the cylinder $\partial R \times \mathbb{R}$. If we slice such a shape by cross-sections which are parallel to the yz plane then we may form the volume as the union of slices each with thickness dx. In particular, at fixed $x = x_o$ we obtain $dV = A(x_o)dx$ where $A(x_o)$ is the area of the slice of the volume by $x = x_o$. Observe, $g_1(x_o) \leq y \leq g_2(x_o)$ and $dA = zdy = f(x_o, y)dy$ thus

$$A(x_o) = \int_{g_1(x_o)}^{g_2(x_o)} f(x_o, y)dy$$

Now, replace x_o with x and note to find the net-volume we simply sum the volumes of each slice which amounts to integrating $dV = A(x)dx$ from $x = a$ to $x = b$:

$$V = \int_a^b \left(\int_{g_1(x)}^{g_2(x)} f(x, y)dy \right) dx$$

But, the volume V is also given by $\iint_R f(x, y)dA$ therefore we find the theorem true for TYPE I regions where $f(x, y) \geq 0$. A similar argument supports the theorem for TYPE II regions.

Of course, we also must consider functions which take negative values. We extend our argument thus far to functions which take negative values. We just chop the given region into smaller TYPE I and II regions on which f is nonzero on the interior of each region. Then apply the argument we already offered to the positive value regions and likewise apply the same argument to $-f$ on the subregions on which $f < 0$ hence $-f > 0$ and we may yet again recycle the argument above. Finally, sum the integrals on each subregion to obtain the desired result. Technically, there could be infinitely many regions on which f is negative so we omit a nontrivial analysis here. Indeed, this assertion that the calculation by cross-sections must yield the same value also hides a nontrivial analysis. Anytime we make some argument involving a rearrangement of infinitely many things we should pause to check our assertions. Unfortunately, the refined, technically correct analysis is beyond this course. Indeed, the argument I present here you'll find in many calculus texts.

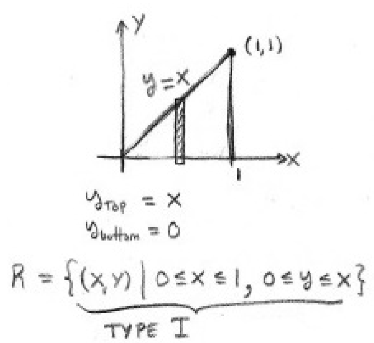
Example 6.2.2. *Let* $R = \{(x,y) \mid 0 \le x \le 1, 0 \le y \le x\}$

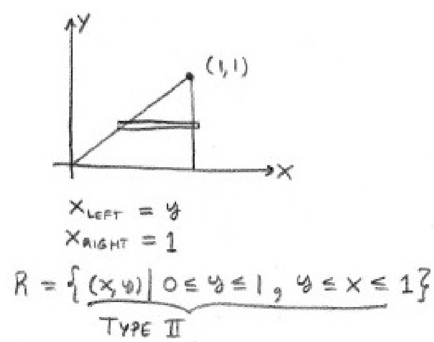
$$\begin{aligned}
\iint_R e^{x^2} &= \int_0^1 \int_0^x e^{x^2}\,dy\,dx \\
&= \int_0^1 (e^{x^2} y|_0^x)dx \\
&= \int_0^1 xe^{x^2}\,dx \\
&= \frac{1}{2}e^{x^2}\Big|_0^1 \\
&= \frac{1}{2}(e^1 - e^0) \\
&= \boxed{\frac{1}{2}(e-1)}
\end{aligned}$$

Remark 6.2.3.

We could just as well describe R as a TYPE II region. However, then we'd be forced with $\int e^{x^2} dx$. This is not an elementary integral.

Example 6.2.4. *Using* R *from Example 6.2.2 calculate* $\iint_R e^{y^2}dA$. *Since treating* R *as TYPE I leads us to* $\int e^{y^2} dy$ *we need to make* dx *appear first in the integration. Thus, convert* R *to a TYPE II region. A picture helps:*





Thus, note the second equality below follows as the integral of the constant e^{y^2} *is the product of the integration region length* $(1-y)$ *and the constant,*

$$\begin{aligned}
\iint e^{y^2}dA &= \int_0^1 \int_y^1 e^{y^2}\,dx\,dy \\
&= \int_0^1 (1-y)e^{y^2}\,dy \\
&= \int_0^1 e^{y^2}\,dy - \int_0^1 ye^{y^2}\,dy \\
&= \int_0^1 e^{y^2}\,dy - \frac{1}{2}(e-1) \qquad : \textit{curses, I had hoped for better} \\
&\approx 1.463 - \frac{1}{2}(e-1)
\end{aligned}$$

In the last step, $\int_0^1 e^{y^2} dy$ required a numerical integration. Sometimes, even swapping the order of the bounds does not make the integral accessible to elementary integration techniques.

Remark 6.2.5.

Not all integrals result in pretty sums and products, if we just make up some examples on a hunch then it can get ugly. Incidentally while indefinites integrals of e^{x^2} are not known in terms of elementary functions, there are improper integrals of e^{-x^2} which do come out quite nicely. See Ex ?? . We need a few ways to make it easier.

Example 6.2.6. *Another application of double integrals is finding the area of a region. For example, $S = \{(x, y) \mid 0 \leq x \leq R,\ 0 \leq y \leq \sqrt{R^2 - x^2}\ \}$*

$$\begin{aligned} A(R) = \iint_S dA &= \int_0^R \int_0^{\sqrt{R^2-x^2}} dydx \\ &= \int_0^R \sqrt{R^2 - x^2} dx \end{aligned}$$

The integration technique to tackle integrals which contain a squareroot is known as trigonometric substitution[5]*. We set $x = R\cos\theta$ so $dx = -R\sin\theta d\theta$ and $\sqrt{R^2 - x^2} = R\sin\theta$. We also must change the bounds. In particular, $x = R \to \theta = 0$ and $x = 0 \to \theta = \pi/2$. Thus,*

$$\begin{aligned} A(R) &= \int_{\pi/2}^0 -R^2 \sin^2\theta d\theta \\ &= R^2 \int_0^{\pi/2} \frac{1}{2}\big(1 - \cos 2\theta\big) d\theta \qquad \text{:we know } \sin^2\theta = \frac{1}{2}\big(1 - \cos 2\theta\big) \\ &= \frac{R^2}{2}\left(\theta - \frac{1}{2}\sin 2\theta\right)\Bigg|_0^{\pi/2} \\ &= \frac{R^2}{2}\left(\frac{\pi}{2}\right) \\ &= \boxed{\frac{\pi R^2}{4}} \end{aligned}$$

If you realized S is a quarter-circle then you should have expected this result.

Remark 6.2.7.

It would be a good exercise to rework this example using polar coordinates. We learn how to change variables in multiple integrals towards the end of this chapter.

Example 6.2.8. *We nay also define the average of a function over R as*

$$\boxed{f^R_{avg} \equiv \frac{1}{A(R)} \iint_R f(x, y) dA}$$

[5]hyperbolic substitution also solves most of the same problems.

Consider $f(x, y) = xy$. If $R = [0, 1] \times [0, 1]$ and S the quarter circle with R=1 from Example 6.2.6. Do you think $f^R_{avg} > f^S_{avg}$ or vice-versa?

$$\iint_R xy dA = \int_0^1 \int_0^1 xy dx\, dy = \int_0^1 x dx \int_0^1 y dy = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

whereas,

$$\begin{aligned}
\iint_S xy dA &= \int_0^1 \int_0^{\sqrt{1-x^2}} xy\, dy\, dx \\
&= \int_0^1 \left(\frac{1}{2} xy^2 \Big|_0^{\sqrt{1-x^2}} \right) dx \\
&= \int_0^1 \frac{1}{2}(x - x^3) dx \\
&= \frac{1}{2}\left(\frac{1}{2} - \frac{1}{4}\right) = \frac{1}{8}
\end{aligned}$$

Thus

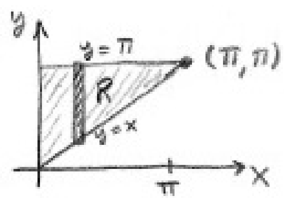
$$f^R_{avg} = \frac{1/4}{A(R)} = \frac{1/4}{1} = \frac{1}{4} \qquad \text{whereas} \qquad f^R_{avg} = \frac{1/8}{A(S)} = \frac{1/8}{\frac{\pi}{4}} = \frac{1}{2\pi}.$$

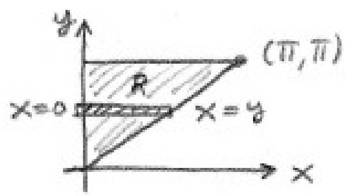
In conclusion, the average of xy is larger on the unit-square since $\frac{1}{4} > \frac{1}{2\pi}$.

Remark 6.2.9.

Notice $\iint_R xy dA$ was considerably easier than $\iint_S dA$.

Example 6.2.10. *Calculate $\int_0^\pi \int_x^\pi \frac{\sin y}{y}\, dydx$. Notice we need reverse the order of integration to do a TYPE II integration (has $dx\, dy$ instead). Our given integral suggests $0 \leq x \leq \pi$ and $x \leq y \leq \pi$.*





It is graphically clear the region can be recast as type II with $0 \leq y \leq \pi$ and $0 \leq x \leq y$. Thus,

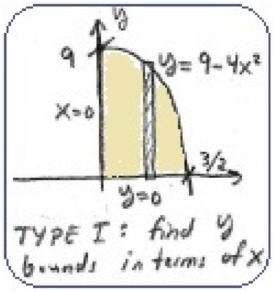
$$\iint_R \frac{\sin y}{y} dA = \int_0^\pi \int_0^y \frac{\sin y}{y} dxdy = \int_0^\pi \sin y\, dy = -\cos y \Big|_0^\pi = \boxed{2}.$$

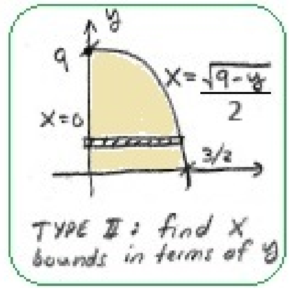
Example 6.2.11. *Calaulate,*

$$\begin{aligned}
\int_0^{3/2} \int_0^{9-4x^2} 16x\, dy\, dx &= \int_0^{3/2} 16x(9 - 4x^2) dx \qquad \star \\
&= \int_0^{3/2} (144x - 64x^3) dx \\
&= 72x^2|_0^{3/2} - 16x^4|_0^{3/2} \\
&= 72(3/2)^2 - 16(3/2)^4 \\
&= 162 - 81 = \boxed{81}
\end{aligned}$$

At ($\star$) we noted that $16x$ is a constant w.r.t. the dy integration. Hence, we simply multiply by the length of the integration region which is $9-4x^2$.

Next, let us reverse the order of integration for fun. Consider the graph:





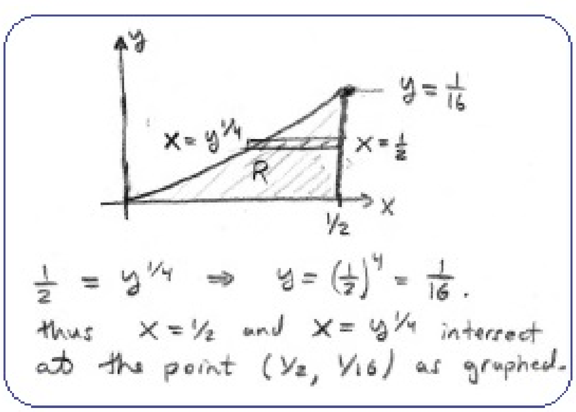
Note $9-4^2=0 \Rightarrow x^2=\frac{9}{4}, y=9-4x^2$ is a parabola with x-intercepts $x \pm 3/2$ and y-intercept 9. Solve $x^2=\frac{1}{4}(9-y)$ for x and keep positive root: $x=\frac{1}{2}\sqrt{9-y}$. Thus,

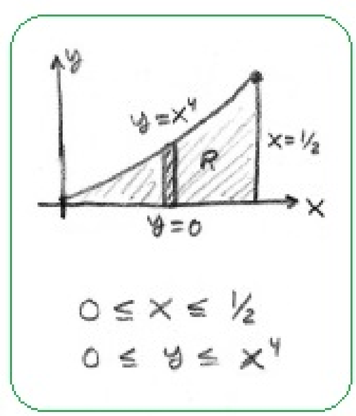
$$\begin{aligned}
\iint_R 16x\,dA &= \int_0^9 \int_0^{\frac{1}{2}\sqrt{9-y}} 16x\,dx\,dy \\
&= \int_0^9 \left(8x^2 \Big|_0^{\frac{1}{2}\sqrt{9-y}} \right) dy \\
&= \int_0^9 2(9-y)\,dy \\
&= (18y-y^2)\Big|_0^9 \\
&= 18(9)-81 \\
&= \boxed{81}
\end{aligned}$$

Example 6.2.12. *Calculate*

$$\int_0^{1/16} \int_{y^{1/4}}^{1/2} \cos(16\pi x^5)\,dx\,dy.$$

Seems changing bounds may be helpful here. To begin $0 \le y \le 1/16$ and $y^{1/4} \le x \le \frac{1}{2}$ which is type II, lets graph to guide our conversion to TYPE I,





$$
\begin{aligned}
\iint_R \cos(16\pi x^5)dA &= \int_0^{1/2}\int_0^{x^4} \cos(16\pi x^5)dydx \\
&= \int_0^{1/2} \cos(16\pi x^5)\left(\int_0^{x^4} dy\right)dx \\
&= \int_0^{1/2} x^4\cos(16\pi x^5)dx \\
&= \frac{1}{80\pi}\sin(16\pi x^5)\Big|_0^{1/2} \\
&= \frac{1}{80\pi}\left(\sin\left(\frac{16\pi}{32}\right) - \sin(0)\right) \\
&= \boxed{\frac{1}{80\pi}.}
\end{aligned}
$$

Remark 6.2.13.

The geometric arguments to set-up TYPE I or II should be familiar from your study of areas bounded by curves in single variable calculus. We said TYPE II regions needed horizontal slicing whereas TYPE I were vertically sliced. Notice, for TYPE I: where $y_{base}(x) \leq y \leq y_{top}(x)$ for $a \leq x \leq b$

$$A = \int_a^b \int_{y_{base}(x)}^{y_{top}(x)} dy\,dx = \int_a^b \Big(y_{top}(x) - y_{base}(x)\Big)dx$$

Whereas for TYPE II: where $x_{left}(y) \leq x \leq x_{right}(y)$ for $c \leq y \leq d$

$$A = \int_c^d \int_{x_{left(y)}}^{x_{right(y)}} dx\,dy = \int_c^d (x_{right}(y) - x_{left}(y))dy$$

Thus, in retrospect, we calculated double integrals in disguise in our previous course.

Example 6.2.14.

$$
\begin{aligned}
\int_{\pi/6}^{\pi/2}\int_{-1}^{5} \cos(y)dx\,dy &= \int_{\pi/6}^{\pi/2}\left(\cos(y)x\Big|_{-1=x}^{5=x}\right)dy \\
&= \int_{\pi/6}^{\pi/2} 6\cos(y)dy \\
&= 6\sin(y)\Big|_{\pi/6}^{\pi/2} \\
&= 6\sin(\pi/2) - 6\sin(\pi/6) = 6 - 3 = \boxed{3.}
\end{aligned}
$$

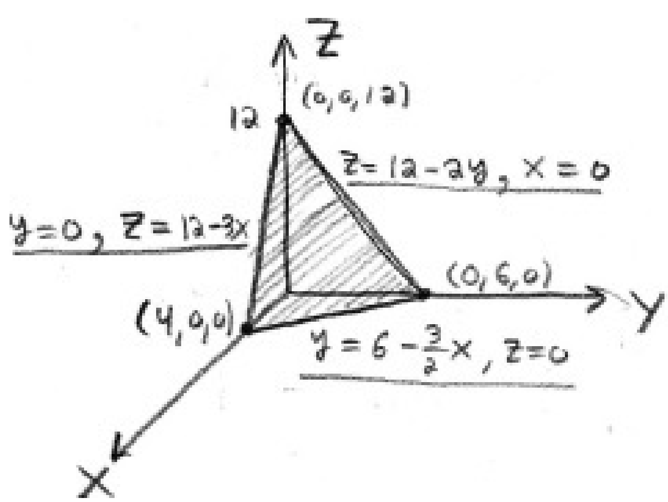
Example 6.2.15.

$$\begin{aligned}
\int_0^1\int_0^3 e^{x+3y}dxdy &= \int_0^1\int_0^3 e^x e^{3y}dxdy \\
&= \int_0^1 e^{3y}\int_0^3 e^x dxdy \\
&= \int_0^1 e^{3y}\left(e^x\Big|_{x=0}^{x=3}\right)dy \\
&= \int_0^1 (e^3-1)e^{3y}dy \\
&= \frac{e^3-1}{3}e^{3y}\Big|_0^1 \\
&= \frac{e^3-1}{3}(e^3-1) \\
&= \boxed{\frac{1}{3}(e^3-1)^2}
\end{aligned}$$

Example 6.2.16. *Find volume of the solid under the plane $3x+2y+z=12$ and above the rectangle*

$$R = \{(x,y,0) \mid 0 \leq x \leq 1, -2 \leq y \leq 3\}.$$

Solution: *We ought to integrate $z = 12-3x-2y \equiv f(x,y)$ on R. This gives the sum of volumes with height Z. Well, let's be careful, it gives the signed volume hopefully $f(x,y) \geq 0$ for $(x,y) \in R$. Let's pause to verify the geometry is arranged as the problem statement suggests.*



As you can see the graph $z = 12-3x-2y$ is entirely above the xy-plane for the given region. In particular, $0 \leq x \leq 1$ with $-2 \leq y \leq 3$ puts $z = f(x,y) > 0$. Therefore, to find the volume of the *solid we integrate:*

$$\begin{aligned}
V &= \iint_R f\,dA \\
&= \int_{-2}^{3}\int_0^1 (12-3x-2y)dx\,dy \\
&= \int_{-2}^{3} \left[12x - \frac{3}{2}x^2 - 2yx\right]\Big|_0^1 dy \\
&= \int_{-2}^{3} \left[12 - \frac{3}{2} - 2y\right] dy \\
&= \int_{-2}^{3} \left(\frac{21}{2} - 2y\right) dy \\
&= \left(\frac{21}{2}y - y^2\right)\Big|_{-2}^{3} \\
&= \frac{105}{2} - (9-4) = \frac{105-10}{2} = \boxed{\frac{95}{2}}.
\end{aligned}$$

Example 6.2.17. *Find the volume of the tetrahedron enclosed by coordinate planes $x = 0$, $y = 0$, $z = 0$ and the plane $2x + y + z = 4$.*

Solution: *the plane $z = 4 - 2x - y$ intersects the xy-plane along the line given by $z = 0$ and $y = 4 - 2x$. Thus the tetrahedron has $0 \leq z \leq 4 - 2x - y$ and $0 \leq y \leq 4 - 2x$. Finally, when $z = y = 0$ we obtain $0 = 4 - 2x$ hence $x = 2$. It follows $0 \leq x \leq 2$ for the tetrahedron. Therefore, the integration below gives the volume:*

$$\begin{aligned}
V &= \iint_D (4-2x-y)dA \\
&= \int_0^2 \int_0^{4-2x} (4-2x-y)dy\,dx \\
&= \int_0^2 \left((4-2x)y - \frac{1}{2}y^2\Big|_0^{4-2x}\right) dx \\
&= \int_0^2 \left((4-2x)(4-2x) - \frac{1}{2}(4-2x)^2\right) dx \\
&= \int_0^2 \frac{1}{2}(4-2x)^2 dx \\
&= \int_0^2 \frac{1}{2}(16-16x+4x^2)dx \\
&= \int_0^2 (8-8x+2x^2)dx \\
&= \left(8x - 4x^2 + \frac{2}{3}x^3\right)\Big|_0^2 \\
&= \boxed{\frac{16}{3}}
\end{aligned}$$

Example 6.2.18. *Calculate the double integral below:*

$$\begin{aligned}
\int_1^4 \int_0^2 (x+\sqrt{y})dx\,dy &= \int_1^4 \left(\frac{1}{2}x^2\Big|_0^2 + x\sqrt{y}\Big|_0^2 \right) dy \\
&= \int_1^4 (2+2\sqrt{y})\,dy \\
&= \left(2y + \frac{4y^{3/2}}{3}\right)\Big|_1^4 \\
&= \left[(2(4) + \frac{4}{3}(\sqrt{4})^3\right] - \left[2+\frac{4}{3}\right] \\
&= \boxed{\frac{46}{3}.}
\end{aligned}$$

Example 6.2.19. *Integrate.*

$$\begin{aligned}
\int_1^2 \int_0^1 \frac{1}{(x+y)^2} dxdy &= \int_1^2 \left(\frac{-1}{(x+y)}\Big|_0^1 \right) dy \\
&= \int_1^2 \left(\frac{-1}{y+1} + \frac{1}{y} \right) dy \\
&= (-\ln|y+1| + \ln|y|)\Big|_1^2 \\
&= \ln\left(\frac{|y|}{|y+1|}\right)\Big|_1^2 \\
&= \ln(2/3) - \ln(1/2) \\
&= \boxed{\ln(4/3)}
\end{aligned}$$

Example 6.2.20. *Integrate* $\int_0^1 \int_0^1 xy\sqrt{x^2+y^2}\,dy\,dx$. *To begin I make a* $u = x^2+y^2$ *substitution for which* $y=0$ *gives* $u=x^2$ *whereas* $y=1$ *gives* $u=x^2+1$. *Also,* $du = 2ydy$ *as we hold* x*-fixed in the initial integration.*

$$\begin{aligned}
\int_0^1 \int_0^1 xy\sqrt{x^2+y^2}\,dy\,dx &= \int_0^1 \left(\int_{x^2}^{x^2+1} \frac{1}{2}x\sqrt{u}\,du \right) dx \\
&= \int_0^1 \frac{1}{2}x \left(\frac{2}{3}u^{3/2}\Big|_{x^2}^{x^2+1} \right) dx \\
&= \frac{1}{3}\int_0^1 \left(x(x^2+1)^{3/2} - x(x^2)^{3/2} \right) dx \\
&= \frac{1}{3}\int_0^1 x(x^2+1)^{3/2}dx - \frac{1}{3}\int_0^1 x^4 dx \\
&= \frac{1}{6}\int_1^2 (x^2+1)^{3/2} d(x^2+1) - \frac{1}{15} \qquad \star \\
&= \frac{1}{6}\cdot\frac{2}{5}(2^{5/2}-1) - \frac{1}{15} = \boxed{\frac{2}{15}\big(2\sqrt{2}-1\big).}
\end{aligned}$$

At $\star$ *I made a* $w = x^2+1$ *substitution.*

Remark 6.2.21.

You may find it easier to go off to the side and calculate difficult integrals indefinitely. Otherwise, you do need to change bounds as I have in the example above.

Example 6.2.22. *Let $R = \{(x, y) \mid 0 \leq x \leq \pi, 0 \leq y \leq \frac{\pi}{2}\}$. Calculate:*

$$\begin{aligned}
\iint_R \cos(x+2y)dA &= \int_0^{\pi} \left(\int_0^{\pi/2} \cos(x+2y)dy \right) dx \\
&= \int_0^{\pi} \left(\frac{1}{2}\sin(x+2y)\Big|_{0=y}^{\pi/2=y} \right) dx \\
&= \frac{1}{2}\int_0^{\pi} \big(\sin(x+\pi) - \sin(x)\big)dx \\
&= \frac{1}{2}\left(-\cos(x+\pi) + \cos(x)\Big|_0^{\pi} \right) \\
&= \frac{1}{2}\Big[(-\cos(2\pi) + \cos(\pi)) - (-\cos(\pi) + \cos(0)) \Big] \\
&= \boxed{-2.}
\end{aligned}$$

Example 6.2.23. *Integrate,*

$$\begin{aligned}
\int_0^1 \int_y^{e^y} \sqrt{x}\, dx\, dy &= \int_0^1 \left(\frac{2}{3}x^{3/2}\Big|_y^{e^y} \right) dy \\
&= \frac{2}{3}\int_0^1 (e^{3y/2} - y^{3/2})dy \\
&= \frac{2}{3}\left[\frac{2}{3}e^{3y/2} - \frac{2}{5}y^{5/2} \right]\Big|_0^1 \\
&= \frac{4}{3}\left[\left(\frac{1}{3}e^{3/2} - \frac{1}{5} \right) - \left(\frac{2}{3} - 0 \right) \right] \\
&= \frac{4}{3}\left(\frac{1}{3}e^{3/2} - \frac{13}{15} \right) \\
&= \boxed{\frac{4}{9}e^{3/2} - \frac{52}{45}}
\end{aligned}$$

Example 6.2.24.

$$\int_0^1 \int_0^y e^{y^2} dxdy = \int_0^1 \left(xe^{y^2}\Big|_0^y \right) dy = \int_0^1 ye^{y^2} dy = \frac{1}{2}e^{y^2}\Big|_0^1 = \boxed{\frac{1}{2}(e-1)}$$

If we had tried to integrate with respect to y first we would have been stuck since $\int e^{y^2} dy$ is <u>not</u> an elementary integral. Sometimes reversing the order of integration makes the problem easier.

Example 6.2.25. $R = \{(x,y) \mid 0 \le x \le 1, 0 \le y \le 1\}$. *Calculate,*

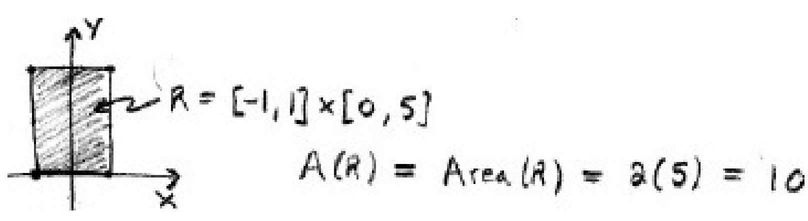
$$\begin{aligned}\iint_R \frac{1+x^2}{1+y^2}dA &= \int_0^1 \int_0^1 \frac{1+x^2}{1+y^2}dydx \\ &= \int_0^1 (1+x^2)dx \int_0^1 \frac{1}{1+y^2}dy \\ &= \left(x + \frac{1}{3}x^3\Big|_0^1\right)\left(\tan^{-1}(y)\Big|_0^1\right) \\ &= \frac{4}{3}\left(\tan^{-1}(1) - \tan^{-1}(0)\right) \\ &= \boxed{\frac{\pi}{3}.}\end{aligned}$$

Example 6.2.26. *Find volume of solid bounded by a top surface of $z = 1 - x^2/4 - y^2/9$ and a base surface given by $R = [-1,1] \times [-2,2]$ on $z = 0$. This makes the sides at $x = -1, x = 1, y = 2, y = -2$. We view R as a subset of xy-plane; that is, we identify R and $R \times \{0\}$ geometrically. In this special case the volune is found by integrating $z = 1 - x^2/4 - y^2/9$,*

$$\begin{aligned}V = \iint_R \left(1 - \frac{1}{4}x^2 - \frac{1}{9}y^2\right) dA &= \int_{-1}^1 \int_{-2}^2 \left(1 - \frac{1}{4}x^2 - \frac{1}{9}y^2\right) dy\, dx \\ &= 4\int_0^1 \int_0^2 \left(1 - \frac{1}{4}x^2 - \frac{1}{9}y^2\right) dy\, dx \qquad \star \\ &= 4\int_0^1 \left(y(1 - \frac{1}{4}x^2) - \frac{1}{27}y^3\right)\Big|_0^2 dx \\ &= 4\int_0^1 \left(2 - \frac{1}{2}x^2 - \frac{8}{27}\right) dx \\ &= 4\left(2 - \frac{1}{6} - \frac{8}{27}\right) \\ &= \boxed{\frac{166}{27}}\end{aligned}$$

At $\star$ I took advantage of the nature of R and the fact that the integrand was even in both x and y.

Example 6.2.27. *The average of $f(x,y) = x^2y$ over some region R is defined to be the $\iint_R f(x,y)dA$ divided by the area of $R = \iint_R dA = A(R)$. Let R be region with vertices $(-1,0), (-1,5), (1,5), (1,0)$.*



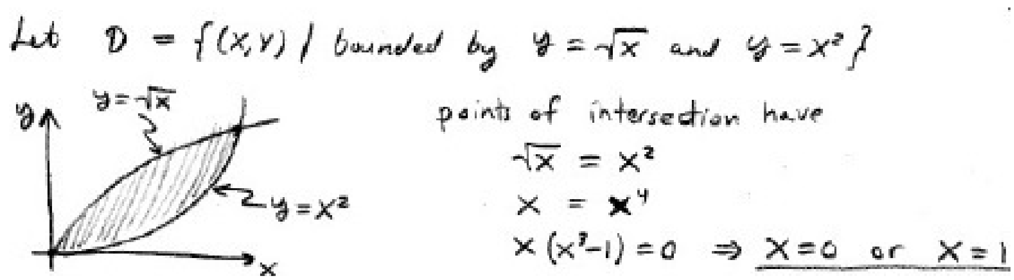
Thus, we calculate:

$$f_{avg} = \frac{1}{10}\int_0^5 \int_{-1}^1 x^2ydxdy = \frac{1}{10}\int_0^5 \frac{1}{3}x^3y\Big|_{-1}^1 dy = \frac{1}{10}\int_0^5 \frac{2}{3}ydy = \frac{2}{30}\frac{y^2}{2}\Big|_0^5 = \frac{25}{30}.$$

Hence, $\boxed{f_{avg} = 5/6.}$

Example 6.2.28. *Let* $D = \{(x, y) \mid \text{bounded by } y = \sqrt{x} \text{ and } y = x^2\}$*. Calculate* $\iint_R (x + y)dA$

Solution: *We study R by graphing paired with algebra:*

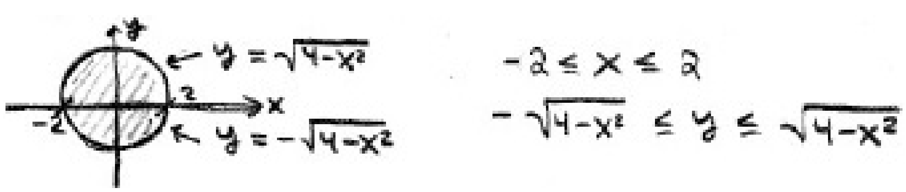


Thus the region $D = \{(x, y)|0 \le x \le 1, x^2 \le y \le \sqrt{x}\}$*. Now we can calculate an integral over the integration as follows:*

$$\begin{aligned}
\iint_D (x+y)dA &= \int_0^1 \int_{x^2}^{\sqrt{x}} (x+y)dy\,dx \\
&= \int_0^1 \left(xy + \frac{1}{2}y^2\right)\Big|_{x^2}^{\sqrt{x}} dx \\
&= \int_0^1 \left[x^{3/2} + \frac{1}{2}x - x^3 - \frac{1}{2}x^4\right] dx \\
&= \left(\frac{2}{5}x^{5/2} + \frac{1}{4}x^2 - \frac{1}{4}x^4 - \frac{1}{10}x^5\right)\Big|_0^1 \\
&= \frac{2}{5} + \frac{1}{4} - \frac{1}{4} - \frac{1}{10} \\
&= \boxed{\frac{3}{10}.}
\end{aligned}$$

Example 6.2.29. $D = \{(x, y)|x^2 + y^2 \le 4\}$*. Calculate* $\iint_D (2x - y)dA$*.*

Solution: *our first task is to describe D via inequalities for a typical point in D:*



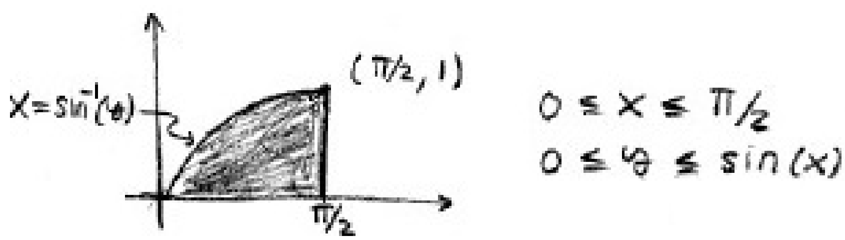
Now we can integrate,

$$\begin{aligned}
\iint_D (2x-y)dA &= \int_{-2}^{2}\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} (2x-y)dy\,dx \\
&= \int_{-2}^{2}\left(2xy - \frac{1}{2}y^2\right)\Big|_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} dx \\
&= \int_{-2}^{2}\left(4x\sqrt{4-x^2} - \frac{1}{2}(4-x^2) + \frac{1}{2}(4-x^2)\right) dx \\
&= \frac{-4}{3}(4-x^2)^{3/2}\Big|_{-2}^{2} \\
&= \frac{-4}{3}(0-0) = \boxed{0.}
\end{aligned}$$

This result is completely unsurprising if we consider the arrangement of values for x and y over the disk of radius 2. There is a perfect balance between positive and negative values.

Example 6.2.30. *Consider the region R defined by $0 \leq y \leq 1$ and $\sin^{-1}(y) \leq x \leq \pi/2$. Let $f(x,y) = \cos(x)\sqrt{1+\cos^2 x}$ and calculate $\iint_R f\,dA$.*

Solution: *we need to reformulate R since integration with respect to x is not obvious. Note:*

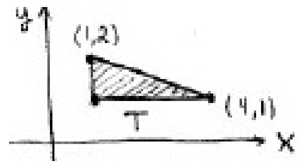


Thus

$$\begin{aligned}
\iint_R f(x,y)\,dA &= \int_0^{\pi/2}\int_0^{\sin(x)} \cos(x)\sqrt{1+\cos^2 x}dy\,dx \\
&= \int_0^{\pi/2} \underbrace{\sqrt{1+\cos^2 x}}_{\sqrt{u}}\ \underbrace{\sin(x)\cos(x)\,dx}_{-du/2} \\
&= \frac{-1}{3}(1+\cos^2 x)^{3/2}\Big|_0^{\pi/2} \\
&= \frac{-1}{3}(1-2^{3/2}) \\
&= \boxed{\frac{1}{3}(2\sqrt{2}-1)}.
\end{aligned}$$

Example 6.2.31. *Find volume under $z = xy$ and above the triangle with vertices $(1,1,0),(4,1,0)\,and(1,2,0)$.*

Solution: *it is natural to identify the given triangle with $T \subset \mathbb{R}^2$ formed by $(1,1)$,$(4,1)$ and $(1,2)$. We picture T below where the top line is $y = \frac{7-x}{3}$ and the bottom line is $y = 1$ for $1 \leq x \leq 4$.*



Notice xy is clearly postive on $T = \{(x,y) \mid 1 \leq x \leq 4,\ 1 \leq y \leq \frac{7-x}{3}\}$ hence the volume is given by:

$$\begin{aligned}
V &= \int_1^4 \int_1^{\frac{1}{3}(7-x)} xy\,dy\,dx \\
&= \int_1^4 \left(\frac{1}{2}xy^2 \bigg|_1^{\frac{1}{3}(7-x)} \right) dx \\
&= \int_1^4 \frac{1}{2}x\left(\frac{1}{9}(7-x)^2 - 1\right) dx \\
&= \frac{1}{18}\int_1^4 x\big[(7-x)^2 - 9\big]dx \\
&= \frac{1}{18}\int_1^4 \big[40x - 14x^2 + x^3\big]dx = \boxed{\frac{31}{8}.}
\end{aligned}$$

I'll let you fill in the last few details above. Notice that we can also write T as

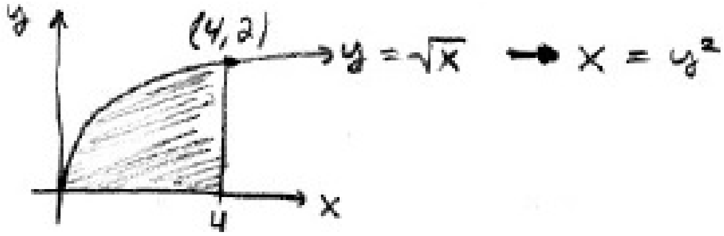
$$T = \{(x,y) \mid 1 \leq y \leq 2, 1 \leq x \leq 7-3y\}$$

then we can integrate x first then y.

$$\begin{aligned}
V &= \int_1^2 \int_1^{7-3y} xydxdy \\
&= \int_1^2 \frac{1}{2}yx^2 \bigg|_1^{7-3y} dy \\
&= \int_1^2 \frac{1}{2}y\left((7-3y)^2 - 1\right) dy \\
&= \int_1^2 \frac{1}{2}y\left(48 - 42y + 9y^2\right) dy \\
&= \frac{1}{2}\int_1^2 \left(48y - 42y^2 + 9y^3\right) dy \\
&= \frac{1}{2}\left(24y^2 - \frac{42}{3}y^3 + \frac{9}{4}y^4 \bigg|_1^2\right) \\
&= \frac{1}{2}\left(24(4-1) - 14(8-1) + \frac{9}{4}(16-1)\right) \\
&= \frac{1}{2}\left(72 - 98 + \frac{135}{4}\right) = \frac{1}{2}\left(-26 + \frac{135}{4}\right) = \frac{1}{2}\left(\frac{-104+135}{4}\right) = \boxed{\frac{31}{8}.}
\end{aligned}$$

Ok, its messy anyway you slice it.

Example 6.2.32. *Consider the following integral,* $\int_0^4 \int_0^{\sqrt{x}} f(x,y)\,dy\,dx$ *this indicates the integral is over* $0 \le y \le \sqrt{x}$ *and* $0 \le x \le 4$.



Equivalently we could say $y^2 \le x \le 4$ *and* $0 \le y \le 2$

$$\int_0^4 \int_0^{\sqrt{x}} f(x,y)\,dy\,dx = \int_0^2 \int_{y^2}^4 f(x,y)\,dx\,dy$$

for problems such as this, you just have to draw the picture and sort it out.

Example 6.2.33. *Let* R *be the region in the* xy*-plane bounded by* $y = 0$, $x = 1$ *and* $y = x^2$. *Calculate* $\iint_R \sqrt{x^3+1}\,dA$.

Solution: *it is clearly unpleasant to integrate with respect to* x *to begin. It follows*[6] *we should view* R *as* $0 \le y \le 1, \sqrt{y} \le x \le 1$. *Hence, integrate:*

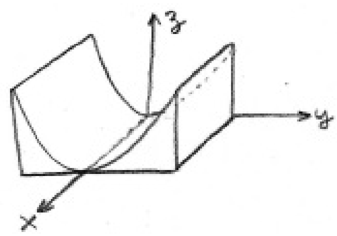
$$\begin{aligned}
\int_0^1 \int_{\sqrt{y}}^1 \sqrt{x^3+1}dx\,dy &= \int_0^1 \int_0^{x^2} \sqrt{x^3+1}dydx \\
&= \int_0^1 y\sqrt{x^3+1}\Big|_0^{x^2} dx \\
&= \int_0^1 x^2\sqrt{x^3+1}dx \\
&= \frac{2}{9}\left(x^3+1\right)^{3/2}\Big|_0^1 \\
&= \frac{2}{9}\left(2^{3/2}-1\right) \\
&= \boxed{\frac{2}{9}(2\sqrt{2}-1)}
\end{aligned}$$

[6]it may be helpful for you to draw a picture to verify this claim

6.3 Triple Integration over General Bounded Regions in $\mathbb{R}^3$

Rather than explicitly stating the Fubini Theorem for triple integrals I will simply illustrate with a few examples. Usually we can bound z in terms of x and y, then we can bound y in terms of x or vice-versa, that gives two orders of integration. Then other problems allow x to be bound in terms of y and z or possibly y in terms of x and z, in total there are six ways to write a particular integral. I don't give general advice on how to rewrite and switch bounds, it's a subtle business and there are far too many cases to enumerate. Generalities aside, let's do a few typical problems.

Example 6.3.1. *Let us find the volume of the region between $z = y^2$ and the xy-plane bounded by $x = 0, x = 1, y = 1$ and $y = -1$. Notice $dV = dxdydz$ so integrating dV gives volume V,*

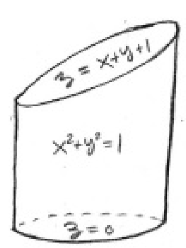


$0 \leq z \leq y^2$
$0 \leq x \leq 1$
$-1 \leq y \leq 1$

- we must integrate w.r.t. z either first or second.

$$
\begin{aligned}
V &= \int_{-1}^{1}\int_{0}^{1}\int_{0}^{y^2} dzdxdy \qquad &&: \textit{we work inside out as usual} \\
&= \int_{-1}^{1}\int_{0}^{1} y^2 dxdy \qquad &&: \textit{back to 2-d integrals} \\
&= \int_{-1}^{1} y^2 dy \qquad &&: \textit{back to 1-d integral} \\
&= \frac{1}{3}y^3\Big|_{-1}^{1} \\
&= \frac{1}{3}(1-(-1)^3) \\
&= \boxed{2/3}
\end{aligned}
$$

Example 6.3.2. *Consider the cylinder $x^2+y^2 = 1$, let $z = 0$ bound it from below and let $z = x+y+1$ bound it above, call this solid B. A sketch of B reveals the inequlities to the right of it.*



$0 \leq z \leq x+y+1$

$0 \leq x^2+y^2 \leq 1 \longrightarrow -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}$, $\quad -1 \leq x \leq 1$

With the geometry settled, we are free to calculate integrals of functions over B. For example, we *can integrate* $f(x,y,z) = x$ *over* B *as follows:*

$$\begin{aligned}\iiint_B x\,dV &= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\int_0^{x+1+y} x\,dzdydx \\ &= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} (x^2+x+xy)dydx\end{aligned}$$

note, the integral of xy *on the symmetric interval* $-\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}$ *is zero since the integral of an odd function of* y *over a symmetric interval in* y *about the origin is zero. Continuing,*

$$\begin{aligned}\iiint_B x\,dV &= \int_{-1}^{1}\left((x^2+x)y\Big|_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\right)dx \\ &= \int_{-1}^{1} 2x^2\sqrt{1-x^2}\;dx \\ &= 4\int_0^1 x^2\sqrt{1-x^2}\;dx\end{aligned}$$

The last step was justified because the integrand is even in x *and the integral of an even function about a symmetric interval of the origin is simply twice the integral of the left or right half of the integral. The remaining intergral can be calculated by a trigonometric subsitution. In particular, let* $x = \sin\theta$ *thus* $dx = \cos\theta\,d\theta$ *and* $1-x^2 = \cos^2\theta$ *and* $\sqrt{1-x^2} = \cos\theta$

$$\begin{aligned}\int x^2\sqrt{1-x^2}dx &= \int \sin^2\theta\cos\theta\cos\theta\,d\theta \\ &= \int (\sin^2\theta - \sin^4\theta)\,d\theta \\ &= \int\left(\frac{1}{2}(1-\cos(2\theta)) - \frac{1}{4}(1-2\cos(2\theta)+\cos^2(2\theta))\right)d\theta \\ &= \int\left(\frac{1}{2} - \frac{1}{2}\cos(2\theta) - \frac{1}{4} + \frac{1}{2}\cos(2\theta) - \frac{1}{4}\cos^2(2\theta)\right)d\theta \\ &= \int\left(\frac{1}{4} - \frac{1}{8}(1-\cos(4\theta))\right)d\theta \\ &= \frac{\theta}{8} + \frac{\sin(4\theta)}{32} + C.\end{aligned}$$

Change bounds on x *from* $x = 0 \to \theta = 0$ *and* $x = 1 \to \theta = \frac{\pi}{2}$

$$\iiint_B xdV = 4\left(\frac{\theta}{8} + \frac{1}{32}\sin(4\theta)\right)\Big|_0^{\pi/2} = \boxed{\frac{\pi}{4}}$$

Example 6.3.3. *Let B be bounded by coordinate plane and the plane possing through $(0,0,1),(0,1,0)$, and $(1,0,0)$. We find the equation of this plane to begin, note $\vec{v}$ and $\vec{w}$ are on the plane*

$$\vec{v} = (0,0,1) - (0,1,0) = \langle 0,-1,1\rangle$$

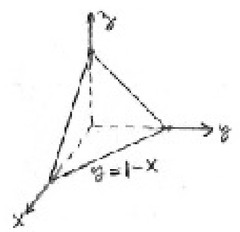
$$\vec{w} = (0,0,1) - (1,0,0) = \langle -1,0,1\rangle$$

$$\vec{v}\times\vec{w} = \begin{vmatrix} \widehat{x} & \widehat{y} & \widehat{z} \\ 0 & -1 & 1 \\ -1 & 0 & 1 \end{vmatrix} = \langle -1,-1,-1\rangle$$

Use normal $\langle -1,-1,-1\rangle$ and basepoint $(0,0,1)$ to give plane equation:

$$-x - y - (z-1) = 0 \quad \Rightarrow \quad z = 1 - x - y$$

Let's plot it. Note $z = 1-x-y$ intersects $z = 0$ on the line $y = 1-x$ in the xy-plane and we find the inequalities to bound each coordinate of (x,y,z) found within the pictured solid:



$$0 \le z \le 1 - x - y$$
$$0 \le y \le 1 - x$$
$$0 \le x \le 1$$

We calculate the volume of B by integrating dV over B:

$$\begin{aligned}
V = \iiint_B dV &= \int_0^1 \int_0^{1-x} \int_0^{1-x-y} dz\, dy\, dx \\
&= \int_0^1 \int_0^{1-x} (1-x-y)dy\, dx \\
&= \int_0^1 \left((1-x)y - \frac{1}{2}y^2 \Big|_0^{1-x} \right) dx \\
&= \int_0^1 \left((1-x)^2 - \frac{1}{2}(1-x)^2 \right) dx \\
&= \int_0^1 \frac{1}{2}(1 - 2x + x^2) dx \\
&= \frac{1}{2}(1 - \frac{2}{2} + \frac{1}{3}) = \boxed{\frac{1}{6}}
\end{aligned}$$

Notice, the bulk of the difficulty is usually in setting-up the integral. The process of calculating an iterated integral is (for most of us) the easy part.

Example 6.3.4. *Find the average value of $f(x,y,z) = x$ on the solid region from Example 6.3.3. The average is defined to be.*

$$f^B_{avg} \equiv \frac{1}{vol(B)} \iiint_B f(x,y,z)dV$$

We just found $vol(B) = 1/6$, let's focus on the $\iiint_B f dV$.

$$\begin{aligned}
\iiint_B f dV &= \int_0^1 \int_0^{1-x} \int_0^{1-x-y} x\,dz\,dy\,dx \\
&= \int_0^1 x \int_0^{1-x} (1-x-y)dy\,dx \\
&= \int_0^1 x\left((1-x)y - \frac{1}{2}y^2\Big|_0^{1-x} \right) dx \\
&= \int_0^1 \frac{1}{2}x(1-x)^2 dx \\
&= \int_0^1 \frac{1}{2}(x - 2x^2 + x^3)dx \\
&= \frac{1}{2}\left(\frac{1}{2} - \frac{2}{3} + \frac{1}{4}\right) = \frac{1}{2}\left(\frac{3}{4} - \frac{2}{3}\right) = \boxed{\frac{1}{24}}.
\end{aligned}$$

Thus, $f^B_{avg} = \frac{1/24}{1/6} = \frac{1}{4}$. Notice, the result seems reasonable in view of the picture in Example 6.3.3.

Example 6.3.5.

$$\begin{aligned}
\int_0^1 \int_0^z \int_0^{x+z} 6xz\,dy\,dx\,dz &= \int_0^1 \int_0^z \left(6xyz\Big|_{0=y}^{x+z=y} \right) dx\,dz \\
&= \int_0^1 \int_0^z 6x(x+z)z dx\,dz \\
&= \int_0^1 \int_0^z \left(6x^2 z + 6xz^2\right) dx\,dz \\
&= \int_0^1 \left(2x^3 z + 3x^2 z^2 \Big|_{x=0}^{x=z} \right) dz \\
&= \int_0^1 (2z^4 + 3z^4)dz \\
&= \int_0^1 5z^4 dz \\
&= z^5 \big|_0^1 \\
&= \boxed{1}
\end{aligned}$$

Example 6.3.6. *Let* $E = \{(x, y, z)|0 \leq x \leq 1, 0 \leq y \leq x, x \leq z \leq 2x\}$

$$\begin{aligned}
\iiint_E yz\cos(x^5)dV &= \int_0^1 \left(\int_x^{2x} \left(\int_0^x yz\cos(x^5)dy\right)dz\right)dx \\
&= \int_0^1 \int_x^{2x} \left(z\cos(x^5)\frac{y^2}{2}\bigg|_0^x\right)dz\,dx \\
&= \int_0^1 \int_x^{2x} \left(\frac{1}{2}x^2 z\cos(x^5)\right)dz\,dx \\
&= \int_0^1 \left(\frac{1}{2}x^2\cos(x^5)\frac{z^2}{2}\bigg|_x^{2x}\right)dx \\
&= \int_0^1 \left(\frac{1}{2}x^2\cos(x^5)\frac{1}{2}(4x^2 - x^2)\right)dx \\
&= \int_0^1 \frac{3}{4}x^4\cos(x^5)dx \qquad \boxed{u = x^5, du = 5x^4dx, u(1) = 1, u(0) = 0} \\
&= \int_0^1 \frac{3}{20}\cos(u)du \\
&= \frac{3}{20}(\sin(1) - \sin(0)) = \boxed{\frac{3\sin(1)}{20}}
\end{aligned}$$

Example 6.3.7. *Evaluate the integral three different ways. The region of integration is* $E = \{(x, y, z) \mid -1 \leq x \leq 1, 0 \leq y \leq 2, 0 \leq z \leq 1\}$-

$$\begin{aligned}
\iiint_E (xz - y^3)dV &= \int_{-1}^1 \int_0^2 \int_0^1 (xz - y^3)dz\,dy\,dx \\
&= \int_{-1}^1 \int_0^2 \left(\frac{1}{2}x - y^3\right)dy\,dx \\
&= \int_{-1}^1 \left(x - \frac{1}{4}(16)\right)dx = -4(2) = \boxed{-8.}
\end{aligned}$$

In what follows below, xz *is an odd-function integrated over symmetric interval about zero vanishes.*

$$\begin{aligned}
\iiint_E (xz - y^3)dV &= \int_0^1 \int_0^2 \int_{-1}^1 (xz - y^3)dx\,dy\,dz \\
&= \int_0^1 \int_0^2 \left(-xy^3\bigg|_{-1}^1\right)dydz \\
&= \int_0^1 \int_0^2 -2y^3dydz \\
&= \int_0^1 \frac{-2}{4}(2)^4dz = -8z\big|_0^1 = \boxed{-8.}
\end{aligned}$$

There are four other ways to interate the integral. Each will yield -8*. This is Fubini's Theorem for* $\iiint$ *in action.*

Example 6.3.8. *The notation on the evaluation bars in this example is optional. You may find it helpful as you begin your study of multivariate integration. You can contrast this notation with the less explicit notationin the example which follows.*

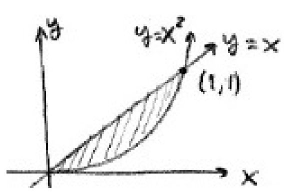
$$\begin{aligned}
\int_0^3 \int_0^1 \int_0^{\sqrt{1-z^2}} ze^y dx\, dz\, dy &= \int_0^3 \int_0^1 ze^y x\Big|_{x=0}^{x=\sqrt{1-z^2}} dz\, dy \\
&= \int_0^3 \int_0^1 z\sqrt{1-z^2}e^y dzdy \\
&= \int_0^3 \left(\frac{-1}{3}(1-z^2)^{3/2} e^y \Big|_{z=0}^{z=1} \right) dy \\
&= \int_0^3 \frac{1}{3} e^y dy = \boxed{\frac{1}{3}(e^3-1)}
\end{aligned}$$

Example 6.3.9. *Calculate the integral below:*

$$\begin{aligned}
\int_0^1 \int_0^z \int_0^y ze^{-y^2} dx\, dz\, dy &= \int_0^1 \int_0^z \left(ze^{-y^2} x \Big|_0^y \right) dy\, dz \\
&= \int_0^1 \int_0^z ze^{-y^2} ydy\, dz \\
&= \int_0^1 \left(\frac{-1}{2} ze^{-y^2} \Big|_0^z \right) dz \\
&= \int_0^1 \left(\frac{-1}{2} ze^{-z^2} + \frac{1}{2} z \right) dz \\
&= \left(\frac{1}{4} e^{-z^2} + \frac{1}{4} z^2 \right) \Big|_0^1 \\
&= \frac{1}{4} \left(e^{-1} + 1 - 1 \right) = \boxed{\frac{1}{4e}.}
\end{aligned}$$

Example 6.3.10. *E be the solid region in $\mathbb{R}^3$ bounded by the parabolic cylinder $y = x^2$ and the planes $x = z$, $x = y$ and $z = 0$. Calculate $\iiint_E (x + 2y) dV$.*

Solution: *we bound z to begin, $0 \leq z \leq x$. Then a two-dimension picture will do:*



$$x^2 \leq y \leq x$$
$$0 \leq x \leq 1$$

Note $(x, y, z) \in E$ implies $0 \leq z \leq x$ for $x^2 \leq y \leq x$ where $0 \leq x \leq 1$. Hence, integrate:

$$\begin{aligned}
\iiint_E (x+2y)dV &= \int_0^1 \int_{x^2}^x \int_0^x (x+2y)dz\,dy\,dx \\
&= \int_0^1 \int_{x^2}^x (x^2+2yx)dy\,dx \\
&= \int_0^1 \left(x^2y + xy^2 \Big|_{x^2}^x \right) dx \\
&= \int_0^1 (x^3+x^3-x^4-x^5)dx \\
&= \left(\frac{2}{4}x^4 - \frac{1}{5}x^5 - \frac{1}{6}x^6 \right) \Big|_0^1 \\
&= \frac{1}{2} - \frac{1}{5} - \frac{1}{6} = \frac{15-6-5}{30} = \boxed{\frac{2}{15}}
\end{aligned}$$

Example 6.3.11. *$E = \{(x, y, z) \mid 0 \leq x \leq 1, 0 \leq y \leq x, x \leq z \leq 2x\}$. We must integrate with respect to dz then dx. That is the natural order here. For example, to integrate $f(x, y, z) = yz\cos(x^5)$ on E we calculate as follows:*

$$\begin{aligned}
\iiint_E yz\cos(x^5)dV &= \int_0^1 \int_0^x \int_x^{2x} yz\cos(x^5)dz\,dy\,dx \\
&= \int_0^1 \cos(x^5) \left(\int_0^x \int_x^{2x} yz\,dz\,dy \right) dx \\
&= \int_0^1 \cos(x^5) \left(\int_0^x \frac{1}{2}\big[(2x)^2 - x^2\big])ydy \right) dx \\
&= \int_0^1 \frac{3}{2}x^2\cos(x^5) \left(\int_0^x ydy \right) dx \\
&= \int_0^1 \frac{3}{2}x^2\cos(x^5) \left(\frac{1}{2}y^2 \Big|_0^x \right) dx \\
&= \int_0^1 \frac{3}{4}x^4\cos(x^5)dx \\
&= \frac{3}{20}\sin(x^5)\Big|_0^1 = \boxed{\frac{3}{20}\sin(1).}
\end{aligned}$$

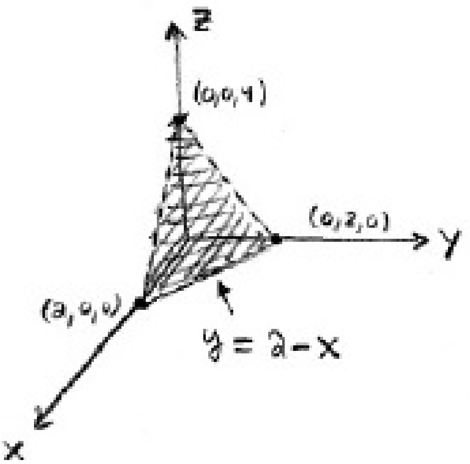
Example 6.3.12. *Calculate* $\iiint_E y dV$ *where* E *is the solid region in the first octant bounded by* $x = 0, y = 0, z = 0$ *and* $2x + 2y + z = 4$.

(i.) (xz)*-plane* $(y = 0) : 2x + z = 4$ *yields* $z = 4 - 2x$,

(ii.) (xy)*-plane* $(z = 0) : 2x + 2y = 4$ *yields* $y = 2 - x$,

(iii.) (yz)*-plane* $(x = 0) : 2y + z = 4$ *yields* $z = 4 - 2y$.

these details are not strictly speaking necessary but sometimes it helps to get some additional details to help insure graph is correct.



So we can describe the region of integration as

$0 \leq z \leq 4-2x-2y$

but what about x & y? Note on (xy)-plane we have

$0 \leq y \leq 2-x$

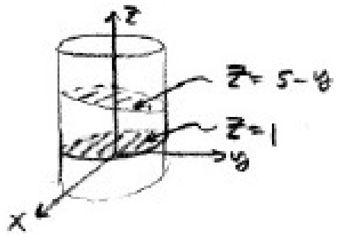
And finally

$0 \leq x \leq 2$

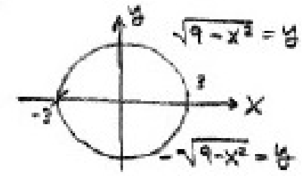
now integrate,

$$
\begin{aligned}
\iiint_E y dV &= \int_0^2 \int_0^{2-x} \int_0^{4-2x-2y} y dz\, dy\, dx \\
&= \int_0^2 \int_0^{2-x} y(4 - 2x - 2y) dy\, dx \\
&= \int_0^2 \int_0^{2-x} \big(2(2 - x)y - 2y^2\big) dy\, dx \\
&= \int_0^2 \left((2 - x)y^2 - \frac{2}{3}y^3 \right) \Big|_0^{2-x} dx \\
&= \int_0^2 \left((2 - x)^3 - \frac{2}{3}(2 - x)^3 \right) dx \\
&= \int_0^2 \frac{1}{3}(2 - x)^3 dx \\
&= \frac{-1}{12}(2 - x)^4 \Big|_0^2 \\
&= \frac{-1}{12}(0 - 16) = \frac{16}{12} = \boxed{\frac{4}{3}}
\end{aligned}
$$

Example 6.3.13. *Find volume enclosed by $x^2 + y^2 = 9$ and $y + z = 5, z = 5 - y$ and $z = 1$*



$$1 \le z \le 5 - y$$
$$-\sqrt{9-x^2} \le y \le \sqrt{9-x^2}$$
$$-3 \le x \le 3$$



In view of the diagrams above,

$$V = \int_{-3}^{3}\int_{-\sqrt{9-x^2}}^{\sqrt{9-x^2}}\int_{1}^{5-y} dz\, dy\, dx = \int_{-3}^{3}\int_{-\sqrt{9-x^2}}^{\sqrt{9-x^2}} (4-y)dy\, dx$$
$$= \int_{-3}^{3} 8\sqrt{9-x^2}dx$$

If we substitute $x = 3\sin\theta$ then $9 - x^2 = 9 - 9\sin^2\theta = 9\cos^2\theta$ and $dx = 3\cos\theta d\theta$. Also $x = 3$ corresponds to $\theta = \pi/2$ whereas $x = -3$ corresponds to $\theta = -\pi/2$. Consequently,

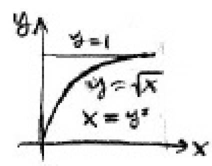
$$V = \int_{-\pi/2}^{\pi/2} 72\cos^2\theta d\theta$$
$$= 72\int_{-\pi/2}^{\pi/2} \frac{1}{2}(1+\cos 2\theta)d\theta$$
$$= \boxed{36\pi.}$$

Example 6.3.14. *Consider $\int_0^1\int_{\sqrt{x}}^1\int_0^{1-y} f(x,y,z)dz\, dy\, dx$. Find five other orders of iterating this integral. Let us consider the cases:*

$$0 \le z \le 1-y, \quad \sqrt{x} \le y \le 1, \quad 0 \le x \le 1$$

OR

$$0 \le z \le 1-y, \quad 0 \le x \le y^2, \quad 0 \le y \le 1$$

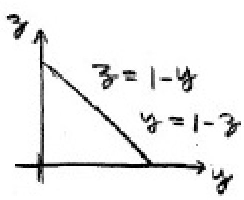


OR

$$0 \le x \le y^2, \quad 0 \le y \le 1-z, \quad 0 \le z \le 1$$

OR

$$0 \le x \le y^2, \quad 0 \le z \le 1-y, \quad 0 \le y \le 1$$

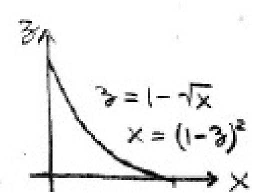


OR

$$\sqrt{x} \le y \le 1-z, \quad 0 \le z \le 1-\sqrt{x}, \quad 0 \le x \le 1$$

OR

$$\sqrt{x} \le y \le 1-z, \quad 0 \le x \le (1-z)^2, \quad 0 \le z \le 1$$



these are equivalent views of the integration region. Therefore, by Fubini's Theorem:

$$\begin{aligned}
\int_0^1 \int_{\sqrt{x}}^1 \int_0^{1-y} f\,dz\,dy\,dx &= \int_0^1 \int_0^{y^2} \int_0^{1-y} f\,dz\,dx\,dy \\
&= \int_0^1 \int_0^{1-z} \int_0^{y^2} f\,dx\,dy\,dz \\
&= \int_0^1 \int_0^{1-y} \int_0^{y^2} f\,dx\,dz\,dy \\
&= \int_0^1 \int_0^{1-\sqrt{x}} \int_{\sqrt{x}}^{1-z} f\,dy\,dz\,dx \\
&= \int_0^1 \int_0^{(1-z)^2} \int_{\sqrt{x}}^{1-z} f\,dy\,dx\,dz.
\end{aligned}$$

Remark 6.3.15.

We have studied how to integrate in Cartesian Coordinates in some detail. It turns out that this is quite limiting. To do many interesting problems with better efficiency it pays to employ cylindrical or spherical coordinates. Before getting to those special choices we consider a general coordinate change briefly and in the process derive what we later use for the cylindrical and spherical coordinates.

6.4 Change of Variables in Multivariate Integration

Our goal in this section is to give partial motivation for the analog of u-substitution for multiple integrals. In the one-dimensional case, we have to substitute the new variables for the old in the integrand, change dx to a corresponding expression with du, and we must change the bounds to the u-domain. We expect all three of these to appear in the multiple integral problem. It turns out the problem of substitution and bound-changing is nearly the same as in one-dimension. However, the method to change the measure offers a little surprise.

6.4.1 coordinate change and transformation

To begin, I encourage the reader to revisit Section 1.6 where we introduced polar, cylindrical and spherical coordinates. In general, a coordinate change on $\mathbb{R}^n$ is given by some mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ which is locally invertible at most points. Sometimes this is written as:

$$x_1 = x_1(u_1, u_2, \dots, u_n),\ x_2 = x_2(u_1, u_2, \dots, u_n),\ \dots\ x_n = x_n(u_1, u_2, \dots, u_n)$$

Invertibility means we can also solve for u_i as a function of x_i:

$$u_1 = u_1(x_1, x_2, \dots, x_n),\ u_2 = u_2(x_1, x_2, \dots, x_n),\ \dots\ u_n = u_n(x_1, x_2, \dots, x_n).$$

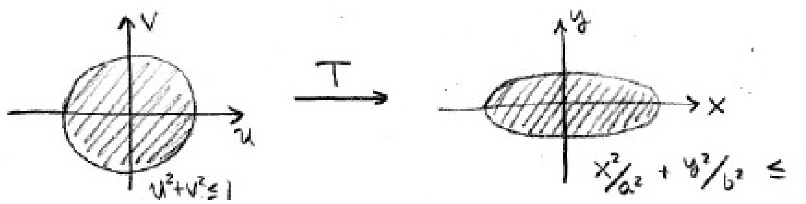
We don't insist on strict injectivity since polar, spherical and cylindrical coordinates all lack injectivity in their usual use. In particular, here are the usual polar coordinate transformations as well as the inverse transformations which apply to the half-space $x > 0$

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

We prefer to identify angles which correspond geometrically and the origin has $\rho = 0$ yet the polar and azmuthial angle of the origin is undefined. Fortunately, we are primarily interested in integration and the ambiguities of these standard coordinate systems do not change integrals. For a double integral, the addition or removal of a one-dimensional space does not change the integral. For a triple integral, the addition or removal of a two-dimensional space does not change the integral. To verify these assertions in general is beyond this course and properly belongs to the topic of **measure theory**. In addition, the more careful concept of a **coordinate chart** belongs to manifold theory where injectivity is required and some examples we consider here no longer fit the abstract definition of coordinate system.

We primarily discuss the problem of coordinate change in $\mathbb{R}^2$. It is a fortunate fact that the higher-dimensional problem admits almost the same analysis so little is lost by focusing on this readily visualized case. Let us begin with an example:

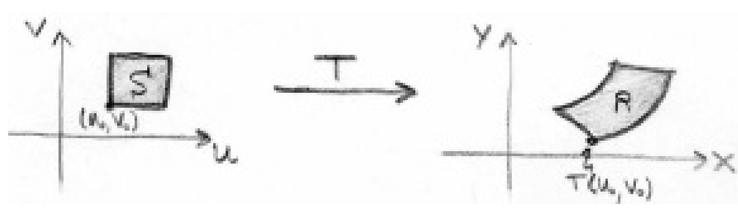
Example 6.4.1. *$S = \{(u, v) \mid u^2 + v^2 \leq 1\}$ and $x = au, y = bv$. Notice $x^2/a^2 + y^2/b^2 = a^2u^2/a^2 + b^2v^2/b^2 = u^2 + v^2 \leq 1$. Thus if we define $T(u, v) = (au, bv)$ we find*



the transformations T deforms a disk to an oval.

6.4.2 determinants for good measure

Consider two planes, the (x, y)-plane and the (u, v)-plane, the coordinate change map T takes (u, v) to $T(u, v) = (x(u, v), y(u, v))$. In particular, we study $T : S \subset \mathbb{R}^2 \to R \subset \mathbb{R}^2$ and we insist that T be invertible, except possibly on the boundaries. This means the equation relating x, y and u, v can be solved for either x, y or u, v locally. In the diagram below I illustrate how T might map a rectangle in the uv-space to a curved region in xy-space.



However, if we focus on a very small region then a little rectangle is essentially sent to a little parallelogram. Relating the areas of the rectangle and parallelogram will provide the relation between the measure in xy versus uv-coordinates. Furthermore, since we consider very small rectangle the first order approximation of T suffices. Recall, in our discussion of differentiability we learned T may be approximated by the linearization[7] of T

$$h(u, v) = T(u_o, v_o) + \begin{bmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{bmatrix} \begin{bmatrix} u - u_o \\ v - v_o \end{bmatrix}$$

[7]technically, this is an affine approximation of T

where the 2×2 Jacobian Matrix is evaluated at (u_o, v_o). A parallelogram at (u_o, v_o) is transported by h to a new parallelogram at $T(u_o, v_o)$ whose area is distorted according to the structure of the Jacobian matrix. You can prove the following for some extra credit if you wish:

Proposition 6.4.2. *(based on 5.1 of Colley's Vector Calculus)*

Let $h(u,v) = \begin{bmatrix} x_o \\ y_o \end{bmatrix} + A \begin{bmatrix} u \\ v \end{bmatrix}$ where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $\det A \neq 0$ and x_o, y_o are constants, then if D^* is a parallelogram then $h(D^*) = D$ is also a parallelogram and

$$\text{Area}(D) = |det\ A|\text{Area}(D^*).$$

If T maps $\triangle u, \triangle v$ at (u_o, v_o) to $\triangle x, \triangle y$ at $T(u_o, v_o)$ then as $\triangle u, \triangle v \to 0$ the proposition above applied to the linearization yields:

$$\boxed{\triangle x\, \triangle y = \det \begin{bmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{bmatrix} \triangle u\, \triangle v}$$

For finite changes this is an approximation since the real changes in x and y respective are based on the generally non-linear nature of T. However, as the size of the rectangles is reduced the approximation improves. We ultimately apply the boxed formula inside an integral where the approximating rectangles are made arbitrarily small and consequently the result above is made exact. Of course, I have omitted some careful analysis here. There are many excellent advanced calculus texts which justify the multivariate change of variables theorem. For example, you might look at Munkres *Calculus on Manifolds*. It contains a lengthy justification of multivariate u-substitution. Finally, let us conclude with the generalization to n-dimensions. We again find the determinant of the Jacobian matrix gives the necessary volume rescaling factor. In particular, if $T(u_1, \dots u_n) = (x_1, \dots, x_n)$ is an invertible coordinate map which maps $\triangle u_1, \triangle u_2, \dots, \triangle u_n$ to $\triangle x_1, \triangle x_2, \dots, \triangle x_n$ then the volumes[8] of the parallel-n-pipeds are related by:

$$\triangle x_1 \triangle x_2 \cdots \triangle x_n = \det(DT) \triangle u_1 \triangle u_2 \cdots \triangle u_n$$

where $\det(DT)$ is the the determinant of the Jacobian matrix. For example, if $n = 3$ then

$$\triangle x \triangle y \triangle z = \det \begin{bmatrix} \partial_u x & \partial_v x & \partial_w x \\ \partial_u y & \partial_v y & \partial_w y \\ \partial_u z & \partial_v z & \partial_w z \end{bmatrix} \triangle u \triangle v \triangle w.$$

You may find the rows and columns of the matrix above reversed in some calculus texts. That operation of changing rows into corresponding columns is called **transposition**. You should learn in your linear algebra course that $\det A = \det A^T$ where A^T is the **transpose** of A. It follows the formula above can also be written with the transpose of the Jacobian matrix. In any event, you should appreciate this section gives (without proof) an indication of the geometric significance of the determinant; the determinant quantifies generalized volume. The work which follows from here is merely a synthesis of the geometry of determinants with calculus.

[8] to be careful, the magnitude of these expressions are volumes, it is possible the expression is negative in which case the volume is given by the absolute value of the expression.

6.5 double integrals involving coordinate change

It is convenient to set some standard notation which helps us set-up the calculation of the Jacobian.

Definition 6.5.1.

The Jacobian of the transformation $T(u,v) = (x(u,v), y(u,v))$ is,

$$\frac{\partial(x,y)}{\partial(u,v)} = \det \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} = x_u y_v - x_v y_u$$

The "Jacobian" is the determinant of what I called the "Jacobian matrix".

Example 6.5.2. *Consider polar coordinates: $T(r,\theta) = (r\cos\theta, r\sin\theta)$. Let's calculate the Jacobian, note $x = r\cos\theta$ and $y = r\sin\theta$*

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = \boxed{r}$$

Example 6.5.3. *Find Jacobian of the following transformation*

$$\begin{aligned} x &= 5u - v \quad &\Rightarrow \quad & x_u = 5, \; x_v = -1 \\ y &= u + 3v \quad &\Rightarrow \quad & y_u = 1, \; y_v = 3 \end{aligned}$$

Therefore,

$$\frac{\partial(x,y)}{\partial(u,v)} = det \begin{bmatrix} x_u & x_v \\ y_u & y_v \end{bmatrix} = \det \begin{bmatrix} 5 & -1 \\ 1 & 3 \end{bmatrix} = 15 + 1 = \boxed{16.}$$

Example 6.5.4. *Find the Jacobian of the transformation $x = u + 4v$ and $y = 3u - 2v$*
By definition,

$$\frac{\partial(x,y)}{\partial(u,v)} = \det \begin{bmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{bmatrix} = \det \begin{bmatrix} 1 & 4 \\ 3 & -2 \end{bmatrix} = -2 - 12 = \boxed{-14}$$

Example 6.5.5. *Let $x = \alpha\sin\beta$ and $y = \alpha\cos\beta$ find the Jacobian of $(x,y) \mapsto (\alpha,\beta)$*
By definition,

$$\frac{\partial(x,y)}{\partial(\alpha,\beta)} = \det \begin{bmatrix} \partial x/\partial\alpha & \partial x/\partial\beta \\ \partial y/\partial\alpha & \partial y/\partial\beta \end{bmatrix} = \det \begin{bmatrix} \sin\beta & \alpha\cos\beta \\ \cos\beta & -\alpha\sin\beta \end{bmatrix} = -\alpha\sin^2\beta - \alpha\cos^2\beta = \boxed{-\alpha.}$$

The proof of the following is a simple consequence of the general chain rule. It is also very useful for certain problems where finding the inverse transformations is troublesome.

Proposition 6.5.6.

Given the transformation $T(u,v) = (x(u,v), y(u,v))$ is differentiable and invertible,

$$\frac{\partial(x,y)}{\partial(u,v)} \cdot \frac{\partial(u,v)}{\partial(x,y)} = 1 \;\; \text{which gives} \;\; \frac{\partial(x,y)}{\partial(u,v)} = \left[\frac{\partial(u,v)}{\partial(x,y)}\right]^{-1}.$$

Now that we have a little experience with Jacobians, let us return to the problem of integration.

In terms of our new-found notation the central point of the last section reads:

$$\triangle x\, \triangle y = \frac{\partial(x,y)}{\partial(u,v)} \triangle u\, \triangle v$$

As $\triangle x\, \triangle y \to 0$ this formula above intuitively tells us that:

$$dx\, dy = \frac{\partial(x,y)}{\partial(u,v)} du\, dv.$$

This is how we change the *measure* of integration in double integrals. Naturally, we also have to modify the integrand and bounds. The absolute value bars are needed as the sign of the integral arises from the integrand not dA for a double integral[9]. In particular:

Theorem 6.5.7. *Changing variables in double integrals:*

Suppose $T : S \to R$ is a differentiable mapping that is mostly invertible (except possibly on the boundary) from TYPE I or II region S to TYPE I or II region R and suppose that f is a continuous function whose domain includes R,

$$\iint_R f(x,y) dx\, dy = \iint_S f(x(u,v), y(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| du\, dv$$

where the $|\cdot|$ on the Jacobian are absolute value bars.

Example 6.5.8. *Let's apply this Theorem to Polar Coordinates, suppose f is continuous etc...*

$$\iint_R f(x,y) dx\, dy = \iint_S f(r\cos\theta, r\sin\theta) \left| \frac{\partial(x,y)}{\partial(r,\theta)} \right| dr\, d\theta = \iint_S f(r\cos\theta, r\sin\theta) r dr\, d\theta$$

Example 6.5.9. *Using Ex.1.4.5 calculate the area of a circle of radius A, call it R*

$$\iint_R dx\, dy = \iint_S r dr\, d\theta = \int_0^{2\pi} \int_0^{A} r dr\, d\theta = \int_0^{2\pi} \frac{1}{2} A^2 d\theta = \frac{1}{2} A^2 \cdot 2\pi = \boxed{\pi A^2}$$

Remark 6.5.10.

This is considerably easier than the direct Cartesian calculation of area, although the same geometry makes both solutions work. Notice "mostly invertible" is a needed qualifier since the angle θ doubles up on $\theta = 0$ and 2π given (x,y) along $\theta = 0$ should we say it corresponds to $\theta = 0$ or $\theta = 2\pi$? Fortunately a curve or two will not change double integral's result.

Example 6.5.11. *Evaluate the integral by performing an appropriate coordinate change,*

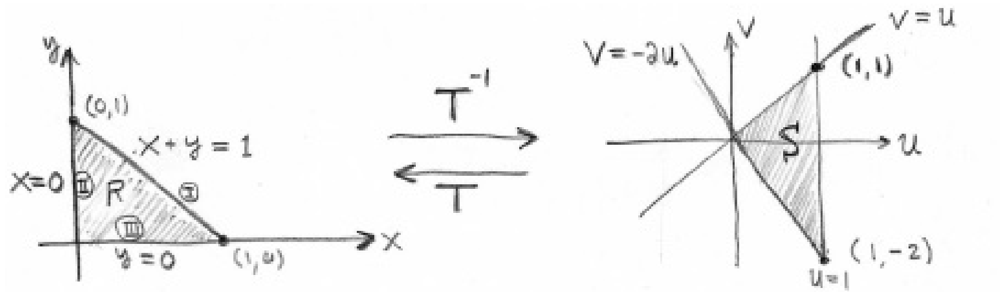
$$I \equiv \int_0^1 \int_0^{1-x} \sqrt{x+y}(y-2x)^2 dy\, dx$$

This suggests we choose $u = x + y$ and $v = y - 2x$. Solving for x,y yields $x = \frac{u}{3} - \frac{v}{3}$ and $y = \frac{2u}{3} + \frac{v}{3}$, Thus

$$T(u,v) = \left(\frac{1}{3}(u-v), \frac{1}{3}(2u+v) \right).$$

If $T : S \to R$ than what is S in this case? We are interested in R that is indicated by the integral I, namely $R = \{(x,y) \in \mathbb{R}^2 \mid 0 \le y \le 1 - x \ \&\ 0 \le x \le 1\}$. The graph is justified by the analysis below the graph:

[9] you may recall, as we calculated TYPE I or TYPE II regions the construction of dA requires it be positive. This was implicit within the definitions of TYPE I and II.



To figure out the boundaries in the uv-triangle we are guided by the knowledge that for a simple linear T as we have here triangles go to triangles, vertices to vertices.

(I.) $x + y = \frac{1}{3}u - \frac{1}{3}v + \frac{2}{3}u + \frac{1}{3}v = \boxed{u = 1}$

(II.) $0 = x = \frac{1}{3}u - \frac{1}{3}v \Rightarrow \boxed{u = v}$

(III.) $0 = y = \frac{2}{3}u + \frac{1}{3}v \Rightarrow \boxed{v = -2u}$.

Thus, $S = \{(u,v) \in \mathbb{R}^2 \mid -2u \leq v \leq u \ \& \ 0 \leq u \leq 1\}$. *Notice then, for* $x = \frac{1}{3}(u - v)$ *and* $y = \frac{1}{3}(2u + v)$ *we calculate the Jacobian:*

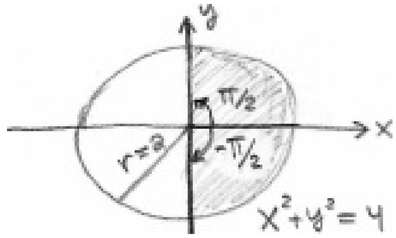
$$\frac{\partial(x,y)}{\partial(u,v)} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} = \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) - \left(\frac{-1}{3}\right)\left(\frac{2}{3}\right) = \frac{3}{9} = \frac{1}{3}$$

Apply what we've learned.

$$\begin{aligned}
I &= \int_0^1 \int_0^{1-x} \sqrt{x+y}(y-2x)^2 dy\, dx \\
&= \int_0^1 \int_{-2u}^{u} \sqrt{u}\, v^2 \left| \frac{\partial(x,y)}{\partial(u,v)} \right| dv\, du \\
&= \int_0^1 \int_{-2u}^{u} \frac{1}{3} v^2 \sqrt{u}\, dv\, du \\
&= \int_0^1 \frac{\sqrt{u}}{9} v^3 \Big|_{-2u}^{u} du \\
&= \int_0^1 \frac{\sqrt{u}}{9}(u^3 - (-2u)^3) du \\
&= \int_0^1 u^{3+1/2} du \\
&= \frac{2}{9} u^{9/2}\Big|_0^1 \\
&= \boxed{2/9}
\end{aligned}$$

Acknowledgment: this example borrowed from Thomas' Calculus 10th Ed, pg 1040.

Example 6.5.12. *Let $R = \{(x,y) | x^2 + y^2 \leq 4, x \geq 0\}$. Convert this region to Polars and integrate $f(x,y) = \sqrt{4 - x^2 - y^2}$.*



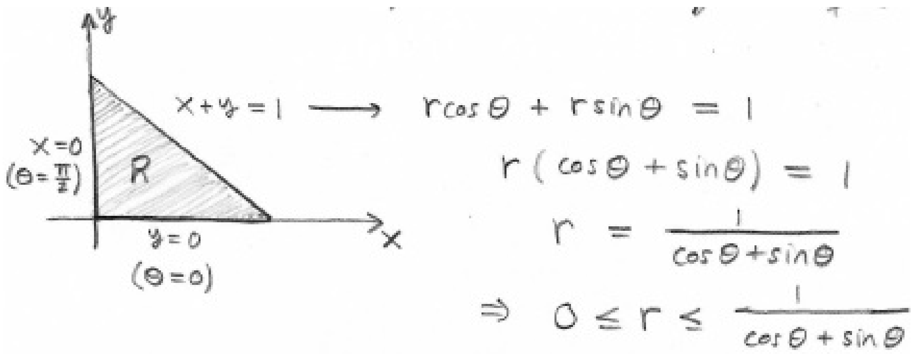
It is clear from the plot above that $S = \{(r,\theta) \mid 0 \leq r \leq 2,\ -\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}\}$. Note, we allow θ to range outside $[0, 2\pi)$ for our convenience. If we insisted on using only $[0, 2\pi)$ then the θ-domain would be built from disconnected intervals $[0, \pi/2]$ and $[3\pi/2, 2\pi)$. Thankfully, for problems of integration we consider, we are free to use the more convenient domain. I should warn the reader this difficulty cannot be avoided in complex analysis and ultimately leads to some rather interesting results. I digress, let's get back to the integration:

$$\begin{aligned}
\iint_R \sqrt{4 - x^2 - y^2}\, dA &= \int_0^2 \int_{-\pi/2}^{\pi/2} \sqrt{4 - r^2\cos^2\theta - r^2\sin^2\theta}\, r\, d\theta\, dr \\
&= \int_0^2 \int_{-\pi/2}^{\pi/2} r\sqrt{4 - r^2}\, d\theta\, dr \\
&= \int_0^2 \pi r\sqrt{4 - r^2} dr \\
&= -\frac{\pi}{2}\frac{2}{3}(4 - r^2)^{3/2}\Big|_0^2 \\
&= -\frac{\pi}{3}[0 - (2^2)^{3/2}] \\
&= \boxed{\frac{8\pi}{3}.}
\end{aligned}$$

Example 6.5.13. *Let $f(x,y) = \frac{1}{\sqrt{x^2+y^2}}$ find $\iint_R f(x,y) dA$ where R is the region in xy-plane with $1 \leq r \leq 2$ and $0 \leq \theta \leq \pi/3$.*

$$\begin{aligned}
\iint_R \frac{1}{\sqrt{x^2 + y^2}} dA &= \int_0^{\pi/3} \int_1^2 \frac{1}{r} r dr\, d\theta \\
&= \left(\int_0^{\pi/3} d\theta\right)\left(\int_1^2 dr\right) \\
&= \boxed{\frac{\pi}{3}.}
\end{aligned}$$

Example 6.5.14. The wrong way to calculate the area of a triangle: *Find the area of the triangle bounded by $x = 0, y = 0$ and $x + y = 1$ using polar coordinates. It's fairly easy to see that $0 \leq \theta \leq \pi/2$ on R, however bounding r requires thought,*



This is a less trivial polor region, we <u>*must*</u> *put the integration over dr first since its bounds are θ-dependent.*

$$
\begin{aligned}
Area(R) &= \int_0^{\pi/2} \int_1^{\frac{1}{\cos\theta+\sin\theta}} r\, dr\, d\theta \\
&= \int_0^{\pi/2} \frac{1}{2} \left(r^2 \Big|_0^{1/[\cos\theta+\sin\theta]} \right) d\theta \\
&= \frac{1}{2} \int_0^{\pi/2} \frac{d\theta}{(\cos\theta+\sin\theta)^2} \\
&= \frac{1}{2} \int_0^{\pi/2} \frac{1}{2} \csc^2(\theta+\pi/4) \Big|_0^{\pi/2} \\
&= \frac{-1}{4} \cot(\theta+\pi/4) \Big|_0^{\pi/2} \\
&= \frac{-1}{4} \Big(\cot(3\pi/4) - \cot(\pi/4) \Big) \\
&= \frac{-1}{4}(-1-1) \\
&= \boxed{\frac{1}{2}}
\end{aligned}
$$

At $\star$ we noticed $\sin(\theta + \pi/4) = \sin\theta\cos\pi/4 + \sin\pi/4\cos\theta = \frac{1}{\sqrt{2}}(\sin\theta + \cos\theta)$ thus $\sin\theta + \cos\theta = \sqrt{2}(\sin(\theta+\pi/4)$. This is a horrible method to find the area of a triangle. But, it illustrates a general principle which is that coordinates should be chosen to fit the problem. Obviously this problem is far more natural in Cartesians coordinates.

Example 6.5.15. *Let $R = \{(x,y) \,|\, 9x^2 + 4y^2 \leq 36\}$. Calculate $\iint_R x^2 dA$.*

Solution: *Let $x = 2u$ and $y = 3v$ and observe that $36 = 9x^2 + 4y^2 = 9(2u)^2 + 4(3v)^2 = 36u^2 + 36v^2$ hence the equation of the ellipse is transformed to the unit-circle $u^2 + v^2 = 1$ in uv-space. Note,*

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & 3 \end{vmatrix} = 6$$

Let $S = \{(u,v) \mid u^2 + v^2 \leq 1\}$ and calculate:

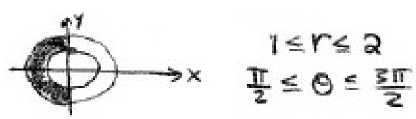
$$\begin{aligned} \iint_R x^2 dA &= \iint_S (2u)^2 \, \frac{\partial(x,y)}{\partial(u,v)} \, du\, dv \\ &= \iint_S 24u^2 du\, dv \end{aligned}$$

It is useful to change coordinates once more: let $u = r\cos\theta$ and $v = r\sin\theta$ then S transforms to $0 \leq \theta \leq 2\pi$ and $0 \leq r \leq 1$ in $r\theta$-space.

$$\begin{aligned} \iint_R x^2 dA &= \int_0^1 \int_0^{2\pi} 24r^2 \cos^2\theta r d\theta\, dr \\ &= \int_0^1 24r^3 dr \int_0^{2\pi} \frac{1}{2}(1 + \cos(2\theta))d\theta \\ &= \left(6r^4 \Big|_0^1 \right) \left(\frac{1}{2}\left(1 + \frac{1}{2}\sin(2\theta)\right) \Big|_0^{2\pi} \right) \\ &= (6-0)\left[\frac{1}{2}\left(2\pi + \frac{\sin(4\pi)}{2}\right) - \frac{1}{2}\left(0 + \frac{1}{2}\sin(0)\right) \right] \\ &= \boxed{6\pi.} \end{aligned}$$

Example 6.5.16. *Let R be the region in the xy-plane with $x \leq 0$ and $1 \leq x^2 + y^2 \leq 4$. Calculate $\iint_R (x+y)dA$ by changing the integral to polar coordinates.*

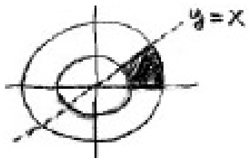
Solution: *to begin R is easily seen to have $1 \leq r \leq 2$ and $\frac{\pi}{2} \leq \theta \leq \frac{3\pi}{2}$. See the crude sketch below:*



$$\begin{aligned} \iint_R (x+y)dA &= \int_{\pi/2}^{3\pi/2} \int_1^2 (r\cos\theta + r\sin\theta) r dr\, d\theta \\ &= \int_{\pi/2}^{3\pi/2} (\cos\theta + \sin\theta)d\theta \int_1^2 r^2 dr \\ &= \left([\sin\theta - \cos\theta] \Big|_{\pi/2}^{3\pi/2} \right) \frac{1}{3}r^3 \Big|_1^2 \\ &= (-1-1)\left(\frac{1}{3}(8-1)\right) = \boxed{-\frac{14}{3}.} \end{aligned}$$

Example 6.5.17. *Let $R = \{(x, y) \mid 1 \leq x^2 + y^2 \leq 4, 0 \leq y \leq x\}$. Calculate $\iint_R \tan^{-1}(y/x)dA$ by changing the integration to polar coordinates.*

Solution: *usually, the best approach is to draw a picture. Consider,*



$$R = \{(r, \theta) \mid 1 \leq r \leq 2, \; 0 \leq \theta \leq \tfrac{\pi}{4}\}$$

$$\tan^{-1}(y/x) = \tan^{-1}\left(\frac{r\sin\theta}{r\cos\theta}\right) = \tan^{-1}(\tan\theta) = \theta$$

Therefore,

$$\begin{aligned}
\iint_R \tan^{-1}(y/x)dA &= \int_0^{\pi/4}\int_1^2 \theta r\, dr\, d\theta \\
&= \frac{1}{2}\theta^2\Big|_0^{\pi/4}\frac{1}{2}r^2\Big|_1^2 \\
&= \frac{1}{4}\left(\frac{\pi^2}{16}\right)(4-1) \\
&= \boxed{\frac{3\pi^2}{64}.}
\end{aligned}$$

Example 6.5.18. *Find volume bounded by $z = 18 - 2x^2 - 2y^2$ and $z = 0$.*

Solution: *The double integral will yield the volume. First convert to polars,*

$$z = 18 - 2r^2 \quad\Rightarrow\quad 9 = r^2 \quad\Rightarrow\quad r = 3.$$

the surface intersection with $z = 0$ in a circle $r = 3$ and $R = \{(r, \theta) \mid 0 \leq r \leq 3, 0 \leq \theta \leq 2\pi\}$. Therefore, the volume is found by calculating

$$\begin{aligned}
V = \iint_R z dA &= \int_0^{2\pi}\int_0^3 (18 - 2r^2) r dr\, d\theta \\
&= (2\pi)\left(9r^2 - \frac{2}{4}r^4\right)\Big|_0^3 \\
&= (2\pi)(81 - \frac{1}{2}(81)) \\
&= \boxed{81\pi.}
\end{aligned}$$

Example 6.5.19. *Find volume bounded by $z = \sqrt{x^2+y^2}$ and $x^2+y^2+z^2 = 1$. This is a cone bounded by a sphere. The equations are symmetric in x and y hence we deduce $0 \leq \theta \leq 2\pi$. The intersection of these surfaces has $x^2+y^2+(\sqrt{x^2+y^2})^2 = 1$ hence $2r^2 = 1$ from which we find $r = 1/\sqrt{2}$ on the intersection. It follows that $0 \leq r \leq 1/\sqrt{2}$. Naturally, you could use a three-dimensional graphing utility for further confirmation of our claim. To set up the volume, we need to identify that the shape has $\sqrt{x^2+y^2} \leq z \leq \sqrt{1-x^2-y^2}$ for*

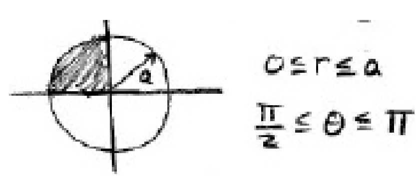
$$\begin{aligned} dV &= (z_{top} - z_{base})\, dA \leftarrow \text{(typital infinitesimal volume.)} \\ &= \left(\sqrt{1-x^2-y^2} - \sqrt{x^2+y^2}\right) dx\, dy \\ &= \left(\sqrt{1-r^2} - r\right) r\, dr\, d\theta \end{aligned}$$

Thus,

$$\begin{aligned} V = \iiint dV &= \int_0^{2\pi}\int_0^{1/\sqrt{2}} \big(r\sqrt{1-r^2} - r^2\big) dr\, d\theta \\ &= 2\pi\left(\frac{-1}{3}(1-r^2)^{3/2} - \frac{1}{3}r^3\Big|_0^{\frac{\sqrt{2}}{2}}\right) \\ &= \frac{-2\pi}{3}\left(\left(\frac{1}{2}\right)^{3/2} + \left(\frac{1}{\sqrt{2}}\right)^3 - 1\right) \\ &= \boxed{\frac{\pi}{3}(2-\sqrt{2})} \end{aligned}$$

Example 6.5.20. *Convert $\int_0^a \int_{-\sqrt{a^2-y^2}}^0 x^2 y\, dx\, dy$ to polar coordinates and calculate the integral.*

Solution: *from the given integral we deduce the integration region is given by $0 \leq y \leq a$ and $-\sqrt{a^2-y^2} \leq x \leq 0$. It follows the region is the top left quarter of the disk $x^2+y^2 \leq a^2$:*

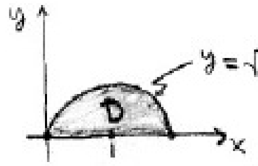


Therefore,

$$\begin{aligned} \int_{\pi/2}^{\pi}\int_0^a (r^2\cos^2\theta)(r\sin\theta) r dr\, d\theta &= \int_{\pi/2}^{\pi} \cos^2\theta \sin\theta d\theta \int_0^a r^4 dr \\ &= \left(\frac{-1}{3}\cos^3\theta\Big|_{\pi/2}^{\pi}\right)\left(\frac{a^5}{5}\right) \\ &= \frac{-1}{3}(-1)(a^5/5) = \boxed{a^5/15.} \end{aligned}$$

Example 6.5.21. *Convert* $\int_0^2 \int_0^{\sqrt{2x-x^2}} \sqrt{x^2+y^2}\,dy\,dx$ *to polar coordinates and calculate the integral.*

Solution: *let D be the region of integration. If $(x,y) \in D$ then we may read from the given integral that $0 \leq y \leq \sqrt{2x-x^2} = \sqrt{x(2-x)}$ and $0 \leq x \leq 2$ for the integration region. This is pictured below:*



$\Rightarrow y^2 = 2x - x^2$
$\Rightarrow x^2 - 2x + y^2 = 0$
$\Rightarrow (x-1)^2 + y^2 = 1$ (it's a circle of radius one at $(1,0)$)

it should be clear that D has $0 \leq \theta \leq \pi/2$. It is also clear that the bound on r must depend on θ since we have differing radii for differing θ (for example, $r = 2$ gives $\theta = 0$ while $r = 0$ gives $\theta = \pi/2$). We need to convert $y = \sqrt{2x-x^2}$ to a more useful form. Note that $x^2 - 2x + y^2 = 0$ yields $x^2 + y^2 = 2x$. However, this yields $r^2 = 2r\cos\theta$ hence we obtain the equation of the half-circle as $r = 2\cos\theta$ for $0 \leq \theta \leq \pi/2$. Therefore,

$$\begin{aligned}
\iint_D \sqrt{x^2+y^2}\,dA &= \int_0^{\pi/2} \int_0^{2\cos\theta} r^2\,dr\,d\theta \\
&= \int_0^{\pi/2} \frac{8}{3}\cos^3\theta\,d\theta \\
&= \int_0^{\pi/2} \frac{8}{3}(1-\sin^2\theta)\cos\theta\,d\theta \\
&= \frac{8}{3}\left(\sin\theta - \frac{1}{3}\sin^3\theta\right)\Big|_0^{\pi/2} \\
&= \frac{8}{3}\left(1 - \frac{1}{3}\right) = \boxed{\frac{16}{9}}
\end{aligned}$$

Example 6.5.22. *Let D_a be disk of radius a centered at origin, define*

$$\begin{aligned} I = \iint_{\mathbb{R}^2} e^{-x^2-y^2} dA &= \lim_{a\to\infty} \int_{D_a} e^{-x^2-y^2} dA \\ &= \lim_{a\to\infty} \int_0^{2\pi} \int_0^a e^{-r^2} r dr d\theta \\ &= \lim_{a\to\infty} \left((2\pi) \frac{-1}{2} e^{-r^2} \Big|_0^a \right) \\ &= \lim_{a\to\infty} \pi(-e^{-a^2} + 1) \\ &= \pi \end{aligned}$$

Alternatively, we could calculate the integral of $e^{-x^2-y^2}$ by a limit of rectangular integrals: let $S_a = [-a, a] \times [-a, a]$

$$\begin{aligned} I = \iint_{\mathbb{R}^2} e^{-x^2-y^2} dA &= \lim_{a\to\infty} \iint_{S_a} e^{-x^2-y^2} dA \\ &= \lim_{a\to\infty} \int_{-a}^{a} \int_{-a}^{a} e^{-x^2} e^{-y^2} dx dy \\ &= \lim_{a\to\infty} \left(\int_{-a}^{a} e^{-x^2} dx \right) \left(\int_{-a}^{a} e^{-y^2} dy \right) \\ &= \lim_{a\to\infty} \left(\int_{-a}^{a} e^{-x^2} dx \right) \lim_{a\to\infty} \left(\int_{-a}^{a} e^{-y^2} dy \right) \\ &= \left[\lim_{a\to\infty} \left(\int_{-a}^{a} e^{-x^2} dx \right) \right]^2 = \pi \end{aligned}$$

Therefore, we discover an interesting result. Assuming both limiting schemes agree,

$$\lim_{a\to\infty} \left(\int_{-a}^{a} e^{-x^2} dx \right) = \sqrt{\pi}.$$

Example 6.5.23. *A slight modification of the previous examples shows $\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}$. Assume that $\frac{\partial}{\partial a} \int_{-\infty}^{\infty} e^{-ax^2} dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial a} e^{-ax^2} dx$ to derive many nice formulas for integrals of the form:*

$$\int_{-\infty}^{\infty} z^n e^{-az^2} dz.$$

If you want to know a lot more about the idea of this example, you might search for the excellent paper by Keith Conrad DIFFERENTIATING UNDER THE INTEGRAL SIGN. In fact, you'll find many interesting expository papers at Professor Conrad's website.

6.6 triple integrals involving coordinate change

The change of variables theorem for triple integrals and higher integrals is essentially the same as we just saw for double integrals. We begin by extending the definition of the Jacobian to three variables:

Definition 6.6.1.

Let $T(u, v, w) = (x(u, v, w), y(u, v, w), z(u, v, w))$ be a differentiable function from $\mathbb{R}^3 \to \mathbb{R}^3$, the Jacobian of T is

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$$

Example 6.6.2. *Cylindrical coordinates:* $x = r\cos\theta$, $y = r\sin\theta$, $z = z$.

$$\frac{\partial(x, y, z)}{\partial(r, \theta, z)} = \det\begin{bmatrix} x_r & x_\theta & x_z \\ y_r & y_\theta & y_z \\ z_r & z_\theta & z_z \end{bmatrix} = \det\begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = (r\cos^2\theta + r\sin^2\theta) = \boxed{r}$$

Example 6.6.3. *Spherical coordinates:* $x = \rho\cos\theta\sin\phi$, $y = \rho\sin\theta\sin\phi$, $z = \rho\cos\phi$. *where* $0 \leq \theta \leq 2\pi$ *and* $0 \leq \phi \leq \pi$ *and* $\rho^2 = x^2 + y^2 + z^2$*:*

$$\frac{\partial(x, y, z)}{\partial(\rho, \theta, \phi)} = \det\begin{bmatrix} x_\rho & x_\theta & x_\phi \\ y_\rho & y_\theta & y_\phi \\ z_\rho & z_\theta & z_\phi \end{bmatrix}$$

$$= \det\begin{bmatrix} \cos\theta\sin\phi & -\rho\sin\theta\sin\phi & \rho\cos\theta\cos\phi \\ \sin\theta\sin\phi & \rho\cos\theta\sin\phi & \rho\sin\theta\cos\phi \\ \cos\phi & 0 & -\rho\sin\phi \end{bmatrix}$$

$$\begin{aligned} &= \cos\theta\sin\phi(\rho\cos\theta\sin\phi)(-\rho\sin\phi) - \sin\theta\sin\phi(\rho^2\sin\theta\sin^2\phi) \\ &\quad + \cos\phi(-\rho^2\sin^2\theta\sin\phi\cos\phi - \rho^2\cos^2\theta\sin\phi\cos\phi) \\ &= -\rho^2\sin^3\phi(\cos^2\theta + \sin^2\theta) - \rho^2\cos^2\phi(\sin^2\theta + \cos^2\theta)\sin\phi \\ &= -\rho^2\sin\phi(\sin^2\phi + \cos^2\phi) \\ &= \boxed{-\rho^2\sin\phi} \end{aligned}$$

Notice that $\frac{\partial(x,y,z)}{\partial(\rho,\phi,\theta)} = \rho^2\sin\phi$ *thanks to the fact that swapping a pair of columns in a determinant changes the sign of the result.*

Example 6.6.4. *Let $x = e^{u-v}, y = e^{u+v}, z = e^{u+v+w}$ find Jacobian. By definition,*

$$\begin{aligned}
\frac{\partial(x,y,z)}{\partial(u,v,w)} &= \det \begin{bmatrix} x_u & x_v & x_w \\ y_u & y_v & y_w \\ z_u & z_v & z_w \end{bmatrix} \\
&= \det \begin{bmatrix} e^{u-v} & -e^{u-v} & 0 \\ e^{u+v} & e^{u+v} & 0 \\ e^{u+v+w} & e^{u+v+w} & e^{u+v+w} \end{bmatrix} \\
&= e^{u+v+w}(e^{u-v}e^{u+v} + e^{u+v}e^{u-v}) \\
&= e^{u+v+w}(2e^{2u}) \\
&= 2e^{3u+v+w}
\end{aligned}$$

I should mention, Proposition 6.5.6 generalizes to three or more variables.

Example 6.6.5. *Let $x = \frac{u}{v}, y = \frac{v}{w}, z = \frac{w}{u}$ then*

$$\begin{aligned}
\frac{\partial(x,y,z)}{\partial(u,v,w)} &= det \begin{bmatrix} x_u & x_v & x_w \\ y_u & y_v & y_w \\ z_u & z_v & z_w \end{bmatrix} \\
&= det \begin{bmatrix} 1/v & -u/v^2 & 0 \\ 0 & 1/w & -v/w^2 \\ -w/u^2 & 0 & 1/u \end{bmatrix} \\
&= \frac{1}{uvw} - \frac{uvw}{u^2v^2w^2} \\
&= \boxed{0.}
\end{aligned}$$

The transformation studied in the example above would not be allowed if we wish to use the change of variables integration Theorem. The everywhere vanishing Jacobian suggests the transformation is not invertible. The theorem below is completely analogous to what we saw already in $n = 2$:

Theorem 6.6.6. *Coordinate Change in Triple Integrals:*

Suppose $T : R \subset \mathbb{R}^3 \to S \subset \mathbb{R}^3$ is a differentiable. mostly invertible mapping where $T(u,v,w) = (x(u,v,w),\, y(u,v,w),\, z(u,v,w))$. Also, suppose f is continuous on S, and $T(S) = R$ then

$$\iiint_S f(x,y,z)\, dx\, dy\, dz = \iiint_R f(T(u,v,w)) \left| \frac{\partial(x,y,z)}{\partial(u,v,w)} \right| du\, dv\, dw$$

Notice the notation $f(T(u,v,w))$ is simply notation for saying that f is to be written in terms of u, v, w as indicated by the formulas $x = x(u,v,w)$, $y = y(u,v,w)$ and $z(u,v,w)$. The following pair of examples give the most common applications of the change of variables theorem for triple integrals in this course:

Example 6.6.7. Cylindrical change of variables: *in view of Example 6.6.2 and the change of variables theorem:*

$$\iiint_R f(x,y,z)dV = \iiint_S f(r\cos\theta, r\sin\theta, z) \cdot r dr\, d\theta\, dz$$

The set S is simply R expressed in cylindrical coordinates.

Example 6.6.8. Spherical change of variables: *in view of Example 6.6.3 and the change of variables theorem:*

$$\iiint_R f(x,y,z)dV = \iiint_S f(\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)\cdot\rho^2\sin\phi\, d\rho\, d\theta\, d\phi$$

The set S is simply R expressed in spherical coordinates.

Example 6.6.9. *Evaluate $\iiint_E (x^3+xy^2)dV$ where E is the solid in the first octant which lies beneath the paraboloid $z = 1-x^2-y^2$.*

Solution: *This problem suggests a cylindrical approach, notice*

$$x^3+xy^2 = x(x^2+y^2) = xr^2 = r^3\cos\theta \qquad \textit{whereas} \qquad z = 1-x^2-y^2 = 1-r^2.$$

Also, the "first octant" is defined by $x \geq 0, y \geq 0$ and $z \geq 0$. On $z=0$ we find the intersection of $z = 1-r^2 = 0 \Rightarrow r^2 = 1$. Recall $dV = r\,dr\,d\theta\,dz$ for cylindrical coordinates. Also, we arrange the iterated bounds to reflect the description we found for E in cylindrical coordinates: $0 \leq z \leq 1-r^2$ for $0 \leq r \leq 1$ where $0 \leq \theta \leq \pi/2$. Therefore,

$$\begin{aligned}\iiint_E (x^3+xy^2)dV &= \int_0^{\pi/2}\int_0^1\int_0^{1-r^2}(r^2\cos\theta)r\,dz\,dr\,d\theta \\ &= \int_0^{\pi/2}\cos\theta d\theta\int_0^1 r^3(1-r^2)dr \\ &= \sin\theta\Big|_0^{\pi/2}\left(\frac{r^4}{4}-\frac{r^6}{6}\right)\Big|_0^1 \\ &= (1-0)\left(\frac{1}{4}-\frac{1}{6}\right) = \frac{2}{24} = \boxed{\frac{1}{12}.}\end{aligned}$$

Example 6.6.10. *Consider $f(x,y,z) = \frac{\exp(\sqrt{x^2+y^2+z^2})}{x^2+y^2+z^2}$. Find $\iiint_E f dV$ where $E = \{(x,y,z) | 1 \leq x^2+y^2+z^2 \leq 9\}$.*

Solution: *clearly this problem is best approached in spherical coordinates. Begin by noting that E may be described in spherical coordinates by $1 \leq \rho \leq 3$ and $0 \leq \theta \leq 2\pi$ with $0 \leq \phi \leq \pi$. The formula for the integrand is likewise transformed to e^ρ/ρ^2. Finally, the Jacobian for spherical coordinates tells us how the volume element transforms to $dV = \rho^2\sin\phi\,d\rho\,d\theta\,d\phi$*

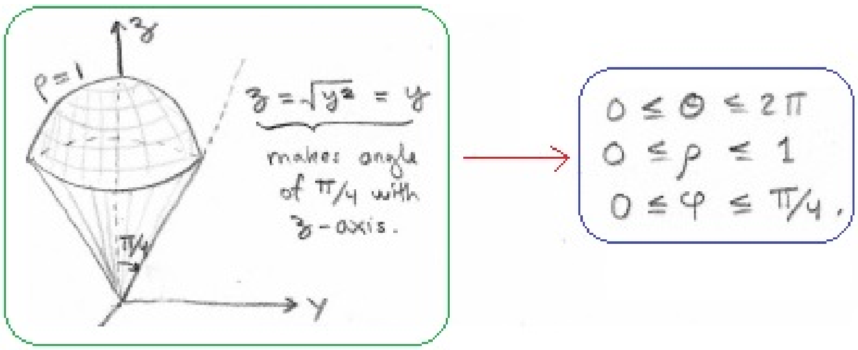
$$\begin{aligned}\iiint_E \frac{\exp(\sqrt{x^2+y^2+z^2})}{x^2+y^2+z^2}dx\,dy\,dz &= \int_0^\pi\int_0^{2\pi}\int_1^3 \frac{1}{\rho^2}e^\rho\rho^2\sin\phi\,d\rho\,d\theta\,d\phi \\ &= \left(\int_0^\pi \sin\phi d\phi\right)\left(\int_0^{2\pi} d\theta\right)\left(\int_1^3 e^\rho d\rho\right) \\ &= (-\cos\pi+\cos 0)(2\pi)(e^3-e^1) \\ &= \boxed{4\pi(e^3-e)}\end{aligned}$$

Example 6.6.11. *Find the volume and center of mass of the region E which is above the cone $z = \sqrt{x^2+y^2}$ and below the sphere $x^2+y^2+z^2=1$. We assume that E has a constant density*[10] δ.

Solution: *the center of mass is found by taking a total-mass-weighted average of x, y, z over the given solid. In particular, essentially by definition, the center of mass at (x_{cm}, y_{cm}, z_{cm}) is given by:*

$$x_{cm} = \frac{1}{M}\iiint_E x\delta\, dV \qquad y_{cm} = \frac{1}{M}\iiint_E y\delta\, dV, \qquad z_{cm} = \frac{1}{M}\iiint_E z\delta\, dV$$

where $\delta = \frac{dM}{dV}$ gives $M = \int dM = \iiint_E \delta dV$. In all four of the integrals we face, a spherical coordinate choice is convenient. Let us begin by picturing the solid:



Thus, pulling out the constant δ,

$$\begin{aligned} M &= \delta \int_0^{\pi/4}\int_0^1\int_0^{2\pi} \rho^2 \sin\phi \, d\theta\, d\rho\, d\phi \\ &= \delta \int_0^1 \rho^2 d\rho \int_0^{\pi/4} \sin\phi\, d\phi \int_0^{2\pi} d\theta \\ &= \delta\left(\frac{1}{3}\right)\Big(-\cos(-\pi/4)+1\Big)2\pi \\ &= \boxed{\frac{2\pi\delta}{3}\left(1-\frac{1}{\sqrt{2}}\right)} \end{aligned}$$

Since δ is a constant density it follows that $\delta = \frac{dM}{dV} = \frac{M}{V}$ hence the volume of E is given by $V = M/\delta = \frac{2\pi}{3}\left(1-\frac{1}{\sqrt{2}}\right)$. If δ is not constant then the volume and mass are not proportional in general. That said, let us continue to the calculation of the coordinates of the center of mass. Begin by using symmetry to see that $x_{cm} = y_{cm} = 0$ hence all we need to explicitly calculate is z_{cm}.

$$\begin{aligned} z_{cm} = \frac{1}{M}\iiint_E z\delta dV &= \frac{\delta}{M}\int_0^{2\pi}\int_0^1\int_0^{\pi/4} (\rho\cos\phi)(\rho^2\sin\phi\, d\phi\, d\rho\, d\theta) \\ &= \frac{\delta}{M}\theta\Big|_0^{2\pi}\left(\frac{1}{4}\rho^4\Big|_0^1\right)\left(\frac{1}{2}\sin^2\phi\Big|_0^{\pi/4}\right) \\ &= \frac{\delta}{M}(2\pi)\left(\frac{1}{8}\right)\left(\frac{1}{\sqrt{2}}\right)^2 = \frac{\delta}{M}\frac{\pi}{8}. \end{aligned}$$

[10] I often use ρ for density in physics, but, it seems that would be a bad notation here

Thus,

$$z_{cm} = \frac{\pi}{8} \cdot \left[\frac{2\pi}{3} \left(1 - \frac{1}{\sqrt{2}} \right) \right]^{-1} = \frac{\pi}{8} \cdot \frac{3}{2\pi} \cdot \frac{\sqrt{2}}{\sqrt{2}-1} = \frac{3\sqrt{2}}{16(\sqrt{2}-1)}.$$

In summary, the center of mass is at $\left(0, 0, \frac{3\sqrt{2}}{16(\sqrt{2}-1)}\right)$.

Sometimes we are interested in the **geometric center** or **centroid** of an object. This can be found in the same manner as the preceding example. We simply set $\delta = 1$ and find the center of mass. This is merely a convenience, if we set δ to any nonzero constant then the resulting center of mass would still be the same. Only in the case that the mass density is variable do we find a possible distinction between the **centroid** and **center of mass**.

Example 6.6.12. *Find the kinetic energy of a ball with radius R and mass m that spins with angular velocity ω and moves linearly with speed v.*

Solution: *It is shown in physics that $KE_{net} = KE_{rot} + KE_{trans}$. Where $KE_{trans} = \frac{1}{2}mv^2$ and $KE_{rot} = \frac{1}{2}I\omega^2$ where I is the moment of inertia measured with respect to the axis of rotation. We take the (moving) z-axis to be the rotation axis. From physics,*

$$\begin{aligned}
I_z &= \iint_E \delta(x,y,z)(x^2+y^2)dV \qquad \textit{constant density has } \delta = \frac{m}{\frac{4}{3}\pi R^3} \\
&= \int_0^{2\pi}\int_0^{\pi}\int_0^R \frac{3m}{4\pi R^3}\left[(\rho\cos\theta\sin\phi)^2 + (\rho\sin\theta\sin\phi)^2\right]\rho^2\sin\phi\, d\rho\, d\phi\, d\theta \\
&= \frac{3m}{4\pi R^3}\int_0^{2\pi} d\theta \int_0^{\pi}\int_0^R \rho^4\sin^3\phi\, d\rho\, d\phi \qquad : \int \sin^3\phi d\phi = \int (1-\cos^2\phi)\sin\phi d\phi \\
&= \frac{3m(2\pi)}{4\pi R^3}\left(\frac{1}{3}\cos^3\phi - \cos\phi \Big|_0^{\pi} \left(\frac{\rho^5}{5}\Big|_0^R\right)\right) \\
&= \frac{3m}{4\pi R^3}(2\pi)\left(\frac{-1}{3} + 1 - \frac{1}{3} + \cos(0)\right)\left(\frac{R^5}{5}\right) \\
&= \frac{3m(2\pi)}{4\pi R^3}\cdot\frac{4}{3}\cdot\frac{R^5}{5} \\
&= \boxed{\frac{2}{5}mR^2 = I_{sphere}}
\end{aligned}$$

Hence, we find:

$$KE_{total} = \frac{1}{2}mv^2 + \frac{1}{5}mR^2\omega^2$$

If the ball is rolling without slipping then $\omega = v/R$

$$KE_{total} = \frac{1}{2}mR^2\omega^2 + \frac{1}{5}mR^2\omega^2 = \frac{7}{10}mR^2\omega^2.$$

The formulas for center of mass or moment of inertia for a given solid all required multivariate integration. In physics, I usually begin by describing the concepts for a finite set of particles then we imagine how as the number of particles increases we can pass to the continuum. In that smearing process the finite sums elevate to continuous sums which we know and love as integrals. I forego this argument from the finite to the continuum as the content is mostly physical. In this course, in these notes, I merely supply some formulas given to us by the subject of physics and we use them as interesting problems to hone our integration skill.

Example 6.6.13. *Let $E = \{(x,y,z) \mid x^2/a^2 + y^2/b^2 + z^2/c^2 \leq 1\}$ calculate the volume of E.*

Solution: *by making the change of coordintes $x = au$, $y = bv$ and $z = cw$. Under this change of coordinates E morphs from an ellipsoid to a sphere B;*

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = (au)^2/a^2 + (bv)^2/b^2 + (cw)^2/c^2 = u^2 + v^2 + w^2$$

Let $B = \{(u,v,w) | u^2 + v^2 + w^2 \leq 1\}$ and calculate:

$$\begin{aligned}
\iiint_E dV &= \iiint_E dx\,dy\,dz \\
&= \iiint_B \frac{\partial(x,y,z)}{\partial(u,v,w)} du\,dv\,dw \qquad : \frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} x_u & x_v & x_w \\ y_u & y_v & y_w \\ z_u & z_v & z_w \end{vmatrix} = \begin{vmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{vmatrix} = abc \\
&= \iiint_B abc\,du\,dv\,dw \\
&= abc\Big(\underbrace{\iiint_B du\,dv\,dw}_{\text{volume of sphere } = \frac{4}{3}\pi} \Big) \\
&= \boxed{\frac{4\pi abc}{3}}
\end{aligned}$$

The volume of the unit sphere is calculated as follows,

$$\iiint_B du\,dv\,dw = \int_0^1 \int_0^\pi \int_0^{2\pi} \left| \frac{\partial(u,v,w)}{\partial(\rho,\theta,\phi)} \right| d\theta\,d\phi\,d\rho \qquad : \begin{cases} u = \rho\cos\theta\sin\phi \\ v = \rho\sin\theta\sin\phi \\ w = \rho\cos\phi \end{cases}$$

Calculate the Jacobian,

$$\begin{aligned}
\frac{\partial(u,v,w)}{\partial(\rho,\theta,\phi)} &= \begin{vmatrix} u_\rho & u_\theta & u_\phi \\ v_\rho & v_\theta & v_\phi \\ w_\rho & w_\theta & w_\phi \end{vmatrix} = \begin{vmatrix} \cos\theta\sin\phi & -\rho\sin\theta\sin\phi & \rho\cos\theta\cos\phi \\ \sin\theta\sin\phi & \rho\cos\theta\sin\phi & \rho\sin\theta\cos\phi \\ \cos\phi & 0 & -\rho\sin\phi \end{vmatrix} \\
&= -\rho^2\cos^2\theta\sin^3\phi + \rho^2\sin\theta\sin\phi[-\sin^2\phi\sin\theta - \cos^2\phi\sin\theta] - \rho^2\cos^2\theta\cos^2\phi\sin\phi \\
&= -\rho^2\cos^2\theta\sin\phi - \rho^2\sin^2\theta\sin\phi \\
&= -\rho^2\sin\phi
\end{aligned}$$

Thus,

$$\iiint_B du\,dv\,dw = \int_0^1 \int_0^\pi \int_0^{2\pi} \rho^2\sin\phi\,d\theta\,d\phi\,d\rho = 2\pi \left(\frac{\rho^3}{3}\Big|_0^1 \right) \left(-\cos\phi\Big|_0^\pi \right) = \frac{4\pi}{3}.$$

In general, the detail given at the end of the previous example is not required. The volume of a sphere is known for future reference. However, you ought to be able to work out the details.

Example 6.6.14. *Let E be the finite cylinder defined by $x^2+y^2 \leq 16$ and $-5 \leq z \leq 4$. Calculate $\iiint_E \sqrt{x^2+y^2}\, dV$.*

Solution: *observe E is $0 \leq r \leq 4$ and $-5 \leq z \leq 4$. Moreover, as there is no restriction on θ we have $0 \leq \theta \leq 2\pi$ for E.*

$$\begin{aligned}
\iiint_E \sqrt{x^2+y^2}\, dV &= \int_0^{2\pi}\int_{-5}^{4}\int_0^4 r^2 dr\, dz\, d\theta && : \text{ recall } dV = r dr\, dz\, d\theta \\
&= \int_0^{2\pi} d\theta \int_{-5}^{4} dz \int_0^4 r^2 dr && : \text{ since } \left|\frac{\partial(x,y,z)}{\partial(r,z,\theta)}\right| = r \\
&= (2\pi)(9)(64/3) \\
&= \boxed{384\pi.}
\end{aligned}$$

Example 6.6.15. *Let $E = \{(x,y,z) \mid 0 \leq z \leq x+y+5,\ 4 \leq x^2+y^2 \leq 9\}$. Calculate $\iiint_E x\, dV$*

Solution: *We can convert E to cylindrical coordinates via $x = r\cos\theta, y = r\sin\theta$ and of course $z = z$. Observe $x^2+y^2 = r^2$, E becomes:*

$$\begin{aligned}
&0 \leq \theta \leq 2\pi \\
&4 \leq r^2 \leq 9 \Rightarrow 2 \leq r \leq 3 \\
&0 \leq z \leq r\cos\theta + r\sin\theta + 5.
\end{aligned}$$

Thus motivating,

$$\begin{aligned}
\iiint_E x\, dV &= \int_0^{2\pi}\int_2^3\int_0^{r(\cos\theta+\sin\theta)+5} (r^2\cos\theta) dz\, dr\, d\theta, \qquad dV = r\, dz\, dr\, d\theta \\
&= \int_0^{2\pi}\int_2^3 \left(z\Big|_0^{r(\cos\theta+\sin\theta)+5} \right) r^2\cos\theta\, dr\, d\theta \\
&= \int_0^{2\pi}\int_2^3 \left[r^3(\cos^2\theta + \sin\theta\cos\theta) + 5r^2\cos\theta\right] dr\, d\theta \\
&= \int_0^{2\pi} \left[\left(\frac{r^4}{4}\Big|_2^3 \right)(\cos^2\theta + \sin\theta\cos\theta) + \left(\frac{5}{3}r^3\Big|_2^3 \cos\theta \right) \right] d\theta \\
&= \int_0^{2\pi} \left[\frac{65}{4}\left(\frac{1}{2} + \frac{1}{2}\cos(2\theta) + \frac{1}{2}\sin(2\theta) \right) + \frac{95}{3}\cos\theta \right] d\theta \qquad \star \\
&= \int_0^{2\pi} \frac{65}{8} d\theta \\
&= \frac{65}{8}\theta\Big|_0^{2\pi} \\
&= \boxed{\frac{65\pi}{4}}
\end{aligned}$$

In the $\star$ step I observe all terms, except the first constant term, integrate away due as a sinusoidal wave-form has equal areas above and below the x-axis during any integer multiple of periods. This observation can be very labor saving in problems such as this.

Example 6.6.16. *Find mass M of a ball B which is the set of all points (x, y, z) such that $x^2 + y^2 + z^2 \leq a^2$ where a is some constant and the mass-density at (x, y, z) is given by:*

$$\rho = k\sqrt{x^2 + y^2} = kr \qquad \textit{(proportionality constant is } k\textit{)}$$

Note $x^2 + y^2 + z^2 \leq a^2$ implies $r^2 + z^2 \leq a^2$ hence $z^2 \leq a^2 - r^2$ and consequently the sphere $\rho \leq a$ in cylindrical coordinates is:

$$0 \leq \theta \leq 2\pi, \qquad 0 \leq r \leq a, \qquad -\sqrt{a^2 - r^2} \leq z \leq \sqrt{a^2 - r^2}.$$

Recall $\rho = \frac{dm}{dV}$ thus $dm = \rho dV$. Therefore, we may calculate M by integrating ρdV:

$$\begin{aligned} M = \int dm &= \iiint_B \rho dV \\ &= \int_0^{2\pi} \int_0^a \int_{-\sqrt{a^2-r^2}}^{\sqrt{a^2-r^2}} kr^2 dz\, dr\, d\theta \\ &= \int_0^a 2\pi kr^2 \left(z \Big|_{-\sqrt{a^2-r^2}}^{\sqrt{a^2-r^2}} \right) dr \qquad : \textit{ the } \theta \textit{ integral yields } 2\pi \\ &= 4\pi k \int_0^a \underbrace{r^2\sqrt{a^2 - r^2} dr}_{\textit{not an obvious integral}} \end{aligned}$$

We can make a $r = a \sin b$ substitution then $a^2 - r^2 = a^2(1 - \sin^2 b) = a^2 \cos^2 b$. Moreover, $dr = a \cos b\, db$ hence $r^2\sqrt{a^2 - r^2} dr = a^4 \sin^2 b \cos^2 b\, db$. I like to use imaginary exponentials to work out the trigonometry here. Note $\sin b = \frac{1}{2i}(e^{ib} - e^{-ib})$ and $\cos b = \frac{1}{2}(e^{ib} + e^{-ib})$ hence

$$\begin{aligned} \sin^2 b \cos^2 b &= \left(\frac{1}{2}\right)^2 \left(\frac{1}{2i}\right)^2 (e^{ib} - e^{-ib})^2 (e^{ib} + e^{-ib})^2 \\ &= \frac{-1}{16}(e^{2ib} - 2 + e^{-2ib})(e^{2ib} + 2 + e^{-2ib}) \\ &= \frac{-1}{16}(e^{4ib} + 2e^{2ib} + 1 - 2e^{2ib} - 4 - 2e^{-2ib} + 1 + 2e^{-2ib} + e^{-4ib}) \\ &= -\frac{1}{8}\frac{1}{2}(e^{4ib} + e^{-4ib}) + \frac{1}{8} \\ &= \frac{1}{8}(1 - \cos(4b)) \end{aligned}$$

Furthermore, note $r = a$ gives $b = \frac{\pi}{2}$ whereas $r = 0$ yields $b = 0$. Therefore,

$$M = 4\pi k \int_0^{\frac{\pi}{2}} \frac{a^4}{8}(1 - \cos(4b)) db = \frac{\pi k a^4}{2} \cdot \frac{\pi}{2} = \boxed{\frac{k\pi^2 a^4}{4}}.$$

Naturally, you might use a table of integrals or a computer algebra system to tackle the integral in the precious example if you did not know how to calculate it from base principles. It is often the case that a given integral is easier if you **choose natural bounds**. The last example had a integrand which was manifestly cylindrical and bounds which were simplest in spherical coordinates. The example below shows how the integration simplifies if we change the integrand to sphericals rather than changing the bounds to cylindricals as in the preceding example.

Example 6.6.17. *Calculate M from Example 6.6.16 via spherical coordinates: note that*[11] $r = \rho \sin\phi$ *connects the spherical radius ρ and the cylindrical radius r.*

$$\begin{aligned}
m &= \iiint_B \rho dV \\
&= \int_0^{\pi}\int_0^{2\pi}\int_0^{a} k\rho\sin\phi\rho^2\sin\phi\, d\rho\, d\theta\, d\phi \\
&= k\int_0^{\pi}\sin^2\phi\, d\phi \int_0^{2\pi} d\theta \int_0^{a}\rho^3 d\rho \\
&= k\cdot\frac{\pi}{2}\cdot(2\pi)\cdot\frac{a^4}{4} \\
&= \boxed{\frac{k\pi^2 a^4}{4}.}
\end{aligned}$$

Moral of story: **coordinate choice matters.**

Example 6.6.18. *Let $B : (x, y, z)$ with $x^2 + y^2 + z^2 \leq 25$. Calculate $\iiint_B (x^2 + y^2 + z^2)^2 dV$.*

Solution: *Recall $dV = \rho^2 \sin\phi d\theta\, d\phi\, d\rho$ and note B is clearly $\rho \leq 5$ in spherical coordinates where θ and ϕ are free to range over their entire domains:*

$$\begin{aligned}
\iiint_B (x^2+y^2+z^2)^2 dV &= \int_0^{5}\int_0^{\pi}\int_0^{2\pi}\rho^6\sin\phi\, d\theta\, d\phi\, d\rho \\
&= \int_0^{2\pi} d\theta \int_0^{\pi}\sin\phi\, d\phi\int_0^{5}\rho^6 d\rho \\
&= (2\pi)(2)\left(\frac{5^7}{7}\right) = \boxed{\frac{312,500\pi}{7}.}
\end{aligned}$$

The integration of the above example is a fairly typical of integration over spheres. When the integrand has no angular dependence the θ and ϕ integrations produce a factor of 4π in total over a spherical region. The **solid angle** Ω measures both the polar and azmuthial angular displacements. In particular, $d\Omega = \sin\phi d\phi d\theta$ and over a sphere we find the total solid angle is 4π.

Example 6.6.19. *Let E be the region described in spherical coordinates by $1 \leq \rho \leq 2$ and $0 \leq \theta \leq \frac{\pi}{2}$ where $0 \leq \phi \leq \frac{\pi}{2}$. Calculate $\iiint_E z dV$.*

[11] geometrically obvious, algebraically follows from $r^2 = x^2 + y^2 = \rho^2\cos^2\theta\sin^2\phi + \rho^2\sin^2\theta\sin^2\phi = \rho^2\sin^2\phi$.

Solution: *we use spherical coordinates to calculate the integral.*

$$\begin{aligned}
\iiint_E z dV &= \int_0^{\pi/2}\int_0^{\pi/2}\int_1^2 (\rho\cos\phi)(\rho^2\sin\phi\, d\rho\, d\phi\, d\theta) \\
&= \int_0^{\pi/2} d\theta \int_0^{\pi/2} \sin\phi\cos\phi d\phi \int_1^2 \rho^3 d\rho \\
&= \left(\theta\Big|_0^{\pi/2}\right)\left(\frac{1}{2}\sin^2\phi\Big|_0^{\pi/2}\right)\left(\frac{1}{4}\rho^4\Big|_1^2\right) \\
&= \left(\frac{\pi}{2}\right)\left(\frac{1}{2}\right)\left(\frac{16-1}{4}\right) \\
&= \boxed{\frac{15\pi}{16}}
\end{aligned}$$

Example 6.6.20. *Find volume of part of ball $\rho \leq a$ that lies between $\phi = \pi/6$ and $\phi = \pi/3$.*

$$\begin{aligned}
V = \underbrace{\iiint dV}_{\textit{part of ball}} &= \int_0^a \int_0^{2\pi}\int_{\pi/6}^{\pi/3} \rho^2 \sin\phi\, d\phi\, d\theta\, d\rho \\
&= \int_0^a \rho^2 d\rho \int_0^{2\pi} d\theta \int_{\pi/6}^{\pi/3} \sin\phi d\phi \\
&= \left(\frac{1}{3}\rho^3\Big|_0^a\right)\left(\theta\Big|_0^{2\pi}\right)\left(-\cos\phi\Big|_{\pi/6}^{\pi/3}\right) \\
&= \left(\frac{1}{3}a^3\right)(2\pi)\left(-\cos(\frac{\pi}{3}) + \cos(\frac{\pi}{6})\right) \\
&= \frac{2\pi a^3}{3}\left(\frac{-1}{2} + \frac{\sqrt{3}}{2}\right) \\
&= \boxed{\frac{\pi a^3}{3}(\sqrt{3}-1)}
\end{aligned}$$

Example 6.6.21. *Let E be the region with $0 \leq \rho \leq 1$ and $0 \leq \phi \leq \pi/2$ with $0 \leq \theta \leq 2\pi$,*

$$\begin{aligned}
\iiint_E (x^2+y^2) dV &= \int_0^1 \int_0^{\pi/2}\int_0^{2\pi} (\rho^2\sin^2\phi\cos^2\theta + \rho^2\sin^2\phi\sin^2\theta)\rho^2\sin\phi d\theta d\phi d\rho \\
&= \int_0^1 \int_0^{\pi/2}\int_0^{2\pi} \rho^4 \sin^3\phi d\theta d\phi d\rho \\
&= \int_0^1 \rho^4 d\rho \int_0^{2\pi} d\theta \int_0^{\pi/2} (1-\cos^2\phi)\sin\phi d\phi \\
&= \frac{1}{5}\cdot 2\pi \cdot \left(\frac{1}{3}\cos^3\phi - \cos\phi\right)\Big|_0^{\pi/2} \\
&= \boxed{\frac{4\pi}{15}}
\end{aligned}$$

Example 6.6.22. *Let* $I = \int_{-3}^{3}\int_0^{\sqrt{9-x^2}}\int_0^{9-x^2-y^2}\sqrt{x^2+y^2}\,dz\,dy\,dx$. *Calculate* I *by changing to cylindrical coordinates. Left to reader for the time being, the picture I had here was a bit off. Perhaps we will work this out in class.*

6.7 integration in n variables

Let $\mathbb{R}^n$ have Cartesian Coordinates$(x_1, x_2, \cdots, x_n)$ and suppose $T(x_1, x_2, \cdots, x_n) = (y_1, y_2, \cdots, y_n)$ where y_i is a function of $x_1, x_2, \cdots, x_n$. For each i and we suppose T has DT invertible over the domain of integration below,

$$\iint_S \cdots \int f(x_1, x_2, \cdots, x_n) d^n x = \iint_{T(S)} \cdots \int f(y_1, y_2, \cdots, y_n) |det(DT)| d^n y$$

where $d^n x \equiv dx_1 dx_2 \cdots dx_n$ and $d^n y = dy_1 dy_2 \cdots dy_n$. The meaning of the n-fold integration should be an easy generalization of $n = 2$ and $n = 3$ which we have already treated in some depth.

Example 6.7.1. *The Hypersphere:* $(x, y, z, w) \in \mathbb{R}^4$ *such that* $x^2 + y^2 + z^2 + w^2 \leq R^2$. *Generalized Spherical Coordinates are*

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

Where $0 \leq \theta \leq 2\pi$ *and* $0 \leq \phi, \psi \leq \pi$. *You can check that* $x^2 + y^2 + z^2 + w^2 = r^2$. *Then it's a long but straightforward calculation,* $\left|\frac{\partial(x,y,z,w)}{\partial(r,\theta,\phi,\psi)}\right| = r^3\sin^2\psi\sin\phi$ *And consequently if we integrate* $d^4x = dx\,dy\,dz\,dw$ *in generalized spherical coordinates then we find:*

$$Vol_4(Hypersphere) = \int_0^R \int_0^{2\pi} \int_0^{\pi} \int_0^{\pi} r^3 \sin^2\psi\, \sin\phi\, d\psi\, d\phi\, d\theta\, dr = \frac{\pi^2 R^4}{2}.$$

I used this result to help calculate Guass' Law in 4 spatial dimensions in Section 8.6.1.

Example 6.7.2. *Find the volume bounded by $x^2 + y^2 = R^2$ and $0 \leq z, w \leq h$. This gives a finite hypercylinder. We calculate $d^4x = dx\,dy\,dz\,dw = r\,dr\,d\theta\,dz\,dw$ by introducing cylindrical coordinates $x = r\cos\theta$ and $y = r\sin\theta$ whereas $z = z$ and $w = w$. Thus, the hypervolume of the generalized cylinder is calculated by the following integral:*

$$\begin{aligned} V_4 &= \int_0^h \int_0^h \int_0^{2\pi} \int_0^R r\,dr\,d\theta\,dz\,dw \\ &= \int_0^h dw \int_0^h dz \int_0^{2\pi} d\theta \int_0^R r\,dr \\ &= \pi h^2 R^2. \end{aligned}$$

Example 6.7.3. *Consider $E = \{(x, y, z, t) \mid x^2 + y^2 + z^2 \leq R^2,\ 0 \leq t \leq h\}$. This hypervolume consists of a solid sphere of radius R attached at each t along $[0, h]$. The hypervolume is found by changing the x, y, z to spherical coordinates where are earlier work still applies:*

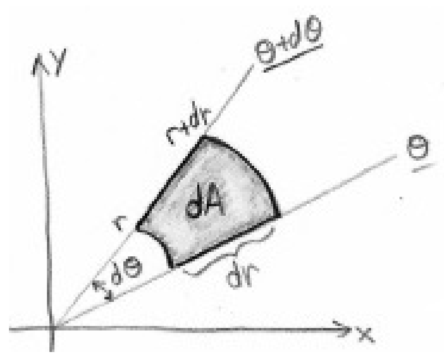
$$\begin{aligned} V_4 &= \int\int\int\int_E d^4x \\ &= \int_0^h \int_0^{2\pi} \int_0^{\pi} \int_0^R \rho^2 \sin\phi\,d\rho\,d\phi\,d\theta dt \\ &= \int_0^h dt \int_0^{2\pi} \int_0^{\pi} \int_0^R \rho^2 \sin\phi\,d\rho\,d\phi\,d\theta \\ &= \frac{4\pi h R^3}{3}. \end{aligned}$$

You might find my examples here a bit contrived. However, if application to the real world is what you seek then be assured that integrals over spaces with more than three variables are found in physics. In particular, if you study statistical mechanics of n-particles then you will face such integrals.

6.8 algebra and geometry of area and volume elements

In this section we take another look at how we can understand dA and dV. I present two methods which complement the Jacobian-based derivation we have thus far considered. Geometrically, we consider infinitesimal area elements or volume elements and direct geometric investigation yields the volume which correspond to small changes in the curvelinear coordinates. Algebraically, we introduce the **wedge product** and show by example how it also allows us to calculate the Jacobian factor. We do not attempt a general treatment here so we investigate only polar, cylindrical and spherical coordinates.

Let us begin with the two-dimensional case. For polar coordinates, we can derive $dA = r\,dr\,d\theta$ by examining the area of the sector pictured below:



The region above is an infinitesimal **polar rectangle** (not to scale !). Formally,

$$dA = \frac{1}{2}(r+dr)^2 d\theta - \frac{1}{2}r^2 d\theta = \frac{1}{2}(r^2 + 2rdr + (dr)^2 - r^2)d\theta = \boxed{r\,dr\,d\theta = dA}$$

We neglect the term $\frac{1}{2}dr^2 d\theta$ as it is much smaller than the other terms as we consider $dr, d\theta \ll 1$. Of course, this notation is just a formalism which we use in place of a careful argument involving finite differences. You'll find the finite argument and how it passes to an integral in any number of standard texts. Here, we embrace the infinitesimal method and forge ahead.

The algebraic method to derive these is given by differential forms and the wedge product. The basic rules are that d on a function gives the total differential and the **wedge product** is denoted $\wedge$ which is an associative product with the usual algebraic rules except, it is not commutative in general. For differentials, $df \wedge dg = -dg \wedge df$. In particular, this means $dx \wedge dx = 0$ and $dy \wedge dy = 0$ as well as $dr \wedge dr = 0$ and $d\theta \wedge d\theta = 0$. Let us proceed with these basic ideas in mind. We use $x = r\cos\theta$ and $y = r\sin\theta$ in the second equality:

$$\begin{aligned}
d\vec{A} &= dx \wedge dy \\
&= d(r\cos\theta) \wedge d(r\sin\theta) \\
&= [\cos\theta dr - r\sin\theta d\theta] \wedge [\sin\theta dr + r\cos\theta d\theta] \\
&= \sin\theta\cos\theta dr \wedge dr + r\cos^2\theta dr \wedge d\theta - r\sin^2\theta d\theta \wedge dr - r^2\sin\theta\cos\theta d\theta \wedge d\theta \\
&= r\cos^2\theta dr \wedge d\theta + r\sin^2\theta dr \wedge d\theta \\
&= \boxed{rdr \wedge d\theta = d\vec{A}.}
\end{aligned}$$

Here $d\vec{A} = -r\,d\theta \wedge dr$ also, in contrast to our usual $dA = r\,dr\,d\theta = r\,d\theta\,dr$ the minus sign encodes the **orientation**. The distinction between $dx \wedge dy$ and $dx\,dy$ is important. The scalar quantity

$dx\,dy$ has no sense of up or down whereas the fact that $dx \wedge dy = -dy \wedge dx$ allows us to identify $dx \wedge dy$ as the area element for the upwards oriented xy-plane. A little later in this course we discuss the **vector area element**. Essentially my comment here is that $dx \wedge dy$ is naturally identified with the vector area element of the plane oriented with $\widehat{z}$.

Wedge products and determinants are closely related. In particular:

$$A\,\widehat{x}_1 \wedge A\,\widehat{x}_2 \wedge \cdots \wedge A\,\widehat{x}_n = \det(A)\,\widehat{x}_1 \wedge \widehat{x}_2 \wedge \cdots \wedge \widehat{x}_n.$$

The identity above allows us to calculate determinants implicitly from wedge products. Some texts take this as the definition of the determinant of a matrix. Let us work out the 2×2 case. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then observe:

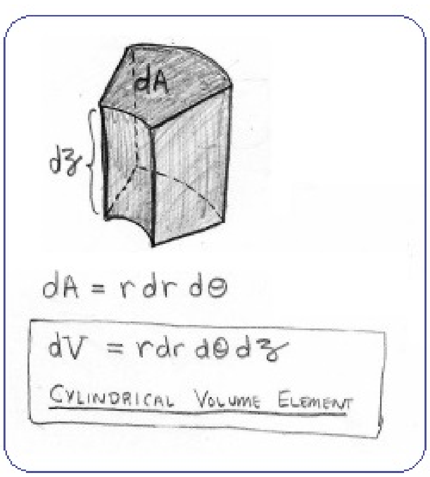
$$A\,\widehat{x}_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix} = a\,\widehat{x}_1 + c\,\widehat{x}_2.$$
$$A\,\widehat{x}_2 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix} = b\,\widehat{x}_1 + d\,\widehat{x}_2.$$

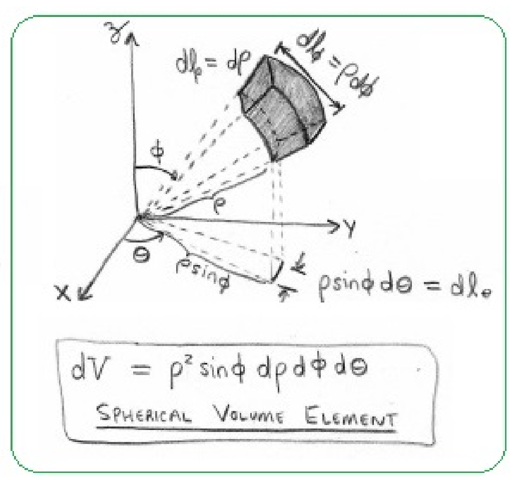
Therefore,

$$\begin{aligned} A\,\widehat{x}_1 \wedge A\,\widehat{x}_2 &= (a\,\widehat{x}_1 + c\,\widehat{x}_2) \wedge (b\,\widehat{x}_1 + d\,\widehat{x}_2) \\ &= ac\,\widehat{x}_1 \wedge \widehat{x}_1 + ad\,\widehat{x}_1 \wedge \widehat{x}_2 + cb\,\widehat{x}_2 \wedge \widehat{x}_1 + cd\,\widehat{x}_2 \wedge \widehat{x}_2 \\ &= (ad - cb)\,\widehat{x}_1 \wedge \widehat{x}_2 \end{aligned}$$

Thus, $(ad - cb)\,\widehat{x}_1 \wedge \widehat{x}_2 = \det(A)\,\widehat{x}_1 \wedge \widehat{x}_2$ thus comparing the coefficient of $\widehat{x}_1 \wedge \widehat{x}_2$ we find $\boxed{\det(A) = ad - cb}$. This is the usual formula for the 2×2 determinant. There is much more to say about the algebraic significance of the wedge product beyond this course. We simply introduce the reader to some of the basic benefits of the $\wedge$-product.

We now turn to the three dimensional case. I attempt to illustrate typical infinitesimal solid regions for cylindrical and spherical coordinates below:





In each case, the infinitesimal volume is obtained by multiplying the coordinate arclengths.

The algebraic method for volume elements follows the natural pattern we already saw for area elements. Again the wedge product yields something a bit different than a simple dV. I will denote the object calculated by $d\vec{V}$ to emphasize there is a distinction. For example, $dV = dx\,dy\,dz = dy\,dx\,dz$ and the distinction is merely the indicated order of an iterated integration. In contrast, by definition, $d\vec{V} = dx \wedge dy \wedge dz = -dy \wedge dx \wedge dz$ and the sign indicates that $dx \wedge dy \wedge dz$ orients a volume in an opposite sense to $dy \wedge dx \wedge dz$. If we could visualize four dimensions (x, y, z, w), then we could envision $dx \wedge dy \wedge dz$ as giving the hyperplane $w = k$ an upward orientation whereas $dy \wedge dx \wedge dz$ gives the hyperplane $w = k$ a downward-pointing orientation. In any event, we will think more on orientation when we study line and surface integrals which are both defined with respect to oriented objects. That said, I hope you can appreciate the following calculations. They are a large part of what sparked my initial interest in the topic of **differential forms**.

The cylindrical volume form follows from the work we already did to calculate the polar area form:

$$d\vec{V} = dx \wedge dy \wedge dz = d(r\cos\theta) \wedge d(r\sin\theta) \wedge dz = \boxed{rdr \wedge d\theta \wedge dz.}$$

In contrast, the spherical volume element requires some effort[12]:

$$\begin{aligned}
d\vec{V} &= dx \wedge dy \wedge dz \\
&= d(\rho\sin\phi\cos\theta) \wedge d(\rho\sin\phi\sin\theta) \wedge d(\rho\cos\phi) \\
&= [\sin\phi\cos\theta d\rho + \rho\cos\phi\cos\theta d\phi - \rho\sin\phi\sin\theta d\theta] \\
&\qquad \wedge [\sin\phi\sin\theta d\rho + \rho\cos\phi\sin\theta d\phi + \rho\sin\phi\cos\theta d\theta] \\
&\qquad \wedge [\cos\phi d\rho - \rho\sin\phi d\phi] \\
&= [\sin\phi\cos\theta d\rho + \rho\cos\phi\cos\theta d\phi - \rho\sin\phi\sin\theta d\theta] \wedge \\
&\qquad [\rho\sin\theta d\phi \wedge d\rho + \rho^2\sin\phi\cos\phi\cos\theta d\theta \wedge d\rho - \rho^2\sin^2\phi\cos\theta d\theta \wedge d\phi] \\
&= -\rho^2\sin^3\phi\cos^2\theta d\rho \wedge d\theta \wedge d\phi + \rho^2\cos^2\phi\sin\phi\cos^2\theta d\phi \wedge d\theta \wedge d\rho + \\
&\qquad \rho^2\sin^3\phi\sin^2\theta d\theta \wedge d\rho \wedge d\phi - \rho^2\sin\phi\cos^2\phi\sin^2\theta d\theta \wedge d\phi \wedge d\rho \\
&= [-\rho^2\sin^3\phi\cos^2\theta - \rho^2\cos^2\phi\sin\phi\cos^2\theta - \rho^2\sin^3\phi\sin^2\theta - \rho^2\sin\phi\cos^2\phi\sin^2\theta]d\rho \wedge d\theta \wedge d\phi \\
&= \rho^2\sin\phi[\sin^2\phi\cos^2\theta + \cos^2\theta\cos^2\phi + \sin^2\phi\sin^2\theta + \sin^2\theta\cos^2\theta]d\rho \wedge d\phi \wedge d\theta \\
&= \rho^2\sin\phi[\cos^2\theta(\sin^2\phi + \cos^2\phi) + \sin^2\theta(\sin^2\phi + \cos^2\phi)]d\rho \wedge d\phi \wedge d\theta \\
&= \boxed{\rho^2\sin\phi d\rho \wedge d\phi \wedge d\theta}
\end{aligned}$$

If you would like to learn more about differential forms then you might read my Advanced Calculus notes. In my current formulation, I spend about half the course developing and applying differential forms. A treatment with just calculus III in mind can be found in Susan Colley's *Vector Calculus*. She devotes a whole chapter to the exposition of basic differential forms.

6.9 Problems

Problem 151 Calculate

$$\int_0^2 \int_0^4 (3x+4y)\,dx\,dy$$

[12]you could try to follow this line by line, but, if you do attempt it, it is better to try your own path

Problem 152

$$\int_0^{\pi/2} \int_0^{\pi/2} \sin(x)\cos(y)\, dx\, dy$$

Problem 153

$$\int_{-1}^{1} \int_0^{1} \sin^3(x)\cos^{42}(y)\, dy\, dx$$

Problem 154 Calculate the average of $f(x, y) = x^2 + y^2$ on the unit-square.

Problem 155 Calculate the average of $f(x, y) = x^2 + y^2$ on the region bounded by $x^2 + y^2 = R^2$.

Problem 156 Calculate the average of $f(x, y) = xy$ on $[1, 2] \times [3, 4]$.

Problem 157 Show that

$$\lim_{n\to\infty} \int_0^1 \int_0^1 x^n\, y^n\, dx\, dy = 0.$$

Problem 158 Calculate

$$\int_0^{\ln(2)} \int_0^{\ln(3)} e^{x+y} dx\, dy$$

Problem 159 Suppose $\int\int_R f\, dA = \int_0^1 \int_{x^2}^{\sqrt{x}} (1 + x)dy\, dx$. Calculate the given integral.

Problem 160 For the integral given in the previous problem, explicitly write R as a subset of $\mathbb{R}^2$ using set-builder notation. In addition, calculate the integral once more with the interation of the integrals beginning with dx. Draw a picture to explain the inequalities which form the basis for your new set-up to the integral.

Problem 161 Reverse the order of integration in order to calculate the following integral:

$$\int_0^{\sqrt{\pi}} \int_y^{\sqrt{\pi}} \sin(x^2)\, dx\, dy.$$

Problem 162 Reverse the order of integration in order to calculate the following integral:

$$\int_0^1 \int_y^1 \frac{1}{1 + x^4}\, dx\, dy.$$

Problem 163 Find the average of $f(x, y) = xy$ over the triangle with vertices $(0, 0), (3, 1)$ and $(-2, 4)$.

Problem 164 Find volume bounded by $z = y + e^x$ and the xy-plane for $(x, y) \in [0, 1] \times [0, 2]$.

Problem 165 Find the volume bounded inside the cylinder $x^2 + y^2 = 1$ and the planes $z = x + 1$ and $z = y - 3$.

Problem 166 Find the volume bounded by the coordinate planes and the plane $3x + 2y + z = 6$.

Problem 167 Calculate the integral (use polars):

$$\int_0^2 \int_0^{\sqrt{4-x^2}} (x^2+y^2)^{3/2}\, dy\, dx.$$

Problem 168 Calculate the integral (use polars):

$$\int_0^1 \int_x^1 (x^2+y^2)^3\, dy\, dx.$$

Problem 169 Suppose R is the region bounded by $y + |x|$ and $x^2 + (y-1)^2 = 1$. Express R in polar coordinates. In other words, draw a picture and indicate how the points in R are reached by particular ranges of r and θ.

Problem 170 Find volume bounded by the paraboloid $x = y^2 + 2z^2$ and the parabolic cylinder $x = 2 - y^2$.

Problem 171 Find the volume bounded by the cylinder $x^2 + y^2 = 1$ and $z = 2 + x + y$ and $z = 1$.

Problem 171 Find the volume bounded by the cones $z = \sqrt{x^2+y^2}$ and $z = 2\sqrt{x^2+y^2}$ and the sphere $\rho = 3$.

Problem 172 Let B be a ball of radius R centered at the origin. Calculate $\int\int\int_B e^{-\rho^3}\, dV$

Problem 173 Let $u = \frac{2x}{x^2+y^2}$ and $v = \frac{-2y}{x^2+y^2}$ calculate $\frac{\partial(x,y)}{\partial(u,v)}$.

Problem 174 Suppose $\delta(x,y,z) = 1 = dM/dV$ for $x,y,z > 0$. Find center of mass for a sphere with this density δ centered at $(1,2,3)$.

Problem 175 Suppose $\delta(x,y,z) = xyz = dM/dV$ for $x,y,z > 0$. Find center of mass for a sphere with this density centered at $(1,2,3)$.

Problem 175+i Suppose you have a cylindrical oil tank which is placed on a hill close to your house. After some time the land settles and the oil tank is not level. Suppose you read a dip-stick which is designed for a level-set-up and find the tank is half-full. Suppose the tank is slanted at 20 degrees relative to the true horizontal. In other words, suppose the axis of the cylinder makes and angle of 70 degrees with the vertical. Fortunately, the tank is only tilted along that direction and the perpendicular direction to the central-axis remains at a right angle to the vertical. If you have a 1000gallon tank then how much oil do you really have?
numerical integration is totally fine here, although, this may have a closed-form solution.

Problem 176 Calculate $\int_R \sqrt{x+2y}\sin(x-y)\, dA$ where $R = [0,1]\times[0,1]$ by making an appropriate change of variables.

Problem 177 Find the center of mass for a laminate of variable density $\delta(r,\theta) = r\sin^2(\theta)$ which is bounded by $r = \sin(2\theta)$

Chapter 7

vector calculus

Vector calculus is the study of vector fields and related scalar functions. For the most part we focus our attention on two or three dimensions in this study. However, certain theorems are easily extended to $\mathbb{R}^n$. We explore these concepts in both Cartesian and the standard curvelinear coordiante systems. I also discuss the dictionary between the notations popular in math and physics [1]

The importance of vector calculus is nicely exhibited by the concept of a force field in mechanics. For example, the gravitational field is a field of vectors which fills space and somehow communicates the force of gravity between massive bodies. Or, in electrostatics, the electric field fills space and somehow communicates the influence between charges; like charges repel and unlike charges attract all through the mechanism propagated by the electric field. In magnetostatics constant magnetic fields fill space and communicate the magnetic force between various steady currents. All the examples above are in an important sense *static*. The source charges[2] are fixed in space and they cause a motion for some test particle immersed in the field. This is an idealization. In truth, the influence of the test particle on the source particle cannot be neglected, but those sort of interactions are far too complicated to discuss in elementary courses.

Often in applications the vector fields also have some time-dependence. The differential and integral calculus of time-dependent vector fields is not much different than that of static fields. The main difference is that what was once a constant is now some function of time. A time-dependent vector field is an assignment of a vectors at each point at each time. For example, the electric and magnetic fields that together make light. Or, the velocity field of a moving liquid or gas. Many other examples abound.

The calculus for vector fields involves new concepts of differentiation and new concepts of integration.

For differentiation, we study gradients, curls and divergence. The gradient takes a scalar field and generates a vector field (actually, this is not news for us). The curl takes a vector field and generates a new vector field which says how the given vector field curls about a point. The divergence takes a given vector field and creates a scalar function which quantifies how the given vector field diverges from a point. Many novel product rules exist for these operations and the algebra which links these

[1] this version of my notes uses the inferior math conventions as to be consistent with earlier math courses etc...

[2] the concept of a charge really allows for electric, magnetic or gravitational although isolated charges exist only for two of the aformentioned

operations is rich and interesting as we already saw at the beginning of this course as we studied vectors at a point.

On the topic of integration there are two types of integration we naturally consider for a vector field. We can integrate along an oriented curve, this type of integration is called a **line integral** even when the curve is not a line. Also, for a given surface, we can calculate a **surface integral** of a vector field. The line integral measures how the vector field lines up along the curve of integration, this is known as the **circulation** of the field along the curve. The surface integrals value depends on how the vector field pokes through the surface of integration, it measures something called **flux**. Critical to both line and surface integrals are parametric equations for curves and surfaces. We'll see that we need parameters to calculate anything yet the answers are completely independent of the parameters utilized. In other words, the surface and line integrals are coordinate free objects. This is in considerable contrast to the types of integrals we studied in the previous part of this course.

Between the study of differentiation and integration of vector fields we find the unifying theorems of Green, Gauss and Stokes. We study Green's theorem to begin since it is simple and merely a two-dimensional result. However, it is evidence of something deeper. Both Gauss and Stokes reduce to Green's in certain context. There is also a fundamental theorem of line integrals which helps validate my claim that intuitively the ∇f is basically the derivative for a function of several variables. In addition, we will learn about how these integral theorems relate to the path-independence of vector fields. We saw earlier that certain vector fields could not be gradients because they violate Clairaut's Theorem on their components. On the other hand, just because a vector field has components consistent with Clairaut's theorem we also saw that it was not necessarily the case they were the gradient of some scalar function on their whole domain. We'll find sufficient conditions to make the Clairaut test work. We will find how to say with certainty a given vector field is the gradient of some scalar function. More than just that, we'll find a method to calculate the scalar function.

The end of the chapter studies potential theory in three dimensions. We find mathematical justification for some of the well-known potential formulas seen in the physics of electricity and magnetism. In short, we show how charge is tied to singularity and how integration of charge (or current) gives rises to the electric and magnetic fields. Of course, the application to electromagnetics is merely a choice, the identities we discover for vector fields in three dimensions are general mathematical results.

I would like to add many pictures in this chapter. I hope to add these in lecture this semester and ideally, this chapter will be revised with many additional pictures at the end of the Fall 2014 semester.

7.1 vector fields

Definition 7.1.1.

A vector field on $S \subseteq \mathbb{R}^n$ is an assignment of a n-dimensional vector to each point in S. If $\vec{F} = \langle F_1, F_2, \dots, F_n \rangle = \sum_{j=1}^{n} F_j \widehat{x}_j$ then the multivariate functions $F_j : S \to \mathbb{R}$ are called the **component functions** of $\vec{F}$. In the cases of $n = 2$ and $n = 3$ we sometimes use the popular notations

$$\vec{F} = \langle P, Q \rangle \qquad \& \qquad \vec{F} = \langle P, Q, R \rangle.$$

We have already encountered examples of vector fields earlier in this course.

Example 7.1.2. *Let $\vec{F} = \widehat{x}$ or $\vec{G} = \widehat{y}$. These are constant vector fields on $\mathbb{R}^2$. At each point we attach the same fixed vector; $\vec{F}(x,y) = \langle 1, 0 \rangle$ or $\vec{G}(x,y) = \langle 0, 1 \rangle$. In contrast, $\vec{H} = \widehat{r}$ and $\vec{I} = \widehat{\theta}$ are non-constant and technically are only defined on the punctured plane $\mathbb{R}^2 - \{(0,0)\}$. In particular,*

$$\vec{H}(x,y) = \left\langle \frac{x}{\sqrt{x^2+y^2}} \ , \ \frac{y}{\sqrt{x^2+y^2}} \right\rangle \qquad \& \qquad \vec{I}(x,y) = \left\langle \frac{-y}{\sqrt{x^2+y^2}} \ , \ \frac{x}{\sqrt{x^2+y^2}} \right\rangle.$$

Of course, given any differentiable $f : S \subseteq \mathbb{R}^2 \to \mathbb{R}$ we can create the vector field $\nabla f = \langle \partial_x f, \partial_y f \rangle$ which is normal to the level curves of f.

Example 7.1.3. *Let $\vec{F} = \widehat{x}$ or $\vec{G} = \widehat{y}$ or $\vec{H} = \widehat{z}$. These are constant vector fields on $\mathbb{R}^3$. At each point we attach the same fixed vector; $\vec{F}(x,y,z) = \langle 1, 0, 0 \rangle$ or $\vec{G}(x,y,z) = \langle 0, 1, 0 \rangle$ and $\vec{H}(x,y,z) = \langle 0, 1, 0 \rangle$. In contrast, $\vec{I} = \widehat{r}$ and $\vec{J} = \widehat{\theta}$ are non-constant and technically are only defined on $\mathbb{R}^3 - \{(0,0,z) \mid z \in \mathbb{R}\}$. In particular,*

$$\vec{I}(x,y,z) = \left\langle \frac{x}{\sqrt{x^2+y^2}} \ , \ \frac{y}{\sqrt{x^2+y^2}} \right\rangle \qquad \& \qquad \vec{J}(x,y,z) = \left\langle \frac{-y}{\sqrt{x^2+y^2}} \ , \ \frac{x}{\sqrt{x^2+y^2}} \right\rangle.$$

Similarly, the spherical coodinate frame $\widehat{\rho}, \widehat{\phi}, \widehat{\theta}$ are vector fields on the domain of their definition. Of course, given any differentiable $f : S \subseteq \mathbb{R}^3 \to \mathbb{R}$ we can create the vector field $\nabla f = \langle \partial_x f, \partial_y f, \partial_z f \rangle$ which is normal to the level surfaces of f.

The **flow-lines** or **streamlines** are paths for which the velocity field matches a given vector field. Sometimes these paths which line up with the vector field are also called **integral curves**.

Definition 7.1.4.

In particular, given a vector field $\vec{F} = \langle P, Q, R \rangle$ we say $\vec{r}(t) = \langle x(t), y(t), z(t) \rangle$ is an **integral curve** of $\vec{F}$ iff

$$\vec{F}(\vec{r}(t)) = \left\langle \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right\rangle \qquad a.k.a \qquad \frac{dx}{dt} = P, \ \frac{dx}{dt} = Q, \ \frac{dz}{dt} = R.$$

We need to integrate each component of the vector field to find this curve. Of course, given that P, Q, R are typically functions of x, y, z the "integration" requires thought. Even in differential equations(334) the general problem of finding integral curves for vector fields is beyond our standard techniques for all but a handful of well-behaved vector fields. That said, the streamlines for the examples below are geometrically obvious so we can reasonably omit the integration.

Example 7.1.5. *Suppose $\vec{F} = \widehat{x}$ then obviously $x(t) = x_o + t, y = y_o, z = z_o$ is the streamline of $\widehat{x}$ through (x_o, y_o, z_o). Likewise, I think you can calculate the steamlines for $\widehat{y}$ and $\widehat{z}$ without much trouble. In fact, any constant vector field $\vec{F} = \vec{v}_o$ simply has streamlines which are lines with direction vector $\vec{v}_o$.*

Example 7.1.6. *The magnetic field around a long steady current in the postive z-direction is conveniently written as $\vec{B}(r, \theta, z) = \frac{\mu_o I}{2\pi r}\widehat{\theta}$. The streamlines are circles which are centered on the axis and point in the $\widehat{\theta}$ direction.*

Example 7.1.7. *If a charge Q is distributed uniformly through a sphere of radius R then the electric field can be show to me a function of the distance from the center of the sphere alone. Placing that center at the origin gives*

$$\vec{E}(\rho) = \widehat{\rho}\begin{cases} \frac{k\rho}{R^3} & 0 \leq \rho \leq R \\ \frac{k}{\rho^2} & \rho \geq R \end{cases}$$

The streamlines are simply lines which flow radially out from the origin in all directions.

Challenge: in electrostatics the density of streamlines (often called fieldlines in physics) is used to measure the magnitude of the electric field. Why is that reasonable?

Example 7.1.8. *The other side of the thinking here is that given a differential equation we could use the plot of the vector field to indicate the flow of solutions. We can solve numerically by playing a game of directed connect the dots which is the multivariate analog of Euler's method for solving $dy/dx = f(x, y)$.*

$$\frac{dx}{dt} = 2x(y^2 + z^2), \qquad \frac{dy}{dt} = 2x^2y, \qquad \frac{dz}{dt} = 2x^2z$$

We'd look to match the curve up with the vector-field plot of $\vec{F} = \langle 2x(y^2 + z^2), 2x^2y, 2x^2z\rangle$. This particular field is a gradient field with $\vec{F} = \nabla f$ for $f(x, y, z) = x^2(y^2 + z^2)$. Solutions to the differential equations describe paths which are orthogonal to the level surfaces of f since the paths are parallel to ∇f.

Perhaps you can see how this way of thinking might be productive towards analyzing otherwise intractable problems in differential equations. I merely illustrate here to give a bit more breadth to the concept of a vector field. Of course most modern texts have pretty pictures with real world jutsu so you should read that if this is not *real* to you without those comments. There are many websites to help visualize vector fields, I'll probably demo one during an appropriate lecture.

7.2 grad, curl and div

In this section we investigate a few natural derivatives we can construct with the operator ∇. Later we will explain what these derivatives *mean*. First, the computation:

Definition 7.2.1.

Suppose f is a scalar function on $\mathbb{R}^3$ then we defined the gradient vector field

$$grad(f) = \nabla f = \langle \partial_x f, \partial_y f, \partial_z f \rangle$$

We studied this before, recall that we can compactly express this by

$$\nabla f = \sum_{i=1}^{3} (\partial_i f)\widehat{x}_i$$

where $\partial_i = \partial/\partial x_i$ and $x_1 = x, x_2 = y$ and $x_3 = z$. Moreover, we have also shown previously in notes or homework that the gradient has the following important properties:

$$\nabla(f+g) = \nabla f + \nabla g, \quad \& \quad \nabla(cf) = c\nabla f, \quad \& \quad \nabla(fg) = (\nabla f)g + f(\nabla g)$$

Together these say that ∇ is a *derivation* of differentiable functions on $\mathbb{R}^n$.

Definition 7.2.2.

Suppose $\vec{F} = \langle F_1, F_2, F_3 \rangle$ is a vector field. We define:

$$Div(\vec{F}) = \nabla \bullet \vec{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}.$$

More compactly, we can express the divergence by

$$\nabla \bullet \vec{F} = \sum_{i=1}^{3} \partial_i F_i.$$

You can prove that the divergence satisfies the following important properties:

$$\nabla \bullet (\vec{F} + \vec{G}) = \nabla \bullet \vec{F} + \nabla \bullet \vec{G} \quad \& \quad \nabla \bullet (c\vec{F}) = c\nabla \bullet \vec{F}, \quad \& \quad \nabla \bullet (f\vec{G}) = \nabla f \bullet \vec{G} + f\nabla \bullet \vec{G}$$

For example,

$$\nabla \bullet (\vec{F} + c\vec{G}) = \sum_{i=1}^{3} \partial_i (F_i + cG_i) = \sum_{i=1}^{3} \partial_i F_i + c\sum_{i=1}^{3} \partial_i G_i = \nabla \bullet \vec{F} + c\nabla \bullet \vec{G}.$$

Linearity of the divergence follows naturally from linearity of the partial derivatives.

Definition 7.2.3.

Suppose $\vec{F} = \langle F_1, F_2, F_3 \rangle$ is a vector field. We define:

$$Curl(\vec{F}) = \nabla \times \vec{F} = \left\langle \frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}, \frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x}, \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right\rangle$$

More compactly, using the antisymmetric symbol ϵ_{ijk}[3],

$$\nabla \times \vec{F} = \sum_{i,j,k=1}^{3} \epsilon_{ijk}(\partial_i F_j)\widehat{x}_k.$$

You can prove that the curl satisfies the following important properties:

$$\nabla \times (\vec{F}+\vec{G}) = \nabla \times \vec{F} + \nabla \times \vec{G} \quad \& \quad \nabla \times (c\vec{F}) = c\nabla \times \vec{F}, \quad \& \quad \nabla \times (f\vec{G}) = \nabla f \times \vec{G} + f\nabla \times \vec{G}.$$

For example,

$$\begin{aligned}
\nabla \times (\vec{F} + c\vec{G}) &= \sum_{i,j,k=1}^{3} \epsilon_{ijk}\partial_i(F_j + cG_j)\widehat{x}_k \\
&= \sum_{i,j,k=1}^{3} \epsilon_{ijk}(\partial_i F_j)\widehat{x}_k + c\sum_{i,j,k=1}^{3} \epsilon_{ijk}(\partial_i G_j)\widehat{x}_k \\
&= \nabla \times \vec{F} + c\nabla \times \vec{G}.
\end{aligned}$$

Linearity of the curl follows naturally from linearity of the partial derivatives.

It is fascinating how many of the properties of ordinary differentiation generalize to the case of vector calculus. The main difference is that we now must take more care to not commute things that don't commute or confuse functions with vector fields. For example, while it is certainly true that $\vec{A}\cdot\vec{B} = \vec{B}\cdot\vec{A}$ it is not even sensible to ask the question does $\nabla \cdot \vec{A} = \vec{A}\cdot\nabla$? Notice $\nabla\cdot\vec{A}$ is a function while $\vec{A}\cdot\nabla$ is an operator, apples and oranges.

The proposition below lists a few less basic identities which are at times useful for differential vector calculus.

Proposition 7.2.4.

Let f, g, h be real valued functions on $\mathbb{R}$ and $\vec{F}, \vec{G}, \vec{H}$ be vector fields on $\mathbb{R}$ then (assuming all the partials are well defined)

$$\begin{aligned}
&(i.)\ \nabla\cdot(\vec{F}\times\vec{G}) = \vec{G}\cdot(\nabla\times\vec{F}) - \vec{F}\cdot(\nabla\times\vec{G}) \\
&(ii.)\ \nabla(\vec{F}\cdot\vec{G}) = \vec{F}\times(\nabla\times\vec{G}) + \vec{G}\times(\nabla\times\vec{F}) + (\vec{F}\cdot\nabla)\vec{G} + (\vec{G}\cdot\nabla)\vec{F} \\
&(iii.)\ \nabla(\vec{F}\times\vec{G}) = (\vec{G}\cdot\nabla)\vec{F} - (\vec{F}\cdot\nabla)\vec{G} + \vec{F}(\nabla\cdot\vec{G}) - \vec{G}(\nabla\cdot\vec{F})
\end{aligned}$$

Proof: Consider (i.), let $\vec{F} = \sum F_i e_i$ and $\vec{G} = \sum G_i e_i$ as usual,

$$\begin{aligned}
\nabla\cdot(\vec{F}\times\vec{G}) &= \sum \partial_k[(\vec{F}\times\vec{G})_k] \\
&= \sum \partial_k[\epsilon_{ijk}F_i G_j] \\
&= \sum \epsilon_{ijk}[(\partial_k F_i)G_j + F_i(\partial_k G_j)] \\
&= \sum \epsilon_{ijk}(\partial_i F_j)G_k + -F_j\epsilon_{ikj}(\partial_i G_k) \\
&= \sum G_k(\nabla\times\vec{F}) - F_j(\nabla\times\vec{G}) \\
&= \vec{G}\cdot(\nabla\times\vec{F}) - \vec{F}\cdot(\nabla\times\vec{G}).
\end{aligned} \tag{7.1}$$

[3] recall we used this before to better denote harder calculations involving the cross-product

where the sums above are taken over the indices which are repeated in the given expressions. In physics the $\sum$ is often removed and the *einstein index convention* or *implicit summation convention* is used to free the calculation of cumbersome summation symbols. The proof of the other parts of this proposition can be handled similarly, although parts (viii) and (ix) require some thought so I may let you do those for homework[4]. $\square$

Proposition 7.2.5.

If f is a differentiable $\mathbb{R}$-valued function and $\vec{F}$ is a differentiable vector field then

$$\begin{array}{l} (i.)\ \ \nabla \cdot (\nabla \times \vec{F}) = 0 \\ (ii.)\ \ \nabla \times \nabla f = 0 \\ (iii.)\ \ \nabla \times (\nabla \times \vec{F}) = \nabla(\nabla \cdot \vec{F}) - \nabla^2 \vec{F} \end{array}$$

Before the proof, let me briefly indicate the importance of $(iii.)$ to physics. We learn that in the absence of charge and current the electric and magnetic fields are solutions of

$$\nabla \bullet \vec{E} = 0, \qquad \nabla \times \vec{E} = -\partial_t \vec{B}, \qquad \nabla \bullet \vec{B} = 0, \qquad \nabla \times \vec{B} = \mu_o \epsilon_o \partial_t \vec{E}$$

If we consider the curl of the curl equations we derive,

$$\nabla \times (\nabla \times \vec{E}) = \nabla \times (-\partial_t \vec{B}) \quad \Rightarrow \quad \nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} = -\partial_t (\nabla \times \vec{B}) \quad \Rightarrow \quad \nabla^2 \vec{E} = \mu_o \epsilon_o \partial_t^2 \vec{E}.$$

$$\nabla \times (\nabla \times \vec{B}) = \nabla \times (\mu_o \epsilon_o \partial_t \vec{E}) \quad \Rightarrow \quad \nabla(\nabla \cdot \vec{B}) - \nabla^2 \vec{B} = \mu_o \epsilon_o \partial_t (\nabla \times \vec{E}) \quad \Rightarrow \quad \nabla^2 \vec{B} = \mu_o \epsilon_o \partial_t^2 \vec{B}.$$

These are **wave equations**. If you study the physics of waves you might recognize that the speed of the waves above is $v = 1/\sqrt{\mu_o \epsilon_o}$. This is the speed of light. We have shown that the speed of light apparently depends only on the basic properties of space itself. It is independent of the x, y, z coordinates so far as we can see in the usual formalism of electromagnetism. This math was only possible because Maxwell added a term called the displacement current in about 1860. Not many years later radio and TV was invented and all because we knew to look for the possiblility thanks to this mathematics. That said, the notation used above was not common in Maxwell's time. His original presentation of what we now call Maxwell's Equations was given in terms of 20 scalar partial differential equations. Now we enjoy the clarity and precision of the vector formalism.

Proof: I like to use parts (i.) and (ii.) for test questions at times. They're pretty easy, I leave them to the reader. The proof of (iii.) is a bit deeper. We need the well-known identity

[4]relax fall 2011 students this did not happen to you

$\sum_{j=1}^{3} \epsilon_{ikj}\epsilon_{lmj} = \delta_{il}\delta_{km} - \delta_{kl}\delta_{im}$[5]

$$\begin{aligned}
\nabla \times (\nabla \times \vec{F}) &= \sum_{i,j,k=1}^{3} \epsilon_{ijk}\partial_i(\nabla \times \vec{F})_j \widehat{x}_k \\
&= \sum_{i,j,k=1}^{3} \epsilon_{ijk}\partial_i(\sum_{l,m=1}^{3} \epsilon_{lmj}\partial_l F_m)\widehat{x}_k \\
&= \sum_{i,j,k,l,m=1}^{3} \epsilon_{ijk}\epsilon_{lmj}(\partial_i\partial_l F_m)\widehat{x}_k \\
&= \sum_{i,j,k,l,m=1}^{3} -\epsilon_{ikj}\epsilon_{lmj}(\partial_i\partial_l F_m)\widehat{x}_k \\
&= \sum_{i,k,l,m=1}^{3} (-\delta_{il}\delta_{km} + \delta_{kl}\delta_{im})(\partial_i\partial_l F_m)\widehat{x}_k \\
&= \sum_{i,k,l,m=1}^{3} (-\delta_{il}\delta_{km}\partial_i\partial_l F_m)\widehat{x}_k + \sum_{i,k,l,m=1}^{3} (\delta_{kl}\delta_{im}\partial_i\partial_l F_m)\widehat{x}_k \\
&= \sum_{i,k=1}^{3} -\partial_i\partial_i(F_k\widehat{x}_k) + \sum_{i,k=1}^{3} (\partial_i\partial_k F_i)\widehat{x}_k \\
&= -\sum_{i=1}^{3} \partial_i\partial_i(\sum_{k=1}^{3} F_k\widehat{x}_k) + \sum_{k=1}^{3} \partial_k(\sum_{i=1}^{3} \partial_i F_i)\widehat{x}_k \\
&= -\nabla^2 \vec{F} + \nabla(\nabla \bullet \vec{F}). \qquad \square
\end{aligned}$$

7.3 line integrals

For those of you who know a little physics, the motivation to define this integral follows from our desire to calculate the work done by a variable force $\vec{F}$ on some particle as it traverses C. In particular, we expect the little bit of work dW done by $\vec{F}$ as the particle goes from $\vec{r}$ to $\vec{r} + d\vec{r}$ is given by $\vec{F} \bullet d\vec{r}$. Then, to find total work, we integrate:

Definition 7.3.1.

Let $\vec{\gamma} : [a, b] \to C \subset \mathbb{R}^3$ be a differentiable path which covers the oriented curve C and suppose that $C \subset dom(\vec{F})$ for a continuous vector field $\vec{F}$ on $\mathbb{R}^3$ then the **vector line integral of $\vec{F}$ along** C is denoted and defined as follows:

$$\int_C \vec{F} \bullet d\vec{r} = \int_a^b \vec{F}(\vec{\gamma}(t)) \bullet \frac{d\vec{\gamma}}{dt}\, dt.$$

This integral measures the **work** done by $\vec{F}$ over C. Alternatively, this is also called the **circulation** of $\vec{F}$ along C, however that usage tends to appear in the case that C is a loop. A closed curve is

[5]this is actually just the first in a whole sequence of such identities linking the antisymmetric symbol and the kronecker deltas... ask me in advanced calculus, I'll show you the secret formulas

defined to be a curve which has the same starting and ending points. We can indicate the line-integral is taken over a loop by the notation $\oint_C \vec{F} \bullet d\vec{r}$. As with the case of the scalar line integral we ought to examine the dependence of the definition on the choice of parametrization for C. If we were to find a dependence then we would have to modify the definition to make it reasonable. Once more consider the reparametrization $\vec{\gamma_2}$ of $\vec{\gamma_1}$ by a strictly monotonic differentiable function $u : [a_1, b_1] \to [a_2, b_2]$ where we have $\vec{\gamma_1}(t) = \vec{\gamma_2}(u(t))$. Consider,

$$\begin{aligned}\int_{a_1}^{b_1} \vec{F}(\vec{\gamma_1}(t)) \bullet \frac{d\vec{\gamma_1}}{dt}\, dt &= \int_{a_1}^{b_1} \vec{F}(\vec{\gamma_2}(u(t))) \frac{d}{dt}\Big[\vec{\gamma_2}(u(t))\Big]\, dt \\ &= \int_{a_1}^{b_1} \vec{F}(\vec{\gamma_2}(u(t))) \frac{d\vec{\gamma_2}}{dt}(u(t)) \cdot \frac{du}{dt}\, dt \\ &= \int_{u(a_1)}^{u(b_1)} \vec{F}(\vec{\gamma_2}(u) \frac{d\vec{\gamma_2}}{du}\, du\end{aligned}$$

If u is orientation preserving then $u(a_1) = a_2$ and $u(b_1) = b_2$ and we find the integral $\int_{a_1}^{b_1} \vec{F}(\vec{\gamma_1}(t)) \bullet \frac{d\vec{\gamma_1}}{dt}\, dt = \int_{a_2}^{b_2} \vec{F}(\vec{\gamma_2}(u) \frac{d\vec{\gamma_2}}{du}\, du$. However, if u is orientation reversing then we find $u(a_1) = b_2$ and $u(b_1) = a_2$ hence $\int_{a_1}^{b_1} \vec{F}(\vec{\gamma_1}(t)) \bullet \frac{d\vec{\gamma_1}}{dt}\, dt = -\int_{b_2}^{a_2} \vec{F}(\vec{\gamma_2}(u) \frac{d\vec{\gamma_2}}{du}\, du$. Therefore, we find that

$$\boxed{\int_{-C} \vec{F} \bullet d\vec{r} = -\int_C \vec{F} \bullet d\vec{r}}$$

This is actually not surprising if you think about the motivation for the integral. The integral measures how the vector field $\vec{F}$ points in the same direction as C. The curve $-C$ goes in the opposite direction thus it follows the sign should differ for the line-integral. Long story short, we must take line-integrals with respect to **oriented curves**.

It is instructive to relate the line-integral and the integral with respect to arclength[6],

$$\int_C \vec{F} \bullet d\vec{r} = \int_a^b \vec{F}(\vec{\gamma}(t)) \bullet \frac{d\vec{\gamma}}{dt}\, dt = \int_a^b \vec{F}(\vec{\gamma}(t)) \bullet \left[\frac{\vec{\gamma}\,'(t)}{||\vec{\gamma}\,'(t)||}\right] ||\vec{\gamma}\,'(t)||\, dt = \int_C (\vec{F} \bullet \vec{T}) ds.$$

As the last equality indicates, the vector line integral of $\vec{F}$ is given by the scalar line integral of the tangential component $\vec{F} \cdot \vec{T}$. Thus the vector line integral of $\vec{F}$ along C gives us a measure of how much $\vec{F}$ points in the same direction as the oriented curve C. If the vector field always cuts the path perpendicularly (if it was normal to the curve) then the vector line integral would be zero.

Example 7.3.2. *Suppose $\vec{F}(x, y, z) = \langle y, z - 1 + x, 2 - x\rangle$ and suppose C is an ellipse on the plane $z = 1 - x - y$ where $x^2 + y^2 = 4$ and we orient C in the CCW direction relative to the xy-plane with positive normal (imagine pointing your right hand above the xy-plane and your fingers curl around the ellipse in the CCW direction). Our goal is to calculate $\int_C \vec{F} \bullet d\vec{r}$. To do this we must first understand how to parametrize the ellipse:*

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

[6] think about this equality with $-C$ in place of C, why is this not a contradiction? On first glance you might think only the lhs is orientation dependent.

gives $x^2 + y^2 = 4$ *and the CCW direction. To find* z *we use the plane equation,*

$$z = 1 - x - y = 1 - 2\cos(t) - 2\sin(t)$$

Therefore,

$$\vec{r}(t) = \langle 2\cos(t), 2\sin(t), 1 - 2\cos(t) - 2\sin(t)\rangle$$

thus

$$\frac{d\vec{r}}{dt} = \Big\langle -2\sin(t), 2\cos(t), 2[\sin(t) - \cos(t)]\Big\rangle$$

Evaluate $\vec{F}(x, y, z) = \langle y, z - 1 + x, 2 - x\rangle$ *at* $x = 2\cos(t)$, $\;y = 2\sin(t)$ *and* $z = 1 - 2\cos(t) - 2\sin(t)$ *to find*

$$\vec{F}(\vec{r}(t)) = \Big\langle 2\sin(t), -2\sin(t), 2[1 - \cos(t)]\Big\rangle$$

Now put it together,

$$\begin{aligned}
\oint_C \vec{F} \bullet d\vec{r} &= \int_0^{2\pi} \Big\langle 2\sin(t), -2\sin(t)), 2[1 - \cos(t)]\Big\rangle \bullet \Big\langle -2\sin(t), 2\cos(t), 2[\sin(t) - \cos(t)]\Big\rangle\, dt \\
&= \int_0^{2\pi} \Big[-4\sin^2(t) - 4\sin(t)\cos(t) + 4[1 - \cos(t)][\sin(t) - \cos(t)]\Big]\, dt \\
&= \int_0^{2\pi} \Big[-4\sin^2(t) - 4\sin(t)\cos(t) + 4\sin(t) - 4\cos(t) - 4\cos(t)\sin(t) + 4\cos^2(t)\Big]\, dt \\
&= 4\int_0^{2\pi} \big[\cos^2(t) - \sin^2(t)\big]\, dt \\
&= 4\int_0^{2\pi} \big[\tfrac{1}{2}(1 + \cos(2t) - \tfrac{1}{2}(1 - \cos(2t))\big]\, dt \\
&= 4\int_0^{2\pi} \big[\tfrac{1}{2}(1 + \cos(2t) - \tfrac{1}{2}(1 - \cos(2t))\big]\, dt \\
&= 4\int_0^{2\pi} \big[\cos(2t)\big]\, dt \\
&= 0.
\end{aligned}$$

The example above indicates how we apply the definition of the line-integral directly. Sometimes it is convenient to use differential notation. If C is parametrized by $\vec{r} = \langle x, y, z\rangle$ for $a \leq t \leq b$ we define the integrals of the differential forms Pdx, Qdy and Rdz in the following way:

Definition 7.3.3.

Let $\vec{r} : [a, b] \to C \subset \mathbb{R}^3$ be a differentiable path which covers the oriented curve C and suppose that $C \subset dom(\langle P, Q, R\rangle)$ for a continuous vector field $\langle P, Q, R\rangle$ on $\mathbb{R}^3$ then we define

$$\int_C Pdx = \int_a^b P(\vec{r}(t))\frac{dx}{dt}\, dt, \quad \int_C Qdx = \int_a^b Q(\vec{r}(t))\frac{dy}{dt}\, dt, \quad \int_C Rdx = \int_a^b R(\vec{r}(t))\frac{dz}{dt}\, dt,$$

These are not basic calculations and in and of themselves they are not terribly interesting. I suppose that the $\int_C Pdx$ measures the work done by the x-vector-component of $\vec{F} = \langle P, Q, R\rangle$ whereas the

$\int_C Qdy$ and the $\int_C Rdz$ measure the work done by the y and z vector components of $\vec{F} = \langle P, Q, R\rangle$. Primarily, these are interesting since when we add them we obtain the full line-integral:

$$\boxed{\int_C \langle P, Q, R\rangle \bullet d\vec{r} = \int_C \big(Pdx + Qdy + Rdz\big)}$$

I invite the reader to verify the formula above. I will illustrate its use in many examples to follow. It should be emphasized that these are just notation to organize the line integral.

Example 7.3.4. *Calculate $\int_C \vec{F} \bullet d\vec{r}$ for $\vec{F}(x, y, z) = \langle y, z-1+x, 2-x\rangle$ given that C is parametrized by $x = \cos(t), y = \sin(t), z = 1$ for $0 \le t \le 2\pi$. Note that*

$$dx = -\sin(t)dt, \quad dy = \cos(t)dt, \quad dz = 0$$

Thus, for $P = y = \sin(t)$, $Q = z - 1 + x = \cos(t)$ and $R = 2 - x = 2 - \cos(t)$ we find

$$\oint_C \vec{F} \bullet d\vec{r} = \int_C \big(Pdx + Qdy + Rdz\big) = \int_0^{2\pi} -\sin^2(t)dt + \cos(t)\cos(t)dt = \int_0^{2\pi} \cos(2t)dt = 0.$$

You might wonder if the integral around a closed curve is always zero.

Example 7.3.5. *Let $\vec{F} = \langle y, -x\rangle$ and suppose $x = R\cos(t), y = R\sin(t)$ parametrizes C for $0 \le t \le 2\pi$. Calculate,*

$$Pdx + Qdy = -yR\sin(t)dt - xR\cos(t)dt = -R^2\sin^2(t)dt - R^2\cos^2(t)dt = -R^2dt$$

Thus, $\oint_C \vec{F} \bullet d\vec{r} = -\int_0^{2\pi} R^2dt = -2\pi R^2$.

Apparently just because we integrate around a loop it does not mean the answer is zero. I suspect that there are loops for which $\vec{F}(x, y, z) = \langle y, z - 1 + x, 2 - x\rangle$ has nonzero circulation. We will return to that example once more in the next section after we learn a test to determine if the $\oint_C \vec{F} \bullet d\vec{r} = 0$ without direct calculation. In conclusion, I should mention that the properties below are easily proved by direct calculation on the defintion,

$$\int_C (\vec{F} + c\vec{G}) \bullet d\vec{r} = \int_C \vec{F} \bullet d\vec{r} + c\int_C \vec{G} \bullet d\vec{r} \qquad \& \qquad \int_{C \cup \bar{C}} \vec{F} \bullet d\vec{r} = \int_C \vec{F} \bullet d\vec{r} + \int_{\bar{C}} \vec{F} \bullet d\vec{r}.$$

7.4 conservative vector fields

In this section we discuss how to identify a conservative vector field and how to use it. There are about 5 equivalent ideas and our job in this section is to explore how these concepts are connected. We also make a few connections with physics and it should be noted that part of the terminology is borrowed from classical mechanics. Let us begin with the fundamental theorem for line integrals.

Theorem 7.4.1. *Fundamental Theorem of Calculus for Line Integrals.*

Suppose f is differentiable on some open set containing the oriented curve C from P to Q then

$$\int_C \nabla f \bullet d\vec{r} = f(Q) - f(P).$$

Proof: let $\vec{r} : [a,b] \to C \subset \mathbb{R}^n$ parametrize C and calculate:

$$\begin{aligned}\int_C \nabla f \bullet d\vec{r} &= \int_a^b \nabla f(\vec{r}(t)) \bullet \frac{d\vec{r}}{dt}\, dt \\ &= \int_a^b \frac{d}{dt}\Big[f(\vec{r}(t))\Big] dt \\ &= f(\vec{r}(b)) - f(\vec{r}(a)) \\ &= f(Q) - f(P).\end{aligned}$$

The two critical steps above are the application of the multivariate chain-rule and then in the next to last step we apply the FTC from single-variable calculus. $\square$

Definition 7.4.2. *conservative vector field*

Suppose $U \subseteq \mathbb{R}^n$ then we say $\vec{F}$ is **conservative** on U iff there exists a **potential function** f such that $\vec{F} = \nabla f$ on U. Moreover, if $\vec{F}$ is conservative on $dom(\vec{F})$ then we say $\vec{F}$ is a conservative vector field.

The beauty of a conservative vector field is we trade computation of a line-integral for evaluation at the end-points.

Example 7.4.3. *Suppose $\vec{F}(x,y,z) = \langle 2x, 2y, 3\rangle$ for all $(x,y,z) \in \mathbb{R}^3$. Suppose C is a curve from $(0,0,0)$ to (a,b,c). Calculate $\int_C \vec{F} \bullet d\vec{r}$. Observe that*

$$f(x,y,z) = x^2 + y^2 + 3z \quad \Rightarrow \quad \vec{F} = \nabla f.$$

Therefore,

$$\int_C \vec{F} \bullet d\vec{r} = f(a,b,c) - f(0,0,0) = a^2 + b^2 + 3c.$$

Notice that we did not have to know where the curve C went since the FTC applies and only the endpoints of the curve are needed. I invite the reader to check this result by explicit computation along some path.

Why "conservative"? Let me address that. The key is a little identity, if m is a constant,

$$\frac{d}{dt}\left[\frac{1}{2}mv^2\right] = \frac{d}{dt}\left[\frac{1}{2}m\vec{v} \bullet \vec{v}\right] = \frac{1}{2}m\left[\frac{d\vec{v}}{dt} \bullet \vec{v} + \vec{v} \bullet \frac{d\vec{v}}{dt}\right] = m\vec{a} \bullet \vec{v}.$$

If $\vec{F}$ is the net-force on a mass m then Newton's Second Law states $\vec{F} = m\vec{a}$ therefore, if C is a curve from $\vec{r}_1$ to $\vec{r}_2$

$$\int_C \vec{F} \bullet d\vec{r} = \int_{t_1}^{t_2} \vec{F} \bullet \frac{d\vec{r}}{dt} dt = \int_{t_1}^{t_2} \big(\, m\vec{a} \bullet \vec{v}\, \big) dt = \int_{t_1}^{t_2} \frac{d}{dt}\left[\frac{1}{2}mv^2\right] dt = K(t_2) - K(t_1)$$

where $K = \frac{1}{2}mv^2$ is the kinetic energy. This result is known as the **work-energy** theorem. It does not require that $\vec{F}$ be conservative. If $\vec{F}$ is conservative then it is traditional to choose a potential energy function U such that $\vec{F} = -\nabla U$. In this case we can use the FTC for line-integrals to once more calculate the work done by the net-force,

$$\int_C \vec{F} \bullet d\vec{r} = -\int_C \nabla U \bullet d\vec{r} = -U(\vec{r}_2) + U(\vec{r}_1)$$

It follows that we have, for a conservative force, $K_2 - K_1 = -U_2 + U_1$ hence $K_1 + U_1 = K_2 + U_2$. The quantity $E = U + K$ is the total mechanical energy and it is a constant of the motion when only conservative forces comprise the net-force. This is the reason I call a vector field which is a gradient field of some pontential a conservative vector field. When viewed as a net-force it provides the conservation of energy[7]. It turns out that usually we can find portions of the domain of an arbitrary vector field on which the vector field is conservative. The obstructions to the existence of a global potential are the interesting part.

Definition 7.4.4. *path-independence*

Suppose $U \subseteq \mathbb{R}^n$ then we say $\vec{F}$ is **path-independent** on U iff $\int_{C_1} \vec{F} \bullet d\vec{r} = \int_{C_2} \vec{F} \bullet d\vec{r}$ for each pair of curves $C_1, C_2 \subset U$ beginning at P and terminating at Q .

It is useful to examine a web of concepts which all serve to characterize conservative vector fields.

Proposition 7.4.5.

Suppose U is an open connected subset of $\mathbb{R}^n$ then the following are equivalent

1. $\vec{F}$ is conservative; $\vec{F} = \nabla f$ on all of U
2. $\vec{F}$ is path-independent on U
3. $\oint_C \vec{F} \bullet d\vec{r} = 0$ for all closed curves C in U
4. (add precondition $n = 3$ and U be simply connected) $\nabla \times \vec{F} = 0$ on U.

Proof: We postpone the proof of (4.) $\Rightarrow$ (1.). However, we can show that (1.) $\Rightarrow$ (4.). Suppose $\vec{F} = \nabla f$. Note that $\nabla \times \vec{F} = \nabla \times \nabla f = 0$. I included this here since we can quickly test to see if $Curl(\vec{F}) \neq 0$. When the curl is nontrivial then we can be certain the given vector field is not conservative. On the other hand, vanishing curl is only useful if it occurs over a *simply connected* domain[8]

(1.) $\Rightarrow$ (2.). Assume $\vec{F} = \nabla f$. Suppose C_1, C_2 are two curves which both start at P and end at Q in the set U. Apply the FTC for line-integrals in what follows:

$$\int_{C_1} \vec{F} \bullet d\vec{r} = \int_{C_1} \nabla f \bullet d\vec{r} = f(Q) - f(P).$$

Likewise, $\int_{C_2} \vec{F} \bullet d\vec{r} = \int_{C_2} \nabla f \bullet d\vec{r} = f(Q) - f(P)$. Therefore (2.) holds true.

(2.) $\Rightarrow$ (1.). Assume $\vec{F}$ is path-independent. Pick some point $A \in U$ and let C be any curve in U from A to $B = (x, y, z)$. We define $f(x, y, z) = \int_C \vec{F} \bullet d\vec{r}$. This is single-valued since we assume $\vec{F}$ is path-independent. We need to show that $\nabla f = \vec{F}$. Denote $\vec{F} = \langle P, Q, R \rangle$. We begin by isolating the x-component. We need to show $\frac{\partial}{\partial x} \int_C \vec{F} \bullet d\vec{r} = P(x, y, z)$. We can write C as curve C_x from A to $B_x = (x_o, y, z)$ with $x_o < x$ pasted togther with the line-segment L_x from B_x to B. Observe that the curve C_x has no dependence on x (of the B point)

$$\frac{\partial}{\partial x} \int_C \vec{F} \bullet d\vec{r} = \frac{\partial}{\partial x} \left[\int_{C_x} \vec{F} \bullet d\vec{r} + \int_{L_x} \vec{F} \bullet d\vec{r} \right] = \frac{\partial}{\partial x} \left[\int_{L_x} \vec{F} \bullet d\vec{r} \right]$$

[7] it is worth noticing that while physically this is most interesting to three dimensions, the math allows for more

[8] a simply connected domain is a set with no holes, any loop can be smoothly shrunk to a point, it has a boundary which is a simple curve. A simple curve is a curve with no self-intersections but perhaps one in the case it is closed. A circle is simple a figure 8 is not.

The line segment L_x has parametrization $\vec{r}(t) = \langle t, y, z\rangle$ for $x_o \leq t \leq x$. We calculate that

$$\int_{L_x} \vec{F} \bullet d\vec{r} = \int_{x_o}^{x} \vec{F}(t, y, z) \bullet \langle 1, 0, 0\rangle dt = \int_{x_o}^{x} P(t, y, z)dt$$

Therefore,

$$\frac{\partial}{\partial x}\int_C \vec{F} \bullet d\vec{r} = \frac{\partial}{\partial x}\int_{x_o}^{x} P(t, y, z)dt = P(x, y, z).$$

We can give similar arguments to show that

$$\frac{\partial}{\partial y}\int_C \vec{F} \bullet d\vec{r} = Q \qquad \& \qquad \frac{\partial}{\partial z}\int_C \vec{F} \bullet d\vec{r} = R.$$

We find $\vec{F}$ is conservative.

(2.) $\Rightarrow$ (1.). Assume $\vec{F}$ is path-independent. Consider a closed curve C in U. Notice we can pick any pair of points P, Q on C and write C_1 from P to Q and C_2 from Q to P such that $C = C_1 \cup C_2$. Furthermore, note that $-C_2$ also goes from P to Q. Path independence yields

$$\int_{C_1} \vec{F} \bullet d\vec{r} = \int_{-C_2} \vec{F} \bullet d\vec{r} \;\;\Rightarrow\;\; 0 = \int_{C_1} \vec{F} \bullet d\vec{r} - \int_{-C_2} \vec{F} \bullet d\vec{r} = \int_{C_1} \vec{F} \bullet d\vec{r} + \int_{C_2} \vec{F} \bullet d\vec{r} = \oint_C \vec{F} \bullet d\vec{r}.$$

Conequently (3.) is true.

(3.) $\Rightarrow$ (2.). Suppose $\oint_C \vec{F} \bullet d\vec{r} = 0$ for all closed curves C in U. Suppse C_1 and C_2 start at P and end at Q. Observe that $C = C_1 \cup (-C_2)$ is a closed curve hence

$$0 = \oint_C \vec{F} \bullet d\vec{r} \;\;\Rightarrow\;\; 0 = \oint_C \vec{F} \bullet d\vec{r} = \int_{C_1} \vec{F} \bullet d\vec{r} + \int_{-C_2} \vec{F} \bullet d\vec{r} = \int_{C_1} \vec{F} \bullet d\vec{r} - \int_{C_2} \vec{F} \bullet d\vec{r}.$$

Clearly (2.) follows.

To summarize we have shown (1.) $\Leftrightarrow$ (2.) $\Leftrightarrow$ (3.) and (1.) $\Rightarrow$ (4.). We postpone the proof that (3.) $\Rightarrow$ (4.) and (4.) $\Rightarrow$ (1.) . $\square$

The point A where $f(A) = 0$ is known as the zero for the potential. You should notice that the choice for f is not unique. If we add a constant c to the potential function f then we obtain the same gradient field; $\nabla f = \nabla(f + c)$. In physics this is the freedom to set the potential energy to be zero at whichever point is convenient.

Example 7.4.6. *In electrostatics, the potential energy per unit charge is called the voltage or simply the electric potential. For finite, localized charg distributions the electric potential is defined by*

$$V(x, y, z) = -\int_{\infty}^{(x,y,z)} \vec{E} \bullet d\vec{r}$$

The electric field of a charge at the origin is given by $\vec{E} = \frac{k}{\rho^2}\widehat{\rho}$. We take the line from the origin to spatial infinity[9] *to calculate the potential.*

$$V(\rho) = -\int_{\infty}^{\rho} \frac{k}{\rho^2} d\rho = \frac{k}{\rho}.$$

The notation $\int_P^Q$ indicates the line integral is taken over a path from P to Q. This notation is only unambiguous if we are working with a conservative vector field.

[9] the claim implicit within such a convention is that it matters not which unbounded direction the path begins, for convenience we usually just use a line which extends to ∞

7.5 green's theorem

The fundamental theorem of calculus shows that integration and differentiation are inverse processes in a certain sense. It is natural to seek out similar theorems for functions of several variables.

We begin our search by defining the flux through a simple closed planar curve[10]. It is just the scalar integral of the outward-facing normal component to the vector field. Then we examine how a vector field flows out of a little rectangle. This gives us reason to define the divergence. In some sense this little picture will derive the first form of Green's Theorem.

Being discontent with just one interpretation, we turn to analyze how the vector field circulates around a given CCW curve. We again look at a little rectangle and quantify how a given vector field twists around the square loop. This leads us to another derivation of Green's Theorem. Moreover, it gives us the reason to define the curl of a vector field.

Finally, we offer a proof which extends the toy derivations to a general Type I & II curve. Past that, properties of the line-integral extend our result to general regions in the plane. Applications to the calculation of area and the analysis of conservative vector fields are given. I conclude this section with a somewhat formal introduction to two-dimensional electrostatics, I show how Green's Theorem naturally supports Gauss' Law for the plane.

7.5.1 geometry of divergence in two dimensions

A curve is said to be **simple** if it has no self-intersections except perhaps one. For example, a circle is a simple curve whereas a figure 8 is not. Both circles and figure 8's are closed curves since they have no end points (or you could say they have the same starting and ending points). In any event, we define the number of field lines which cut through a simple curve by the geometrically natural definition below:

Definition 7.5.1. *flux of $\vec{F}$ through a simple curve C.*

Suppose $\vec{F}$ is continuous on an open set containing the closed simple curve C. Define:

$$\Phi_C = \oint_C (\vec{F} \bullet \vec{n})ds$$

Where $\vec{n}$ is the outward-facing unit-normal to C.

Recall that if $\vec{r}(s) = \langle x(s), y(s)\rangle$ is the arclength parametrization of C then the unit-tangent vector of the Frenet frame was defined by $\vec{T}(s) = \frac{d\vec{r}}{ds}$.[11] For the sake of visualization suppose C is CCW oriented calculate $\vec{n}$. Since $\vec{T} \bullet \vec{n} = 0$ there are only two choices once we calculate $\vec{n}$[12]. We choose the $\vec{n}$ which points outward.

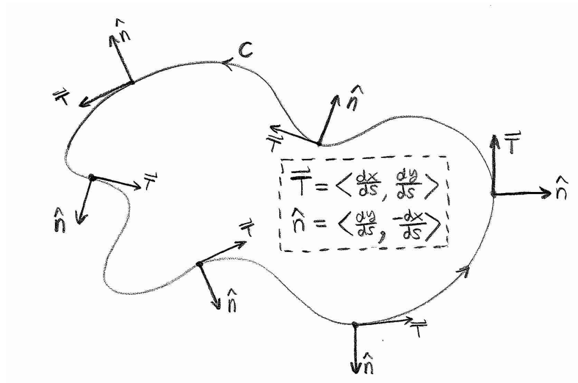
$$\vec{T} = \left\langle \frac{dx}{ds}, \frac{dy}{ds} \right\rangle \qquad \Rightarrow \qquad \vec{n} = \left\langle \frac{dy}{ds}, -\frac{dx}{ds} \right\rangle$$

The picture below helps you see how the outward normal formula works:

[10] these are known as Jordan curves

[11] yes the unit normal is defined by and $\vec{N}(s) = \frac{1}{T'(s)}\frac{d\vec{T}}{ds}$. However, this is not the $\vec{n}$ which we desire because $\vec{N}$ sometimes points inward. Also, direct computation brings us to second derivatives and the geometric argument above avoids that difficulty

[12] if $\hat{u} = \langle a, b\rangle$ then $\hat{v} = \langle b, -a\rangle$ or $\hat{v} = \langle -b, a\rangle$ are the only perpendicular unit-vectors to $\hat{u}$



Let's calculate the flux given this identity. Consider a vector field $\vec{F} = \langle P, Q \rangle$ and once more the Jordan curve C with outward normal $\vec{n}$, suppose length of C is L,

$$\begin{aligned}
\Phi_C &= \oint_C \vec{F} \bullet \vec{n} \\
&= \int_0^L \langle P, Q \rangle \bullet \langle \frac{dy}{ds}, -\frac{dx}{ds} \rangle ds \\
&= \int_0^L \left(P\frac{dy}{ds} - Q\frac{dx}{ds} \right) ds \\
&= \int_0^L \langle -Q, P \rangle \bullet \langle \frac{dx}{ds}, \frac{dy}{ds} \rangle ds \\
&= \oint_C Pdy - Qdx.
\end{aligned}$$

This formula is very nice. It equally well applies to closed simple curves which are only mostly smooth. If we have a few corners on C then we can still calculate the flux by calculating flux through each smooth arc and adding together to find the net-flux. To summarize:

Proposition 7.5.2.

Suppose C is a piecewise-smooth, simple, closed CCW oriented curve. If $\vec{F}$ is continuous on an open set containing C then the flux through C is given by

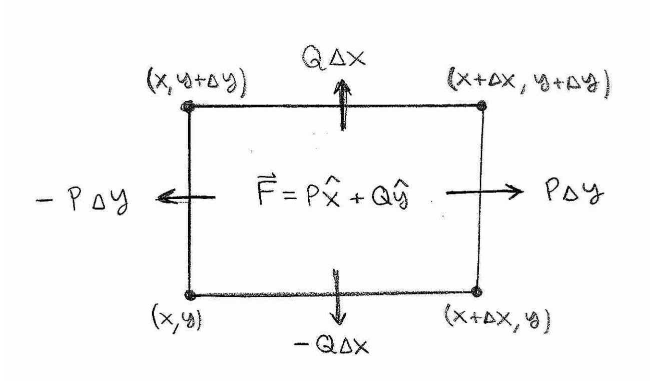
$$\Phi_C = \oint_C \vec{F} \bullet \vec{n} = \oint_C Pdy - Qdx.$$

If you were to consider the CW-oriented curve $-C$ then the outward-normal is given by $\vec{n} = \langle -\frac{dy}{ds}, \frac{dx}{ds} \rangle$ and the formula for flux is

$$\Phi_C = \oint_{-C} \vec{F} \bullet \vec{n} = \oint_{-C} Qdx - Pdy.$$

This formula is in some sense left-handed, hence evil, so we will not use it henceforth.

Now we turn to the task of approximating the flux by direct computation. Consider a little rectangle R with corners at $(x, y), (x + \triangle x, y), (x + \triangle x, y + \triangle y), (x, y + \triangle y)$.



To calculate the flux of $\vec{F} = \langle P, Q \rangle$ through the rectangle we simply find the flux through each side and add it up.

1. **Top:** $(\vec{F} \bullet \widehat{y})\triangle x = Q(x, y + \triangle y)\triangle x$
2. **Base:** $(\vec{F} \bullet [-\widehat{y}])\triangle x = -Q(x, y)\triangle x$
3. **Left:** $(\vec{F} \bullet [-\widehat{x}])\triangle y = -P(x, y)\triangle y$
4. **Right:** $(\vec{F} \bullet \widehat{x})\triangle y = P(x + \triangle x, y)\triangle y$

The net-flux through R is thus,

$$\Phi_R = \Big(Q(x, y + \triangle y) - Q(x, y)\Big)\triangle x + \Big(P(x + \triangle x, y) - P(x, y)\Big)\triangle y$$

Observe that

$$\frac{\Phi_R}{\triangle x \triangle y} = \frac{Q(x, y + \triangle y) - Q(x, y)}{\triangle y} + \frac{P(x + \triangle x, y) - P(x, y)}{\triangle x}$$

In this limit $\triangle x \to 0$ and $\triangle y \to 0$ the expressions above give partial derivatives and we find that:

$$\frac{d\Phi_R}{dA} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}$$

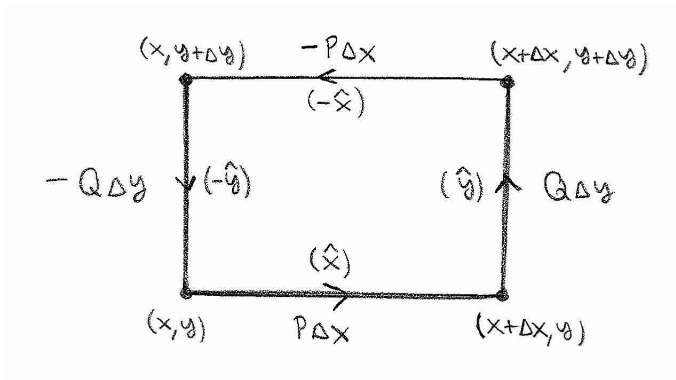
This suggests we define the flux density as $\nabla \bullet \vec{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}$. If we integrate this density over a finite region then we will find the net flux through the region (jump!). In any event, we at least have good reason to suspect that

$$\Phi_{\partial R} = \oint_{\partial R} \vec{F} \bullet \vec{n} = \iint_R \nabla \bullet \vec{F}\, dA \quad \Rightarrow \quad \boxed{\oint_{\partial R} Pdy - Qdx = \iint_R \left[\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right] dA}$$

This is for obvious reasons called the *divergence form of Green's Theorem.* We prove this later in this section. As I mentioned in lecture I found these thoughts in Thomas' Calculus text, however, I suspect we'll find them in many good calculus texts at this time.

7.5.2 geometry of curl in two dimensions

Recall that $\oint_C \vec{F} \bullet d\vec{r}$ calculates the work done by $\vec{F}$ around the loop C. This line-integral is also called the **circulation**. Why? If we think of $\vec{F}$ as the velocity field of some liquid then a positive circulation around a CCW loop suggests that a little paddle wheel placed at the center of the loop will spin in the CCW direction. The greater the circulation the faster it spins. Let us duplicate the little rectangle calculation of the previous section to see what meaning, if any, the circulation per area has: Once more, consider a little rectangle R with corners at $(x, y), (x + \triangle x, y), (x + \triangle x, y + \triangle y), (x, y + \triangle y)$.



To calculate the flow[13] of $\vec{F} = \langle P, Q\rangle$ through the rectangle we simply find the flow through each side and add it up.

1. **Top:** $(\vec{F} \bullet [-\widehat{x}])\triangle x = -P(x, y + \triangle y)\triangle x$
2. **Base:** $(\vec{F} \bullet \widehat{x})\triangle x = P(x, y)\triangle x$
3. **Left:** $(\vec{F} \bullet [-\widehat{y}])\triangle y = -Q(x, y)\triangle y$
4. **Right:** $(\vec{F} \bullet \widehat{y})\triangle y = Q(x + \triangle x, y)\triangle y$

The net-circulation around R is thus,

$$W_R = \Big(Q(x + \triangle x, y) - Q(x, y)\Big)\triangle y - \Big(P(x, y + \triangle y) - P(x, y)\Big)\triangle x$$

Observe that

$$\frac{W_R}{\triangle x\triangle y} = \frac{Q(x + \triangle x, y) - Q(x, y)}{\triangle x} - \frac{P(x, y + \triangle y) - P(x, y)}{\triangle y}$$

In this limit $\triangle x \to 0$ and $\triangle y \to 0$ the expressions above give partial derivatives and we find that:

$$\frac{d\Phi_R}{dA} = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$$

In the case that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 0$ we say the velocity field is *irrotational* since it does not have the tendency to generate rotation at the point in question. The **curl** of a vector field measures

[13]the flow around a closed loop is called circulation, but flow is the term for a curve which is not closed

how a vector field in $\mathbb{R}^3$ rotates about planes with normals $\widehat{x}$, $\widehat{y}$, $\widehat{z}$. In particular, we defined $Curl(\vec{F}) = \nabla \times \vec{F}$. If $\vec{F}(x,y,z) = \langle P(x,y), Q(x,y), 0\rangle$ then

$$Curl(\vec{F}) = \left\langle \; 0,\; 0,\; \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \; \right\rangle$$

and we just derived that nonzero $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$ will spin a little paddle wheel with axis $\widehat{z}$. If $\vec{F} \bullet \widehat{z} \neq 0$ and/or P, Q had nontrivial z-dependence then we would also find nontrivial components of the curl in the $\widehat{x}$ or $\widehat{y}$ directions. If $Curl(\vec{F}) \bullet \widehat{x}$ or $Curl(\vec{F}) \bullet \widehat{y}$ were nonzero at a point then that suggests the vector field will spin a little paddle wheel with axis $\widehat{x}$ or $\widehat{y}$. That is clear from simply generalizing this calculation by replacing x, y with y, z or x, z. Another form of Green's Theorem follows from the curl: since $dW = (\nabla \times \vec{F}) \bullet \widehat{z}\, dA$ we suspect that

$$W_R = \oint_{\partial R} \vec{F} \bullet d\vec{r} = \iint_R (\nabla \times \vec{F}) \bullet \widehat{z}\, dA \quad \Rightarrow \quad \boxed{\oint_{\partial R} P dx + Q dy = \iint_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right] dA}$$

This is the more common form found in calculus texts. Many texts simply state this formula and offer part of the proof given in the next section. Our goal here was to understand why we would expect such a theorem and as an added benefit we have hopefully arrived at a deeper understanding of the differential vector calculus of curl and divergence.

7.5.3 proof of the theorem

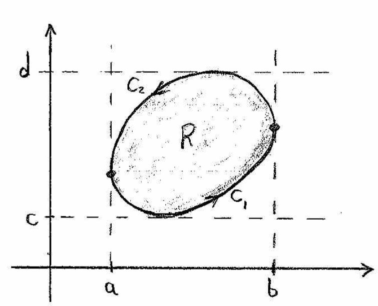
It it is a simple exercise to show that the divergence form of Green's theorem follows from the curl-form we state below.

Theorem 7.5.3. *Green's Theorem for simply connected region:*

Suppose ∂R is a piecewise-smooth, simple, closed CCW oriented curve which bounds the simply connected region $R \subset \mathbb{R}^2$ and suppose $\vec{F}$ is differentiable on an open set containing R then

$$\oint_{\partial R} P dx + Q dy = \iint_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right] dA$$

Proof: we begin by observing that any simply connected region can be subdivided into more basic simply connected regions which are simultaneously Type I and Type II subsets of the plane. Sometimes, it takes a sequence of these basic regions to capture the overall region R and we will return to this point once the theorem is settled for the basic case.



Proof of Green's Theorem for regions which are both type I and II. We assume that there exist constants $a, b, c, d \in \mathbb{R}$ and functions f_1, f_2, g_1, g_2 which are differentiable and describe R as follows:

$$R = \underbrace{\{(x, y) \mid a \le x \le b,\ f_1(x) \le y \le f_2(x)\}}_{\text{type I}} = \underbrace{\{(x, y) \mid c \le y \le d,\ g_1(y) \le x \le g_2(y)\}}_{\text{type II}}$$

Note the boundary $\partial R = C_1 \cup C_2$ can be parametrized in the type I set-up as:

$$C_1 : \vec{r}_1(x) = \langle x, f_1(x) \rangle, \qquad -C_2 : \vec{r}_{-2}(x) = \langle x, f_2(x) \rangle$$

for $a \le x \le b$ (it is easier to think about parametrizing $-C_2$ so I choose to do such). Proof of the theorem can be split into proving two results:

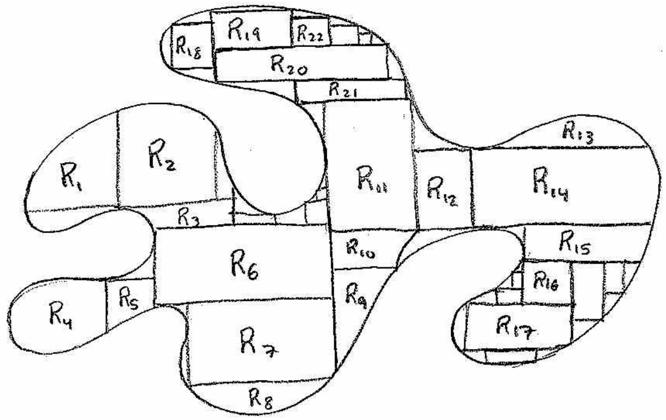
$$(\mathbf{I}.) \quad \oint_{\partial R} P dx = -\iint_R \frac{\partial P}{\partial y} dA \qquad \& \qquad (\mathbf{II}.) \quad \oint_{\partial R} Q dy = \iint_R \frac{\partial Q}{\partial x} dA$$

I prove I. in these notes and I leave II. as a homework for the reader. Consider,

$$\begin{aligned}
\iint_R \frac{\partial P}{\partial y} dA &= \int_a^b \int_{f_1(x)}^{f_2(x)} \frac{\partial P}{\partial y} dy\, dx \\
&= \int_a^b \big[\ P(x, f_2(x)) - P(x, f_1(x))\ \big] dx \\
&= \int_a^b P(x, f_2(x)) dx - \int_a^b P(x, f_1(x)) dx \\
&= \int_{-C_2} P dx - \int_{C_1} P dx \\
&= -\int_{C_2} P dx - \int_{C_1} P dx \\
&= -\int_C P dx
\end{aligned}$$

Hence $\iint_R -\frac{\partial P}{\partial y} dA = \oint_C P dx$. You will show in homework that $\oint_{\partial R} Q dy = \iint_R \frac{\partial Q}{\partial x} dA$ and Green's Theorem for regions which are both type I and II follows. ∇

If the set R is a simply connected subset of the plane then it has no holes and generically the picture is something like what follows. We intend that $R = \sum_k R_k$



Applying Green's theorem to each sub-region gives us the following result.

$$\sum_k \oint_{\partial R_k} Pdx + Qdy = \sum_k \iint_{R_k} \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right] dA$$

It is geometrically natural to suppose the rhs simply gives us the total double integral over R,

$$\sum_k \iint_{R_k} \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right] dA = \iint_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right] dA.$$

I invite the reader to consider the diagram above to see that all the interior *cross-cuts* cancel and only the net-boundary contributes to the line integral over ∂R. Hence,

$$\sum_k \oint_{\partial R_k} Pdx + Qdy = \oint_{\partial R} Pdx + Qdy.$$

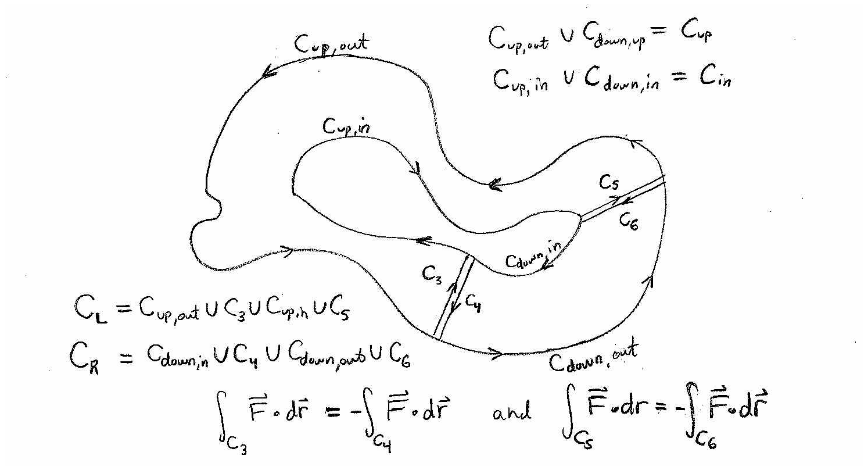
Green's Theorem follows. It should be cautioned that the summations above need not be finite. We neglect some analytical details in this argument. However, I hope the reader sees the big idea here. You can find full details in some advanced calculus texts. □

Theorem 7.5.4. *Green's Theorem for an annulus:*

Suppose ∂R is a pair of simple, closed CCW oriented curve which bounds the connected region $R \subset \mathbb{R}^2$ where $\partial R = C_{in} \cup C_{out}$ and C_{in} is the CW-oriented inner-boundary of R whereas C_{out} is the CCW oriented outer-boundary of R. Furthermore, suppose $\vec{F}$ is differentiable on an open set containing R then

$$\oint_{\partial R} Pdx + Qdy = \iint_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right] dA$$

Proof: See the picture below we can break the annulus into two simply connected regions then apply Green's Theorem for simply connected regions to each piece.



Observe that the cross-cuts cancel (in the diagram above the cancelling pairs are $3,4$ and $5,6$):

$$\begin{aligned}
\int_{C_L} \vec{F} \bullet d\vec{r} + \int_{C_R} \vec{F} \bullet d\vec{r} &= \int_{C_{up,out}} \vec{F} \bullet d\vec{r} + \int_{C_3} \vec{F} \bullet d\vec{r} + \int_{C_{up,in}} \vec{F} \bullet d\vec{r} + \int_{C_5} \vec{F} \bullet d\vec{r} \\
&\quad + \int_{C_{down,in}} \vec{F} \bullet d\vec{r} + \int_{C_4} \vec{F} \bullet d\vec{r} + \int_{C_{down,out}} \vec{F} \bullet d\vec{r} + \int_{C_6} \vec{F} \bullet d\vec{r} \\
&= \int_{C_{up,out}} \vec{F} \bullet d\vec{r} + \int_{C_3} \vec{F} \bullet d\vec{r} + \int_{C_{up,in}} \vec{F} \bullet d\vec{r} + \int_{C_5} \vec{F} \bullet d\vec{r} \\
&\quad + \int_{C_{down,in}} \vec{F} \bullet d\vec{r} - \int_{C_3} \vec{F} \bullet d\vec{r} + \int_{C_{down,out}} \vec{F} \bullet d\vec{r} - \int_{C_5} \vec{F} \bullet d\vec{r} \\
&= \int_{C_{up,out}} \vec{F} \bullet d\vec{r} + \int_{C_{up,in}} \vec{F} \bullet d\vec{r} + \int_{C_{down,in}} \vec{F} \bullet d\vec{r} + \int_{C_{down,out}} \vec{F} \bullet d\vec{r} \\
&= \int_{C_{up}} \vec{F} \bullet d\vec{r} + \int_{C_{down}} \vec{F} \bullet d\vec{r}
\end{aligned}$$

Apply Green's theorem to the regions bounded by C_L and C_R and the theorem follows. $\square$

Notice we can recast this theorem as follows:

$$\oint_{C_{out}} Pdx + Qdy - \oint_{-C_{in}} Pdx + Qdy = \iint_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right] dA.$$

Or, better yet, if C_1, C_2 are two CCW oriented curves which bound R

$$\oint_{C_1} Pdx + Qdy - \oint_{C_2} Pdx + Qdy = \iint_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right] dA.$$

Suppose the vector field $\vec{F} = \langle P, Q \rangle$ passes our *Clairaut Test* on R then we have $\frac{\partial Q}{\partial x} = \frac{\partial P}{\partial y}$ and consequently:

$$\oint_{C_1} Pdx + Qdy = \oint_{C_2} Pdx + Qdy.$$

I often refer to this result as the **deformation theorem** for irrotational vector fields in the plane.

Theorem 7.5.5. *Deformation Theorem for irrotational vector field on the plane:*

Suppose C_1, C_2 are CCW oriented closed simple curves which bound $R \subset \mathbb{R}^2$ and suppose $\vec{F} = \langle P, Q \rangle$ is differentiable on a open set containing R then

$$\oint_{C_1} Pdx + Qdy = \oint_{C_2} Pdx + Qdy.$$

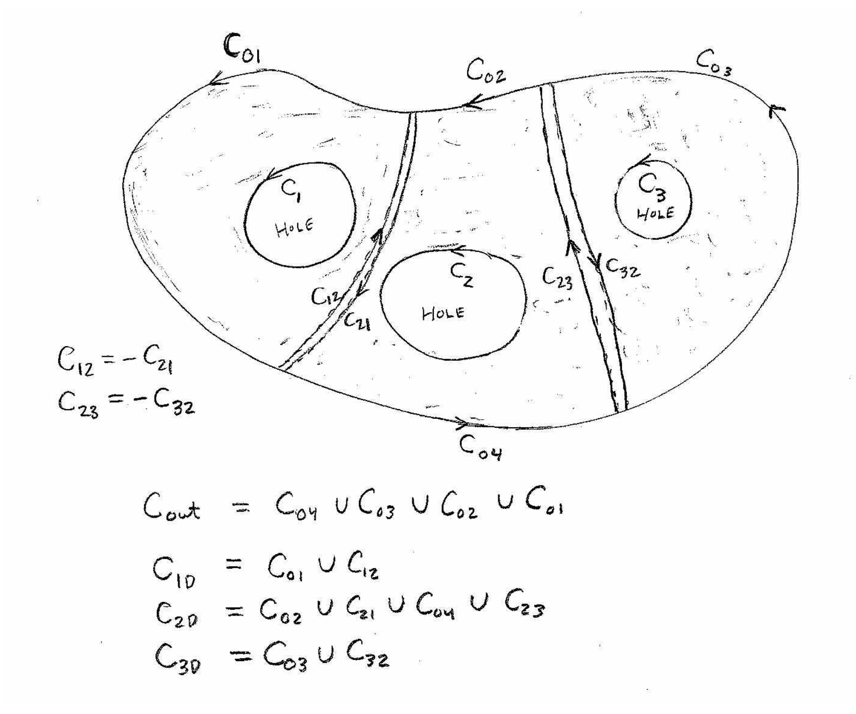
In my view, points where $\nabla \times \vec{F} \neq 0$ are troublesome. This theorem says the line integral is unchanged if we do not enclose any new troubling points as we deform C_1 to C_2. On the flip-side of this, if the integral around some loop is nonzero for a given vector field that must mean that something interesting happens to the curl of the vector field on the interior of the loop.

Theorem 7.5.6. *Green's Theorem for a region with lots of holes.*

Suppose R is a connected subset of $\mathbb{R}^2$ which has boundary ∂R. We orient this boundary curve such that the outer boundary has CCW orientation whereas all the inner-boundaries have CW orientation. Furthermore, suppose $\vec{F}$ is differentiable on an open set containing R then

$$\oint_{\partial R} Pdx + Qdy = \iint_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right] dA$$

Proof: follows from the picture below and a little thinking.



This is more interesting if we state it in terms of the outer loop and CCW oriented inner loops. Denote C_{out} for the outside loop of ∂R and C_k for $k = 1, 2, \dots, N$ for the inner CCW oriented loops. Since $\partial R = C_{out} \cup -C_1 \cup \cdots \cup -C_N$ it follows

$$\oint_{C_{out}} Pdx + Qdy - \sum_{k=1}^{N} \oint_{C_k} Pdx + Qdy = \iint_R \left[\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right] dA$$

If we have $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 0$ throughout R we find the following beautiful generalization of the deformation theorem:

$$\oint_{C_{out}} Pdx + Qdy = \sum_{k=1}^{N} \oint_{C_k} Pdx + Qdy.$$

In words, the net circulation around C is simply a sum of the circulations around each singularity contained within C. A **singularity** is a point at which the field $\langle P, Q \rangle$ obtains a nontrivial circulation around any small loop containing the point.

7.5.4 examples

Example 7.5.7. *Use Green's theorem to calculate $\oint_C x^3 dx + yxdy$ where C is the CCW boundary of the oriented rectangle R: $[0,1] \times [0,1]$. Identify that $P = x^3$ and $Q = xy$. Applying Green's Theorem,*

$$\oint_C x^3 dx + yxdy = \iint \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA = \int_0^1 \int_0^1 y \, dx \, dy = \frac{1}{2}.$$

One important application of Green's theorem involves the calculation of areas. Note that if we choose $\vec{F} = \langle P, Q \rangle$ such that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1$ then the double integral in Green's theorem represents the area of R. In particular, it is common to use

$$\vec{F} = \langle 0, x \rangle, \qquad \text{or} \qquad \vec{F} = \langle -y, 0 \rangle, \qquad \text{or} \qquad \vec{F} = \langle -y/2, x/2 \rangle$$

in Green's theorem to obtain the identities:

$$\boxed{A_R = \oint_{\partial R} xdy = -\oint_{\partial R} ydx = \frac{1}{2} \oint_{\partial R} xdy - ydx}$$

Example 7.5.8. *Find the area of the ellipse bounded by $x^2/a^2 + y^2/b^2 = 1$. Observe that the ellipse ∂R is parametrized by $x = a\cos(t)$ and $y = b\sin(t)$ hence $dx = -a\sin(t)dt$ and $dy = b\cos(t)dt$ hence*

$$A_R = \frac{1}{2} \oint_{\partial R} xdy - ydx = \frac{1}{2} \int_0^{2\pi} 1a\cos(t)b\cos(t)dt - b\sin(t)(-a\sin(t)dt) = \frac{1}{2} \int_0^{2\pi} ab \, dt = \pi ab.$$

When $a = b = R$ we obtain the famous πR^2.

Example 7.5.9. *You can show (perhaps you will in a homework) that for the line-segment L from (x_1, y_1) to (x_2, y_2) we have the following excellent identity:*

$$\frac{1}{2} \int_L xdy - ydx = \frac{1}{2}(x_1 y_2 - x_2 y_1).$$

Consider that if P is a polygon with vertices $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ with sides $L_{1,2}, L_{2,3}, \dots, L_{N-1,N}, L_{N,1}$ then the area of P is given by the line-integral of $\vec{F} = \langle -y/2, x/2 \rangle$ thanks to Green's Theorem:

$$\begin{aligned} A_P = \iint_P dA &= \frac{1}{2} \int_{\partial P} xdy - ydx \\ &= \frac{1}{2} \int_{L_{1,2}} xdy - ydx + \cdots + \frac{1}{2} \int_{L_{N,1}} xdy - ydx \\ &= \frac{1}{2} \big[x_1 y_2 - x_2 y_1 + x_2 y_3 - x_3 y_2 + \cdots + x_{N-1} y_N - x_N y_{N-1} + x_N y_1 - x_1 y_N \big] \end{aligned}$$

You can calculate the area of polygon with vertices $(0,0),(-1,1),(0,2),(1,3),(2,1)$ is $9/2$ by applying the formula above. You could just as well calculate the area of a polygon with 100 vertices. The example below is a twist on the ellipse example already given. This time we study an annulus with elliptical edges.

Example 7.5.10. *Find the area bounded by ellipses $x^2/a^2+y^2/b^2=1$ and $x^2/c^2+y^2/d^2=1$ given that $0<c<a$ and $0<d<b$ to insure that the ellipse $x^2/a^2+y^2/b^2=1$ is exterior to the ellipse $x^2/c^2+y^2/d^2=1$. Observe that the ellpitical annulus has boundary $\partial R = C_{in} \cup C_{out}$ where C_{out} is CCW parametrized by $x=a\cos(t)$ and $y=b\sin(t)$ and C_{in} is CW oriented with parametrization $x=c\cos(t)$ and $y=-d\sin(t)$ it follows that:*

$$\begin{aligned} A_R &= \frac{1}{2}\oint_{\partial R} xdy - ydx \\ &= \frac{1}{2}\oint_{C_{out}} xdy - ydx + \frac{1}{2}\oint_{C_{in}} xdy - ydx \\ &= \frac{1}{2}\int_0^{2\pi} ab\,dt - \frac{1}{2}\int_0^{2\pi} cd\,dt \\ &= \pi ab - \pi cd. \end{aligned}$$

Notice the CW orientation is what caused us to subtract the inner area which is missing from the annulus.

Example 7.5.11. *Consider the CCW oriented curve C with parametrization $x=(10+\sin(30t))\cos(t)$ and $y=(10+\sin(30t))\sin(t)$ for $0\le t\le 2\pi$. This is a wiggly circle with mean radius 10. Calculate*

$$\int_C \frac{xdy - ydx}{x^2+y^2}.$$

Let $P=-y/(x^2+y^2)$ and $Q=x/(x^2+y^2)$ you can show that $\partial_x Q - \partial_y P = 0$ for $(x,y)\neq(0,0)$. It follows that we can deform the given problem to the simpler task of calculating the line-integral around the unit circle S_1: $x=\cos(t)$ and $y=\sin(t)$ hence $xdy-ydx=\cos^2(t)dt+\sin^2(t)dt$ and $x^2+y^2=1$ on S_1, calculate,

$$\begin{aligned} \int_C \frac{xdy - ydx}{x^2+y^2} &= \int_{S_1} \frac{xdy - ydx}{x^2+y^2} \\ &= \int_0^{2\pi} \frac{dt}{1} \\ &= 2\pi. \end{aligned}$$

Notice that we could still make this calculation if the specific parametrization of C was not given. Also, generally, when faced with this sort of problem we should try to pick a deformation which makes the integration easier. It was wise to deform to a circle here since the denominator was greatly simplified.

7.5.5 conservative vector fields and green's theorem

Recall that Proposition 7.4.5 gave us a list of ways of thinking about a conservative vector field:

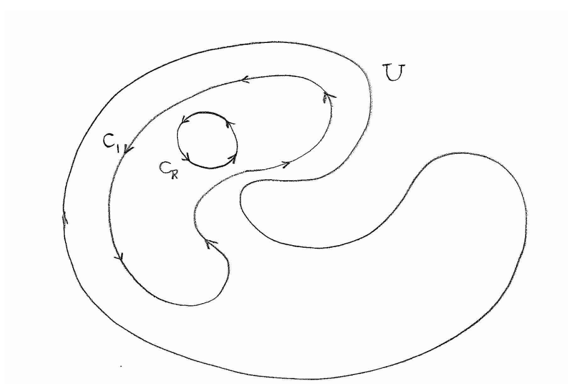
Suppose U is an open connected subset of $\mathbb{R}^n$ then the following are equivalent

1. $\vec{F}$ is conservative; $\vec{F} = \nabla f$ on all of U
2. $\vec{F}$ is path-independent on U
3. $\oint_C \vec{F} \bullet d\vec{r} = 0$ for all closed curves C in U
4. (add precondition U be simply connected) $\nabla \times \vec{F} = 0$ on U.

Proof: We finish the proof by addressing why (4.) $\Rightarrow$ (1.) in light of Green's Theorem. Suppose U is simply connected and $\nabla \times \vec{F} = 0$ on U. Let C_1 be a closed loop in U and let C_R be another loop of radius R inside C_1. Since $\nabla \times \vec{F} = 0$ it follows we can apply the deformation Theorem 7.5.5 on the annulus between C_1 and C_R to obtain $\int_{C_1} \vec{F} \bullet d\vec{r} = \int_{C_R} \vec{F} \bullet d\vec{r}$. Now, as U is simply connected we can smoothly deform C_R to a point C_0. You can show that

$$\lim_{R \to 0} \int_{C_R} \vec{F} \bullet d\vec{r} = 0$$

since C_R becomes a point in this limit. Not convinced? Consider that the integral $\int_{C_R} \vec{F} \bullet d\vec{r}$ is at most the product of $max\{||\vec{F}(\vec{r})|| \mid \vec{r} \in C_R\}$ and the total arclength of C_R. However, the magnitude is bounded as $\vec{F}$ has continuous component functions and the arclength of C_R clearly goes to zero as $R \to 0$. Perhaps the picture below helps communicate the idea of the proof:



We find (4.) $\Rightarrow$ (3.) hence, by our earlier work, (3.) $\Rightarrow$ (2.) $\Rightarrow$ (1.). $\square$

Now, in terms of logical minimalism, to prove that $1, 2, 3, 4$ are equivalent we could just prove the string of implications (1.) $\Rightarrow$ (2.) $\Rightarrow$ (3.) $\Rightarrow$ (4.) $\Rightarrow$ (1.) then any of the reverse implications are easily found by logic. For example, (3.) $\Rightarrow$ (2.) would follow from (3.) $\Rightarrow$ (4.) $\Rightarrow$ (1.) $\Rightarrow$ (2.). That said, I tried to give all directions in the proof to better illustrate how the different views of the conservative vector field are connected.

7.5.6 two-dimensional electrostatics

The fundamental equation of electrostatics is known as *Gauss' Law*. In three dimensions it simply states that the flux through a closed surface is proportional to the charge which is enclosed. We have yet to define flux through a surface, but we do have a careful definition of flux through a simple closed curve. If there was a Gauss' Law in two dimensions then it ought to state that

$$\Phi_E = Q_{enc}$$

In particular, if we denote $\sigma = dQ/dA$ and have in mind the region R with boundary ∂R,

$$\oint_{\partial R} (\vec{E} \bullet \widehat{n}) ds = \iint_R \sigma dA$$

Suppose we have an isolated charge Q at the origin and we apply Gauss law to a circle of radius r centered at the origin then we can argue by symmetry the electric field must be entirely radial in direction and have a magnitude which depends only on r. It follows that:

$$\oint_{\partial R} (\vec{E} \bullet \widehat{n}) ds = \iint_R \sigma dA \quad \Rightarrow \quad (2\pi r)E = Q$$

Hence, the **coulomb field** in two dimensions is as follows:

$$\boxed{\vec{E}(r,\theta) = \frac{Q}{2\pi r}\widehat{r}}$$

Let us calculate the flux of the Coulomb field through a circle C of radius R:

$$\begin{aligned}
\oint_C (\vec{E} \bullet \widehat{n}) ds &= \int_C \left(\frac{Q}{2\pi r}\widehat{r} \bullet \widehat{r} \right) ds \\
&= \int_C \frac{Q}{2\pi R} ds \\
&= \frac{Q}{2\pi R} \int_C ds \\
&= \frac{Q}{2\pi R}(2\pi R) \\
&= Q.
\end{aligned}$$

The circle is complete. In other words, the Coulomb field derived from Gauss' Law does in fact satisfy Gauss Law in the plane. This is good news. Let's examine the divergence of this field. It appears to point away from the origin and as you get very close to the origin the magnitude of E is unbounded. It will be convenient to reframe this formula for the Coulomb field by:

$$\vec{E}(x,y) = \frac{Q}{2\pi(x^2+y^2)}\langle x, y\rangle.$$

Note:

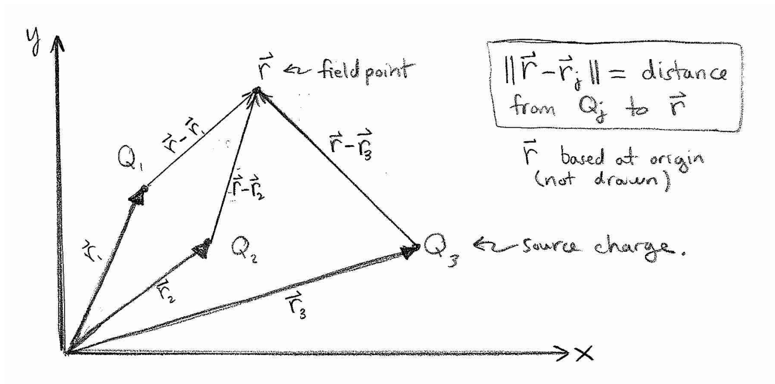
$$\begin{aligned}
\nabla \bullet \vec{E} &= \frac{\partial}{\partial x}\left[\frac{xQ}{2\pi(x^2+y^2)}\right] + \frac{\partial}{\partial y}\left[\frac{yQ}{2\pi(x^2+y^2)}\right] \\
&= \frac{Q}{2\pi}\left[\frac{x^2+y^2-2x^2}{(x^2+y^2)^2} + \frac{x^2+y^2-2y^2}{(x^2+y^2)^2}\right] = 0.
\end{aligned}$$

If we were to carelessly apply the divergence form of Green's theorem this could be quite unsettling: consider,

$$\oint_{\partial R} \vec{E} \bullet \vec{n} = \iint_R \nabla \bullet \vec{E}\, dA \quad \Rightarrow \quad Q = \iint_R (0) dA = 0.$$

But, Q need not be zero hence there is some contradiction? Why is there no contradiction? Can you resolve this paradox?

Moving on, suppose we have N charges placed at source points $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N$ then we can find the total electric field by the principle of superposition.



We simply take the vector sum of all the coulomb fields. In particular,

$$\vec{E}(\vec{r}) = \sum_{j=1}^{N} \vec{E}_j = \sum_{j=1}^{n} \frac{Q_j}{2\pi} \frac{\vec{r}-\vec{r}_j}{||\vec{r}-\vec{r}_j||^2}$$

What is the flux through a circle which encloses just the k-th one of these charges? Suppose C_R is a circle of radius R centered at $\vec{r}_k$. We can calculate that

$$\oint_{C_R} (\vec{E}_k \bullet \widehat{n}) ds = Q_k$$

whereas, since $\vec{E}_j$ is differentiable inside all of C_R for $j \neq k$ and $\nabla \bullet \vec{E}_j = 0$ we can apply the divergence form of Green's theorem to deduce that

$$\oint_{C_R} (\vec{E}_j \bullet \widehat{n}) ds = 0.$$

Therefore, summing these results together we derive for $\vec{E} = \vec{E}_1 + \cdots + \vec{E}_k + \cdots + \vec{E}_N$ that

$$\oint_{C_R} (\vec{E} \bullet \widehat{n}) ds = Q_k$$

Notice there was nothing particularly special about Q_k so we have derived this result for each charge in the distribution. If we take a circle around a charge which contains just one charge then Gauss' Law applies and the flux is simply the charge enclosed. Denote $C_1, C_2, \dots, C_N$ as little circles which each enclose a single charge. In particular, $C_1, C_2, \dots, C_N$ enclose the charges $Q_1, Q_2, \dots, Q_N$ respective. We have

$$Q_1 = \oint_{C_1} (\vec{E} \bullet \widehat{n}) ds, \quad Q_2 = \oint_{C_2} (\vec{E} \bullet \widehat{n}) ds, \quad \dots, \quad Q_N = \oint_{C_N} (\vec{E} \bullet \widehat{n}) ds$$

Now suppose we have a curve C which encloses all N of the charges. The electric field is differentiable and has vanishing divergence at all points except the location of the charges. In fact, the coulomb field passes Clairaut's test everywhere. It just has the isolated singularity where the charge is found. We can apply the general form of the deformation theorem to arrive at Gauss' Law for the distribution of N-charges:

$$\oint_C (\vec{E} \bullet \widehat{n}) ds = \oint_{C_1} (\vec{E} \bullet \widehat{n}) ds + \oint_{C_2} (\vec{E} \bullet \widehat{n}) ds + \cdots + \oint_{C_N} (\vec{E} \bullet \widehat{n}) ds = Q_1 + Q_2 + \cdots + Q_N$$

You can calculate the divergence is zero everywhere except at the location of the source charges. Moral of story: even one point thrown out of a domain can have dramatic and global consequences for the behaviour of a vector field. In physics literature you might find the formula to describe what we found by a *dirac-delta function* these distributions capture certain infinities and let you work with them. For example: for the basic coulomb field with a single point charge at the origin $\vec{E}(r,\theta) = \frac{Q}{2\pi r}\widehat{r}$ this derived from a charge density function σ which is zero everywhere except at the origin. Somehow $\iint_R \sigma dA = Q$ for any region R which contains $(0,0)$. Define $\sigma(\vec{r}) = Q\delta(\vec{r})$. Where we define: for any function f which is continuous near 0 and any region which contains the origin

$$\int_R f(\vec{r})\delta(\vec{r})dA = f(0)$$

and if R does not contain $(0,0)$ then $\iint_R f(\vec{r})\delta(\vec{r})dA = 0$. The dirac delta function turns integration into evaluation. The dirac delta function is not technically a function, in some sense it is zero at all points and infinite at the origin. However, we insist it is manageably infinity in the way just described. Notice that it does at least capture the right idea for density of a point charge: suppose R contains $(0,0)$,

$$\iint_R \sigma dA = \iint_R Q\delta(\vec{r})dA = Q.$$

On the other hand, we can better understand the divegence calculation by the following calculations[14]:

$$\nabla \bullet \frac{\vec{r}}{r^2} = 2\pi\delta(\vec{r}).$$

Consequentlly, if $\vec{E} = \frac{Q}{2\pi}\frac{\vec{r}}{r^2}$ then $\nabla \bullet \vec{E} = \nabla \bullet \left[\frac{Q}{2\pi}\frac{\vec{r}}{r^2}\right] = \frac{Q}{2\pi}\nabla \bullet \frac{\vec{r}}{r^2} = Q\delta(\vec{r})$. Now once more apply Green's theorem to the Coulomb field. Use the divergence form of the theorem and this time appreciate that the divergence of $\vec{E}$ is not strictly zero, rather, the dirac-delta function captures the divergence: recall the RHS of this calculation followed from direct calculation of the flux of the Coloumb field through the circle ∂R,

$$\oint_{\partial R} \vec{E} \bullet \vec{n}\, ds = \iint_R \nabla \bullet \vec{E}\, dA \quad \Rightarrow \quad Q = \iint_R Q\delta(\vec{r})dA = Q.$$

All is well. This is the way to extend Green's theorem for Coulomb fields. You might wonder about other types of singularities. Are there similar techniques? Probably, but that is beyond these notes. I merely wish to sketch the way we think about these issues in electrostatics. In truth, this section is a bit of a novelty. What really matters is three-dimensional Coulomb fields whose magnitude depends on the squared-reciprocal of the distance from source charge to field point. Perhaps I will write an analogus section once we have developed the concepts of flux and the three-dimensional divergence theorem.

[14]I don't intend to explain where this 2π comes from, except to tell you that it must be there in order for the extension of Green's theorem to work out nicely.

7.6 surface integrals

We discuss how to integrate over surfaces in $\mathbb{R}^3$. This section is the natural extension of the work we have already accomplished with curves. We begin by describing the integral with respect to the surface area, naturally this integral calculates surface area, but more generally it allows us to continuously sum some density over a given surface. Next we discuss how to find the flux through a surface. The concept of flux requires we give the surface a direction. The flux of a vector field is the *number*[15] of field lines which cut through the surface. The parametric viewpoint is primary in this section, but we also make an effort to show how to calculate surface integrals from the graphical or level-surface viewpoint.

7.6.1 motivations for surface area integral and flux

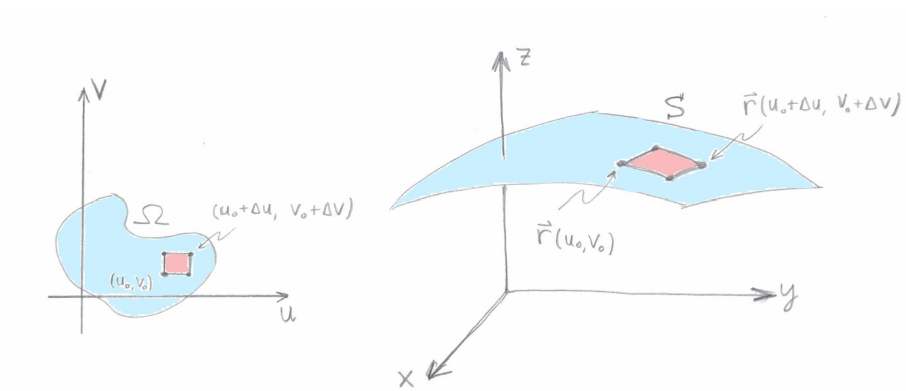
Let us consider a surface S parametrized by smooth $\vec{r}(u,v) = \langle x(u,x), y(u,v), z(u,v)\rangle$ for $(u,v) \in \Omega$. If we wish to approximate the surface area of S then we should partition the parameter space Ω into subregions Ω_{ij} such that $\bigcup_{i=1}^m \bigcup_{j=1}^m \Omega_{ij} = \Omega$ where we assume that Ω_{ij} are mostly disjoint, they might share an edge or point, but not an area. Naturally this partitions S into subsurfaces; $S = \bigcup_{i=1}^m \bigcup_{j=1}^m S_{ij}$ where $S_{ij} = \vec{r}(\Omega_{ij})$. Next, we replace each subsurface S_{ij} with its tangent plane based at some point[16] $\vec{r}(u_i^*, v_j^*) \in S_{ij}$. For convenience of this motivation we may assume that the partition is made so that Ω_{ij} is a little rectangle which is $\triangle u$ by $\triangle v$. The length of the coordinate curves are well-approximated by $\frac{\partial \vec{r}}{\partial u}\triangle u$ and $\frac{\partial \vec{r}}{\partial v}\triangle v$ based at the point $\vec{r}(u_i^*, v_j^*)$. The coordinate lines on S are not necessarily perpendicular so the area is not simply the product of length times width, in fact, we have a little paralellogram to consider. It follows that the area A_{ij} of the i,j-tangent plane is given by

$$A_{ij} = \left|\left|\left(\frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v}\right)(u_i^*, v_j^*)\right|\right| \triangle u \triangle v.$$

A good approximation to surface area is given for $m, n >> 1$ by

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \left|\left|\left(\frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v}\right)(u_i^*, v_j^*)\right|\right| \triangle u \triangle v$$

As we pass to the limit $m \to \infty$ and $n \to \infty$ the double finite sum becomes the double integral taken over the parameter space Ω.



For these reasons we define:

[15]relative to some convention, lines drawn per unit of flux

[16]when I illustrate this idea I usually take these points to the lower left of the partition region, but in principle you could sample elsewhere.

Definition 7.6.1. *scalar surface integral.*

Suppose S is a surface in $\mathbb{R}^3$ paramterized by $\vec{r}$ with domain Ω and with parameters u, v. Furthermore, suppose f is a continuous function on some open set containing S then we define the scalar surface integral of f over S by the following integral (when it exists)

$$\iint_S f\, dS = \iint_\Omega f(\vec{r}(u,v)) \left|\left| \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v} \right|\right| du\, dv.$$

In the case we integrate $f = 1$ over S then we obtain the **surface area** of S. On the other hand, if f was the mass-density $f = \frac{dM}{dS}$ then $dM = f dS$ and the integral $\iint_S f\, dS$ calculates the total mass of S. Clearly it is convenient to think of dS as something on its own, however, it should be remembered that this is just notation to package the careful definition given above,

$$\boxed{dS = \left|\left| \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v} \right|\right| du\, dv}$$ **infinitesimal scalar surface area element**

Moreover, we should recall that the normal vector field to S induced by $\vec{r}$ was given by $\vec{N} = \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v}$ hence we can write $dS = N du dv$. This is nice, but for most examples it does not save us from explicit calculation of the cross-product.

Next, consider a vector field $\vec{F}$ defined on some open set containing S. Suppose that S is a regular surface and as such has a well-defined normal vector field $\vec{N}$. If we define $\widehat{n} = \frac{1}{N}\vec{N}$ then $\vec{F} \bullet \widehat{n}$ gives the component of $\vec{F}$ which points in the normal direction of the surface. It is customary to draw field-lines to illustrate both the direction and magnitude of a vector field. The number of lines crossing a particular surface illustrates the magnitude of the vector field relative to the given area. For example, if we had an area A_1 which had 4 field lines of $\vec{F}$ and another area A_2 which had 8 field lines of $\vec{F}$ then the magnitude of $\vec{F}$ on these areas is proportional to $\frac{4}{A_1}$ and $\frac{8}{A_2}$ respectively. If the vector fields are constant, then the flux through A_1 is $F_1 A_1$ whereas the flux through A_2 is $F_2 A_2$. Generally, the flux of a vector field through a surface depends both on the size of the surface and the magnitude of the vector field in the normal direction of the surface. This is the natural generalization of the flux-integral we discussed previously for curves in the plane.

Definition 7.6.2. *surface integral of vector field; the flux through a surface.*

Suppose S is an oriented surface in $\mathbb{R}^3$ paramterized by $\vec{r}$ with domain Ω and with parameters u, v which induce unit-normal vector field $\widehat{n}$. Furthermore, suppose $\vec{F}$ is a continuous vector field on some open set containing S then we define the **surface integral of** $\vec{F}$ over S by the following integral (when it exists)

$$\iint_S \vec{F} \bullet d\vec{S} = \iint_S \left(\ \vec{F} \bullet \widehat{n}\ \right) dS$$

In practice, the formula we utilize for direct computation is not the one given above. Let us calculate,

$$\iint_S \left(\ \vec{F} \bullet \widehat{n}\ \right) dS = \iint_\Omega \left(\ \vec{F} \bullet \frac{1}{N}\vec{N}\ \right) N\, du\, dv = \iint_\Omega \left(\ \vec{F} \bullet \vec{N}\ \right) du\, dv.$$

Hence, recalling once more that $\vec{N}(u,v) = \partial_u \vec{r} \times \partial_v \vec{r}$ we find

$$\boxed{\iint_S \vec{F} \bullet d\vec{S} = \iint_\Omega \vec{F}(\vec{r}(u,v)) \bullet \left(\frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v} \right) du\, dv.}$$

The equation boxed above is how we typically calculate flux of $\vec{F}$ through S. I should mention that is often convenient to calculate $d\vec{S}$ separately before computation of the integral, this quantity is called the **infinitesimal vector surface area element** and is defined by

$$\boxed{d\vec{S} = \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v} du\, dv = \widehat{n} dS} \qquad \textbf{vector area element } d\vec{S}$$

Once more, dS is the scalar area element. Both of these are only meaningful when viewed in connection with the parametric set-up described in this section.

The uninterested reader may skip to the examples, however, there is some unfinished business theoretically here. We should demonstrate that the definitions given in this section are independent of the parametrization. If this fails to be true then the concepts of surface area, total mass etc... and flux are in doubt. We must consider a **reparametrization** of S by $\vec{X} : D \to S$ where a, b are the typical parameters in D and the normal vector field induced by $\vec{X}$ is $\vec{N}_X(a,b) = \partial_a \vec{X} \times \partial_b \vec{X}$. Let us, in contrast denote $\vec{N}_r(u,v) = \partial_u \vec{r} \times \partial_v \vec{r}$. Since each point on S is covered smoothly by both $\vec{r} = \langle x, y, z \rangle$ and $\vec{X} = \langle X_1, X_2, X_3 \rangle$ there exist functions which transition between the two parametrizations. In particular, we can find $\vec{T} : \Omega \to D$ such that $\vec{T} = \langle h, g \rangle$ and

$$\vec{r}(u,v) = \vec{X}(h(u,v), g(u,v))$$

We need to sort through the partial derivatives so we can understand how the normal vector fields $\vec{N}_X$ and $\vec{N}_r$ are related, let $a = h(u)$ and $b = g(v)$ hence $\vec{r}(u,v) = \vec{X}(a,b)$. I'll expand $\vec{X}$ into its component function notation to make sure we understand what we're doing here:

$$\frac{\partial \vec{r}}{\partial u} = \frac{\partial}{\partial u} \Big\langle X_1(a,b), X_2(a,b), X_3(a,b) \Big\rangle = \Big\langle \frac{\partial}{\partial u}[X_1(a,b)], \frac{\partial}{\partial u}[X_2(a,b)], \frac{\partial}{\partial u}[X_3(a,b)] \Big\rangle$$

We calculate by the chain-rule, (omitting the (a,b) dependence on the lhs, technically we should write $\frac{\partial}{\partial u}[X_1(a,b)]$ etc...)

$$\frac{\partial X_1}{\partial u} = \frac{\partial X_1}{\partial a}\frac{\partial h}{\partial u} + \frac{\partial X_1}{\partial b}\frac{\partial g}{\partial u}, \quad \frac{\partial X_2}{\partial u} = \frac{\partial X_2}{\partial a}\frac{\partial h}{\partial u} + \frac{\partial X_2}{\partial b}\frac{\partial g}{\partial u}, \quad \frac{\partial X_3}{\partial u} = \frac{\partial X_3}{\partial a}\frac{\partial h}{\partial u} + \frac{\partial X_3}{\partial b}\frac{\partial g}{\partial u}$$

Likewise, for the derivative with respect to v we calculate,

$$\frac{\partial X_1}{\partial v} = \frac{\partial X_1}{\partial a}\frac{\partial h}{\partial v} + \frac{\partial X_1}{\partial b}\frac{\partial g}{\partial v}, \quad \frac{\partial X_2}{\partial v} = \frac{\partial X_2}{\partial a}\frac{\partial h}{\partial v} + \frac{\partial X_2}{\partial b}\frac{\partial g}{\partial v}, \quad \frac{\partial X_3}{\partial v} = \frac{\partial X_3}{\partial a}\frac{\partial h}{\partial v} + \frac{\partial X_3}{\partial b}\frac{\partial g}{\partial v}$$

We find,

$$\begin{aligned}
\frac{\partial \vec{r}}{\partial u} &= \Big\langle \frac{\partial X_1}{\partial a}\frac{\partial h}{\partial u} + \frac{\partial X_1}{\partial b}\frac{\partial g}{\partial u}, \ \frac{\partial X_2}{\partial a}\frac{\partial h}{\partial u} + \frac{\partial X_2}{\partial b}\frac{\partial g}{\partial u}, \ \frac{\partial X_3}{\partial a}\frac{\partial h}{\partial u} + \frac{\partial X_3}{\partial b}\frac{\partial g}{\partial u} \Big\rangle \\
&= \Big\langle \frac{\partial X_1}{\partial a}, \ \frac{\partial X_2}{\partial a}, \ \frac{\partial X_3}{\partial a} \Big\rangle \frac{\partial h}{\partial u} + \Big\langle \frac{\partial X_1}{\partial b}, \ \frac{\partial X_2}{\partial b}, \ \frac{\partial X_3}{\partial b} \Big\rangle \frac{\partial g}{\partial u} \\
&= \frac{\partial \vec{X}}{\partial a}\frac{\partial h}{\partial u} + \frac{\partial \vec{X}}{\partial b}\frac{\partial g}{\partial u}
\end{aligned}$$

Similarly,

$$\frac{\partial \vec{r}}{\partial v} = \frac{\partial \vec{X}}{\partial a}\frac{\partial h}{\partial v} + \frac{\partial \vec{X}}{\partial b}\frac{\partial g}{\partial v}$$

Calculate, by the antisymmetry of the cross-product,

$$\begin{aligned}
\frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v} &= \left(\frac{\partial \vec{X}}{\partial a}\frac{\partial h}{\partial u} + \frac{\partial \vec{X}}{\partial b}\frac{\partial g}{\partial u}\right) \times \left(\frac{\partial \vec{X}}{\partial a}\frac{\partial h}{\partial v} + \frac{\partial \vec{X}}{\partial b}\frac{\partial g}{\partial v}\right) \\
&= \left[\frac{\partial \vec{X}}{\partial a} \times \frac{\partial \vec{X}}{\partial b}\right]\frac{\partial h}{\partial u}\frac{\partial g}{\partial v} + \left[\frac{\partial \vec{X}}{\partial b} \times \frac{\partial \vec{X}}{\partial a}\right]\frac{\partial g}{\partial u}\frac{\partial h}{\partial v} \\
&= \left[\frac{\partial \vec{X}}{\partial a} \times \frac{\partial \vec{X}}{\partial b}\right]\left[\frac{\partial h}{\partial u}\frac{\partial g}{\partial v} - \frac{\partial g}{\partial u}\frac{\partial h}{\partial v}\right]
\end{aligned}$$

It follows that $\vec{N}_r(u,v) = \vec{N}_X(h(u,v), g(u,v))\frac{\partial(h,g)}{\partial(u,v)}$ hence $N_r(u,v) = N_X(h(u,v), g(u,v))\left|\frac{\partial(h,g)}{\partial(u,v)}\right|$. The vertical bars denote absolute value; when we pull a scalar out of a magnitude of a vector it gets absolute value bars; $||c\vec{v}|| = |c|\,||\vec{v}||$. Consider the surface area integral of f over S which is parametrized by both $\vec{r}(u,v)$ and $\vec{X}(a,b)$ as discussed above. Observe,

$$\begin{aligned}
\iint_S f\, dS &= \iint_\Omega f(\vec{r}(u,v))N_r(u,v)du\, dv \\
&= \iint_\Omega f(\vec{X}(h(u,v), g(u,v)))N_X(h(u,v), g(u,v))\left|\frac{\partial(h,g)}{\partial(u,v)}\right|du\, dv \\
&= \iint_D f(\vec{X}(a,b))N_X(a,b)da\, db.
\end{aligned}$$

In the last line I applied the multivariate change of variables theorem. Notice that the absolute value bars are important to the calculation. We will see in the corresponding calculation for flux the absolute value bars are absent, but this is tied to the orientation-dependence of the flux integral. In the scalar surface integral the direction (outward or inward) of the normal vector field does not figure into the calculation. We have shown

$$\boxed{\iint_S f\, dS = \iint_{-S} f\, dS}.$$

Next, turn to the reparametrization invariance of the flux integral. Suppose once more $\vec{r}$ and $\vec{X}$ both parametrize S. Calculate the flux of a continuous vector field $\vec{F}$ defined on some open set containing S (via the boxed equation following the definition)

$$\begin{aligned}
\iint_S \vec{F} \bullet d\vec{S} &= \iint_\Omega \vec{F}(\vec{r}(u,v)) \bullet \vec{N}_r(u,v)\, du\, dv \\
&= \iint_\Omega \vec{F}(\vec{X}(h(u,v), g(u,v))) \bullet \vec{N}_X(h(u,v), g(u,v))\frac{\partial(h,g)}{\partial(u,v)}\, du\, dv
\end{aligned}$$

If $\frac{\partial(h,g)}{\partial(u,v)} > 0$ for all $(u,v) \in \Omega$ then it follows that $\vec{N}_r$ and $\vec{N}_X$ point on the same side of S and we say they are **consistently oriented** parametrizations of S. Clearly if we wish for the flux to be meaningful we must choose a side for S and insist that we use a consistently oriented parametrization to calculate the flux. If $\vec{r}$ and $\vec{X}$ are consistently oriented then $\frac{\partial(h,g)}{\partial(u,v)} > 0$ for all

$(u,v) \in \Omega$ and hence $\left|\frac{\partial(h,g)}{\partial(u,v)}\right| = \frac{\partial(h,g)}{\partial(u,v)}$ for all $(u,v) \in \Omega$ and we find

$$\begin{aligned}\iint_S \vec{F} \bullet d\vec{S} &= \iint_\Omega \vec{F}(\vec{X}(h(u,v), g(u,v))) \bullet \vec{N}_X(h(u,v), g(u,v)) \Big| frac\partial(h,g)\partial(u,v) \Big| \, du \, dv \\ &= \iint_D \vec{F}(\vec{X}(a,b) \bullet \vec{N}_X(a,b) \, da \, db\end{aligned}$$

by the change of variable theorem for double integrals. On the other hand, if $\vec{r}$ is oppositely oriented from $\vec{X}$ then we say $\vec{r}$ parametrizes S whereas $\vec{X}$ parametrizes $-S$. In the case $\vec{X}$ points in the direction opposite $\vec{r}$ we find the coefficient $\left|\frac{\partial(h,g)}{\partial(u,v)}\right| = -\frac{\partial(h,g)}{\partial(u,v)}$ and it follows that:

$$\boxed{\iint_{-S} \vec{F} \bullet d\vec{S} = -\iint_S \vec{F} \bullet d\vec{S}}$$

7.6.2 standard surface examples

In this section we derive $d\vec{S}$ and dS for the sphere, cylinder, cone, plane and arbitrary graph. You can add examples past these, but these are essential. I also derive the surface area where appropriate.

1. S_R the **sphere of radius** R centered at the origin. We have spherical equation $\rho = R$ which suggests the natural parametric formulas: for $0 \leq \theta \leq 2\pi$ and $0 \leq \phi \leq \pi$

$$x = R\cos(\theta)\sin(\phi), \qquad y = R\sin(\theta)\sin(\phi), \qquad z = R\cos(\phi)$$

$$\vec{r}(\phi,\theta) = \big\langle R\cos(\theta)\sin(\phi),\ R\sin(\theta)\sin(\phi),\ R\cos(\phi)\big\rangle$$

We can limit the parameter space $[0,\pi] \times [0, 2\pi]$ to select subsets of the sphere if need arises. The normal vector field is calculated from partial derivatives of $\vec{r}(\phi,\theta)$;

$$\frac{\partial \vec{r}}{\partial \phi} = \Big\langle R\cos(\theta)\cos(\phi),\ R\sin(\theta)\cos(\phi),\ -R\sin(\phi)\Big\rangle$$

$$\frac{\partial \vec{r}}{\partial \theta} = \Big\langle -R\sin(\theta)\sin(\phi),\ R\cos(\theta)\sin(\phi),\ 0\Big\rangle$$

I invite the reader to calculate the cross-product above and derive that

$$\frac{\partial \vec{r}}{\partial \phi} \times \frac{\partial \vec{r}}{\partial \theta} = R^2\sin(\phi)\Big\langle \cos(\theta)\sin(\phi),\ \sin(\theta)\sin(\phi),\ \cos(\phi)\Big\rangle = R^2\sin(\phi)\widehat{\rho}.$$

We will find is useful to note that $\vec{N}(\phi,\theta) = R^2\sin(\phi)\widehat{\rho}$ for S_R. This is an outward pointing normal vector field. To summarize, for the sphere S_R with outward orientation we find

$$\boxed{d\vec{S} = R^2\sin(\phi)d\phi d\theta\widehat{\rho} \qquad \text{and} \qquad dS = R^2\sin(\phi)d\phi d\theta}$$

The surface area of the sphere S_R is given by

$$Area(S_R) = \int_0^{2\pi}\int_0^{\pi} R^2\sin(\phi)d\phi d\theta = 4\pi R^2.$$

It is interesting to note that $d/dR(\frac{4}{3}\pi R^3) = 4\pi R^2$ just like $d/dR(\pi R^2) = 2\pi R$.

2. **Right circular cylinder** of radius R with axis along z. In cylindrical coordinates we have the simple formulation $r = R$ which gives the natural parametrization:

$$x = R\cos(\theta), \qquad y = R\sin(\theta), \qquad z = z$$

$$\vec{r}(\theta, z) = \big\langle R\cos(\theta),\ R\sin(\theta),\ z\big\rangle$$

for $0 \leq \theta \leq 2\pi$ and $z \in \mathbb{R}$. Calculate $\frac{\partial \vec{r}}{\partial \theta} = \langle -R\sin(\theta),\ R\cos(\theta), 0\rangle = R\widehat{\theta}$ and $\frac{\partial \vec{r}}{\partial z} = \widehat{z}$ thus

$$\frac{\partial \vec{r}}{\partial \theta} \times \frac{\partial \vec{r}}{\partial z} = R\widehat{\theta} \times \widehat{z} = R\widehat{r}$$

Consequently, $\vec{N}(\theta, z) = R\widehat{r} = R\langle \cos(\theta), \sin(\theta), 0\rangle$ and we find

$$\boxed{d\vec{S} = Rd\theta\, dz\, \widehat{r} \qquad \text{and} \qquad dS = Rd\theta dz}$$

The surface area of the cylinder for $0 \leq z \leq L$ is given by

$$Area = \int_0^{2\pi}\int_0^L Rdz\, d\theta = 2\pi RL.$$

Of course, the whole cylinder with unbounded z has infinite surface area.

3. **Cone** at angle ϕ_o. In cylindrical coordinates $r = \rho\sin(\phi_o)$ thus the cartesian equation of this cone is easily derived from $r^2 = \rho^2\sin^2(\phi_o)$ gives $x^2 + y^2 = \sin^2(\phi_o)(x^2 + y^2 + z^2)$ hence, for $\phi_o \neq \pi/2$, we find $x^2 + y^2 = \tan^2(\phi_o)z^2$. In cylindrical coordinates this cone has equation $r = \tan(\phi_o)z$. From spherical coordinates we find a natural parametrization,

$$x = \rho\cos(\theta)\sin(\phi_o), \qquad y = \rho\sin(\theta)\sin(\phi_o), \qquad z = \rho\cos(\phi_o)$$

For convenience denote $a = \sin(\phi_o)$ and $b = \cos(\phi_o)$ thus

$$\vec{r}(\theta, \rho) = \big\langle a\rho\cos(\theta),\ a\rho\sin(\theta),\ b\rho\big\rangle$$

for $0 \leq \theta \leq 2\pi$ and $\rho \in [0, \infty)$. Differentiate to see that

$$\frac{\partial \vec{r}}{\partial \theta} = \big\langle -a\rho\sin(\theta),\ a\rho\cos(\theta),\ 0\big\rangle \qquad \& \qquad \frac{\partial \vec{r}}{\partial \rho} = \big\langle a\cos(\theta),\ a\sin(\theta),\ b\big\rangle.$$

Calculate,

$$\frac{\partial \vec{r}}{\partial \theta} \times \frac{\partial \vec{r}}{\partial \rho} = \langle ab\rho\cos(\theta), ab\rho\sin(\theta), -a^2\rho\rangle = a\rho\big\langle b\cos(\theta), b\sin(\theta), -a\big\rangle$$

Note that $\widehat{\phi} = \big\langle \cos(\theta)\cos(\phi), \cos(\theta)\cos(\phi), -\sin(\phi)\big\rangle$ and $a = \sin(\phi_o)$ and $b = \cos(\phi_o)$ hence

$$\frac{\partial \vec{r}}{\partial \theta} \times \frac{\partial \vec{r}}{\partial \rho} = \rho\sin(\phi_o)\widehat{\phi}$$

Consequently, $\vec{N}(\theta, \rho) = \rho\sin(\phi_o)\widehat{\phi}$. We find for the cone $\phi = \phi_o$,

$$\boxed{d\vec{S} = \rho\sin(\phi_o)d\theta\, d\rho\widehat{\phi} \qquad \text{and} \qquad dS = \rho\sin(\phi_o)d\theta\, d\rho.}$$

where $\widehat{\phi} = \big\langle \cos(\theta)\cos(\phi_o), \cos(\theta)\cos(\phi_o), -\sin(\phi_o)\big\rangle$. The surface area of the cone $\phi = \phi_o$ for $0 \leq \rho \leq R$ is given by

$$Area = \int_0^R \int_0^{2\pi} \rho \sin(\phi_o) d\theta\, d\rho = \sin(\phi_o)\pi R^2.$$

Of course, the whole cone with unbounded ρ has infinite surface area. On the other hand, the result above is quite reasonable in the case $\phi_o = \pi/2$.

4. **Plane containing vectors $\vec{A}$ and $\vec{B}$ and base-point $\vec{r}_o$.** We parametrize by

$$\vec{r}(u,v) = \vec{r}_o + u\vec{A} + v\vec{B}$$

for $(u,v) \in \mathbb{R}^{\,n\times n}$. Clearly $\partial_u \vec{r} = \vec{A}$ and $\partial_v \vec{r} = \vec{B}$ hence $\vec{N}(u,v) = \vec{A}\times\vec{B}$. The plane has a constant normal vector field. We find:

$$\boxed{d\vec{S} = \vec{A}\times\vec{B} du\, dv \qquad \text{and} \qquad dS = ||\vec{A}\times\vec{B}|| du\, dv.}$$

If we select a compact subset Ω of the plane then $\vec{r}(\Omega)$ has surface area

$$Area = \iint_\Omega ||\vec{A}\times\vec{B}|| du\, dv = \iint_\Omega ||\vec{A}\times\vec{B}||\, dA.$$

In the last equation I mean to emphasize that the problem reduces to an ordinary double integral of a constant over the parameter space Ω. Usually there is some parameter dependence in dS, but the plane is a very special case.

5. **Graph** $z = f(x,y)$. Naturally we take parameters x, y and form

$$\vec{r}(x,y) = \big\langle x, y, f(x,y)\big\rangle$$

this is the *Monge patch* on the surface formed by the graph. Differentiate,

$$\partial_x \vec{r} = \big\langle 1, 0, \partial_x f\big\rangle \qquad \& \qquad \partial_y \vec{r} = \big\langle 0, 1, \partial_y f\big\rangle$$

Calculate the normal vector field,

$$\vec{N}(x,y) = \big\langle 1, 0, \partial_x f\big\rangle \times \big\langle 0, 1, \partial_y f\big\rangle = \big\langle -\partial_x f, -\partial_y f, 1\big\rangle$$

We find:

$$\boxed{d\vec{S} = \big\langle -\partial_x f, -\partial_y f, 1\big\rangle\, dx\, dy \qquad \text{and} \qquad dS = \sqrt{1 + (\partial_x f)^2 + (\partial_y f)^2}\; dx\, dy.}$$

If Ω is a compact subset of $dom(f)$ then we can calculate the surface area by

$$Area = \iint_\Omega \sqrt{1 + (\partial_x f)^2 + (\partial_y f)^2}\; dx\, dy.$$

7.6.3 scalar surface integrals

I examine several examples to further illustrate the construction of the scalar surface integral. We use the surface area elements developed in the previous section.

Example 7.6.3. *Calculate the total mass of a sphere of radius R which has mass-density $\sigma(x, y, z) = \tan^{-1}(y/x)$. Identify that $\sigma = dM/dS$ hence $dM = \sigma dS$. Moreover, in spherical coordinates $\sigma = \theta$ hence thus the integral below gives the total mass:*

$$M = \iint_{S_R} \sigma \, dS = \int_0^{2\pi} \int_0^{\pi} R^2 \theta \sin(\phi) d\phi d\theta = R^2 \left(-\cos(\phi) \Big|_0^{\pi} \right) \left(\frac{\theta^2}{2} \Big|_0^{2\pi} \right) = \boxed{4\pi^2 R^2.}$$

Example 7.6.4. *Calculate the average of $f(x, y, z) = z^2$ over the circular cylinder S: $x^2 + y^2 = R^2$ for $0 \leq z \leq L$ (assume the caps are open, just find the average over the curved side). By logic,*

$$f_{avg} = \frac{1}{2\pi RL} \iint_S z^2 dS = \frac{1}{2\pi RL} \int_0^L \int_0^{2\pi} Rz^2 d\theta \, dz = \frac{1}{2\pi RL}(2\pi R)(L^3/3) = \boxed{\frac{L^2}{3}.}$$

Generally if we wish to calculate the average of a function over a surface of finite total surface area we define f_{avg} to be the value such that $\iint_S f \, dS = f_{avg} Area(s)$.

Example 7.6.5. *Find the centroid of the cone $\phi = \pi/4$ for $0 \leq \rho \leq R$. The centroid is the geometric center of the object with regard to the density. In other words, calculate the center of mass under the assumption $dM/dS = 1$. However you like to think of it, the centroid $(\bar{x}, \bar{y}, \bar{z})$ is given by*

$$\bar{x} = \frac{1}{Area(S)} \iint_S x dS, \qquad \bar{y} = \frac{1}{Area(S)} \iint_S y \, dS, \qquad \bar{z} = \frac{1}{Area(S)} \iint_S z \, dS$$

For the cone S it is clear by symmetry that $\bar{x} = \bar{y} = 0$. Once more building off (3.) of the previous section we calculate: $Area(S) = \sin(\phi_o)\pi R^2 = \pi R^2/\sqrt{2}$ hence as $z = \rho \cos(\phi_o) = \rho/\sqrt{2}$ and $dS = \frac{\rho}{\sqrt{2}} d\theta \, d\rho$

$$\bar{z} = \frac{\sqrt{2}}{\pi R^2} \int_0^R \int_0^{2\pi} \frac{\rho}{\sqrt{2}} \frac{\rho}{\sqrt{2}} d\theta \, d\rho = \frac{\sqrt{2}}{\pi R^2} \frac{R^3}{3} \frac{2\pi}{2} = \boxed{\frac{R\sqrt{2}}{3}.}$$

We can also calculate the moment of inertia about the z-axis for the cone S (assume constant mass density 1 for this example) . The moment of inertia is defined by $I_z = \iint_S r^2 dM$ and as the equation of this cone is simply $r = z$ we find $r^2 = \rho^2 \cos^2(\pi/4) = \rho^2/2$ thus

$$I_z = \iint_S r^2 \delta \, dS = \int_0^R \int_0^{2\pi} \frac{\rho^2}{2} \frac{\rho}{\sqrt{2}} d\theta \, d\rho = \frac{R^4(2\pi)}{8\sqrt{2}} = \boxed{\frac{\pi R^4}{4\sqrt{2}}.}$$

Example 7.6.6. *Find the scalar surface integral of $f(x,y,z) = xyz$ on the graph S: $z = 6+x+y$ for $0 \leq y \leq x^2$ and $0 \leq x \leq 1$ (this is just a portion of the total graph $z = 6+x+y$ which is an unbounded plane). Observe that $f_x = 1$ and $f_y = 1$ thus $dS = \sqrt{1+f_x^2+f_y^2}\ dx\,dy = \sqrt{3}\ dx\,dy$*

$$\begin{aligned}
\iint_S xyz\,dS &= \int_0^1 \int_0^{x^2} xy(6+x+y)\sqrt{3}\,dy\,dx \\
&= \sqrt{3}\int_0^1 \int_0^{x^2} \Big(\ 6xy + 6x^2y + 6xy^2\ \Big)\,dy\,dx \\
&= \sqrt{3}\int_0^1 \Big(\ 3xy^2 + 3x^2y^2 + 2xy^3\ \Big)\Big|_0^{x^2}\ dx \\
&= \sqrt{3}\int_0^1 \Big(\ 3x^5 + 3x^6 + 2x^7\ \Big)\Big|_0^{x^2}\ dx \\
&= \sqrt{3}\left(\ \frac{3}{6} + \frac{3}{7} + \frac{2}{8}\ \right) \\
&= \boxed{\frac{33\sqrt{3}}{28}.}
\end{aligned}$$

7.6.4 flux integrals

Once more I build off the examples from Section 7.6.2.

Example 7.6.7. *Calculate the flux of $\vec{F} = \langle 1,2,3\rangle$ through the part of the sphere $x^2+y^2+z^2 = 4$ which is above the xy-plane. Recall $d\vec{S} = R^2\sin(\phi)d\theta\,d\phi\widehat{\rho}$ and note for $z \geq 0$ we need no restriction of the polar angle θ ($0 \leq \theta \leq 2\pi$) however the azmuthial angle ϕ falls into the interval $0 \leq \phi \leq \pi/2$. Thus, as $R = 2$ for this example,*

$$\begin{aligned}
\Phi = \iint_S \vec{F}\bullet d\vec{S} &= \int_0^{\pi/2}\int_0^{2\pi} \langle 1,2,3\rangle \bullet \Big\langle \cos(\theta)\sin(\phi), \sin(\theta)\sin(\phi), \cos(\phi)\Big\rangle 4\sin(\phi)d\theta\,d\phi \\
&= 4\int_0^{\pi/2}\int_0^{2\pi}\Big(\cos(\theta)\sin^2(\phi) + 2\sin(\theta)\sin^2(\phi) + 3\cos(\phi)\sin(\phi)\Big)d\theta\,d\phi \\
&= 6\int_0^{\pi/2}\int_0^{2\pi}\Big(\sin(2\phi)\Big)d\theta\,d\phi \\
&= 12\pi\left(\frac{-1}{2}\cos(2\phi)\right)\Big|_0^{\pi/2} \\
&= \boxed{12\pi.}
\end{aligned}$$

Example 7.6.8. *Let $n \in \mathbb{Z}$ and calculate the flux of $\vec{F}(x,y,z) = (x^2+y^2+z^2)^{n/2-1}\langle x,y,z\rangle$ through the sphere S_R. Observe that $\vec{F} = \rho^n\,\widehat{\rho}$ and recall*

$$\vec{r}(\phi,\theta) = \big\langle R\cos(\theta)\sin(\phi), R\sin(\theta)\sin(\phi), R\cos(\phi)\big\rangle = R\widehat{\rho}$$

Thus calculate, $\vec{F}(\vec{r}(\phi,\theta)) = R^n\widehat{\rho}$

$$\begin{aligned}\Phi = \iint_S \vec{F}\bullet d\vec{S} &= \int_0^\pi \int_0^{2\pi} \big(\; R^n\,\widehat{\rho}\;\big)\bullet\big(\;\widehat{\rho}\,R^2\sin(\phi)d\theta\,d\phi\big)\\ &= R^{n+2}\int_0^\pi\int_0^{2\pi}\sin(\phi)d\theta\,d\phi\\ &= \boxed{4\pi R^{n+2}.}\end{aligned}$$

Believe it! Notice that the case $n=-2$ is very special. In that case the flux is independent of the radius of the sphere. The flux spreads out evenly and is neither created nor destroyed for $n=-2$.

Example 7.6.9. *Calculate the flux of $\vec{F} = \langle x^2, z, y\rangle$ through the closed cylinder $x^2+y^2=R^2$ with $0\le z\le L$. Notice that $S = S_1\cup S_2\cup S_3$ where I mean to denote the top by S_1, the base by S_2 and the side by S_3. The parametrizations and normal vectors to these faces are naturally given by*

$$\begin{aligned}&\vec{X}_1(r,\theta) = \langle r\cos(\theta), r\sin(\theta), L\rangle && \vec{N}_1(r,\theta) = \partial_r\vec{X}_1\times\partial_\theta\vec{X}_1 = \widehat{r}\times r\widehat{\theta} = r\,\widehat{z}\\ &\vec{X}_2(\theta,r) = \langle r\cos(\theta), r\sin(\theta), L\rangle && \vec{N}_2(\theta,r) = \partial_\theta\vec{X}_2\times\partial_r\vec{X}_2 = r\widehat{\theta}\times\widehat{r} = -r\,\widehat{z}\\ &\vec{X}_3(\theta,z) = \langle R\cos(\theta), R\sin(\theta), z\rangle && \vec{N}_3(\theta,z) = \partial_\theta\vec{X}_3\times\partial_z\vec{X}_3 = R\widehat{\theta}\times\,\widehat{z} = R\widehat{r}\end{aligned}$$

where $0\le r\le R$, $0\le\theta\le 2\pi$ and $0\le z\le L$. I'll calculate the flux through each face separately. Begin with S_1:

$$\vec{F}(\vec{X}_1(r,\theta)) = \vec{F}(r\cos(\theta), r\sin(\theta), L) = \big\langle r^2\cos^2(\theta), L, r\sin(\theta)\big\rangle$$

note that $d\vec{S}_1 = rdr\,d\theta\;\widehat{z}$ and we find

$$\Phi_{S_1} = \iint_{S_1}\vec{F}\bullet d\vec{S} = \int_0^R\int_0^{2\pi}\langle r^2\cos^2(\theta), L, r\sin(\theta)\rangle\bullet\widehat{z}\big(\,rd\theta\,dr\,\big) = \int_0^R\int_0^{2\pi} r^2\sin(\theta)d\theta\,dr = 0.$$

Through a similar calculation we find $\Phi_{S_2} = 0$. To calculate the flux through S_3 we should evaluate the vector field on the parametrization,

$$\vec{F}(\vec{X}_3(r,\theta)) = \vec{F}(R\cos(\theta), R\sin(\theta), z) = \big\langle R^2\cos^2(\theta), z, R\sin(\theta)\big\rangle$$

also recall that $\widehat{r} = \langle\cos(\theta),\sin(\theta),0\rangle$ thus $d\vec{S} = R\langle\cos(\theta),\sin(\theta),0\rangle d\theta\,dz$. Thus,

$$\begin{aligned}\Phi_{S_3} &= \int_0^{2\pi}\int_0^L \big\langle R^2\cos^2(\theta), z, R\sin(\theta)\big\rangle\bullet R\big\langle\cos(\theta),\sin(\theta),0\big\rangle d\theta\,dz\\ &= \int_0^{2\pi}\int_0^L\Big(R^3\cos^3(\theta) + zR\sin(\theta)\Big)\,d\theta\,dz\\ &= LR^3\int_0^{2\pi}\cos^3(\theta)\,d\theta\\ &= LR^3\int_0^{2\pi}\big[1-\sin^2(\theta)\big]\cos(\theta)\,d\theta\\ &= LR^3\int_{\sin(0)}^{\sin(2\pi)}\big[1-u^2\big]du\\ &= \boxed{0.}\end{aligned}$$

Example 7.6.10. *Find the flux of $\vec{F} = \vec{C}$ on the subset of the plane $\vec{r}(u,v) = \vec{r}_o + u\vec{A} + v\vec{B}$ defined by $1 \leq u^2 + v^2 \leq 4$. Denote $\Omega = dom(\vec{r})$. You can calculate $d\vec{S} = \vec{A} \times \vec{B}\ du\,dv$ hence*

$$\iint_S \vec{F} \bullet d\vec{S} = \iint_\Omega \vec{C} \bullet \left(\vec{A} \times \vec{B} \right) du\,dv = \vec{C} \bullet \left(\vec{A} \times \vec{B} \right) \iint_\Omega du\,dv = \boxed{15\pi\vec{C} \bullet \left(\vec{A} \times \vec{B} \right).}$$

Since the area of Ω is clearly $\pi(4)^2 - \pi(1)^2 = 15\pi$.

Finally, we conclude by developing a standard formula which is the focus of flux calculations in texts such as Stewart's.

Example 7.6.11. *Find the flux of $\vec{F} = \langle P, Q, R\rangle$ through the upwards oriented graph S: $z = f(x,y)$ with domain Ω. We derived that $d\vec{S} = \langle -\partial_x f, -\partial_y f, 1\rangle dx\,dy$ relative to the Monge parametrization $\vec{r}(x,y) = \langle x, y, f(x,y)\rangle$ for $(x,y) \in \Omega$. We calculate, from the definition,*

$$\begin{aligned} \iint_S \langle P, Q, R\rangle \bullet d\vec{S} &= \iint_\Omega \langle P, Q, R\rangle \bullet \langle -\partial_x f, -\partial_y f, 1\rangle dx\,dy \\ &= \iint_\Omega \left(-P\frac{\partial f}{\partial x} - Q\frac{\partial f}{\partial y} + R \right) dx\,dy \end{aligned}$$

which is technically incorrect, we really mean the following:

$$\iint_S \langle P, Q, R\rangle \bullet d\vec{S} = \iint_\Omega \left(-P(x,y,f(x,y))\frac{\partial f}{\partial x} - Q(x,y,f(x,y))\frac{\partial f}{\partial y} + R(x,y,f(x,y)) \right) dA$$

For example, to calculate the flux of $\vec{F}(x,y,z) = \langle -x, -y, e^{x^2+y^2}\rangle$ on $z = x^2+y^2$ for $0 \leq x^2+y^2 \leq 1$ we calculate $\partial_x f = 2x$ and $\partial_y f = 2y$

$$\iint_S \langle -x, -y, e^{x^2+y^2}\rangle \bullet d\vec{S} = \iint_\Omega \left(-2x^2 - 2y^2 + e^{x^2+y^2} \right) dA$$

In polar coordinates Ω is described by $0 \leq r \leq 1$ and $0 \leq \theta \leq 2\pi$ so we calculate

$$\iint_S \langle -x, -y, e^{x^2+y^2}\rangle \bullet d\vec{S} = \int_0^{2\pi} \int_0^1 \left(-2r^2 + e^{r^2} \right) r dr\,d\theta = 2\pi \left[\frac{-r^4}{2} + \frac{1}{2}e^{r^2} \right] \Bigg|_0^1 = \boxed{\pi(e-2).}$$

7.7 stokes' theorem

We have already encountered a simple version of Stokes' Theorem in the two-dimensional context. Recall that

$$\oint_{\partial R} P dx + Q dy = \iint_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA$$

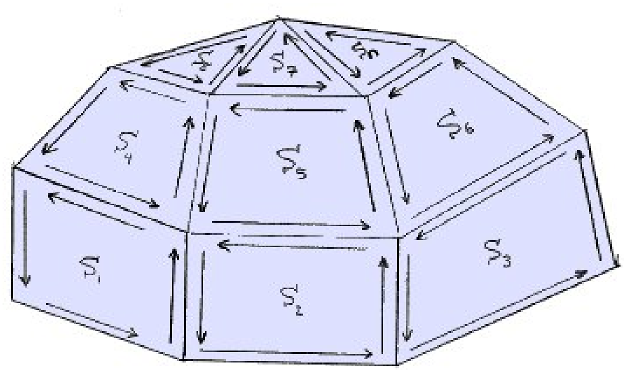
could be stated in terms of the z-component of the curl for $\vec{F} = \langle P, Q, 0\rangle$:

$$\oint_{\partial R} \vec{F} \bullet d\vec{r} = \iint_R \left(\nabla \times \vec{F}\right) \bullet \widehat{z}\, dx\, dy$$

However, notice that the double integral above is actually the surface integral of $\nabla \times \vec{F}$ over the planar surface R where $d\vec{S} = \widehat{z} dx\, dy$. Let's generalize this idea a little. Suppose S is some simply connected planar region with unit-normal $\widehat{n}$ which is consistently oriented[17] with ∂S then we can derive Green's Theorem for S and by the arguments of the earlier section we have that

$$\oint_{\partial S} \vec{F} \bullet d\vec{r} = \iint_S \left(\nabla \times \vec{F}\right) \bullet \widehat{n}\, du\, dv = \iint_S \left(\nabla \times \vec{F}\right) \bullet d\vec{S}. \qquad \star$$

Suppose that $S = S_1 \cup S_2 \cup \cdots \cup S_n$ is a simply connected surface where each S_j is a planar region with unit normal $\widehat{n}$ and consistent boundary ∂S_j. The planar regions S_j are called **faces** and we call such a surface a **polyhedron**.



Theorem 7.7.1. *baby Stokes' for piece-wise flat surfaces.*

Suppose S is a polyhedron S with consistently oriented boundary ∂S and suppose $\vec{F}$ is differentiable on some open set containing S then

$$\oint_{\partial S} \vec{F} \bullet d\vec{r} = \iint_S \left(\nabla \times \vec{F}\right) \bullet d\vec{S}.$$

Proof: The proof is mostly by the picture before the theorem. The key is that because S is composed of flat faces we can apply $\star$ to each face and obtain for $j = 1, 2, \ldots, n$:

$$\oint_{\partial S_j} \vec{F} \bullet d\vec{r} = \iint_{S_j} \left(\nabla \times \vec{F}\right) \bullet d\vec{S}.$$

[17] this means as we travel around the boundary the surface is on the left

Add these equations together and identify the surface integral,

$$\sum_{j=1}^{n} \oint_{\partial S_j} \vec{F} \bullet d\vec{r} = \sum_{j=1}^{n} \iint_{S_j} (\nabla \times \vec{F}) \bullet d\vec{S} = \iint_{S_1 \cup S_2 \cup \cdots \cup S_n} (\nabla \times \vec{F}) \bullet d\vec{S}.$$

To simplify the sum of the circulations we need to realize that all the edges of faces which are interior cancel against oppositely oriented adjacent face edges. The only edge which leads to an uncancelled flow are those outer edges which are not common to two faces. This is best seen in the picture. It follows that

$$\sum_{j=1}^{n} \oint_{\partial S_j} \vec{F} \bullet d\vec{r} = \oint_{\partial S} \vec{F} \bullet d\vec{r}$$

which completes the proof of the theorem. □

The theorem above naturally extends to a theorem for mostly regular surfaces. I say *mostly* regular since we do allow for surfaces which have edges and corners. The normal vector field may vanish at such edges, however, it is assumed to be nonvanishing elsewhere. There are surfaces where the normal vector field vanishes at points other than the edge or corner. For example, the mobius band. Such a surface is **non-orientable**. Generally, we only wish to consider oriented surfaces. I implicitly assume S is oriented by stating it has consistently oriented boundary ∂S.

Theorem 7.7.2. *Stokes' Theorem for simply connected surface.*

Suppose S is a simply connected surface S with consistently oriented boundary ∂S and suppose $\vec{F}$ is differentiable on some open set containing S then

$$\oint_{\partial S} \vec{F} \bullet d\vec{r} = \iint_{S} (\nabla \times \vec{F}) \bullet d\vec{S}.$$

Proof: If S is a mostly regular, simply connected, surface then it can be approximately modeled by a simply connected polyhedron with n-faces. As we take $n \to \infty$ this approximation becomes exact and we obtain Stokes' Theorem. □.

The reader should find the limit above geometrically obvious, but analytically daunting. We will not pursue the full analysis of the limit imiplicit within the proof above. However, we will offer another proof of Stokes' Theorem for a curved surface of a simple type at the end of this section. This should help convince the reader of the generality of the theorem. That said, the reader probably just wants to see some examples at this point:

Example 7.7.3. *Let $\vec{F} = \langle -y, x, z \rangle$. Find the flux of $\nabla \times \vec{F}$ over the half of the outward-oriented sphere $\rho = R$ with $z \geq 0$. Denote the hemisphere S_{R+}. The hemisphere is simply connected and the boundary of the outward-oriented hemisphere is given by $x = R\cos(\theta)$, $y = R\sin(\theta)$ and $z = 0$.*

Apply Stokes' Theorem:

$$\begin{aligned}\iint_{S_{R+}} (\nabla \times \vec{F}) \bullet d\vec{S} &= \oint_{\partial S_{R+}} \vec{F} \bullet d\vec{r} \\ &= \int_0^{2\pi} \big\langle -R\sin(\theta), R\cos(\theta), 0\big\rangle \bullet \big\langle -R\sin(\theta), R\cos(\theta), 0\big\rangle d\theta \\ &= \int_0^{2\pi} R^2 d\theta \\ &= \boxed{2\pi R^2.}\end{aligned}$$

Example 7.7.4. *Let $\vec{F} = \langle -y, x, z\rangle$. Find the flux of $\nabla \times \vec{F}$ over the half of the outward-oriented sphere $\rho = R$ with $z < 0$. Denote the lower hemisphere by S_{R-}. To solve this we can use the result of the previous problem. Notice that S_{R+} and S_{R-} share the same set of points as a boundary, however, $\partial S_{R+} = -\partial S_{R-}$. Apply Stokes' Theorem and the orientation-swapping identity for line-integrals:*

$$\iint_{S_{R-}} (\nabla \times \vec{F}) \bullet d\vec{S} = \oint_{\partial S_{R-}} \vec{F} \bullet d\vec{r} = -\oint_{\partial S_{R+}} \vec{F} \bullet d\vec{r} = \boxed{-2\pi R^2.}$$

Example 7.7.5. *Once more think about the vector field $\vec{F} = \langle -y, x, z\rangle$. Notice that $\vec{F}$ is differentiable on $\mathbb{R}^3$. We can apply Stokes' Theorem to any simply connected surface. If the consistently-oriented boundary of that surface is ∂S_{R+} then the flux of $\nabla \times \vec{F}$ is $2\pi R^2$.*

Stokes' Theorem allows us to deform the flux integral of $\nabla \times \vec{F}$ over a family of surfaces which share a common boundary. What about a closed surface? A sphere, ellipsoid, or the faces comprising a cube are all examples of closed surfaces. If S is an closed surface then $\partial S = \emptyset$. Does Stokes' Theorem hold in this case?

Example 7.7.6. *Suppose S is a simply connected closed surface S and suppose $\vec{F}$ is differentiable on some open set containing S then I claim that*

$$\iint_S \big(\ \nabla \times \vec{F}\big) \bullet d\vec{S} = 0.$$

To see why this is true simply cut S in halves S_1 and S_2. Notice that to consistently oriented S_1 and S_2 we must have that $\partial S_1 = -\partial S_2$. Apply Stokes' Theorem to each half to obtain:

$$\oint_{\partial S_1} \vec{F} \bullet d\vec{r} = \iint_{S_1} \big(\ \nabla \times \vec{F}\big) \bullet d\vec{S} \qquad \& \qquad \oint_{\partial S_2} \vec{F} \bullet d\vec{r} = \iint_{S_2} \big(\ \nabla \times \vec{F}\big) \bullet d\vec{S} \qquad (\star).$$

Note that,

$$\iint_{S_1} \big(\ \nabla \times \vec{F}\big) \bullet d\vec{S} + \iint_{S_2} \big(\ \nabla \times \vec{F}\big) \bullet d\vec{S} = \iint_S \big(\ \nabla \times \vec{F}\big) \bullet d\vec{S}$$

But, $\oint_{\partial S_2} \vec{F} \bullet d\vec{r} = -\oint_{\partial S_1} \vec{F} \bullet d\vec{r}$. Therefore, adding the eq. in $\star$ yields that $\iint_S \big(\ \nabla \times \vec{F}\big) \bullet d\vec{S} = 0$.

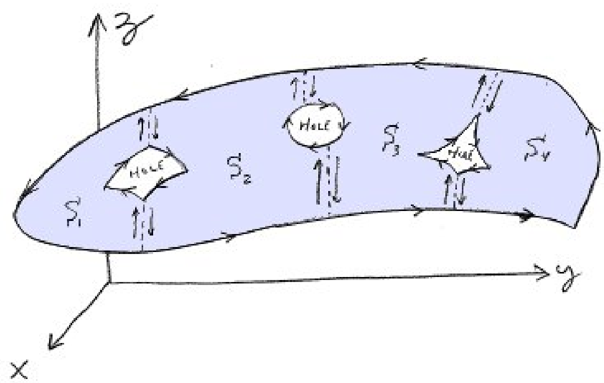
We can easily include the result of the example above by defining the integral over the empty set to be zero. Another interesting extension of the basic version of the theorem is the case that the surface has a few holes. The justification for this theorem will be a simple extension of the argument we already offered in the case of Green's Theorem.

Theorem 7.7.7. *Stokes' Theorem for connected surface possibly including holes.*

Suppose S is a connected surface S with consistently oriented boundary ∂S. If S has holes then we insist that the boundaries of the holes be oriented such that S is toward the left of the curve as we travel along the edge of the hole. Likewise, the outer boundary curve must also be oriented such that the surface if on the left as we traverse the boundary in the direction of its orientation. Suppose $\vec{F}$ is differentiable on some open set containing S then

$$\oint_{\partial S} \vec{F} \bullet d\vec{r} = \iint_S \left(\nabla \times \vec{F} \right) \bullet d\vec{S}.$$

Proof: Mostly by a picture. If S is connected with holes then we can cut S into pieces which are simply connected. We then apply Stokes' Theorem to each simply connected component surface. Finally, **sum** these equations together to obtain the surface integral of the flux over the whole surface and the line-integral around the boundary.



The key feature revealed by the picture is that all the interior cuts will cancel in this **sum** since any edge which is shared by two simply connected components must be oppositely oriented when viewed as the consistent boundary of the simply connected components. $\square$.

Example 7.7.8. *Suppose S is a pyramid with square-base on the xy-plane (do not include the square-base in the surface S so the boundary of S is the square at the base of the pyramid). Find the flux of $\nabla \times \vec{F}$ through the pyramid if $\vec{F}(x,y,z) = \langle 1, 3, z^3 \rangle$. Apply Stokes' Theorem,*

$$\iint_S \nabla \times \vec{F} \bullet d\vec{S} = \oint_{\partial S} \vec{F} \bullet d\vec{r}.$$

Note that $\vec{F}|_{\partial S} = \langle 1, 3, 0 \rangle$ since $z = 0$ on the boundary of the pyramid. Define $g(x,y,z) = x + 3y$ and note that $\nabla g = \langle 1, 3, 0 \rangle$ thus $\vec{F}|_{\partial S} = \langle 1, 3, 0 \rangle$ is conservative on the xy-plane and it follows that the integral around the closed square loop ∂S is zero. Thus,

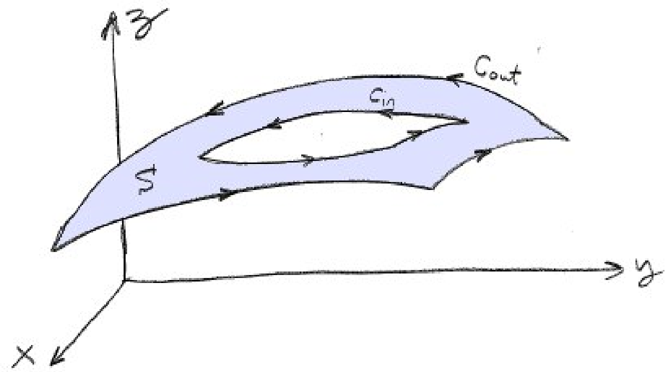
$$\iint_S \nabla \times \vec{F} \bullet d\vec{S} = 0.$$

This is not terribly surprising since direct computation easily shows that $\nabla \times \vec{F} = 0$.

Suppose S is a connected surface which has outer boundary C_{out} and inner boundary C_{in} where we have consistently oriented C_{out} but oppositely oriented C_{in}; $\partial S = C_{out} \cup (-C_{in})$. Applying Stokes' Theorem with holes to a vector field $\vec{F}$ which is differentiable on an open set containing S[18],

$$\oint_{C_{out}} \vec{F} \bullet d\vec{r} - \oint_{C_{in}} \vec{F} \bullet d\vec{r} = \iint_S \big(\nabla \times \vec{F} \big) \bullet d\vec{S}$$

This is a very interesting formula in the case $\nabla \times \vec{F} = 0$ on some connected annular surface:



Theorem 7.7.9. *Deformation Theorem for connected surfaces*

> Suppose S is connected with inner boundary C_{in} (oriented such that the surface's normal side is to the right of C_{in}) and outer boundary C_{out} oriented such that S is on the left of the curve then if $\vec{F}$ is differentiable on an open set containing S and has $\nabla \times \vec{F} = 0$ on S,
>
> $$\oint_{C_{out}} \vec{F} \bullet d\vec{r} = \oint_{C_{in}} \vec{F} \bullet d\vec{r}.$$

Proof: follows from the formula above the theorem $\square$.

Application to conservative vector fields: How does Stokes' Theorem help us understand conservative vector fields in $\mathbb{R}^3$? Recall we have a list of equivalent characterizations for a simply connected space U as given in Proposition 7.4.5: Suppose U is an open connected subset of $\mathbb{R}^n$ then the following are equivalent

1. $\vec{F}$ is conservative; $\vec{F} = \nabla f$ on all of U
2. $\vec{F}$ is path-independent on U
3. $\oint_C \vec{F} \bullet d\vec{r} = 0$ for all closed curves C in U
4. (add precondition $n = 3$ and U be simply connected) $\nabla \times \vec{F} = 0$ on U.

We argued before how $\big[\nabla \times \langle P, Q, 0\rangle\big] \bullet \widehat{z} = \partial_x Q - \partial_y P = 0$ paired with the deformation version of Green's Theorem allowed us to shrink loop integrals to a point hence estabishing that $\vec{F} = \langle P, Q, 0\rangle$

[18] if S were a donut this does not necessitate that $\vec{F}$ be differentiable in the center of the big-circle of the donut, it merely means $\vec{F}$ is differentiable near where the actual donut is found

is conservative if it passed Clairaut's test ($\partial_y P = \partial_x Q$ on a simple connected two-dimensional domain. Let us continue to three dimensions now that we have the needed technology.

Suppose $\vec{F} = \langle P, Q, R\rangle$ has vanishing curl ($\nabla \times \vec{F} = 0$) on some simply connected subset U of $\mathbb{R}^3$. Suppose C_1 and C_2 are any two paths from P to Q in U. Observe that $C_1 \cup (-C_2)$ bounds a simply connected surface S on which $\nabla \times \vec{F} = 0$ (since S sits inside U, and U has no holes). Apply Stokes' Theorem to S:

$$\oint_{C_1 \cup (-C_2)} \vec{F} \bullet d\vec{r} = \iint_S (\nabla \times \vec{F}) \bullet d\vec{S} = 0$$

Therefore, $\int_{C_1} \vec{F} \bullet d\vec{r} = \int_{C_2} \vec{F} \bullet d\vec{r}$ and path-independence on U follows.

Discussion: criteria 1,2 and 3 are n-dimensional results since the arguments we gave apply equally well in higher dimensions. However, item 4 is only worthwhile in its application to $n = 3$ or with proper specialization $n = 2$ because we have no cross-product and hence no curl in dimension $n = 4, 5 \dots$ etc.... You might wonder what is the generalization of (4.) for vector fields in $\mathbb{R}^n$. The answer involves differential forms. The exterior derivative allows us to properly extend vector differentiation to n-dimensions. This is not just an *academic*[19] comment, in the study of differential equations we enjoy solving exact differential equation. If $Pdx + Qdy + Rdz = 0$ then we can solve by $f(x, y, z) = 0$ if we can find f with $\partial_x f = P$, $\partial_y f = Q$ and $\partial_z f = R$. But, this problem is one we have already solved:

$$Pdx + Qdy + Rdz = 0 \ \text{ is exact} \qquad \leftrightarrows \qquad \vec{F} = \langle P, Q, R\rangle = \nabla f$$

Thus $\nabla \times \langle P, Q, R\rangle = 0$ on simply connected $U \subset \mathbb{R}^3$ implies existence of solutions for the given differential equation $Pdx + Qdy + Rdz = 0$. What about the case of additional independent variables; suppose w, x, y, z are variables

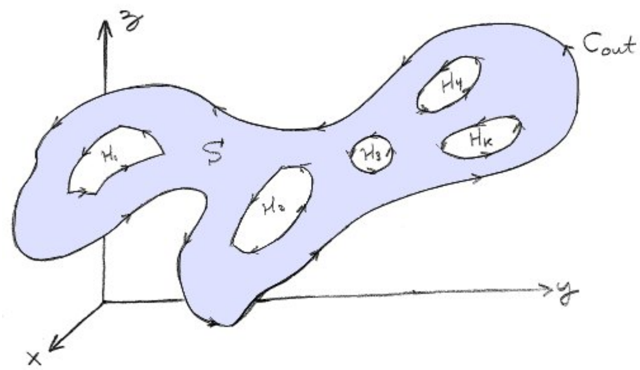
$$Idw + Pdx + Qdy + Rdz = 0 \ \text{ is exact} \qquad \leftrightarrows \qquad \vec{F} = \langle I, P, Q, R\rangle = \nabla f$$

If we could find the condition analogue to $\nabla \times \vec{F} = 0$ then we would have a nice test for the existence of solutions to the given DEqn. It turns out that the test is simply given by the exterior derivative of the given DEqn; [20] If $P_1dx_1 + P_2dx_2 + \cdots + P_ndx_n = 0$ then this DEQn is exact on a simply connected domain in $\mathbb{R}^n$ iff $d[P_1dx_1 + P_2dx_2 + \cdots + P_ndx_n] = 0$. Consult my advanced calculus notes and/or ask me for details about what "d" means in the context above.

Naturally, we can extend the annular result to the more general case that the surface as finitely-many holes:

[19] sad comment on our culture that this is an insult!

[20] example: $x^2dy - ydx = 0$ is not exact since $d[x^2dy - ydx] = 2xdx \wedge dy - dy \wedge dx \neq 0$. In contrast $ydx + xdy = 0$ is exact since $d[ydx + xdy] = dy \wedge dx + dx \wedge dy = 0$. We discussed the wedge product and exterior derivative in lecture, ask if interested and missed it...



Theorem 7.7.10. *Stokes' Theorem for connected surface possibly including holes.*

Suppose S is a connected surface S with consistently oriented boundary ∂S. If S has k holes $H_1, H_2, \dots, H_k$ then $\partial S = C_{out} \cup (-\partial H_1) \cup (-\partial H_2) \cdots (-\partial H_k)$. Suppose $\vec{F}$ is differentiable on some open set containing S then

$$\oint_{C_{out}} \vec{F} \bullet d\vec{r} - \oint_{\partial H_1} \vec{F} \bullet d\vec{r} - \oint_{\partial H_2} \vec{F} \bullet d\vec{r} - \cdots - \oint_{\partial H_k} \vec{F} \bullet d\vec{r} = \iint_S (\nabla \times \vec{F}) \bullet d\vec{S}.$$

Remark 7.7.11.

Electrostatics for a curved surface is an interesting problem. Imagine that the electric field was confined to the surface. We considered this problem in some depth for the plane. It is worth mentioning that we could just as well repeat those arguments here if we wished to model some field which is bound to flow along a surface. The details of the theory would depend on the particulars of the surface. I leave this as an open problem for the interested reader. You might even think about what surface you could pick to force the field to have certain properties... this is a prototype for the idea used in string theory; the geometry of the underlying space derives the physics. At least, this is one goal, sometimes realized...

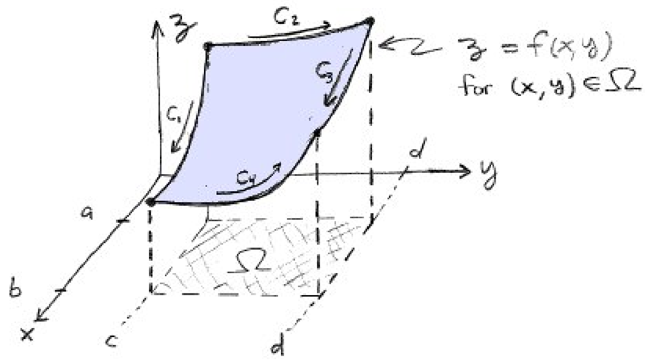
7.7.1 proof of Stokes' theorem for a graph

Our goal is to show that $\oint_{\partial S} \vec{F} \bullet d\vec{r} = \iint_S (\nabla \times \vec{F}) \bullet d\vec{S}$ for a simply connected surface S which can be expressed as a graph. Suppose $z = f(x, y)$ for $(x, y) \in \Omega$ In particular, as a starting point, let $\Omega = [a, b] \times [c, d]$[21]. It is easily calculated that $\vec{r}(x, y) = \langle x, y, f(x, y) \rangle$ induces normal vector field $\vec{N}(x, y) = \langle -\partial_x f, -\partial_y f, 1 \rangle$. The boundary of S consists of 4 line segments: $C_1 \cup C_2 \cup (-C_3) \cup (-C_4)$ where

1. C_1: for $a \leq t \leq b$ we set $x = t,\ y = c,\ z = f(t, c)$ hence $dx = dt,\ dy = 0,\ dz = \partial_x f(t, c) dt$
2. C_2: for $c \leq t \leq d$ we set $x = a,\ y = t,\ z = f(a, t)$ hence $dx = 0,\ dy = dt,\ dz = \partial_y f(a, t) dt$
3. C_3: for $a \leq t \leq b$ we set $x = t,\ y = d,\ z = f(t, d)$ hence $dx = dt,\ dy = 0,\ dz = \partial_x f(t, d) dt$
4. C_4: for $c \leq t \leq d$ we set $x = b,\ y = t,\ z = f(b, t)$ hence $dx = 0,\ dy = dt,\ dz = \partial_y f(b, t) dt$.

[21] In my first attempt I tried Ω as a type-I region given by functions f_1, f_2 such that $\Omega = \{(x, y) \mid f_1(x) \leq y \leq f_2(x),\ a \leq x \leq b\}$, however, this is too technical, it is clearer to show how this works for a rectangular domain.

We could visualize it as follows:



Consider a vector field $\vec{F} = \langle P, Q, R\rangle$ which is differentiable on some open set containing S. Calculate, for reference in the calculations below,

$$\nabla \times \vec{F} = \Big\langle\ \partial_y R - \partial_z Q,\ \partial_z P - \partial_x R,\ \partial_x Q - \partial_y P\ \Big\rangle$$

To calculate the flux of $\nabla \times \vec{F}$ we need to carefully compute $(\nabla \times \vec{F}) \bullet \vec{N}$;

$$\begin{aligned}(\nabla \times \vec{F}) \bullet \vec{N} &= \Big\langle\ \partial_y R - \partial_z Q,\ \partial_z P - \partial_x R,\ \partial_x Q - \partial_y P\ \Big\rangle \bullet \langle -\partial_x f, -\partial_y f, 1\rangle \\ &= \big[\partial_z Q \partial_x f + \partial_x Q\big] - \big[\partial_z P \partial_y f + \partial_y P\big] + \big[\partial_x R \partial_y f - \partial_y R \partial_x f\big]\end{aligned}$$

To proceed we break the problem into three. In particular $\vec{F} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3$ where we let $\vec{F}_1 = \langle P, 0, 0\rangle$, $\vec{F}_2 = \langle 0, Q, 0\rangle$ and $\vec{F}_3 = \langle 0, 0, R\rangle$. For $\vec{F}_1 = \langle P, 0, 0\rangle$ we calculate:

$$\begin{aligned}\iint_S \big(\nabla \times \vec{F}_1\big) \bullet d\vec{S} &= \int_a^b \int_c^d \big(-\partial_z P \partial_y f - \partial_y P\big) dy\, dx \\ &= -\int_a^b \int_c^d \partial_y \big[P(x, y, f(x, y))\big] dy\, dx \qquad \text{(chain-rule)} \\ &= \int_a^b \big[P(x, c, f(x, c)) - P(x, d, f(x, d))\big] dx. \quad (\star)\end{aligned}$$

On the other hand, we calculate the circulation of $\vec{F}_1 = P\,\widehat{x}$ around ∂S note that $dx = 0$ for C_2 and C_4 hence only C_1 and C_3 give nontrivial results,

$$\begin{aligned}\int_{\partial S} \vec{F}_1 \bullet d\vec{r} = \int_{\partial S} P dx &= \int_{C_1} P dx - \int_{C_3} P dx \\ &= \int_a^b P(t, c, f(t, c)) dt - \int_a^b P(t, d, f(t, d)) dt \\ &= \int_a^b \big[P(x, c, f(x, c)) - P(x, d, f(x, d))\big] dx.\end{aligned}$$

Consquently we have established Stokes' Theorem for $\vec{F}_1$ over our rather simple choice of surface. Continuing, consider $\vec{F}_2 = Q\,\widehat{y}$. Calculate, given our experience with the Pdx integrals we need not meet in the middle this time, I offer a direct computation:

$$\begin{aligned}
\iint_S (\nabla \times \vec{F}_2) \bullet d\vec{S} &= \int_a^b \int_c^d \big(\partial_z Q \partial_x f - \partial_x Q\big) dy\, dx \\
&= \int_c^d \int_a^b \partial_x \big[Q(x,y,f(x,y))\big] dy\, dx \qquad \text{(chain-rule, \& swapped bounds)} \\
&= \int_c^d Q(b,y,f(b,y))dy - \int_c^d Q(a,y,f(a,y))dy \\
&= \int_c^d Q(b,t,f(b,t))dt - \int_c^d Q(a,t,f(a,t))dt \\
&= \int_{C_2} Qdy - \int_{-C_4} Qdy \\
&= \oint_{\partial S} Qdy \qquad \text{(integrals along } C_1 \text{ and } C_3 \text{ are zero)} \\
&= \oint_{\partial S} \vec{F}_2 \bullet d\vec{r}.
\end{aligned}$$

Next, we work out Stokes' Theorem for $\vec{F}_3 = R\,\widehat{z}$. I'll begin with the circulation this time,

$$\begin{aligned}
\oint_{\partial S} \vec{F}_3 \bullet d\vec{r} &= \int_{C_1} Rdz - \int_{C_3} Rdz + \int_{C_2} Rdz - \int_{C_4} Rdz \\
&= \int_a^b R(t,c,f(t,c)) \frac{\partial f}{\partial x}(t,c)\ dt - \int_a^b R(t,d,f(t,d)) \frac{\partial f}{\partial x}(t,d)\ dt \\
&\quad + \int_c^d R(b,t,f(b,t)) \frac{\partial f}{\partial y}(b,t) dt - \int_c^d R(a,t,f(a,t)) \frac{\partial f}{\partial y}(a,t) dt \\
&= -\int_a^b \left(R(x,d,f(x,d)) \frac{\partial f}{\partial x}(x,d) - R(x,c,f(x,c)) \frac{\partial f}{\partial x}(x,c) \right) dx \\
&\quad + \int_c^d \left(R(b,y,f(b,y)) \frac{\partial f}{\partial y}(b,y) - R(a,y,f(a,y)) \frac{\partial f}{\partial y}(a,y) \right) dy \\
&= -\int_a^b \left[R(x,y,f(x,y)) \frac{\partial f}{\partial x}(x,y) \right] \Bigg|_c^d dx + \int_c^d \left[R(x,y,f(x,y)) \frac{\partial f}{\partial y}(x,y) \right] \Bigg|_a^b dy \\
&= -\int_a^b \int_c^d \frac{\partial R}{\partial y} \frac{\partial f}{\partial x}\, dy\ dx + \int_c^d \int_a^b \frac{\partial R}{\partial x} \frac{\partial f}{\partial y}\, dx\ dy \\
&= \int_a^b \int_c^d \left(\frac{\partial R}{\partial x} \frac{\partial f}{\partial y} - \frac{\partial R}{\partial y} \frac{\partial f}{\partial x} \right) dy\, dx \\
&= \iint_S (\nabla \times \vec{F}_3) \bullet d\vec{S}.
\end{aligned}$$

Therefore, by linearity of the curl and line and surface integrals we find that

$$\oint_{\partial S} \vec{F} \bullet d\vec{r} = \iint_S (\nabla \times \vec{F}) \bullet d\vec{S}.$$

Notice that the choice of rectangular bounds for Ω allowed us to freely exchange the order of integration since x and y bounds were independent. If Ω was a less trivial type I or type II region, then the arguments given in this section need some modification since swapping bounds is in general a somewhat involved process. That said, the result just proved is quite robust when paired with the earlier polyhedral proof to make a general argument. If surface consists of a graph with a curved domain then we can break it into rectangular subdomains and apply the result of this section to each piece. Once more when we sum those results together the nature of the adjoining regions is to cancel all line integrals modulo the boundary of the overall surface.[22] If the surface does not admit presentation as a graph $z = f(x, y)$ then generally we can patch it together with several graphs[23]. We apply the result of this section to each such patch and the sum the results to obtain Stokes' Theorem for a general simply connected surface.

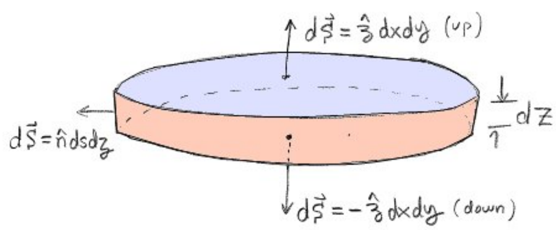
7.8 gauss theorem

Warning: the argument that follows is an infinitesimal argument. To properly undestand the meaning of this discussion we should remember this is simply notation for a finite approximation where a certain limit is taken. That said, I will not clutter this argument with limits. I leave those to the reader here. Also... I will offer another less heuristic argument towards the end of this section, I prefer this one since it connects with our previous discussion about the divergence of a two-dimensional vector field and Green's Theorem.

Green's Theorem in the plane quantifies the divergence of the vector field $P\,\widehat{x} + Q\,\widehat{y}$ through the curve ∂D;

$$\int_{\partial D} (P\,\widehat{x} + Q\,\widehat{y}) \bullet \widehat{n}\,ds = \iint_D \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dx\,dy$$

Suppose we consider a three-dimensional vector field $\vec{F} = P\,\widehat{x} + Q\,\widehat{y} + R\,\widehat{z}$. Furthermore, suppose we consider an infinitesimal cylinder $E = D \times dz$.



What is the flux of $\vec{F}$ out of the cylinder? Apply Green's Theorem to see that the flux through the vertical faces of the cylinder are simply given by either of the expressions below:

$$\Phi_{hor} = \iint_{\partial D \times [z, z+dz]} (P\,\widehat{x} + Q\,\widehat{y}) \bullet \widehat{n}\,ds\,dz = \iiint_{D \times [z, z+dz]} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dx\,dy\,dz$$

Since $d\vec{S} = \widehat{n} ds\,dz$ is clearly the vector area element of the vertical face(s) of the cylinder it is clear that the double integral above is simply the surface integral over the vertical faces of the cylinder.

[22]Technically, we'll have to form a sequence of such regions for some graphs and then take the limit as the rectangular net goes to infinitely many sub-divisions, however, the details of such analysis are beyond the scope of these notes. If this seems similar to the proof we presented for Green's Theorem then your intuition may serve you well in the remainder of this course.

[23]could be $z = f(x, y)$ type, or $y = g(x, z)$ or $x = h(y, z)$, the implicit function theorem of advanced calcululus will give a general answer to how this is done for a level surface

We identify,

$$\Phi_{hor} = \iint_{\partial D\times[z,z+dz]} \vec{F} \bullet d\vec{S} = \iiint_{D\times[z,z+dz]} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}\right) dx\, dy\, dz.$$

On the other hand, the flux through the horizontal caps of the cylinder $D \times \{z\}$ and $D \times \{z + dz\}$ involve only the z-component of $\vec{F}$ since $d\vec{S} = \widehat{z} dx\, dy$ for the upper cap and $d\vec{S} = -\widehat{z} dx\, dy$ for the lower cap hence the fluxes are

$$\Phi_{up} = \iint_D R(x, y, z + dz)\, dx\, dy \qquad \& \qquad \Phi_{down} = \iint_D -R(x, y, z)\, dx\, dy$$

The sum of these gives the net vertical flux:

$$\Phi_{vert} = \iint_D (\ R(x, y, z + dz) - R(x, y, z)\ )\, dx\, dy = \iiint_{D\times[z,z+dz]} \frac{\partial R}{\partial z} dx\, dy\, dz.$$

where in the last step we used the FTC to rewrite the difference as an integral. To summarize,

$$\Phi_{vert} = \iint_{caps} \vec{F} \bullet d\vec{S} = \iiint_{D\times[z,z+dz]} \frac{\partial R}{\partial z} dx\, dy\, dz.$$

The net-flux through the cylinder is the sum $\Phi_{vert} + \Phi_{hor}$. We find that,

$$\iint_{\partial E} \langle P, Q, R\rangle \bullet d\vec{S} = \iiint_E \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}\right) dx\, dy\, dz.$$

I prefer to write the result as follows:

$$\boxed{\iint_{\partial E} \vec{F} \bullet d\vec{S} = \iiint_E \nabla \bullet \vec{F}\, dV.}$$

This is the celebrated theorem of Gauss[24]. We often refer to it as the *divergence theorem*. It simply says that the net flux through a surface is portional to the continuous sum of the divergence throughout the solid. In other words, the divergence of a vector field $\vec{F}$ measures the number of field lines flowing from a particular volume. We found the two-dimensional analogue of this in our analysis of Green's Theorem and this is the natural three-dimensional extension of that discussion. For future reference: (this is also called Gauss' Theorem)

Theorem 7.8.1. *divergence theorem for a simple solid*

Suppose E a simple solid (has no interior holes) with consistently oriented outward facing boundary ∂E. If $\vec{F}$ is differentiable on an open set containing E then,

$$\iint_{\partial E} \vec{F} \bullet d\vec{S} = \iiint_E \nabla \bullet \vec{F}\, dV.$$

Discussion: But, I only proved it for a cylinder? Is this needed? Does it apply to other shapes? Yes. Consider the case that D is a rectangular region. We can use the argument offered above to obtain Gauss' theorem for a rectangular solid. Take any other simple solid (one with no holes) and

[24] although, apparently this was known to Lagrange some 50 years earlier, and to Russians as Ostrogradsky's theorem, see this Wikipedia article for discussion

note that you can obtain the solid as a union (possibly infinite!) of rectangular solids. Positively orient each rectangular solid and apply Gauss' Theorem to each member of the partition. Next, add the results together. On the one side we obtain the volume integral of the divergence. On the other side we get a sum of flux over many rectangular solids, some with adjacent faces. Think about this, any interior face of a particular rectangular solid will share a face with another member of the partition. Moreover, the common faces must be oppositely oriented in the distinct, but adjacent, rectangular solids. Thus, the interior flux all cancels leaving only the outside faces. The sum of the flux over all outside faces is simply the surface integral over the boundary of the simple solid. In this way we extend Gauss' Theorem to any solid without holes. Naturally, this leaves something to be desired analytically, but you can also appreciate this argument is very much the same we gave for Green's Theorem. This would seem to be part of some larger story... but, that is a story for another day[25].

Example 7.8.2. *Suppose $\vec{F}(x,y,z) = \langle x+y, y+x, z+y\rangle$ and you wish to calculate the flux of $\vec{F}$ through a set of stairs which has width 3 and 10 steps which are each height and depth 1. Let E be the set of stairs and ∂E the outward-oriented surface. Clearly the calculation of the flux over the surface of the stairs would be a lengthy and tedious computation. However, note that $\nabla \bullet \vec{F} = \frac{\partial(x+y)}{\partial x} + \frac{\partial(y+x)}{\partial y} + \frac{\partial(z+y)}{\partial z} = 3$ hence we find by Gauss' Theorem:*

$$\iint_{\partial E} \vec{F} \bullet d\vec{S} = \iiint_E \nabla \bullet \vec{F}\, dV = 3 \iiint_E dV = 3Vol(E).$$

Elementary school math shows:

$$Vol(E) = 3(1+2+3+4+5+6+7+8+9+10) = 165$$

Hence, $\boxed{\iint_{\partial E} \vec{F} \bullet d\vec{S} = 495.}$

Challenge: work the previous example for n-steps.

Example 7.8.3. *Suppose $\vec{F}$ is a differentiable at all points near a simple solid E. Calculate the flux of the curl of $\vec{F}$ through ∂E:*

$$\iint_{\partial E} (\nabla \times \vec{F}) \bullet d\vec{S} = \iiint_E \nabla \bullet (\nabla \times \vec{F}) dV = \iiint_E (0) dV = 0.$$

I used the identity $\nabla \bullet (\nabla \times \vec{F}) = 0$. You can contrast this argument with the one given in Example 7.7.6. Both examples are worth study.

Example 7.8.4. Problem: Consider the cube E with side-length L and one corner at the origin. Calculate the flux of $\vec{F} = \langle x,y,z\rangle$ through the upper face of the cube.

Solution: *Note that we cannot use a simple symmetry argument to see it is 1/6 of the given cube since the face in question differs from the base face (for example) in its relation to the vector field $\vec{F}$. On the other hand, if we imagine a larger cube of side-length $2L$ which is centered at the origin*

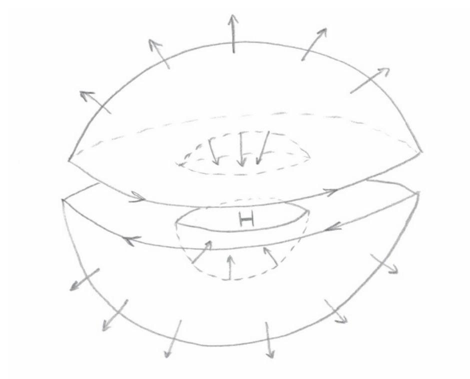
[25] look-up a proof of the generalized Stokes' Theorem in an advanced calculus text if you are interested. The key construction involves generalizing the polyhedral decomposition to something called an n-chain or perhaps an n-simplex depending what you read. Basically, you need some way of breaking the space into oriented parts with nice oriented boundaries, you prove the theorem for one such item and extrapolate via face-cancelling arguments as we have seen here in this case

then the vector field is symmetric with respect to the faces of $[-L, L]^3$. Call this larger cube E' and observe that we can easily calculate the net-flux through $\partial E'$ by the divergence theorem.

$$\Phi_{\partial E'} = \iint \vec{F} \bullet d\vec{S} = \iiint_{E'} \nabla \bullet \vec{F} dV = \iiint_{E'} 3\, dV = 3Vol(E') = 3(2L)^3 = 24L^3.$$

Notice that the face $[0, L] \times [0, L] \times \{L\}$ is 1/4 of the upper face of E' and it is symmetric with respect to the other 3/4 of the face $[-L, L] \times [-L, L] \times \{L\}$ with regard to $\vec{F}$. It follows the flux through the upper face of E is $1/4$ of the flux through the upper face of E'. Moreover, since the faces of E' are symmetric with regard to $\vec{F}$ we find that 1/6 of the total flux through $\partial E'$ passes through that upper face of E'. In summary, the flux through the face in question is simply $1/24$ of the total flux through $\partial E'$ and the flux through the upper face of E is $\boxed{L^3.}$

It should come as no suprise that there is a simple argument to extend the divergence theorem to a solid with a hole(s) in it. Suppose E is a solid which has a hole H in it. Denote the boundary of E by $\partial E = S_{out} \cup S_{in}$ where these surfaces are oriented to point out of E. Notice we can do surgery on E and slice it in half so that the remaining parts are simple solids (with no holes). The picture below illustrates this basic cut.



More exotic holes require more cutting, but the idea remains the same, we can cut any solid with a finite number of holes into a finite number of simple solids. Apply the divergence theorem, for an appropriately differentiable vector field, to each piece. Then add these together, note that the adjacent face's flux cancel leaving us the simple theorem below:

Theorem 7.8.5. *divergence theorem for a solid with k-interior holes.*

Suppose E a solid with interior holes $H_1, H_2, \dots, H_k$. Orient the surfaces $S_1, S_2, \dots, S_k$ of the holes such that the normals point into the holes and orient the outer surface S_{out} of E to point outward; hence $\partial E = S_{out} \cup S_1 \cup S_2 \cup \cdots \cup S_k$ gives E an outward oriented boundary. If $\vec{F}$ is differentiable on an open set containing E then,

$$\iint_{\partial E} \vec{F} \bullet d\vec{S} = \iiint_E \nabla \bullet \vec{F}\, dV.$$

This is perhaps more interesting if we take the holes as solids on their own right with outward oriented surfaces $\partial H_1, \partial H_2, \dots, \partial H_k$ (this makes $\partial H_i = -S_i$ for $i = 1, 2, \dots, k$. It follows that:

$$\iint_{\partial S_{out}} \vec{F} \bullet d\vec{S} - \iint_{\partial H_1} \vec{F} \bullet d\vec{S} - \iint_{\partial H_2} \vec{F} \bullet d\vec{S} - \cdots - \iint_{\partial H_k} \vec{F} \bullet d\vec{S} = \iiint_E \nabla \bullet \vec{F}\, dV.$$

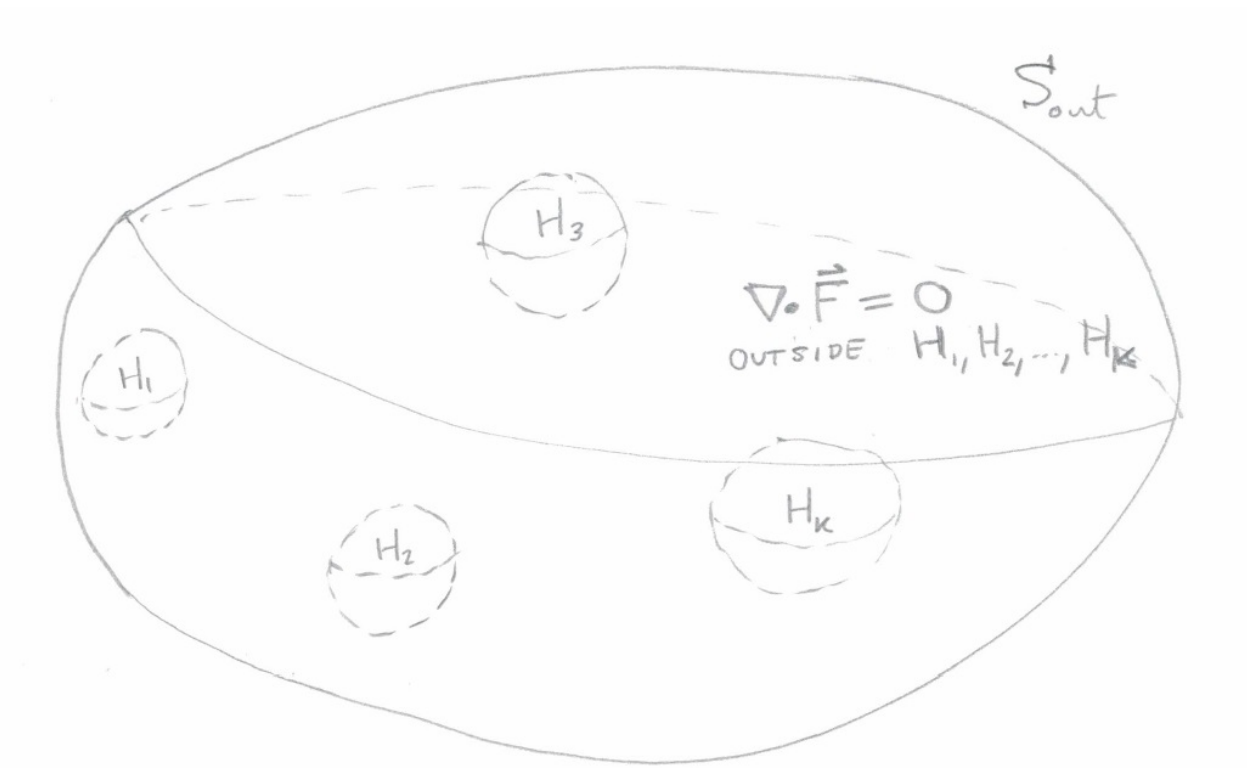
Note that we have an interesting result if $\nabla \bullet \vec{F} = 0$ on the E described above. In that case we obtain a deformation theorem for the flux spreading between surfaces:

Theorem 7.8.6. *deformation theorem for a solid with k-interior holes.*

Suppose E a solid with interior holes $H_1, H_2, \dots, H_k$. Let the outer surface S_{out} of E be oriented to point outward and give the hole surfaces an orientation which points out of the hole; If $\vec{F}$ is differentiable on an open set containing E and $\nabla \bullet \vec{F} = 0$ then,

$$\iint_{\partial S_{out}} \vec{F} \bullet d\vec{S} = \iint_{\partial H_1} \vec{F} \bullet d\vec{S} + \iint_{\partial H_2} \vec{F} \bullet d\vec{S} + \cdots + \iint_{\partial H_k} \vec{F} \bullet d\vec{S}.$$

This theorem forms the basis for the three-dimensional electrostatics and much more. Basically it says that if a field is mostly divergence free then the flux comes from only those places where the divergence is non-vanishing. Everywhere else the field lines just spread out. Given $\vec{F}$ with $\nabla \bullet \vec{F} = 0$ most places, I think of the holes as where the charge for the field is, either **sinks** or **sources**. From these mysterious holes the field lines flow in and out.



7.8.1 three-dimensional electrostatics

The fundamental equation of electrostatics is known as *Gauss' Law.* In three dimensions it simply states that the flux through a closed surface is proportional to the charge which is enclosed.

$$\Phi_E = Q_{enc}$$

In particular, if we denote $\sigma = dQ/dV$ and have in mind the solid E with boundary ∂E,

$$\oint_{\partial E} (\vec{E} \bullet d\vec{S} = \iiint_E \sigma dV$$

Suppose we have an isolated charge Q at the origin and we apply Gauss law to a sphere of radius ρ centered at the origin then we can argue by symmetry the electric field must be entirely radial in direction and have a magnitude which depends only on ρ. It follows that:

$$\oint_{\partial E} (\vec{E} \bullet d\vec{S} = \iiint_E \delta dV \quad \Rightarrow \quad (4\pi\rho^2)E = Q$$

Hence, the **coulomb field** in three dimensions is as follows:

$$\boxed{\vec{E}(\rho, \phi, \theta) = \frac{Q}{4\pi\rho^2}\widehat{\rho}}$$

Let us calculate the flux of the Coulomb field through a sphere S_R of radius R:

$$\begin{aligned} \oint_{S_R} (\vec{E} \bullet d\vec{S} &= \int_{S_R} \Bigg(\frac{Q}{4\pi\rho^2}\widehat{\rho} \bullet \widehat{\rho} dS \\ &= \int_{S_R} \frac{Q}{4\pi R^2} dS \\ &= \frac{Q}{4\pi R^2} \int_{S_R} dS \\ &= \frac{Q}{4\pi R^2}(4\pi R^2) \\ &= Q. \end{aligned} \tag{7.2}$$

The sphere is complete. In other words, the Coulomb field derived from Gauss' Law does in fact satisfy Gauss Law in the plane. This is good news. Let's examine the divergence of this field. It appears to point away from the origin and as you get very close to the origin the magnitude of E is unbounded. It will be convenient to reframe this formula for the Coulomb field by

$$\vec{E}(x, y, z) = \frac{Q}{4\pi(x^2 + y^2 + z^2)^{3/2}} \langle x, y, z \rangle = \frac{Q}{4\pi\rho^3} \langle x, y, z \rangle.$$

Note that as $\rho = \sqrt{x^2+y^2+z^2}$ it follows that $\partial_x \rho = x/\rho$ and $\partial_y \rho = y/\rho$ and $\partial_z \rho = z/\rho$. Consequently:

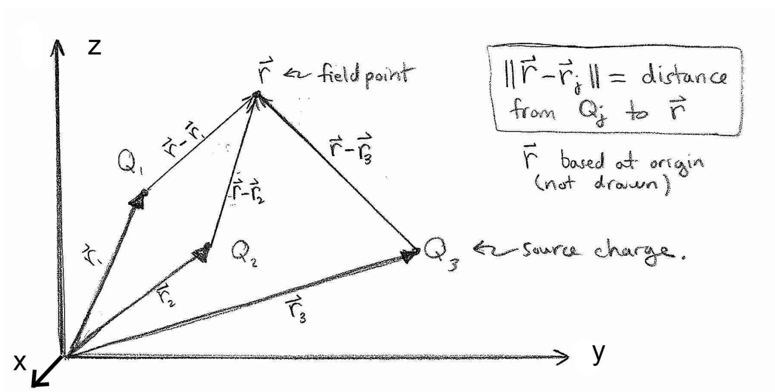
$$\begin{aligned}
\nabla \bullet \vec{E} &= \frac{Q}{4\pi}\left(\frac{\partial}{\partial x}\left[\frac{x}{\rho^3}\right] + \frac{\partial}{\partial y}\left[\frac{y}{\rho^3}\right] + \frac{\partial}{\partial z}\left[\frac{z}{\rho^3}\right]\right) \\
&= \frac{Q}{4\pi}\left(\frac{\rho^3 - 3x\rho^2\partial_x\rho}{\rho^6} + \frac{\rho^3 - 3y\rho^2\partial_y\rho}{\rho^6} + \frac{\rho^3 - 3z\rho^2\partial_z\rho}{\rho^6}\right) \\
&= \frac{Q}{4\pi}\left(\frac{\rho^3 - 3x^2\rho}{\rho^6} + \frac{\rho^3 - 3y^2\rho}{\rho^6} + \frac{\rho^3 - 3z^2\rho}{\rho^6}\right) \\
&= \frac{Q}{4\pi}\left(\frac{3\rho^3 - 3\rho(x^2+y^2+z^2)}{\rho^6}\right) \\
&= 0.
\end{aligned}$$

If we were to carelessly apply the divergence theorem this could be quite unsettling: consider,

$$\iint_{\partial E} \vec{E} \bullet d\vec{S} = \iiint_E \nabla \bullet \vec{E}\, dV \qquad \Rightarrow \qquad Q = \iiint_E (0) dV = 0.$$

But, Q need not be zero hence there is some contradiction? Why is there no contradiction? Can you resolve this paradox?

Moving on, suppose we have N charges placed at source points $\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N$ then we can find the total electric field by the principle of superposition.



We simply take the vector sum of all the coulomb fields. In particular,

$$\vec{E}(\vec{r}) = \sum_{j=1}^{N} \vec{E}_j = \sum_{j=1}^{n} \frac{Q_j}{4\pi}\frac{\vec{r}-\vec{r}_j}{||\vec{r}-\vec{r}_j||^3}$$

What is the flux through a sphere which encloses just the k-th one of these charges? Suppose S_R is a sphere of radius R centered at $\vec{r}_k$. We can calculate that

$$\iint_{S_R} \vec{E}_k \bullet d\vec{S} = Q_k$$

whereas, since $\vec{E}_j$ is differentiable inside all of S_R for $j \neq k$ and $\nabla \bullet \vec{E}_j = 0$ we can apply the divergence theorem to deduce that

$$\iint_{S_R} \vec{E}_j \bullet d\vec{S} = 0.$$

Therefore, summing these results together we derive for $\vec{E} = \vec{E}_1 + \cdots + \vec{E}_k + \cdots + \vec{E}_N$ that

$$\iint_{S_R} \vec{E} \bullet d\vec{S} = Q_k$$

Notice there was nothing particularly special about Q_k so we have derived this result for each charge in the distribution. If we take a sphere around a charge which contains just one charge then Gauss' Law applies and the flux is simply the charge enclosed. Denote $S_1, S_2, \dots, S_N$ as little spheres which enclose the charges $Q_1, Q_2, \dots, Q_N$ respective. We have

$$Q_1 = \iint_{S_1} \vec{E} \bullet d\vec{S}, \quad Q_2 = \iint_{S_2} \vec{E} \bullet d\vec{S}, \quad \dots, \quad Q_N = \iint_{S_N} \vec{E} \bullet d\vec{S}$$

Now suppose we have a surface S which encloses all N of the charges. The electric field is differentiable and has vanishing divergence at all points except the location of the charges. In fact, the superposition of the coulomb fields has vanishing divergence ($\nabla \bullet \vec{E} = 0$) everywhere except the location of the charges. It just has the isolated singularities where the charge is found. We can apply deformation theorem version of the divergence theorem to arrive at Gauss' Law for the distribution of N-charges:

$$\iint_{S} \vec{E} \bullet d\vec{S} = \iint_{S_1} \vec{E} \bullet d\vec{S} + \iint_{S_2} \vec{E} \bullet d\vec{S} + \cdots + \iint_{S_N} \vec{E} \bullet d\vec{S} = Q_1 + Q_2 + \cdots + Q_N$$

You can calculate the divergence is zero everywhere except at the location of the source charges. Moral of story: even one point thrown out of a domain can have dramatic and global consequences for the behaviour of a vector field. In physics literature you might find the formula to describe what we found by a *dirac-delta function* these distributions capture certain infinities and let you work with them. For example: for the basic coulomb field with a single point charge at the origin $\vec{E}(\rho, \phi, \theta) = \frac{Q}{4\pi\rho^2}\widehat{\rho}$ this derived from a charge density function σ which is zero everywhere except at the origin. Somehow $\iiint_E \sigma dV = Q$ for any region R which contains $(0,0,0)$. Define $\sigma(\vec{r}) = Q\delta(\vec{r})$. Where we define: for any function f which is continuous near 0 and any solid region E which contains the origin

$$\int_E f(\vec{r})\delta(\vec{r})dV = f(0)$$

and if E does not contain $(0,0,0)$ then $\iint_E f(\vec{r})\delta(\vec{r})dV = 0$. The dirac delta function turns integration into evaluation. The dirac delta function is not technically a function, in some sense it is zero at all points and infinite at the origin. However, we insist it is manageably infinity in the way just described. Notice that it does at least capture the right idea for density of a point charge: suppose E contains $(0,0,0)$,

$$\iint_E \sigma dV = \iint_E Q\delta(\vec{r})dV = Q.$$

On the other hand, we can better understand the divegence calculation by the following calculations[26]:

$$\nabla \bullet \frac{\widehat{\rho}}{\rho^2} = 4\pi\delta(\vec{r}).$$

Consequentlly, if $\vec{E} = \frac{Q}{4\pi}\frac{\widehat{\rho}}{\rho^2}$ then $\nabla \bullet \vec{E} = \nabla \bullet \left[\frac{Q}{4\pi}\frac{\widehat{\rho}}{\rho^2}\right] = \frac{Q}{4\pi}\nabla \bullet \frac{\widehat{\rho}}{\rho^2} = Q\delta(\vec{r})$. Now once more apply Gauss' theorem to the Coulomb field. This time appreciate that the divergence of $\vec{E}$ is not strictly

[26]I don't intend to explain where this 4π comes from, except to tell you that it must be there in order for the extension of Gauss' theorem to work out nicely.

zero, rather, the dirac-delta function captures the divergence: recall the RHS of this calculation followed from direct calculation of the flux of the Coloumb field through the circle ∂R,

$$\iint_{\partial E} \vec{E} \bullet d\vec{S} = \iiint_E \nabla \bullet \vec{E}\, dV \qquad \Rightarrow \qquad Q = \iiint_E Q\delta(\vec{r})dV = Q.$$

All is well. This is the way to extend Gauss' theorem for coulomb fields. The fields discussed in this section are the ones found in nature for the most part. Electric fields do propagate in three-dimensions and that means that isolated charges establish a coulomb field. In a later section of this chapter we seek to describe how a continuous distribution of charge can generate a field. At the base of that discussion are the ideas presented here, although, we will not have need of the dirac-delta for the continuous smeared out charge. Some physicists argue that there is no such thing as a point charge because the existence of such a charge comes with some nasty baggage. For example, if you calculate the total energy of the Coulomb field for a single point charge you find there is infinite energy in the field. Slightly unsettling.

7.8.2 proof of divergence theorem for a rectangular solid

Suppose $\vec{F}$ is differentiable near the solid $E = [x_1, x_2] \times [y_1, y_2] \times [z_1, z_2]$. Denote the faces of the solid as follows:

$$S_1: \qquad x = x_1, (z, y) \in [z_1, z_2] \times [y_1, y_2] \;\text{ has }\; d\vec{S} = -\,\widehat{x}\, dz\, dy$$

$$S_2: \qquad x = x_2, (y, z) \in [y_1, y_2] \times [z_1, z_2] \;\text{ has }\; d\vec{S} = \widehat{x}\, dy\, dz$$

$$S_3: \qquad y = y_1, (x, z) \in [x_1, x_2] \times [z_1, z_2] \;\text{ has }\; d\vec{S} = -\,\widehat{y}\, dx\, dz$$

$$S_4: \qquad y = y_2, (z, x) \in [z_1, z_2] \times [x_1, x_2] \;\text{ has }\; d\vec{S} = \widehat{y}\, dz\, dx$$

$$S_5: \qquad z = z_1, (y, x) \in [y_1, y_2] \times [x_1, x_2] \;\text{ has }\; d\vec{S} = -\,\widehat{z}\, dy\, dx$$

$$S_6: \qquad z = z_2, (x, y) \in [x_1, x_2] \times [y_1, y_2] \;\text{ has }\; d\vec{S} = \widehat{z}\, dx\, dy$$

The nice thing about the rectangular solid is that only one component of $\vec{F} = \langle P, Q, R \rangle$ cuts through a given face of the solid.

Observe that:

$$\begin{aligned}
\Phi_{12} &= \int_{S_1} \vec{F} \bullet d\vec{S} + \int_{S_2} \vec{F} \bullet d\vec{S} \qquad \text{(this defines } \Phi_{12} \text{ for future reference)} \\
&= \int_{y_1}^{y_2} \int_{z_1}^{z_2} \vec{F}(x_1, y, z) \bullet (-\widehat{x} dz\, dy) + \int_{z_1}^{z_2} \int_{y_1}^{y_2} \vec{F}(x_2, y, z) \bullet (\widehat{x} dz\, dy) \\
&= \int_{y_1}^{y_2} \int_{z_1}^{z_2} \big[-P(x_1, y, z)\big] dz\, dy + \int_{z_1}^{z_2} \int_{y_1}^{y_2} \big[P(x_2, y, z)\big] dy\, dz \\
&= \int_{z_1}^{z_2} \int_{y_1}^{y_2} \big[P(x_2, y, z) - P(x_1, y, z)\big] dy\, dz \\
&= \int_{z_1}^{z_2} \int_{y_1}^{y_2} \int_{x_1}^{x_2} \frac{\partial P}{\partial x}\, dx\, dy\, dz \qquad \text{by the FTC.}
\end{aligned}$$

Likewise,

$$\begin{aligned}
\Phi_{34} &= \int_{S_3} \vec{F} \bullet d\vec{S} + \int_{S_4} \vec{F} \bullet d\vec{S} \qquad \text{(this defines } \Phi_{34} \text{ for future reference)} \\
&= \int_{z_1}^{z_2} \int_{x_1}^{x_2} \vec{F}(x, y_1, z) \bullet (-\widehat{y} dx\, dz) + \int_{x_1}^{x_2} \int_{z_1}^{z_2} \vec{F}(x, y_2, z) \bullet (\widehat{y} dz\, dx) \\
&= \int_{z_1}^{z_2} \int_{x_1}^{x_2} \big[-Q(x, y_1, z)\big] dz\, dy + \int_{x_1}^{x_2} \int_{z_1}^{z_2} \big[Q(x, y_2, z)\big] dy\, dz \\
&= \int_{x_1}^{x_2} \int_{z_1}^{z_2} \big[Q(x, y_2, z) - Q(x, y_1, z)\big] dy\, dz \\
&= \int_{x_1}^{x_2} \int_{z_1}^{z_2} \int_{y_1}^{y_2} \frac{\partial Q}{\partial y}\, dx\, dy\, dz.
\end{aligned}$$

Repeating the same argument once more we derive:

$$\begin{aligned}
\Phi_{56} &= \int_{S_5} \vec{F} \bullet d\vec{S} + \int_{S_6} \vec{F} \bullet d\vec{S} \qquad \text{(this defines } \Phi_{12} \text{ for future reference)} \\
&= \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} \frac{\partial R}{\partial z}\, dz\, dy\, dx.
\end{aligned}$$

The flux of $\vec{F}$ over the entire boundary ∂E is found by summing the flux through each face. Therefore, by linearity of the triple integral for the second line,

$$\begin{aligned}
\iint_{\partial E} \vec{F} \bullet d\vec{S} &= \Phi_{12} + \Phi_{34} + \Phi_{56} \\
&= \int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dz\, dy\, dx.
\end{aligned}$$

Which proves the divergence theorem for an arbitrary rectangular solid. □

In contrast to the earlier argument in this section the third dimension of the cylinder was not take as infinitesimal. That said, it wouldn't take much to modify the earlier argument for a finite height. The result just proved extends to more general solids in the way discussed earlier in this section following the cylindrical proof.

7.9 green's identities and helmholtz theorem

This section is mostly developed in your homework, well, except for Green's Third identity and its applications. I think you'll be happy I did not relagate that to your homework. The purpose of this section is severalfold: (1.) we get to see how to use the divergence theorem yields further identities, in some sense these are the generalization of integration by parts to our current context, (2.) we lay some foundational mathematics which is important for the logical consistency of the potential formulation of electromagnetism (3.) we see new and fun calculations.

In your homework I asked you to show the following identities:

Proposition 7.9.1. *Green's First and Second Identities*

Suppose $f, g \in C^2(D')$ where E' is an open set containing the simple solid E which has piecewise smooth boundary ∂E. Then,

$$(1.) \quad \iiint_E \nabla f \bullet \nabla g \, dV + \iiint_E f \nabla^2 g \, dV = \iint_{\partial E} (f \nabla g) \bullet d\vec{S}$$

$$(2.) \quad \iiint_E \big(f \nabla^2 g - g \nabla^2 f\big) \, dV = \iint_{\partial E} \big(f \nabla g - g \nabla f\big) \bullet d\vec{S}$$

The identities above yield important results about harmonic functions. A function f is called **harmonic** on E if $\nabla^2 f = 0$ on E. You are also asked to show in the homework that:

Proposition 7.9.2. *properties of harmonic functions on a simple solid E*

Suppose the simple solid E which has piecewise smooth boundary ∂E and suppose f satisfies $\nabla^2 f = 0$ throughout E. Then,

(1.) $\displaystyle \iint_{\partial E} \nabla f \bullet d\vec{S} = 0$

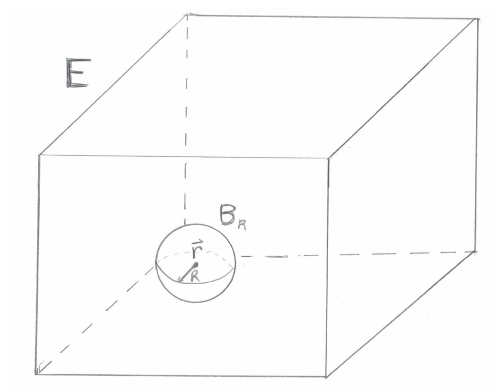
(2.) $\displaystyle \iiint_E \nabla f \bullet \nabla f \, dV = \iint_{\partial E} (f \nabla f) \bullet d\vec{S}$

(3.) $f(x, y, z) = 0$ for all $(x, y, z) \in \partial E \quad \Rightarrow \quad f(x, y, z) = 0$ for all $(x, y, z) \in E$.

(4.) If $\nabla^2 V_1 = b$ and $\nabla^2 V_2 = b$ throughout E and $V_1 = V_2$ on ∂E then $V_1 = V_2$ throughout E.

In words, (3.) states that if the restriction of f to ∂E is identically zero then f is zero throughout E. Whereas, (4.) states the solution to the **Poisson Equation** $\nabla^2 V = b$ is uniquely determined by its values on the boundary of a simple solid region.

This picture should help make sense of the Lemmas use to prove the Third Identity of Green:



Proposition 7.9.3. *Green's Third Identity.*

Suppose E is a simple solid, with piecewise smooth boundary ∂E, and assume f is twice differentiable throughout E. We denote $\vec{r} = \langle x, y, z\rangle$ as a fixed, but arbitrary, point in E and denote the variables of the integration by $\vec{r}\,' = \langle x', y', z'\rangle$, so $dV' = dx'\,dy'\,dz'$ and $\nabla' = \widehat{x}'\frac{\partial}{\partial x'} + \widehat{y}'\frac{\partial}{\partial y'} + \widehat{z}'\frac{\partial}{\partial z'}$. With this notation in mind,

$$f(\vec{r}) = \frac{-1}{4\pi}\iiint_E \frac{\nabla'^2 f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}dV' + \frac{1}{4\pi}\iint_{\partial E}\left(-f(\vec{r}\,')\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] + \frac{\nabla' f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right)\bullet d\vec{S}$$

We partly follow the proof in Colley. Begin with a Lemma,

Lemma 7.9.4. *Let h be a continuous function and S_R is a sphere of radius R centered at $\vec{r}$ then*

$$\lim_{R\to 0^+}\iint_{S_R}\frac{h(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\,dS = 0.$$

The proof relies on the fact that if $\vec{r}\,' \in S_R$ then, by the definition of a sphere of radius R centered at $\vec{r}$, we have $||\vec{r}-\vec{r}\,'|| = R$. Thus,

$$\iint_{S_R}\frac{h(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\,dS = \iint_{S_R}\frac{h(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\,dS = \frac{1}{R}\iint_{S_R}h(\vec{r}\,')\,dS$$

Since h is continuous[27] it follows that there exist $\vec{a}, \vec{b} \in S_R$ such that $h(\vec{a}) \leq h(\vec{r}\,') \leq h(\vec{b})$ for all $\vec{r}\,' \in S_R$. Consequently,

$$\frac{h(\vec{a})}{||\vec{r}-\vec{r}\,'||} \leq \frac{h(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||} \leq \frac{h(\vec{b})}{||\vec{r}-\vec{r}\,'||} \quad\Rightarrow\quad 4\pi R h(\vec{a}) \leq \iint_{S_R}\frac{h(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\,dS \leq 4\pi R h(\vec{b})$$

As $R \to 0^+$ is clear that $\vec{a}, \vec{b} \to \vec{r}$ hence $h(\vec{b}) \to h(\vec{r})$ and $h(\vec{a}) \to h(\vec{r})$. Observe, as $R \to 0^+$ we obtain $4\pi R h(\vec{a}) \to 4\pi R h(\vec{r}) \to 0$ and $4\pi R h(\vec{b}) \to 4\pi R h(\vec{r}) \to 0$. The Lemma above follows by the squeeze theorem ∇

[27] I use the extreme value theorem: any continuous, real-valued, image of a compact domain attains its extrema;

Lemma 7.9.5. *Let h be a continuous function and S_R is a sphere of radius R centered at $\vec{r}$ then*

$$\lim_{R\to 0^+} \iint_{S_R} h(\vec{r}\,')\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] \bullet d\vec{S} = -4\pi h(\vec{r}).$$

The proof of this Lemma is similar to the previous. We begin by simplifying the integral. Note,

$$\vec{r}-\vec{r}\,' = \langle x-x',\ y-y',\ z-z'\ \rangle$$

Let $L = ||\vec{r}-\vec{r}\,'||$ thus $L^2 = (x-x')^2+(y-y')^2+(z-z')^2$. Calculate,

$$\frac{\partial}{\partial x'}[L^2] = 2L\frac{\partial L}{\partial x'} = -2(x-x') \quad \Rightarrow \quad \frac{\partial L}{\partial x'} = \frac{x'-x}{L}.$$

Likewise, a similar calculation shows, $\frac{\partial L}{\partial y'} = \frac{y'-y}{L}$ and $\frac{\partial L}{\partial z'} = \frac{z'-z}{L}$. Thus,

$$\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] = \nabla'\left[\frac{1}{L}\right] = -\frac{1}{L^2}\nabla' L = \frac{1}{L^3}\langle x-x',\ y-y',\ z-z'\rangle = \frac{1}{L^3}(\vec{r}-\vec{r}\,')$$

Continuing, note the normal vector field $\vec{N}$ on S_R points in the $\vec{r}-\vec{r}\,'$ direction at $\vec{r}\,'$ thus

$$d\vec{S} = \frac{\vec{r}-\vec{r}\,'}{||\vec{r}-\vec{r}\,'||}dS = \frac{1}{L}(\vec{r}-\vec{r}\,')dS.$$

To calculate the surface integral, note $L = R$ on S_R thus,

$$\begin{aligned}
\iint_{S_R} h(\vec{r}\,')\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] \bullet d\vec{S} &= \iint_{S_R} h(\vec{r}\,')\frac{1}{R^3}(\vec{r}-\vec{r}\,') \bullet \frac{1}{R}(\vec{r}-\vec{r}\,')dS \\
&= \iint_{S_R} h(\vec{r}\,')\frac{||\vec{r}-\vec{r}\,'||^2}{R^4}\,dS \\
&= \frac{1}{R^2}\iint_{S_R} h(\vec{r}\,')\,dS
\end{aligned}$$

We once again use the squeeze theorem argument of the previous Lemma, since h is continuous it follows that there exist $\vec{a}, \vec{b} \in S_R$ such that $h(\vec{a}) \leq h(\vec{r}\,') \leq h(\vec{b})$ for all $\vec{r}\,' \in S_R$. Consequently,

$$\iint_{S_R} h(\vec{a})dS \leq \iint_{S_R} h(\vec{r}\,')dS \leq \iint_{S_R} h(\vec{b})dS$$

But, the integrals on the edges are easily calculated since $h(\vec{a}), h(\vec{b})$ are just constants and we deduce:

$$4\pi R^2 h(\vec{a}) \leq \iint_{S_R} h(\vec{r}\,')dS \leq 4\pi R^2 h(\vec{b}) \quad \Rightarrow \quad h(\vec{a}) \leq \frac{1}{4\pi R^2}\iint_{S_R} h(\vec{r}\,')dS \leq h(\vec{b}).$$

As $R \to 0^+$ is clear that $\vec{a}, \vec{b} \to \vec{r}$ hence $h(\vec{b}) \to h(\vec{r})$ and $h(\vec{a}) \to h(\vec{r})$ and the lemma follows by the squeeze theorem. ∇.

Green's Second Identity applies to solid regions with holes provided we give the boundary the standard outward orientation. With that in mind, consider $E' = E - B_R$ where B_R is the closed-ball of radius R which takes boundary S_R; $\partial B_R = S_R$. However, we insist that $\partial E' = \partial E \cup (-S_R)$ so the hole at $\vec{r}$ has inward-pointing normals. Apply Green's Second Identity with $g(\vec{r}\,') = \frac{1}{||\vec{r}-\vec{r}\,'||}$:

$$\iiint_{E'}\left(f(\vec{r}\,')\nabla'^2\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] - \frac{\nabla'^2 f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right)dV' = \iint_{\partial E'}\left(f(\vec{r}\,')\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] - \frac{\nabla' f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right)\bullet d\vec{S}$$

Once more, use the notation $L = ||\vec{r} - \vec{r}\,'||$. Observe that

$$\frac{\partial^2 L}{\partial x'^2} = \frac{\partial}{\partial x'}\left[\frac{x'-x}{L}\right] = \frac{L - (x'-x)\left(\frac{x'-x}{L}\right)}{L} = \frac{L^2 - (x'-x)^2}{L^2}$$

similar formulas hold for y and z hence:

$$\nabla'^2 \frac{1}{||\vec{r} - \vec{r}\,'||} = \frac{\partial^2 L}{\partial x'^2} + \frac{\partial^2 L}{\partial y'^2} + \frac{\partial^2 L}{\partial z'^2} = \frac{L^2 - (x'-x)^2 - (y'-y)^2 - (z'-z)^2}{L^2} = 0.$$

Therefore, Green's Second Identity simplifies slightly: (in the second line we use $\partial E' = \partial E \cup (-S_R)$

$$\begin{aligned}\iiint_{E'}\left(-\frac{\nabla'^2 f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right) dV' &= \iint_{\partial E'}\left(f(\vec{r}\,')\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] - \frac{\nabla' f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right)\bullet d\vec{S} \\ &= \iint_{\partial E}\left(f(\vec{r}\,')\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] - \frac{\nabla' f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right)\bullet d\vec{S} \\ &\quad - \iint_{S_R}\left(f(\vec{r}\,')\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] - \frac{\nabla' f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right)\bullet d\vec{S}\end{aligned}$$

Observe that $\nabla' f(\vec{r}\,')$ is continuous as required by the Lemma 7.9.5. Suppose $R \to 0^+$ and apply Lemma 7.9.5 and Lemma 7.9.4 to simplify the surface integrals over S_R. Moreover, as $R \to 0^+$ we see $E' \to E - \{\vec{r}\}$ and it follows:

$$\iiint_{E}\left(-\frac{\nabla'^2 f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right) dV' = \iint_{\partial E}\left(f(\vec{r}\,')\nabla'\left[\frac{1}{||\vec{r}-\vec{r}\,'||}\right] - \frac{\nabla' f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||}\right)\bullet d\vec{S} + 4\pi f(\vec{r}).$$

Green's Third Identity follows by algebra, just solve for $f(\vec{r})$. $\square$

I like the proof of this proposition because it is little more than careful calculation paired with a few natural limits. If you study the Coulomb field and the way it escapes the divergence theorem[28] due to the singularity at the origin then you might be led to these calculations. In any event, we've proved it so we can use it now.

Problem: Solve $\nabla^2 f = b$ for f on E.

Take the Laplacian of Green's Third Identity with respect to $\vec{r} \in E$. It can be shown through relatively straight-forward differentiation the surface integral over ∂E are trivial hence we find the beautiful formula:

$$\nabla^2 f(\vec{r}) = \frac{-1}{4\pi}\nabla^2 \iiint_E \frac{\nabla'^2 f(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||} dV'$$

Let $h(\vec{r}) = \nabla^2 f(\vec{r})$ to see what the formula above really says:

$$\boxed{h(\vec{r}) = \nabla^2 \iiint_E \frac{-h(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||} dV'}$$

[28] I discussed how many physics students are taught to escape the difficulty in Section 7.8.1

We have only provided evidence it is true if h is the Laplacian of another function f, but it is true in more generality[29] The formula boxed above shows how a particular modified triple integration in an inverse process to taking the Laplacian. It's like a second-order FTC for volume integrals. Returning to the problem, and placing faith in the generality of the formula[30], think of $h = f$ and assume $\nabla^2 f = b$: (I leave the details of why ∇^2 can be pulled into the integral and changed to ∇'^2)

$$f(\vec{r}) = \nabla^2 \iiint_E \frac{-f(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||} dV' = \iiint_E \frac{-\nabla'^2 f(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||} dV' = \iiint_E \frac{-b(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||} dV'$$

Therefore, we find the following theorem.

Theorem 7.9.6. *Integral solution to Poisson's Equation:*

If $\nabla^2 f = b$ for some continuous function b on a simple solid region E then

$$\boxed{f(\vec{r}) = \iiint_E \frac{-b(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||} dV'.}$$

If we are given that b tends to zero *fast* enough as we let $||\vec{r}|| \to \infty$ then the domain of integration E may be extended to $\mathbb{R}^3$ and the boxed equation serves to define a global solution to Poisson's Equation. Helmholtz' Theorem is related to this discussion. Let me state the Theorem for reference:

Theorem 7.9.7. *Helmholtz*

Suppose $\vec{F}$ is a vector field for which $\nabla \bullet \vec{F} = D$ and $\nabla \times \vec{F} = \vec{C}$. Furthermore, suppose $\vec{F} \to 0$ as $||\vec{r}|| \to \infty$ and C, D tend to zero faster than $1/||\vec{r}||^2$ then $\vec{F}$ is uniquely given by:

$$\vec{F} = -\nabla U + \nabla \times \vec{W}$$

where

$$U(\vec{r}) = \frac{1}{4\pi} \iiint_{\mathbb{R}^3} \frac{D(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||} dV' \qquad \& \qquad W(\vec{r}) = \frac{1}{4\pi} \iiint_{\mathbb{R}^3} \frac{C(\vec{r}\,')}{||\vec{r}-\vec{r}\,'||} dV'$$

For the interested reader, this does not contradict the multivariate Taylor theorem. In Taylor's theorem we are given all derivatives at a particular point and that data allows us to reconstruct the function. In Helmholtz Theorem we are given two globally defined vector fields and some first order derivative data which is sufficient to reconstruct the vector field $\vec{F}$. The difference is in the domain of the givens. At a point vs. on all points.

[29] Colley points to Kellog's *Foundations of Potential Theory* from 1928, I'd wager you could find dozens of texts to support this point. For example, Flander's text develops these ideas in blinding generality via differential form calculations.

[30] sorry folks, I'd like to fill this gap, but time's up

7.10 maxwell's equations and the theory of potentials

The central equations of electromagnetism are known as **Maxwell's Equations** in honor of James Clerk Maxwell who completed these equations around the time of the Civil War. Parts of these were known before Maxwell's work, however, Maxwell's addition of the term $\mu_o \epsilon_o \frac{\partial E}{\partial t}$ was crucial in the overal consistency and eventual success of the theory in explaining electromagnetic phenomena. In this section we will examine how these equations can either be stated locally as a system of PDEs or as integrals which yield the fields. Potentials for the fields are also analyzed, we see how Green's identities help connect the integral formulations of the potentials and the corresponding PDE which are, in a particular gauge, Poisson-type equations. Please understand I am not attempting to explain the physics here! That would take a course or two, our focus is on the mathematical backdrop for electromagnetism. I'll leave the physics for our junior-level electromagnetism course.

Let me set the stage here: $\vec{E}$ is the electric field, $\vec{B}$ is the magnetic field, both depend on time t and space x, y, z generally. In principle the particular field configuration is due to a given charge density ρ and current density $\vec{J}$. The electric and magnetic fields are solutions to the following set of PDEs[31]:

$$\nabla \bullet \vec{E} = \frac{\rho}{\epsilon_o}, \qquad \nabla \times \vec{B} = \mu_o\left(\vec{J} + \epsilon_o \frac{\partial \vec{E}}{\partial t}\right), \qquad \nabla \bullet \vec{B} = 0, \qquad \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

These are Gauss' Law, Ampere's Law with Maxwell's correction, no magnetic monopoles and Faraday's Law all written as local field equations. In most elementary electromagnetics courses[32] these Laws are presented as integral relations. In those elementary treatments the integral relations are taken as primary, or basic, experimentally verified results. In constrast, to Colley's *Vector Calculus* text, or most introductory physics texts, I take the PDE form of Maxwell's equations as basic. In my opinion, these are the nexus from which all else flows in E & M. For me, Maxwell's equations *define* electromagnetism[33]. Let's see how Stokes' and Gauss' theorems allow us to translate Maxwell's equations in PDE-form to Maxwell's equations in integral form.

7.10.1 Gauss' Law

Suppose M is a simple solid with closed surface $S = \partial M$ where and apply the divergence theorem to Gauss' Law:

$$\nabla \bullet \vec{E} = \frac{\rho}{\epsilon_o} \quad \Rightarrow \quad \iiint_M (\nabla \bullet \vec{E})\, dV = \frac{1}{\epsilon_o} \iiint_M \rho \, dV \quad \Rightarrow \quad \boxed{\iint_{\partial M} \vec{E} \bullet d\vec{S} = \frac{Q_{enc}}{\epsilon_o}}$$

The equation boxed above is what we call *Gauss' Law* in the freshman[34] E & M course. It simply states that the flux through a closed "gaussian" surface is given by the charge enclosed in the surface divided by the permitivity of space.

Example 7.10.1. *Suppose you have a very long line of charge along the z-axis with constant density $\lambda = dQ/dz$. Imagine a Gaussian cylinder S length L centered about the z-axis: only the*

[31] I use SI units and ϵ_o and μ_o are the permitivity and permeability of empty space

[32] for example, Physics 232 at LU

[33] ok, in truth there is a case that escapes Maxwell's equations, but that nonlinear case is considerably more sophisticated than these notes...

[34] because there are three more years of physics past this course... it is to be done in the Freshman year of university.

curved side of the cylinder gets flux and by the geometry it is clear that $\vec{E} = E\hat{r}$. Thus,

$$\iint_S \vec{E} \bullet d\vec{S} = \frac{Q_{enc}}{\epsilon_o} \quad \Rightarrow \quad 2\pi r L E = \frac{\lambda L}{\epsilon_o}$$

and we find $\boxed{\vec{E} = \dfrac{\lambda}{2\pi\epsilon_o r}\hat{r}.}$

It is interesting this is the dependence on distance we saw in two-dimensional electrostatics.

7.10.2 Ampere's Law

Suppose we have steady currents. It turns out that $\frac{\partial E}{\partial t} = 0$ in this case so we need to solve $\nabla \times \vec{B} = \mu_o \vec{J}$ in this **magnetostatic** case. Suppose S is a simply connected surface and use Stokes' Theorem in the calculation that follows:

$$\nabla \times \vec{B} = \mu_o \vec{J} \quad \Rightarrow \quad \iint_S (\nabla \times \vec{B}) \bullet d\vec{S} = \mu_o \iint_S \vec{J} \bullet d\vec{S} \quad \Rightarrow \quad \boxed{\oint_{\partial S} \vec{B} \bullet d\vec{r} = \mu_o I_{enc}.}$$

The boxed equation above is called *Ampere's Law* in the basic course. It states that the circulation of the magnetic field around an "amperian" loop is the product of the current that cuts through the loop and the permeability of space.

Example 7.10.2. *Suppose we are given a constant current I in the $\hat{z}$-direction. We can apply Ampere's Law to a circle centered on the z-axis:*

$$\oint_C \vec{B} \bullet d\vec{r} = \mu_o I \quad \Rightarrow \quad B(2\pi r) = \mu_o I \quad \Rightarrow \quad B = \frac{\mu_o I}{2\pi r}$$

To make the calculation above we used the symmetry of the given current. There was no way to have a z or θ dependence on the magnitude. To anticipate the direction of $\vec{B}$ I used the right-hand-rule for current which states that the magnetic field induced from a current wraps around the wire. If you point your right thumb in the direction of current then your fingers wrap around like the magnetic field. In short, $\vec{B} = \frac{\mu_o I}{2\pi r}\hat{\theta}$. We could derive the directionality of $\vec{B}$ from direct analysis of $\nabla \times \vec{B} = \mu_o \vec{J}$, this says that $\vec{B}$ is changing in a direction which is perpendicular to $\vec{J}$. The cross-product of the curl builds in the right-handedness.

If the electric field has a time-dependence we can use a modified form of Ampere's Law where the psuedo-current $\vec{J}_d = \epsilon_o \frac{\partial \vec{E}}{\partial t}$ is added to the real current. Otherwise the mathematics is the same, we just have to throw in Maxwell's correction to the enclosed current.

7.10.3 Faraday's Law

Suppose S is a simply connected surface with consistently oriented boundary ∂S.

$$\begin{aligned} \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad &\Rightarrow \quad \iint_S (\nabla \times \vec{E}) \bullet d\vec{S} = -\iint_S \frac{\partial \vec{B}}{\partial t} \bullet d\vec{S} \\ &\Rightarrow \quad \oint_{\partial S} \vec{E} \bullet d\vec{r} = -\frac{\partial}{\partial t} \iint_S \vec{B} \bullet d\vec{S} \\ &\Rightarrow \quad \boxed{\mathcal{E} = -\frac{\partial \Phi_B}{\partial t}.} \end{aligned}$$

The boxed equation is called *Faraday's Law* in the basic E & M course. The script E is the voltage around the loop ∂S and Φ_B is the magnetic flux through the loop. In words, a changing magnetic flux induces a voltage in the opposite direction. The minus in this law expresses the fact that nature abhors a change in flux. Of course, we have yet to properly introduce voltage in this section. That is our next project, right after we deal with the $\nabla \bullet \vec{B} = 0$ equation.

7.10.4 no magnetic monopoles?

In the case of the electric field we have $\nabla \bullet \vec{E} = \frac{\rho}{\epsilon_o}$ and the charge density can be nonzero since electric charge does in fact exist. On the other hand, if there was such a thing as magnetic charge[35] it would have a density ρ_m and, depending on our choice of units, some k such that

$$\nabla \bullet \vec{B} = k\rho_m$$

However, since we are taking $\nabla \bullet \vec{B} = 0$ as basic in this section we can conclude that $\rho_m = 0$ throughout all space. There is no *isolated* magnetic charge. North and South poles always come in pairs it seems. For these reasons the equation $\nabla \bullet \vec{B} = 0$ is called the no magnetic monopoles equation. It is a silent participant in the story of basic electromagnetism. Its presence is felt in what is not considered. All of this said, the existence of magnetic monopoles is an ongoing and interesting question in theoretical physics.

7.10.5 potentials

Both the electric and magnetic fields can be derived from potentials. See Theorem 7.9.7 to appreciate that generally we need both types of potentials to produce a given vector field. In the case of $\vec{E}$ and $\vec{B}$ it is customary to derive these from the **scalar potential** ϕ and the **vector potential** $\vec{A}$. In particular: for a given $\vec{E}$ and $\vec{B}$ we insist the potentials satisfy:

$$\boxed{\vec{E} = -\nabla\phi + \frac{\partial A}{\partial t} \qquad \& \qquad \vec{B} = \nabla \times \vec{A}.}$$

Notice this means that the choice of ϕ and $\vec{A}$ are far from unique. Why? Check it for yourself, if $\vec{A}, \phi$ are as above then $\vec{A}' = \vec{A} + \nabla g$ and $\phi' = \phi + \partial_t g$ will yield the same $\vec{E}, \vec{B}$. For convenience in this section we suppose that $\nabla \bullet \vec{A} = 0$ this is called the **Coulomb gauge** condition. Typically we call ϕ the **voltage** whereas $\vec{A}$ is the **vector potential**. Only ϕ is seen in the basic E & M course. I want to connect with some earlier discussion of Laplace's and Poisson's Equations. Let's examine how Gauss' Law translates when put into the potential formulation:

$$\nabla \bullet \vec{E} = \frac{\rho}{\epsilon_o} \quad \Rightarrow \quad \nabla \bullet (-\nabla\phi) = \frac{\rho}{\epsilon_o} \quad \Rightarrow \quad \boxed{\nabla^2\phi = \frac{-\rho}{\epsilon_o}.}$$

Therefore, the potential must be a solution to Poisson's equation. In the case that $\rho = 0$ on some simply connected volume we find the potential must be a solution to Laplace's equation $\nabla^2\phi = 0$. This means in the absence of charge we can state some very interesting results about voltage:

1. if the voltage is constant along the boundary of a charge-free simply connected region then the voltage is constant throughout the region.

[35]if you'd like to read more on magnetic charge there are many sources. One nice discussion is found in *Gravitation* by Misner, Thorne and Wheeler where possible theories of electromagnetism are selected by something called the complexion

2. by Theorem 7.9.6 we can solve for ϕ by calculating the appropriate integral of the charge density! Identify that $b = -\rho/\epsilon_o$ hence

$$\boxed{\phi(\vec{r}) = \iiint_M \frac{\rho(\vec{r}\,')}{4\pi\epsilon_o||\vec{r}-\vec{r}\,'||}dV'.}$$

In electrostatics we arrive at the same formula by applying the superposition principle to the potential due to a continuum of charge $\rho = \frac{dQ}{dV}$. Each infinitesimal charge dQ at $\vec{r}\,'$ gives a voltage of $d\phi = \frac{dQ}{4\pi\epsilon_o||\vec{r}-\vec{r}\,'||}$ then we integrate over the whole region where $\rho \neq 0$ and obtain the net-voltage due to the charge distribution ρ. We have shown that this intuitive approach is correct, it yields the same solution to Gauss' Law for the potential. Moreover, since this is a Poisson equation we even know that once we are given ρ and specify the voltage along a particular boundary then the voltage is uniquely specified elsewhere. I highly reccomend Griffith's *Introduction to Electrodynamics* if you'd like to read the physics which leads to the same conclusion.

Continuing, what about Ampere's Law? How does it translate into potential formalism?

$$\nabla \times \vec{B} = \mu_o\left(\vec{J} + \epsilon_o\frac{\partial \vec{E}}{\partial t}\right) \quad \Rightarrow \quad \nabla \times (\nabla \times \vec{A}) = \mu_o\left(\vec{J} - \epsilon_o\frac{\partial \nabla\phi}{\partial t}\right)$$

$$\Rightarrow \quad \nabla(\nabla \bullet \vec{A}) - \nabla \bullet \nabla\vec{A} = \mu_o\left(\vec{J} - \epsilon_o\nabla\frac{\partial \phi}{\partial t}\right)$$

$$\Rightarrow \quad \boxed{\nabla^2\vec{A} = -\mu_o\left(\vec{J} + \epsilon_o\nabla\frac{\partial \phi}{\partial t}\right).}$$

The process of choosing the Coulomb gauge is an example of *gauge fixing.* These partially adhoc conditions placed on the potentials greatly simplify the derivation of the potentials. Although, here, the Coulomb gauge actually makes the Poisson Equation for the vector potential a bit cluttered. However, if the currents considered are all steady and constant then there is no time-dependence to the electric field and the Coulomb gauge turns Ampere's Law into a set of three Poisson-type equations:

$$\nabla^2 A_1 = -\mu_o J_1, \qquad \nabla^2 A_2 = -\mu_o J_2, \qquad \nabla^2 A_3 = -\mu_o J_3.$$

Perhaps you can also appreciate that the distinction between ϕ and $\vec{A}$ is less than you might think, in fact, these are just two parts of what is properly viewed as the four-potential in relativistic electrodynamics, but, I think I should probably leave that for another course[36]. This much we can say, in the absence of current the vector potential must have component functions which solve Laplace's equation $\nabla^2 A_i = 0$ for $i = 1, 2, 3$. This means:

1. if the vector potential is constant along the boundary of a current-free simply connected region then the vector-potential is constant throughout the region.

2. by Theorem 7.9.6 we can solve for $\vec{A}$ by calculating the appropriate integral of the current density! (in the magnetostatic case where $\frac{\partial \phi}{\partial t} = 0$). Identify that $b = -\mu_o J_k$ hence for $k = 1, 2, 3$ hence, applying the Theorem three times:

$$A_1(\vec{r}) = \iiint_M \frac{\mu_o J_1(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||}dV'$$

[36] those notes are already posted, see my Math 430 notes, the whole goal of that course is to analyze Maxwell's Equations in differential forms

$$A_2(\vec{r}) = \iiint_M \frac{\mu_o J_2(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||} dV'$$

$$A_3(\vec{r}) = \iiint_M \frac{\mu_o J_3(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||} dV'$$

Consequently, we can write the single vector integral to capture the vector potential:

$$\boxed{\vec{A}(\vec{r}) = \iiint_M \frac{\mu_o \vec{J}(\vec{r}\,')}{4\pi||\vec{r}-\vec{r}\,'||.} dV'}$$

7.11 Problems

Problem 178 Let $\vec{F}(x,y,z) = \langle x+z, x+y^2, z^3\rangle$ calculate $\nabla \bullet \vec{F}$ and $\nabla \times \vec{F}$.

Problem 179 Suppose a,b,c are constants. Let $\vec{G}(x,y,z) = \langle x+z+a, x+y^2+b, z^3+c\rangle$ calculate $\nabla \bullet \vec{G}$ and $\nabla \times \vec{G}$.

Problem 180 Find a function f and a vector field $\vec{A}$ such that $\vec{F} = \nabla f + \nabla \times \vec{A}$ where $\vec{F}$ is the vector field studied in problem 178.

Problem 181 Suppose $\nabla \times \vec{F} = \nabla \times \vec{G}$. Does it follow that $\vec{F} = \vec{G}$?

Problem 182 Suppose $\nabla \bullet \vec{F} = \nabla \bullet \vec{G}$. Does it follow that $\vec{F} = \vec{G}$?

Problem 183 Suppose $\nabla \times \vec{F} = \nabla \times \vec{G}$ and $\nabla \bullet \vec{F} = \nabla \bullet \vec{G}$ make a conjecture: does it follow that $\vec{F} = \vec{G} + \vec{c}$ for some constant vector $\vec{c}$? (no work required if your answer is yes, however if your answer is no then I would like for you to provide a counter-example)

Problem 184 Show that

$$\nabla \bullet (\vec{A} \times \vec{B}) = \vec{B} \bullet (\nabla \times \vec{A}) - \vec{A} \bullet (\nabla \times \vec{B}).$$

Problem 185 Given $\nabla \bullet \vec{E} = \rho/\epsilon_o$ and $\nabla \times \vec{B} = \mu_o \vec{J} + \mu_o \epsilon_o \frac{\partial \vec{E}}{\partial t}$ show that $\nabla \bullet \vec{J} = \frac{\partial \rho}{\partial t}$. If $\vec{J}$ is the charge per unit area time following in the area with direction $\widehat{J}$ and ρ is the charge per unit volume then what does this equation mean physically speaking?

Problem 186 Calculate $\nabla(\vec{A} \bullet (\vec{B} \times \vec{C}))$ where $\vec{A}, \vec{B}, \vec{C}$ are all smooth vector fields.

Problem 187 $\displaystyle\int_C 3x^2yz\, ds$ where C is the curve parameterized by $\mathbf{X}(t) = \left(t, t^2, \frac{2}{3}t^3\right)$ and $0 \leq t \leq 1$.

Problem 188 Find the centroid of the curve C: the upper-half of the unit circle plus the x-axis from -1 to 1.
Hint: Use geometry and symmetry to compute 2 of the 3 line integrals.

Problem 189 Find the centroid of the helix C parameterized by $\mathbf{X}(t) = (2\sin(t), 2\cos(t), 3t)$ where $0 \leq t \leq 2\pi$.

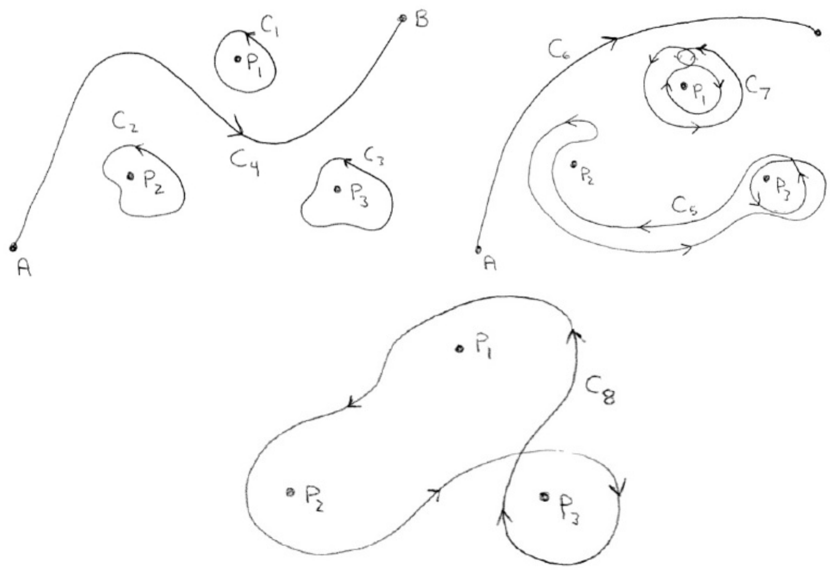
Problem 190 Let $\vec{F} = \langle z, y, x\rangle$. Calculate $\int_C \vec{F} \bullet d\vec{r}$ for the line-segment C from $(1,1,1)$ to $(3,4,5)$.

Problem 191 Let $\vec{F} = \langle 0, 0, -mg\rangle$ where m, g are positive constants. Find the work done as you travel up the helix $\vec{r}(t) = \langle R\cos(t), R\sin(t), t\rangle$ for $0 \leq t \leq 4\pi$.

Problem 192 Let $\vec{F} = \langle 0, 0, -mg \rangle$ where m, g are positive constants and suppsoe $\vec{F}_f = -b\vec{T}$ where v is your speed and b is a constant and $\vec{T}$ is the unit-vector which points along the tangential direction of the path. This is a simple model of the force of kinetic friction, it just acts opposite your motion. Find the work done by $\vec{F}_f + \vec{F}$ as you travel up the helix $\vec{r}(t) = \langle R\cos(t), R\sin(t), t \rangle$ for $0 \leq t \leq 4\pi$.

Problem 193 Let $\vec{F}(x, y) = \langle 1 + y, -x + 2 \rangle$. Let C be the ellipse $x^2/4 + y^2/9 = 1$ given a CCW orientation. Calculate $\int \vec{F} \bullet d\vec{r}$.

Problem 194 Let $\mathbf{F}(x, y) = (P(x, y), Q(x, y))$ be a vector field where $P_y = Q_x$ except at the points P_1, P_2, and P_3.
Suppose that $\displaystyle\int_{C_1} \mathbf{F} \cdot d\mathbf{X} = 1$, $\displaystyle\int_{C_2} \mathbf{F} \cdot d\mathbf{X} = 2$, $\displaystyle\int_{C_3} \mathbf{F} \cdot d\mathbf{X} = 3$, and $\displaystyle\int_{C_4} \mathbf{F} \cdot d\mathbf{X} = 4$ where C_1, C_2, $\ldots C_8$ are pictured below:.



Compute:

(a.) $\displaystyle\int_{C_5} \mathbf{F} \cdot d\mathbf{X}$ [Answer: 6]

(b.) $\displaystyle\int_{C_6} \mathbf{F} \cdot d\mathbf{X}$

(c.) $\displaystyle\int_{C_7} \mathbf{F} \cdot d\mathbf{X}$

(d.) $\displaystyle\int_{C_8} \mathbf{F} \cdot d\mathbf{X}$

Problem 195 Determine if the vector fields below are conservative. Find potential functions where possible.

(a.) $\mathbf{F}(x, y) = \left(x^2 + y^2, xy\right)$

(b.) $\mathbf{F}(x, y) = \left(2x + 3x^2y^2 + 5, 2x^3y\right)$

(c.) $\mathbf{F}(x,y) = \left(e^x, ye^{-y^2}\right)$

Problem 196 Determine if the vector fields below are conservative. Find potential functions where possible.

(a.) $\mathbf{F}(x,y) = \left(e^{xy}, x^4y^3 + y\right)$

(b.) $\mathbf{F}(x,y) = \left(e^x + \dfrac{y}{1+x^2}, \arctan(x) + (1+y)e^y\right)$

(c.) $\mathbf{F}(x,y,z) = (yz + y + 1, xz + x + z, xy + y + 1)$

Problem 197 Let $\vec{E}(x,y,z) = \frac{1}{(x^2+y^2+z^2)^{(3/2)}}\langle x,y,z\rangle$. Calculate $\nabla \bullet \vec{E}$ and $\nabla \times \vec{E}$. (use cartesians or spherical coordinates, your choice (the formulas for divergence and curl in sphericals are contained within the pdf on "VectorFieldDifferentiation" in course content). Plot this vector field (don't have to turn in plot, I trust you to do it), are your calculations surprising?

Problem 198 Let C be the square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$ (oriented counter-clockwise). Compute the line integral: $\displaystyle\int_C y^2\,dx + x^2\,dy$ two ways. First, compute the integral directly by parameterizing each side of the square. Then, compute the answer again using Green's Theorem.

Problem 199 Compute $\displaystyle\int_C \mathbf{F}\cdot d\mathbf{X}$ where $\mathbf{F}(x,y) = (y - \ln(x^2+y^2), 2\arctan(y/x))$ and C is the circle $(x-2)^2 + (y-3)^2 = 1$ oriented counter-clockwise.

Problem 200 prove the other half of Green's theorem. (I only proved half in lecture).

Problem 201 Calculate the integral below. Define C to be the CCW oriented triangle with vertices $(0,0,),(0,1)$ and $(1,1)$.

$$\oint_C (xydx + x^2dy)$$

Problem 202 Let C be the CCW oriented circle $x^2 + y^2 = 1$. Calculate

$$\oint_C (3ydx + 5xdy)$$

Problem 203 Let C be the CCW oriented rectangle with vertices $(0,0),(a,0),(a,b)$ and $(0,b)$. Calculate

$$\oint_C x^2dy$$

Problem 204 Suppose that f,g are continuously differentiable on an simply connected open set R. Show that if C is any piecewise-smooth simple closed curve in R then

$$\oint \left[f(\vec{r})(\nabla g)(\vec{r}) + g(\vec{r})(\nabla f)(\vec{r})\right] \bullet d\vec{r} = 0$$

Problem 205 Show that for a simply connected region R with consistently oriented boundary ∂R if f,g are differentiable on some open set containing R then

$$\iint_R \left(f\nabla^2 g + \nabla f \bullet \nabla g\right)dA = \int_{\partial R} f\nabla g \bullet \widehat{n}\,ds.$$

Problem 206 Show that for a simply connected region R with consistently oriented boundary ∂R if f, g are differentiable on some open set containing R then

$$\iint_R \big(f\nabla^2 g - g\nabla^2 f\big)dA = \int_{\partial R} \big[f\nabla g \bullet \widehat{n} - g\nabla f \bullet \widehat{n}\big]\, ds.$$

Problem 207 Suppose $\nabla^2 f = 0$ on a simply connected region R. If $f|_{\partial R} = 0$ then what can you say about f throughout R?
(here $|_{\partial R}$ denotes restriction of f to the subset ∂R. In particular this means you are given that $f(x,y) = 0$ for all $(x,y) \in \partial R$.)

Problem 208 Suppose $b : \mathbb{R}^2 \to \mathbb{R}$ is a particular function and $\nabla^2 f = b$ on a simply connected region R. If g is another such solution ($\nabla^2 g = b$) on R then show that $f = g$ on R.
The equation $\nabla^2 f = b$ is called Poisson's Equation. When $b = 0$ then it's called Laplace's Equation. You are showing the solution to Poisson's Equation is unique on a simply connected region. **Hint:** *use the last problem's result on $f - g$... hmmm... I guess this is a retroactive hint for Problem 207 if you think about it.*

Problem 209 Find the surface area of $z = xy$ for $x^2 + y^2 \leq 1$.

Problem 210 Find the surface area of the plane $y + 2z = 2$ bounded by the cylinder $x^2 + y^2 = 1$.

Problem 211 Find the surface area of the cone frustrum $z = \frac{1}{3}\sqrt{x^2 + y^2}$ with $1 \leq z \leq 4/3$

Problem 212 Find the surface area of torus with radii $A, R > 0$ and $R \geq A$ parametrized by

$$\vec{X}(\alpha, \beta) = \Big\langle \big[R + A\cos(\alpha)\big]\cos(\beta), \big[R + A\cos(\alpha)\big]\sin(\beta), A\sin(\alpha) \Big\rangle$$

for $0 \leq \alpha \leq 2\pi$ and $0 \leq \beta \leq 2\pi$.

Problem 213 Find an explicit double integral which gives the surface area of the graph $x = g(y,z)$ for $(y,z) \in D$.

Problem 214 Consider a napkin ring which is formed by taking a sphere of radius R and drilling out a circular cylinder of radius B through the center of the sphere. Find the surface area of the napkin ring (include the inner as well as outer surfaces).

Problem 215 Integrate $H(x,y,z) = xyz$ over the surface of the solid $[0,a] \times [0,b] \times [0,c]$ where $a, b, c > 0$.

Problem 216 Integrate $G(x,y,z) = x^2$ over the surface of the unit-sphere.

Problem 217 Integrate $H(x,y,z) = z - x$ on the graph $z = x + y^2$ over the triangular region with vertices $(0,0,0), (1,0,0)$ and $(0,1,0)$.

Problem 218 Consider a thin-shell of constant density δ. Let the shell be cut from the cone $x^2 + y^2 - z^2 = 0$ by the planes $z = 1$ and $z = 2$. Find **(a.)** the center of mass and **(b.)** the moment of inertia with respect to the z-axis.

Problem 219 Find the flux of $\vec{F}(x,y,z) = \langle z^2, x, -3z\rangle$ through the parabolic cylinder $z = 4 - y^2$ bounded by the planes $x = 0$, $x = 1$ and $z = 0$. Assume the orientation of the surface is outward, away from the x-axis.

Problem 220 Find the flux of $\vec{F}(x,y,z) = z\,\widehat{z}$ through the portion the sphere of radius R in the first octant. Give the sphere an orientation which points away from the origin. In other words, assume the sphere is outwardly oriented.

Problem 221 Find the flux of $\vec{F}(x,y,z) = \langle -x, -y, z^2 \rangle$ through the conical frustrum $z = \sqrt{x^2+y^2}$ between the planes $z = 1$ and $z = 2$ with outward orientation.

Problem 222 Let S be the outward oriented paraboloid $z = 6 - x^2 - y^2$ for $x, z \geq 0$. Calculate the flux of $\vec{F} = (x^2+y^2)\,\widehat{z}$

Problem 223 Find the flux of $\vec{F}(x,y,z) = \langle 2xy, 2yz, 2xz \rangle$ upward through the subset of the plane $x + y + z = 2c$ where $(x,y) \in [0,c] \times [0,c]$.

Problem 224 Suppose $\vec{C}$ is a constant vector. Let $\vec{F}(x,y,z) = \vec{C}$ find the flux of $\vec{F}$ through a surface S on plane with nonzero vectors $\vec{A}, \vec{B}$. In particular, the surface S is parametrized by $\vec{r}(u,v) = \vec{r}_o + u\vec{A} + v\vec{B}$ for $(u,v) \in \Omega$.

Problem 225 Let $\vec{F}(x,y,z) = \langle a, b, c \rangle$ for some constants a, b, c. Calculate the flux of $\vec{F}$ through the upper-half of the outward oriented sphere $\rho = R$.

Problem 226 Once more consider the constant vector field $\vec{F}(x,y,z) = \langle a, b, c \rangle$. Calculate the flux of $\vec{F}$ through the downward oriented disk $z = 0$ for $\phi = \pi/2$.

Problem 227 Let $\vec{F} = \langle x^2, y^2, z^2 \rangle$. Calculate the flux of $\vec{F}$ through $z = 4 - x^2 - y^2$ for $z \geq 0$.

Problem 228 Let $\vec{F} = \langle x^2, y^2, z^2 \rangle$. Calculate the flux of $\vec{F}$ through the downward oriented disk $x^2 + y^2 \leq 4$ with $\phi = \pi/2$.

Problem 229 Let $\phi = \pi/4$ define a closed surface S with $0 \leq \rho \leq 2$. Find the flux of

$$\vec{F}(\rho,\phi,\theta) = \phi^2\widehat{\rho} + \rho\widehat{\phi} + \widehat{\theta}$$

through the outward oriented S.

Problem 230 Consider the closed cylinder $x^2 + y^2 = R^2$ for $0 \leq z \leq L$. Find the flux of

$$\vec{F}(r,\theta,z) = \theta\,\widehat{z} + z\widehat{\theta} + r^2\widehat{r}$$

out of the cylinder.

Problem 231 Let S be the pseudo-tetrahedra with vertices $(0,0,0)$, $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$. Three of the faces of S_1 are subsets of the coordinate planes call these S_{xy}, S_{zx}, S_{yz} with the obvious meanings and call S_T the top face. Let $\vec{F} = \langle y, -x, y \rangle$ and define $\vec{G} = \nabla \times \vec{F}$.

(a.) Calculate the circulation of $\vec{F}$ around each face .

(b.) Calculate the flux of $\vec{G}$ through each face.

(c.) do you see any pattern?

Problem 232 Let $\vec{F}(x,y,z) = \langle x^2, 2x, z^2 \rangle$. Calculate $\int_E \vec{F} \bullet d\vec{r}$ where E is the CCW oriented ellipse $4x^2 + y^2 = 4$ with $z = 0$. (use Stokes' Theorem)

Problem 233 Let $\vec{F}(x,y,z) = \langle y^2 + z^2, x^2 + z^2, x^2 + y^2 \rangle$. Find the work done by $\vec{F}$ around the CCW (as viewed from above) triangle formed from the intersection of the plane $x + y + z = 1$ and the coordinate planes. (use Stokes' Theorem)

Problem 234 Let $\vec{F}(x,y,z) = \langle y^2+z^2, x^2+y^2, x^2+y^2\rangle$. Find the work done by $\vec{F}$ around the CCW-oriented square bounded by $x = \pm 1$ and $y = \pm 1$ in the $z = 0$ plane (use Stokes' Theorem).

Problem 235 Consider the elliptical shell $4x^2 + 9y^2 + 36z^2 = 36$ with $z \geq 0$ and let

$$\vec{F}(x,y,z) = \left\langle y, x^2, (x^2+y^4)^{\frac{3}{2}} \sin(exp(\sqrt{xyz}))\right\rangle.$$

Find the flux of $\nabla \times \vec{F}$ through the outwards oriented shell.

Problem 236 Let $\vec{F} = \langle 2x, 2y, 2z\rangle$ and suppose S is a simply connected surface with boundary ∂S a simple closed curve. Show by Stokes' theorem that $\int_{\partial S} \vec{F} \bullet d\vec{r} = 0$

Problem 237 Suppose S is the union of the cylinder $x^2 + y^2 = 1$ for $0 \leq z \leq 1$ and the disk $x^2 + y^2 \leq 1$ at $z = 1$. Suppose $\vec{F}$ is a vector field such that

$$\nabla \times \vec{F} = \left\langle \sinh(z)(x^2+y^2), ze^{xy+\cos(x+y)}, (xz+y)\tan^{-1}(z)\right\rangle.$$

Calculate the flux of $\nabla \times \vec{F}$ though S.

Problem 238 Let E be the cube $[-1,1]^3$. Calculate the flux through ∂E of the vector field

$$\vec{F}(x,y,z) = \langle y-x, z-y, y-x\rangle$$

(please use the divergence theorem!)

Problem 239 Let E be the set of (x,y,z) such that $x^2 + y^2 \leq 4$ and $0 \leq z \leq x^2 + y^2$. Find the flux through ∂E of the vector field $\vec{F}$ given below:

$$\vec{F}(x,y,z) = \langle y, xy, -z\rangle$$

(please use the divergence theorem)

Problem 240 Suppose E is the spherical shell $R_1 \leq \rho \leq R_2$ and suppose $\vec{F}(x,y,z) = \nabla \times \vec{A}$ for some everywhere smooth vector field $\vec{A}$. Show that the flux through $\rho = R_1$ is the same as the flux through $\rho = R_2$ by applying the divergence theorem to the spherical shell.

Note: I assigned 5 problems from Briggs, Cochrane and Gillett in Fall 2013 semester. This is the reason for the number jump here.

Problem 246 We gave definitions for curl and divergence which were based in cartesian coordinates. Some authors actually use the identities below to define curl and divergence. Naturally, if you use these as definitions then the question of what div and curl mean are easily answered. However, on the other hand, in that approach you have no simple formula to calculate curl or div until you have mastered both surface and line integrals. I wanted to talk about curl and div before that point so for that reason I did not take these as definitions.

(a.) Assume E is a volume with piecewise smooth, outward oriented, boundary ∂E where E contains the point P. Then if we shrink the volume down to P we obtain the divergence of a differentiable $\vec{F}$ as follows:

$$div(\vec{F})(P) = \lim_{V \to 0^+} \frac{1}{V} \iint_{\partial E} \vec{F} \bullet d\vec{S}.$$

Show the formula above is true by an argument involving the divergence theorem.

(b.) Assume S is a surface with piecewise smooth, consistently oriented, boundary ∂S where E contains the point P. Then if we shrink the surface to P we obtain the curl of the the vector field in the direction of the normal $\widehat{n}$ to S at P as follows:

$$\left[curl(\vec{F})(P) \right] \bullet \widehat{n} = \lim_{A \to 0^+} \frac{1}{A} \oint_{\partial S} \vec{F} \bullet d\vec{r}$$

Problem 247 Suppose we have a vector field expressed in cylindrical coordinates; $\vec{F} = F_r \widehat{r} + F_\theta \widehat{\theta} + F_z \widehat{z}$. Calculate the formulas for

(a.) $div(\vec{F})$

(b.) $curl(\vec{F})$

By applying the boxed formulas of Problem 246 to appropriate volumes and loops. For the divergence you want to think about a little part of a cylinder which corresponds to a small change in r, θ and z. For the curl you want to think about three loops. To pick out the z-component you should fix z and allow r and θ to sweep out a little sector. I'll draw the pictures for you before class.

Problem 248 Suppose we have a vector field expressed in spherical coordinates; $\vec{F} = F_\rho \widehat{\rho} + F_\phi \widehat{\phi} + F_\theta \widehat{\theta}$. Calculate the formulas for

(a.) $div(\vec{F})$

(b.) $curl(\vec{F})$

By applying the boxed formulas of Problem 246 to appropriate volumes and loops.

Note: if you look at my handwritten calculus III notes then you'll find direct, explicit, arguments which show the non-cartesian formulas for divergence, curl and gradient follow from the cartesian formulas via the chain rule and orthonormal frame arguments. Trust me when I tell you the arguments I outlined in the last three problems are far more efficient.

Problem 249 Suppose $\vec{J} = \sigma \vec{E}$ (this is Ohm's Law for current density, the constant σ is the **conductivity**. Show that Maxwell's equations yield the equation below:

$$\nabla^2 \vec{E} = \mu_o \sigma \frac{\partial \vec{E}}{\partial t} + \mu_o \epsilon_o \frac{\partial^2 \vec{E}}{\partial t^2}$$

also, show the same equation holds for $\vec{B}$. This is called the **telegrapher's equation**.

Problem 250 Fun with differential forms: let $\omega = y^2 z^3 dx + 2xyz^3 dy + 3xy^2 z^2 dz = 0$. Show that $\omega = 0$ is an exact equation on $\mathbb{R}^3$ by showing the **exterior derivative** $d\omega = 0$ on all of $\mathbb{R}^3$. Also, find the potential form for ω; that is, find ϕ such that $\omega = d\phi$. (technically this problem belongs to the next Chapter, but I'll leave it here for now)

Chapter 8

Exterior Calculus

8.1 algebra of wedge products

For the sake of generality let us begin this discussion in $\mathbb{R}^n$ with the standard basis $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ with corresponding differentials $dx_1, dx_2, \dots, dx_n$. A function $f : \mathbb{R}^n \to \mathbb{R}$ has **total differential** df given by

$$df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \cdots + \frac{\partial f}{\partial x_n}dx_n.$$

An object like the one given above is known as a **one-form** or **differential one-form**. More generally,

$$\alpha = \sum_{i=1}^{n} \alpha_i dx_i$$

where $\alpha_i : \mathbb{R}^n \to \mathbb{R}$ are functions is a **one-form**. We introduce an algebra for differential forms known as the **wedge product**. This is similar to the cross product, but differs in several ways. First, the wedge product is defined for $\mathbb{R}^n$ for any $n \in \mathbb{N}$ whereas the cross product is special to $\mathbb{R}^3$. Second, the cross product is not associative whereas the wedge product is associative. Third, the wedge product is an **exterior product** in the sense that when we take the wedge product of differential forms we usually obtain a new type of object. In contrast, the cross product of vector fields is once more a vector field. A **two-form** is formed from a sum of terms of the form $dx_i \wedge dx_j$ for $1 \leq i, j \leq n$. In particular,

$$dx_i \wedge dx_j = -dx_j \wedge dx_i$$

hence $dx_i \wedge dx_i = -dx_i \wedge dx_i$ which reveals $2dx_i \wedge dx_i = 0$. Generally,

$$\beta = \sum_{i,j=1}^{n} \frac{1}{2}\beta_{ij} dx_i \wedge dx_j = \sum_{i<j} \beta_{ij} dx_i \wedge dx_j$$

where $\beta_{ij} = -\beta_{ji}$ so the terms in the first sum are not independent[1]. A **three-form** is an expression of the form

$$\gamma = \sum_{i,j,k} \frac{1}{6}\gamma_{ijk} dx_i \wedge dx_j \wedge dx_k = \sum_{i<j<k} \gamma_{ijk} dx_i \wedge dx_j \wedge dx_k.$$

[1]. Sometimes it is convenient to use components with respect to a **basis** and insisting $i < j$ enforces the desired independence. Showing the equality above holds true is often assigned as an exercise in my Advanced Calculus course.

In general, a p-**form** has the form

$$\sigma = \sum_{i_1,\dots,i_p} \frac{1}{p!}\sigma_{i_1,\dots,i_p} dx_{i_1} \wedge \cdots \wedge dx_{i_p} = \sum_{i_1<\cdots<i_p} \sigma_{i_1,\dots,i_p} dx_{i_1} \wedge \cdots \wedge dx_{i_p}.$$

In $\mathbb{R}^n$ the form of largest degree is known as the **top-form** it is given by a single coefficient function multiplied by the wedge product of all the differentials; $\gamma = g dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$. Any form of higher degree than n is necessarily zero in $\mathbb{R}^n$ since $n+1$ differentials taken from $dx_1, \dots, dx_n$ necessarily has a repeated differential and the wedge product of a differential with itself is zero. Let us present the properties of the wedge product as a definition:

Definition 8.1.1. *let α, β, γ be differential forms of degrees p, q, r respective and c a constant,*

(i.) $\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma$,

(ii.) $(c\alpha + \beta) \wedge \gamma = c\alpha \wedge \gamma + \beta \wedge \gamma$,

(iii.) $\alpha \wedge \beta = (-1)^{pq} \beta \wedge \alpha$.

Example 8.1.2. *Let us examine a calculation in $\mathbb{R}^4_{txyz}$ which has differentials dt, dx, dy, dz. Consider $\alpha = adt \wedge dx + bdy \wedge dz$ and $\beta = cdt \wedge dy + edy \wedge dz$ where a, b, c, e are real valued functions on $\mathbb{R}^4$.*

$$\begin{aligned} \alpha \wedge \beta &= (adt \wedge dx + bdy \wedge dz) \wedge (cdt \wedge dy + edy \wedge dz) \\ &= acdt \wedge dx \wedge dt \wedge dy + aedt \wedge dx \wedge dy \wedge dz + bcdy \wedge dz \wedge dt \wedge dy + bedy \wedge dz \wedge dy \wedge dz \\ &= aedt \wedge dx \wedge dy \wedge dz. \end{aligned}$$

Likewise,

$$\begin{aligned} \beta \wedge \alpha &= (cdt \wedge dy + edy \wedge dz) \wedge (adt \wedge dx + bdy \wedge dz) \\ &= cadt \wedge dy \wedge dt \wedge dx + cbdt \wedge dy \wedge dy \wedge dz + eady \wedge dz \wedge dt \wedge dx + ebdy \wedge dz \wedge dy \wedge dz \\ &= eady \wedge dz \wedge dt \wedge dx \\ &= aedt \wedge dx \wedge dy \wedge dz. \end{aligned}$$

We see that $\alpha \wedge \beta = \beta \wedge \alpha$. Furthermore,

$$\begin{aligned} \alpha \wedge \alpha &= (adt \wedge dx + bdy \wedge dz) \wedge (adt \wedge dx + bdy \wedge dz) \\ &= abdt \wedge dx \wedge dy \wedge dz + bady \wedge dz \wedge dt \wedge dx \\ &= 2abdt \wedge dx \wedge dy \wedge dz. \end{aligned}$$

In contrast, if we take any form of odd degree γ then $\gamma \wedge \gamma = -\gamma \wedge \gamma$ thus $\gamma \wedge \gamma = 0$. For example, $\gamma = dt + dy$ has

$$\gamma \wedge \gamma = (dt + dy) \wedge (dt + dy) = dt \wedge dt + dt \wedge dy + dy \wedge dt + dy \wedge dy = dt \wedge dy - dt \wedge dy = 0.$$

In short, the algebra of the wedge product resembles the usual arithmetic with a slight adjustment to reflect the fact that the wedge product is not generally commutative.

There is a geometric meaning to the wedge product; roughly speaking the wedge product of k-one forms corresponds to the k-volume of the object with sides correpsonding to the one-forms. For example, ldx and wdy are one-forms where $ldx \wedge wdy = lwdx \wedge dy$ and the area $A = lw$ for the rectangle with length l and width w. Or, consider a parallelogram with sides bdx and $adx + hdy$ then $bdx \wedge (adx + hdy) = bhdx \wedge dy$. The area is the base b times the height h of the parallelogram. These examples are a little silly, perhaps we will investigate deeper illustrations in the homework. Ultimately we must understand the wedge product and the determinant are both intricately tied to the calculation of volume in the generalized sense.

Theorem 8.1.3. *Determinants and wedge products:*

Let $v_1, \dots, v_n$ be vectors in $\mathbb{R}^n$ and let us extend the wedge product to vectors then

$$v_1 \wedge \cdots \wedge v_n = det(v_1|\cdots|v_n)\hat{x}_1 \wedge \cdots \wedge \hat{x}_n.$$

The wedge product of vectors or vector fields are known as **multivectors**. Another useful result showcases why wedge products allow more nuance than determinants alone:

Theorem 8.1.4. *Wedge products and linear dependence:*

Let $v_1, \dots, v_k$ be vectors in $\mathbb{R}^n$ then

$$v_1 \wedge \cdots \wedge v_k = 0$$

if and only if the vectors $v_1, \dots, v_k$ are linearly dependent.

For example, in $\mathbb{R}^3$ if $v \wedge w = 0$ then the vectors v, w are **colinear**. Or, again in $\mathbb{R}^3$, if $u \wedge v \wedge w = 0$ then u, v, w are **coplanar**. Much more can be said about multivectors, but we will focus on differential forms in this chapter.

There is a natural correspondence between vector fields in $\mathbb{R}^n$ and one-forms in $\mathbb{R}^n$.

Definition 8.1.5. *work form correspondence*

If $\vec{F} = \langle F_1, \dots, F_n \rangle$ then define the **work form** corresponding to $\vec{F}$ by:

$$\omega_{\vec{F}} = F_1 dx_1 + \cdots + F_n dx_n.$$

It is simple to verify that $\omega_{\vec{F}+\vec{G}} = \omega_{\vec{F}} + \omega_{\vec{G}}$ and $\omega_{g\vec{F}} = g\omega_{\vec{F}}$ for vector fields $\vec{F}, \vec{G}$ and function g. Theorem 8.1.4 immediately implies the following as well:

Theorem 8.1.6. *Determinants and wedge products:*

Let $\vec{F}_1, \dots, \vec{F}_n$ be vector fields in $\mathbb{R}^n$ then

$$\omega_{\vec{F}_1} \wedge \cdots \wedge \omega_{\vec{F}_n} = det\left(\vec{F}_1|\cdots|\vec{F}_n\right) dx_1 \wedge \cdots \wedge dx_n.$$

There is more to say about the wedge product algebra in $\mathbb{R}^n$, but first we should examine how differential forms work in $\mathbb{R}^3$.

8.2 exterior algebra of three dimensional space

The set of all differential forms on $\mathbb{R}^3$ is denoted $\Omega\mathbb{R}^3$ it includes p-forms where $p = 0, 1, 2, 3$. To begin a **zero form** is simply a function $f : \mathbb{R}^3 \to \mathbb{R}$. A **one-form** has the form

$$\alpha = a\,dx + b\,dy + c\,dz$$

where a, b, c are real valued functions on $\mathbb{R}^3$ and $\alpha \in \Lambda^1\mathbb{R}^3$. A **two-form** is is given by

$$\beta = a\,dy \wedge dz + b\,dz \wedge dx + c\,dx \wedge dy$$

and we say $\beta \in \Lambda^2\mathbb{R}^3$. Finally, a **three-form** has the form

$$\gamma = g\,dx \wedge dy \wedge dz$$

and we say $\gamma \in \Lambda^3\mathbb{R}^3$. In short, any differential form in $\mathbb{R}^3$ is built as a linear combination of 1, dx, dy, dz, $dy \wedge dz, dz \wedge dx, dx \wedge dy$ and $dx \wedge dy \wedge dz$ with coefficients which are functions on $\mathbb{R}^3$. There are two natural correspondences with vector fields:

Definition 8.2.1. *work form and flux form correspondences*

Let $\vec{F} = \langle a, b, c\rangle$ be a vector field on $\mathbb{R}^3$ then

$$\omega_{\vec{F}} = a\,dx + b\,dy + c\,dz \qquad \& \qquad \Phi_{\vec{F}} = a\,dy \wedge dz + b\,dz \wedge dx + c\,dx \wedge dy$$

define the **work form** $\omega_{\vec{F}}$ and the **flux form** $\Phi_{\vec{F}}$ corresponding to $\vec{F}$.

Proposition 8.2.2. *Linearity of the work and flux forms:*

Let $\vec{F}, \vec{G}$ be vector fields and c a function on $\mathbb{R}^3$ then

$$\omega_{c\vec{F}+\vec{G}} = c\,\omega_{\vec{F}} + \omega_{\vec{G}} \qquad \& \qquad \Phi_{c\vec{F}+\vec{G}} = c\,\Phi_{\vec{F}} + \Phi_{\vec{G}}.$$

Notice that both ω and Φ defined by $\vec{F} \mapsto \omega_{\vec{F}}$ and $\vec{F} \mapsto \Phi_{\vec{F}}$ are invertible maps. Notice

$$\omega^{-1}(adx + bdy + cdz) = \langle a, b, c\rangle \qquad \& \qquad \Phi^{-1}(a\,dy \wedge dz + b\,dz \wedge dx + c\,dx \wedge dy) = \langle a, b, c\rangle.$$

Thus we have freedom to formulate models based on vector fields either using one-forms or two-forms. Such data is interchangeable. What is the significance of the wedge product here ? Let's calculate:

$$\begin{aligned}
\omega_{\vec{A}} \wedge \omega_{\vec{B}} &= (A_1dx + A_2dy + A_3dz) \wedge (B_1dx + B_2dy + B_3dz) \\
&= A_1B_2dx \wedge dy + A_1B_3dx \wedge dz + A_2B_3dy \wedge dz + A_2B_1dy \wedge dx + A_3B_1dz \wedge dx + A_3B_2dz \wedge dy \\
&= (A_1B_2 - A_2B_1)dx \wedge dy + (A_3B_1 - A_1B_3)dz \wedge dx + (A_2B_3 - A_3B_2)dy \wedge dz \\
&= \Phi_{\langle A_2B_3 - A_3B_2, A_3B_1 - A_1B_3, A_1B_2 - A_2B_1\rangle} \\
&= \Phi_{\vec{A}\times\vec{B}}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\omega_{\vec{A}} \wedge \Phi_{\vec{B}} &= (A_1dx + A_2dy + A_3dz) \wedge (B_1dy \wedge dz + B_2dz \wedge dx + B_3dx \wedge dy) \\
&= A_1B_1dx \wedge dy \wedge dz + A_2B_2dy \wedge dz \wedge dx + A_3B_3dz \wedge dx \wedge dy \\
&= (A_1B_1 + A_2B_2 + A_3B_3)dx \wedge dy \wedge dz \\
&= \vec{A} \bullet \vec{B}\, dx \wedge dy \wedge dz.
\end{aligned}$$

On the other hand, $\Phi_{\vec{A}} \wedge \Phi_{\vec{B}} = 0$ since it necessarily includes a repeated differential in every term. I'll give a transitional definition of Hodge duality on Euclidean three dimensional space as this is the easiest way to begin the discussion:

Definition 8.2.3. *Hodge duality*

In $\mathbb{R}^3$ we define $\star 1 = dx \wedge dy \wedge dz$ and $\star dx \wedge dy \wedge dz = 1$. Furthermore, $\star dx = dy \wedge dz$ and $\star dy = dz \wedge dx$ and $\star dz = dx \wedge dy$. Then we extend $\star$ linear to all other differential forms.

In particular, we have for functions f, g and vector field $\vec{F}$,

$$\star f = f dx \wedge dy \wedge dz \qquad \star(g dx \wedge dy \wedge dz) = g \qquad \star\,\omega_{\vec{F}} = \Phi_{\vec{F}} \qquad \star\,\Phi_{\vec{F}} = \omega_{\vec{F}}$$

In summary, $\star : \Lambda^p\mathbb{R}^3 \to \Lambda^{3-p}\mathbb{R}^3$ for $p = 0, 1, 2, 3$ and evidentally $\star\star\,\sigma = \sigma$ for every differential form on $\mathbb{R}^3$. It turns out it is very helpful to characterize the Hodge dual in relation to the top form $dx \wedge dy \wedge dz$. I'll illustrate with an example.

Example 8.2.4. *Let $\sigma = adx + bdy$ then $\star\sigma = ady \wedge dz + bdz \wedge dx$. Then,*

$$\sigma \wedge \star\sigma = (adx + bdy) \wedge (ady \wedge dz + bdz \wedge dx) = (a^2 + b^2)dx \wedge dy \wedge dz.$$

Likewise, $\beta = adx \wedge dy + bdy \wedge dz$ and $\star\beta = adz + bdx$ hence

$$\beta \wedge \star\beta = (adx \wedge dy + bdy \wedge dz) \wedge (adz + bdx) = (a^2 + b^2)dx \wedge dy \wedge dz.$$

We define the **length** of a differential form via; $\|\omega_{\vec{F}}\| = \|\vec{F}\|$ and $\|\Phi_{\vec{F}}\| = \|\vec{F}\|$ whereas $\|f\| = |f|$ and $\|gdx \wedge dy \wedge dz\| = |g|$. Then notice the preceding example reveals

$$\sigma \wedge \star\sigma = \|\sigma\|^2 dx \wedge dy \wedge dz \qquad \& \qquad \beta \wedge \star\beta = \|\beta\|^2 dx \wedge dy \wedge dz.$$

The formulas above are not accidental and it turns out they suffice to implicitly define the Hodge dual. In other words, an alternative definition of Hodge duality is simply:

Definition 8.2.5. *Hodge duality*

Given a p-form σ there exists a unique $3 - p$ form $\star\sigma$ such that

$$\sigma \wedge \star\sigma = \|\sigma\|^2 dx \wedge dy \wedge dz$$

where $\|\sigma\|$ is the Euclidean length of σ.

We will soon see how the formulation above very useful for the calculation of the Hodge dual in non-Cartesian coordinate systems such as cylindrical or spherical coordinates.

8.2.1 cylindrical and spherical coframes for $\mathbb{R}^3$

Consider cylindrical coordinates r, θ, z given by

$$x = r\cos\theta, \quad y = r\sin\theta, \quad z = z$$

If we define dl_u to denote the infinitesimal change in length when the u-coordinate is changed whilst holding its complementary coordinates fixed then we may discern through geometric argument that:

$$dl_r = dr, \quad dl_\theta = rd\theta, \quad dl_z = dz \tag{8.1}$$

Likewise, for spherical coordinates, ρ, ϕ, θ given by

$$x = \rho\cos\theta\sin\phi, \quad y = \rho\sin\theta\sin\phi, \quad z = \rho\cos\phi.$$

Once more, by geometric argument, notice $r = \rho\sin\phi$ connects spherical and cylindrical coordinates:

$$dl_\rho = d\rho, \quad dl_\phi = \rho d\phi, \quad dl_\theta = \rho\sin\phi d\theta. \tag{8.2}$$

Next, we will see how the calculus of frames can also be used to derive these results[2].

Definition 8.2.6. *contraction of one-forms and vectors fields*

If $\alpha = \alpha_1\, dx_1 + \alpha_2\, dx_2 + \alpha_3\, dx_3$ and $V = V_1\, \hat{x_1} + V_2\, \hat{x_2} + V_3\, \hat{x_3}$ then

$$\alpha(V) = \alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3.$$

In other words, if $\vec{\alpha} = \langle \alpha_1, \alpha_2, \alpha_3\rangle$ and $\vec{V} = \langle V_1, V_2, V_3\rangle$ then $\alpha(V) = \vec{\alpha} \bullet \vec{V}$.

Recall a **frame** or **orthonormal frame** of vector fields was a set of unit-length vector fields which was orthogonal. It is helpful to introduce the concept of a **coframe**:

Definition 8.2.7. *coframe in $\mathbb{R}^3$*

If E_1, E_2, E_3 is an orthonormal frame on $\mathbb{R}^3$ then $\theta_1, \theta_2, \theta_3$ is its **coframe** if $E_i(\theta_j) = \delta_{ij}$.

In other words, the vectors corresponding to E_i and θ_i are orthonormal.

Example 8.2.8. *Consider $\hat{x}, \hat{y}, \hat{z}$ then dx, dy, dz serve as the coframe to the standard Cartesian frame since $\vec{dx} = \langle 1,0,0\rangle$ and $\vec{dy} = \langle 0,1,0\rangle$ and $\vec{dz} = \langle 0,0,1\rangle$ using the notation of Definition 8.2.6. In other words, if $\hat{x} = E_1, \hat{y} = E_2$ and $\hat{z} = E_3$ then $\theta_1 = dx, \theta_2 = dy$ and $\theta_3 = dz$.*

Example 8.2.9. *Let us denote unit vectors in the direction of increasing r, θ, z by $\widehat{r}$, $\widehat{\theta}$, $\widehat{z}$ respectively. You can derive by geometry alone that*

$$\begin{aligned} \widehat{r} &= \cos(\theta)\,\widehat{x} + \sin(\theta)\,\widehat{y} \\ \widehat{\theta} &= -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y} \\ \widehat{z} &= \langle 0,0,1\rangle. \end{aligned}$$

We call $\{\widehat{r},\ \widehat{\theta},\ \widehat{z}\}$ the **unit-frame** *of cylindrical coordinates. Note $E_1 = \widehat{r}, E_2 = \widehat{\theta}, E_3 = \widehat{z}$ forms an orthonormal frame. Furthermore, a standard trick is to use the same coefficients to formulate the coframe: swapping out unit vectors for corresponding differentials,*

$$\begin{aligned} \theta_1 &= \cos(\theta)dx + \sin(\theta)dy \\ \theta_2 &= -\sin(\theta)dx + \cos(\theta)dy \\ \theta_3 &= dz. \end{aligned} \tag{8.3}$$

The formulas above are best formulated in terms of cylindrical coordinates. We wish to replace dx, dy, dz in terms of $dr, d\theta, dz$. Good news, θ_3 is already done. Recall, $x = r\cos\theta$ and $y = r\sin\theta$ thus

$$dx = \cos\theta dr - r\sin\theta d\theta, \qquad \& \qquad dy = \sin\theta dr + r\cos\theta d\theta.$$

[2] these formulas are most natural if we adopt the viewpoint that vector fields are derivations to which one-forms are dual, but I'm trying to avoid that abstraction here, see my Advanced Calculus notes for further refinement here

substituting the results above into Equation 8.3 yields:

$$\begin{aligned} \theta_1 &= dr \\ \theta_2 &= r d\theta \\ \theta_3 &= dz. \end{aligned} \tag{8.4}$$

This is in perfect agreement with Equation 8.1. We can either derive the coframe via geometry or via algebra and calculus.

Example 8.2.10. *In* §1.6 *we geometrically derived* $\{\widehat{\rho},\ \widehat{\phi},\ \widehat{\theta}\}$ *the* **frame** *of spherical coordinates. It has the explicit form:*

$$\begin{aligned} \widehat{\rho} &= \sin(\phi)\cos(\theta)\,\widehat{x} + \sin(\phi)\sin(\theta)\,\widehat{y} + \cos(\phi)\,\widehat{z} \\ \widehat{\phi} &= \cos(\phi)\cos(\theta)\,\widehat{x} + \cos(\phi)\sin(\theta)\,\widehat{y} - \sin(\phi)\,\widehat{z} \\ \widehat{\theta} &= -\sin(\theta)\,\widehat{x} + \cos(\theta)\,\widehat{y}. \end{aligned} \tag{8.5}$$

Let $E_1 = \widehat{\rho}$, $E_2 = \widehat{\phi}$ *and* $E_3 = \widehat{\theta}$ *and note the spherical frame is an orthonormal frame. Once more, we use the standard trick to construct the coframe:*

$$\begin{aligned} \theta_1 &= \sin(\phi)\cos(\theta)dx + \sin(\phi)\sin(\theta)dy + \cos(\phi)dz \\ \theta_2 &= \cos(\phi)\cos(\theta)dx + \cos(\phi)\sin(\theta)dy - \sin(\phi)dz \\ \theta_3 &= -\sin(\theta)dx + \cos(\theta)dy. \end{aligned} \tag{8.6}$$

It is advantageous for spherical problems to reformulate the coframe in terms of $d\rho, d\phi$ *and* $d\theta$. *Recall* $x = \rho\cos\theta\sin\phi$, $y = \rho\sin\theta\sin\phi$, $z = \rho\cos\phi$. *Thus,*

$$dx = \cos\theta\sin\phi\,d\rho - \rho\sin\theta\sin\phi\,d\theta + \rho\cos\theta\cos\phi\,d\phi \tag{8.7}$$

$$dy = \sin\theta\sin\phi\,d\rho + \rho\cos\theta\sin\phi\,d\theta + \rho\sin\theta\cos\phi\,d\phi$$

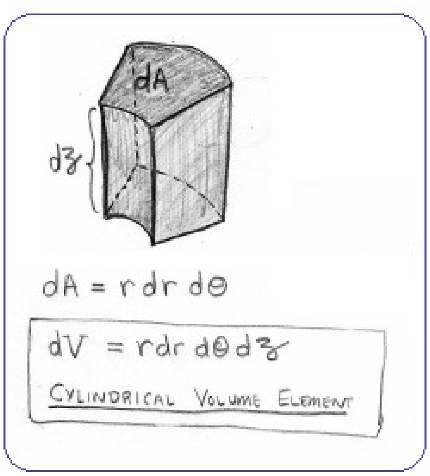
$$dz = \cos\phi\,d\rho - \rho\sin\phi d\phi.$$

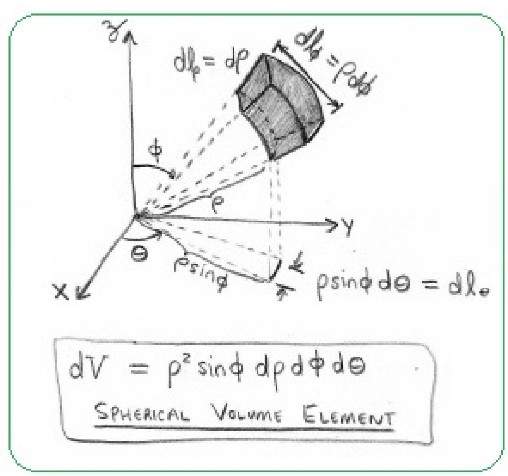
After a bit of calculation, we find from substituting the above into Equation 8.8 that

$$\begin{aligned} \theta_1 &= d\rho \\ \theta_2 &= \rho\,d\phi \\ \theta_3 &= \rho\sin\phi\,d\theta. \end{aligned} \tag{8.8}$$

We find perfect agreement with Equation 8.2.

The calculations in the previous two examples are related to the algebra seen in §6.8 where we studied the related problems of describing the infinitesimal volume element in curvelinear coordinates:





The infinitesimal volume form in $\mathbb{R}^3$ is given by $dV = dx \wedge dy \wedge dz$ or

$$dV = r dr \wedge d\theta \wedge dz$$

in cynlindrical coordinates, or

$$dV = \rho^2 \sin\phi \, d\rho \wedge d\phi \wedge d\theta$$

in spherical coordinates. Hodge duals can be calculated from the following:

$$\boxed{\sigma \wedge \star\sigma = \|\sigma\|^2 dx \wedge dy \wedge dz = \|\sigma\|^2 r dr \wedge d\theta \wedge dz = \|\sigma\|^2 \rho^2 \sin\phi \, d\rho \wedge d\phi \wedge d\theta.}$$

We can use the formula above to calculate Hodge duals in cylindrical or spherical coordinates.

Example 8.2.11. *Notice Equation 8.8 should remind us that $d\rho, \rho d\phi, \rho\sin\phi d\theta$ are all unit length. Therefore, $\|d\rho\| = 1$, $\|d\phi\| = \frac{1}{\rho}$ and $\|d\theta\| = \frac{1}{\rho\sin\phi}$. Consider, $\sigma = d\rho$ then $\|\sigma\|^2 = 1$ hence*

$$d\rho \wedge \star d\rho = \rho^2 \sin\phi \, d\rho \wedge d\phi \wedge d\theta = d\rho \wedge \big(\rho^2 \sin\phi \, d\phi \wedge d\theta\big) \quad \Rightarrow \quad \boxed{\star d\rho = \rho^2 \sin\phi \, d\phi \wedge d\theta.}$$

Next, $\sigma = d\phi$ then $\|\sigma\|^2 = 1/\rho^2$ hence

$$d\phi \wedge \star d\phi = \frac{1}{\rho^2}\rho^2 \sin\phi \, d\rho \wedge d\phi \wedge d\theta = d\phi \wedge (-\sin\phi \, d\rho \wedge d\theta) \quad \Rightarrow \quad \boxed{\star d\phi = \sin\phi \, d\theta \wedge d\rho.}$$

Finally, if $\sigma = d\theta$ then $\|\sigma\|^2 = \frac{1}{\rho^2 \sin^2\phi}$ and

$$d\theta \wedge \star d\theta = \frac{1}{\rho^2 \sin^2\phi}\rho^2 \sin\phi \, d\rho \wedge d\phi \wedge d\theta = d\theta \wedge \left(\frac{1}{\sin\phi} \, d\rho \wedge d\phi\right) \quad \Rightarrow \quad \boxed{\star d\theta = \frac{1}{\sin\phi} \, d\theta \wedge d\rho.}$$

However, if we use the coframe notation $\theta_1 = d\rho$, $\theta_2 = \rho \, d\phi$ and $\theta_3 = \rho\sin\phi \, d\theta$ then we can easily derive from the boxed equations above that

$$\star\theta_1 = \theta_2 \wedge \theta_3, \quad \star\theta_2 = \theta_3 \wedge \theta_1, \quad \star\theta_3 = \theta_1 \wedge \theta_2. \tag{8.9}$$

Example 8.2.12. *Notice Equation 8.4 should remind us that $dr, r\,d\theta, dz$ are all unit length. Therefore, $\|dr\| = 1$, $\|d\theta\| = \frac{1}{r}$ and $\|dz\| = 1$. From $\sigma \wedge \star\sigma = \|\sigma\|^2 dx \wedge dy \wedge dz = \|\sigma\|^2 r dr \wedge d\theta \wedge dz$ we may calculate Hodge duals. Setting $\sigma = dr$,*

$$dr \wedge \star dr = r dr \wedge d\theta \wedge dz = dr \wedge (r \, d\theta \wedge dz) \quad \Rightarrow \quad \boxed{\star dr = r \, d\theta \wedge dz.}$$

With $\sigma = d\theta$,

$$d\theta \wedge \star d\theta = \frac{1}{r^2} r dr \wedge d\theta \wedge dz = d\theta \wedge \left(-\frac{1}{r} \, dr \wedge dz\right) \quad \Rightarrow \quad \boxed{\star d\theta = \frac{1}{r} \, dz \wedge dr.}$$

Next, $\sigma = dz$ hence

$$dz \wedge \star dz = r dr \wedge d\theta \wedge dz = dz \wedge (r \, dr \wedge d\theta) \quad \Rightarrow \quad \boxed{\star dz = r \, dr \wedge d\theta.}$$

Once again, we see Equation 8.9 hold for $\theta_1, \theta_2, \theta_3$.

8.3 exterior derivatives

Given a differential p-form we can form a $(p+1)$-form via exterior differentiation.

Definition 8.3.1. *exterior derivative*

Let $\alpha = \displaystyle\sum_{i_1,\dots,i_p=1}^{n} \frac{1}{p!}\alpha_{i_1,\dots,i_p} dx_{i_1} \wedge \cdots \wedge dx_{i_p}$ be a differential p-form on $\mathbb{R}^n$ then

$$d\alpha = \sum_{i_1,\dots,i_p=1}^{n} \sum_{j=1}^{n} \frac{1}{p!} \frac{\partial \alpha_{i_1,\dots,i_p}}{\partial x_j} dx_j \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p}.$$

If we use multi-index notation it gets a little easier to process; $dx_I = dx_{i_1} \wedge \cdots \wedge dx_{i_p}$ where $I = (i_1, \dots, i_p)$ then if $\sum_I$ denotes the sum over all multi-indices then $\alpha = \sum_I \frac{1}{p!}\alpha_I dx_I$ and $d\alpha = \sum_I \frac{1}{p!} d\alpha_I \wedge dx_I$ where

$$d\alpha_I = \sum_{j=1}^{n} \frac{\partial \alpha_I}{\partial x_j} dx_j$$

is the usual **total differential** of multivariate calculus. I often ask my Advanced Calculus students to prove the following results are true:

Proposition 8.3.2. *linearity and product rule for exterior derivative:*

If α, β be differential forms on $\mathbb{R}^n$ and c a constant then

$$d(c\alpha + \beta) = cd\alpha + d\beta, \qquad \& \qquad d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^{deg(\alpha)} \alpha \wedge d\beta.$$

Calculating exterior derivatives is not too hard, we simply take the total differential of the component functions and wedge that with any differentials which are in the given form.

Example 8.3.3. *Let* $\alpha = xtdx \wedge dy$ *then*

$$d\alpha = (tdx + xdt) \wedge dx \wedge dy = xdt \wedge dx \wedge dy$$

since $dx \wedge dx = 0$. *Also, calculate* $d(d\alpha) = dx \wedge dt \wedge dx \wedge dy = 0$.

Example 8.3.4. *Let* $\alpha = x^2t\ dy + dz$ *and* $\beta = zdx$ *then*

$$d\alpha = 2xtdx \wedge dy + x^2dt \wedge dy \qquad \& \qquad d\beta = dz \wedge dx.$$

Note, $\alpha \wedge \beta = (x^2t\ dy + dz) \wedge (zdx) = x^2tzdy \wedge dx + zdz \wedge dx$. *Consequently,*

$$\begin{aligned} d(\alpha \wedge \beta) &= d(x^2tz) \wedge dy \wedge dx + dz \wedge dz \wedge dx \\ &= (2xtzdx + x^2zdt + x^2tdz) \wedge dy \wedge dx \\ &= x^2zdt \wedge dy \wedge dx + x^2tdz \wedge dy \wedge dx. \end{aligned}$$

On the other hand, since $deg(\alpha) = 1$ *consider:*

$$\begin{aligned} d\alpha \wedge \beta - \alpha \wedge d\beta &= (2xtdx \wedge dy + x^2dt \wedge dy) \wedge (zdx) - (x^2t\ dy + dz) \wedge dz \wedge dx \\ &= x^2zdt \wedge dy \wedge dx - x^2tdy \wedge dz \wedge dx \\ &= x^2zdt \wedge dy \wedge dx + x^2tdz \wedge dy \wedge dx. \end{aligned}$$

Thus $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^{deg(\alpha)}\alpha \wedge d\beta$ *just as was expected. Notice,*

$$d(d\alpha) = 0, \quad d(d\beta) = 0.$$

and since $deg(d\alpha) = 2$*,*

$$d(d(\alpha \wedge \beta)) = d(d\alpha \wedge \beta) - d(\alpha \wedge d\beta) = d(d\alpha) \wedge \beta + d\alpha \wedge d\beta - d\alpha \wedge d\beta - \alpha \wedge d(d\beta) = 0.$$

Of course, $d(d(\alpha \wedge \beta)) = 0$ *can also be seen by direct calculation.*

Given the calculations of the examples above you should not be surprised by the following:

Proposition 8.3.5. *nilpotence of the differential:*

If α is a differential form on $\mathbb{R}^n$ then $d(d\alpha) = 0$. That is, $d^2 = 0$.

Recall a vector field $\vec{F}$ is **conservative** if there exists a potential function f for which $\vec{F} = \nabla f$. Some differential forms are much like conservative vector fields. We should give a precise definition:

Definition 8.3.6. *exact and closed forms in* $\mathbb{R}^n$

Let α be a differential form and suppose $U \subseteq \mathbb{R}^n$ then we say α is a **exact** differential form with **potential form** ϕ if $\alpha = d\phi$ on U. We say α is **closed** on U if $d\alpha = 0$ on U.

Observe the nilpotence of the differential immediately reveals that every exact form is closed. It is natural to ponder if the converse is also true. Is every closed form exact ?

Proposition 8.3.7. *Poincare's Lemma:*

If $U \subseteq \mathbb{R}^n$ is a simply connected then α is closed on U if and only if α is exact on U.

This result should remind us of the result that $\vec{F}$ is conservative if and only if $\nabla \times \vec{F} = 0$ given the context is a simply connected subset of $\mathbb{R}^3$. The proof of Poincare's Lemma involves rudimentary algebraic topology and I usually cover it in Advanced Calculus. Topology is the study of the shape of space in an abstract sense. Topology is interested in questions like whether or not a space is connected, whether or not a loop in the space can be continuously deformed to a point, whether or not there exists holes in the space of various dimension and much more. The fascinating thing is that the existence or lack thereof of exact differential forms on a space reveals topological data about the space.

Example 8.3.8. *Consider* $\theta = \tan^{-1}(y/x)$ *then* $d\alpha = \frac{-ydx+xdy}{x^2+y^2}$*. If we define* $\alpha = \frac{-ydx+xdy}{x^2+y^2}$ *on the punctured plane* $U = \mathbb{R}^2 - \{(0,0)\}$ *then you can check* $d\alpha = 0$ *hence* α *is closed on the puncture plane. However,* α *is not closed because it is impossible to define an angle function which encircles the origin and is continuous all the way around. Notice the exterior derivative requires differentiation and differentiability implies continuity. Hence, due to the necessity of the* 2π*-angle jump,* α *cannot be exact on* U*. In contrast, if we let* $U_R = \{(x,y) \mid x > 0$ *be the right half-plane then* $\theta = \tan^{-1}(y/x)$ *serves as a potential for* α *on* U_R*. Notice* U *is not simply connected whereas* U_R *is simply connected since we can deform any loop in* U_R *to a point without obstruction.*

Proof of the impossibility follows easily from integration of forms and we postpone further discussion until we discuss integration of forms.

8.3.1 exterior dervatives in three dimensions

Consider a function $f : \mathbb{R}^3 \to \mathbb{R}$ then

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz \quad \Rightarrow \quad \boxed{df = \omega_{\nabla f}}.$$

where $\nabla f = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \rangle$. Consider $\beta = \beta_1 dx + \beta_2 dy + \beta_3 dz$ then, using $dx \wedge dx = 0$ etc. to omit three terms,

$$\begin{aligned} d\beta &= d\beta_1 \wedge dx + d\beta_2 \wedge dy + d\beta_3 \wedge dz \\ &= (\partial_2\beta_1 dy + \partial_3\beta_1 dz) \wedge dx + (\partial_1\beta_2 dx + \partial_3\beta_2 dz) \wedge dy + (\partial_1\beta_3 dx + \partial_2\beta_3 dy) \wedge dz \\ &= (\partial_2\beta_3 - \partial_3\beta_2) dy \wedge dz + (\partial_3\beta_1 - \partial_1\beta_3) dz \wedge dx + (\partial_1\beta_2 - \partial_2\beta_1) dx \wedge dy \end{aligned}$$

Notice, if $\beta = \omega_{\vec{F}}$ then the calculation above reveals

$$\boxed{d\omega_{\vec{F}} = \Phi_{\nabla \times \vec{F}}}.$$

Next, consider $\gamma = G_1 dy \wedge dz + G_2 dz \wedge dx + G_3 dx \wedge dy$, we omit terms which are trivial since $dx \wedge dx = 0$, $dy \wedge dy = 0$ and $dz \wedge dz = 0$,

$$\begin{aligned} d\gamma &= dG_1 \wedge dy \wedge dz + dG_2 \wedge dz \wedge dx + dG_3 \wedge dx \wedge dy \\ &= \partial_1 G_1 dx \wedge dy \wedge dz + \partial_2 G_2 dy \wedge dz \wedge dx + \partial_3 G_3 dz \wedge dx \wedge dy \\ &= (\partial_1 G_1 + \partial_2 G_2 + \partial_3 G_3) dx \wedge dy \wedge dz. \end{aligned}$$

Identify $\gamma = \Phi_{\vec{G}}$ where $\vec{G} = \langle G_1, G_2, G_3 \rangle$ and we find

$$\boxed{d\Phi_{\vec{G}} = \nabla \bullet \vec{G}\, dx \wedge dy \wedge dz.}$$

In summary, we have shown that the exterior derivative unifies the gradient, curl and divergence of vector calculus into a single sweeping operation.

Proposition 8.3.9. *Identities for the gradient and curl:*

$$\nabla \times \nabla f = 0, \quad \nabla \bullet (\nabla \times \vec{F}) = 0$$

Proof: let's see how the claims of the proposition are essentially corollaries of $d^2 = 0$.

$$d^2 f = d(df) = d\omega_{\nabla f} = \Phi_{\nabla \times \nabla f} = 0 \quad \Rightarrow \quad \nabla \times \nabla f = 0.$$

Likewise,

$$d^2 \omega_{\vec{F}} = d(\Phi_{\nabla \times \vec{F}}) = \nabla \bullet (\nabla \times \vec{F})\, dx \wedge dy \wedge dz = 0 \quad \nabla \bullet (\nabla \times \vec{F}) = 0. \qquad \square$$

It may be instructive to derive product rules for vector calculus from the product rule for the exterior derivative.

Example 8.3.10. *Suppose f, g are functions,*

$$\omega_{\nabla(fg)} = d(fg) = df \wedge g + f \wedge dg = g\omega_{\nabla f} + f\omega_{\nabla g} = \omega_{g\nabla f + f\nabla g}$$

hence $\nabla(fg) = g\nabla f + f\nabla g$.

Example 8.3.11. *If f is a function and $\vec{G}$ is a vector field ,*

$$\Phi_{\nabla\times(f\vec{G})} = d\omega_{f\vec{G}} = d(f\omega_{\vec{G}}) = df \wedge \omega_{\vec{G}} + f \wedge d\omega_{\vec{G}} = \omega_{\nabla f} \wedge \omega_{\vec{G}} + f\Phi_{\nabla\times\vec{G}} = \Phi_{\nabla f\times\vec{G}+f\nabla\times\vec{G}}$$

hence $\nabla \times (f\vec{G}) = \nabla f \times \vec{G} + f\nabla \times \vec{G}$.

Example 8.3.12. *Let $\vec{A}, \vec{B}$ be vector fields. Since $\omega_{\vec{A}} \wedge \omega_{\vec{B}} = \Phi_{\vec{A}\times\vec{B}}$ we find*

$$\begin{aligned} \nabla \bullet (\vec{A} \times \vec{B}) dx \wedge dy \wedge dz &= d\Phi_{\vec{A}\times\vec{B}} \\ &= d\omega_{\vec{A}} \wedge \omega_{\vec{B}} - \omega_{\vec{A}} \wedge d\omega_{\vec{B}} \\ &= \Phi_{\nabla\times\vec{A}} \wedge \omega_{\vec{B}} - \omega_{\vec{A}} \wedge \Phi_{\nabla\times\vec{B}} \\ &= \Big[(\nabla \times \vec{A}) \bullet \vec{B} - \vec{A} \bullet (\nabla \times \vec{B})\Big] dx \wedge dy \wedge dz \end{aligned}$$

thus $\nabla \bullet (\vec{A} \times \vec{B}) = (\nabla \times \vec{A}) \bullet \vec{B} - \vec{A} \bullet (\nabla \times \vec{B})$.

I tried to find an elegant derivation of the identity for $\nabla \times (\vec{A} \times \vec{B})$. I remain defeated on this point. If you are curious,

$$\nabla \times (\vec{A} \times \vec{B}) = (\vec{B} \bullet \nabla)\vec{A} - (\vec{A} \bullet \nabla)\vec{B} + \vec{A}(\nabla \bullet \vec{B}) - \vec{B}(\nabla \bullet \vec{A}).$$

I can derive this via tensor arithmetic,

$$\begin{aligned} (\nabla \times (\vec{A} \times \vec{B}))_k &= \sum_{i,j} \epsilon_{ijk} \partial_i (\vec{A} \times \vec{B})_j \\ &= \sum_{i,j} \epsilon_{ijk} \partial_i \left[\sum_{l,m} \epsilon_{lmj} A_l B_m \right] \\ &= -\sum_{i,j,l,m} \epsilon_{ikj} \epsilon_{lmj} \partial_i \left[A_l B_m \right] \\ &= -\sum_{i,l,m} (\delta_{il}\delta_{km} - \delta_{im}\delta_{kl}) \partial_i \left[A_l B_m \right] \\ &= \sum_{i,l,m} (\delta_{im}\delta_{kl} - \delta_{il}\delta_{km}) \left[(\partial_i A_l) B_m + A_l \partial_i B_m \right] \\ &= \sum_i \left[B_i \partial_i A_k + A_k (\partial_i B_i) - (\partial_i A_i) B_k - A_i \partial_i B_k \right] \\ &= \Big[(\vec{B} \bullet \nabla)\vec{A} + (\nabla \bullet \vec{B})\vec{A} - (\nabla \bullet \vec{A})\vec{B} - (\vec{A} \bullet \nabla)\vec{B} \Big]_k . \end{aligned}$$

Notice $\vec{B} \bullet \nabla = B_1\partial_1 + B_2\partial_2 + B_3\partial_3$ is an operator, so the formula above involves things we really have not encountered elsewhere. In any event, it is above my current skill level to show this in terms of differential forms.

8.4 integration of differential forms

The general theory of integration for differential forms requires more sophistication than would be appropriate for the current discussion. To keep things relatively simple we focus on the context of integrating a k-form over a k-dimensional subspace of $\mathbb{R}^n$. Furthermore, we suppose M is a

k-dimensional subspace of $\mathbb{R}^n$ such that $M = \vec{X}(D)$ where $\vec{X} : D \subseteq \mathbb{R}^k \to \mathbb{R}^n$ is a smooth **patch** or **parametrization** of M. We use $u_1, \dots, u_k$ for the parameters of M hence

$$\vec{X}(u) = (X_1(u), X_2(u), \dots, X_n(u))$$

where $u = (u_1, u_2, \dots, u_k) \in D$. In this context, the tangent space of M is formed by tangent vectors in k-directions given by the partial velocities of X,

$$\frac{\partial \vec{X}}{\partial u_j} = \left\langle \frac{\partial X_1}{\partial u_j}, \frac{\partial X_2}{\partial u_j}, \dots, \frac{\partial X_n}{\partial u_j} \right\rangle$$

for $j = 1, \dots, k$.

Definition 8.4.1. *exact and closed forms in $\mathbb{R}^n$*

Let β be a differential k-form defined near a k-dimensional subspace M with parameterization X and parameters $u_1, \dots, u_k$ taken from $D \subseteq \mathbb{R}^k$ then

$$\int_M \beta = \sum_{i_1, i_2, \dots, i_k = 1}^{n} \int_D \alpha_{i_1, i_2, \dots, i_k}(X(u)) \frac{\partial X_{i_1}}{\partial u_1} \frac{\partial X_{i_2}}{\partial u_2} \cdots \frac{\partial X_{i_k}}{\partial u_k} du_1\, du_2 \cdots du_k$$

The integral is linear, for c a constant and differential k-forms α, β,

$$\int_M (c\alpha + \beta) = c \int_M \alpha + \int_M \beta.$$

If we swap any two parameters then the resulting parametrization defines $-M$ which is the same point set with opposite **orientation**. Furthermore,

$$\int_{-M} \alpha = -\int_M \alpha.$$

Example 8.4.2. *Let $\alpha = (t^2 + x^2) dx \wedge dy \wedge dz$ define a 3-form on $\mathbb{R}^4_{txyz}$. Define $M = \vec{X}(D)$ where $D = [0, 1]^3$*

$$\vec{X}(u, v, w) = \langle 1, 2u, 3v, 4w^2 \rangle$$

then

$$\alpha_{234}(\vec{X}(u, v, w)) = 1 + 4u^2, \quad \frac{\partial X_2}{\partial u} = 2, \quad \frac{\partial X_3}{\partial v} = 3, \quad \frac{\partial X_4}{\partial w} = 4w$$

and

$$\int_M \alpha = \int_0^1 \int_0^1 \int_0^1 (1 + 4u^2)(2)(3)(4w) du\, dv\, dw = 12 \int_0^1 (1 + 4u^2) du \int_0^1 dv \int_0^1 2w\, dw = 16.$$

8.4.1 integration of differential forms in $\mathbb{R}^3$

In $\mathbb{R}^3$ we can consider integrals of zero, one, two or three-forms.

Integration of 0-forms: The integral over a zero-dimensional space is a bit of an odd case. In particular, if $M = \{p, q\}$ and f is a zero-form (that is, a function) then

$$\boxed{\int_M f = f(q) - f(p).}$$

Admittedly this is a definition.

Integration of 1-forms: Next, consider one-dimensional space, that is a curve C given by $\vec{r} : [t_1, t_2] \to \mathbb{R}^3$ and if $\alpha = \alpha_1 dx + \alpha_2 dy + \alpha_3 dz$ then using $\vec{r} = \langle x, y, z \rangle$

$$\int_C \alpha = \int_{t_1}^{t_2} \left(\alpha_1(\vec{r}(t)) \frac{dx}{dt} + \alpha_2(\vec{r}(t)) \frac{dy}{dt} + \alpha_3(\vec{r}(t)) \frac{dz}{dt} \right) dt$$

We should recognize the formula above. If $\alpha = \omega_{\vec{F}}$ then we see:

$$\boxed{\int_C \omega_{\vec{F}} = \int_C \vec{F} \bullet d\vec{r}.}$$

Integration of 2-forms: Next, suppose S is a surface with parametrization $\vec{X} = \langle x, y, z \rangle$ with parameters u, v taken from $D \subseteq \mathbb{R}^2$. Let $\beta = \beta_1 dy \wedge dz + \beta_2 dz \wedge dx + \beta_3 dx \wedge dy$ then

$$\int_S \beta = \int_D \left(\beta_1 \left[\frac{\partial y}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial z}{\partial u}\frac{\partial z}{\partial v} \right] + \beta_2 \left[\frac{\partial z}{\partial u}\frac{\partial x}{\partial v} - \frac{\partial x}{\partial u}\frac{\partial z}{\partial v} \right] + \beta_3 \left[\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v} \right] \right) du\, dv$$

where $\beta_1, \beta_2, \beta_3$ are evaluated at $\vec{X}(u, v) = (x(u, v), y(u, v), z(u, v))$. If we consider $\beta = \Phi_{\vec{F}}$ then it is clear that the integral of a two-form is naturally connected to the calculation of vector flux through a surface:

$$\boxed{\int_S \Phi_{\vec{F}} = \int_S \vec{F} \bullet d\vec{S}.}$$

Integration of 3-forms: If we consider the three-form $\gamma = g dx \wedge dy \wedge dz$ and a volume $E \subseteq \mathbb{R}^3$ and if we parametrize E via $\vec{X}$ as a function of $(u, v, w) \in D$ then

$$\int_E \gamma = \int_D g \left[\frac{\partial x}{\partial u}\frac{\partial y}{\partial v}\frac{\partial z}{\partial w} + \frac{\partial y}{\partial u}\frac{\partial z}{\partial v}\frac{\partial x}{\partial w} + \frac{\partial z}{\partial u}\frac{\partial x}{\partial v}\frac{\partial y}{\partial w} - \frac{\partial z}{\partial u}\frac{\partial y}{\partial v}\frac{\partial x}{\partial w} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}\frac{\partial z}{\partial w} - \frac{\partial x}{\partial u}\frac{\partial z}{\partial v}\frac{\partial y}{\partial w} \right] du\, dv\, dw$$

where g is evaluated at $\vec{X}(u, v, w) = (x(u, v, w), y(u, v, w), z(u, v, w))$. The six terms in the brackets above are the Jacobian determinant which is usually denoted $\frac{\partial(x,y,z)}{\partial(u,v,w)}$ and hence

$$\int_E \gamma = \int_D g(\vec{X}(u, v, w)) \frac{\partial(x, y, z)}{\partial(u, v, w)} du\, dv\, dw.$$

We assume $\vec{X}$ has positive Jacobian determinant hence $\left| \frac{\partial(x,y,z)}{\partial(u,v,w)} \right| = \frac{\partial(x,y,z)}{\partial(u,v,w)}$ and thus the change of variables theorem for triple integrals reveals

$$\int_E \gamma = \int_E g dx\, dy\, dz$$

Therefore,

$$\boxed{\int_E g dx \wedge dy \wedge dz = \int_E g\, dx\, dy\, dz.}$$

In other words, to integrate a three-form in $\mathbb{R}^3$ we simply drop the wedges and calculate the volume integral of the coefficient function g which defines the three-form. There is a subtle difference between the 3-form integration and standard volume integrals; the integration of a 3-form is an oriented integral. The key is the order of $dx \wedge dy \wedge dz$. If $\gamma = g dx \wedge dy \wedge dz$ and $\beta = g dy \wedge dx \wedge dz$ then $\int_E \gamma = -\int_E \beta$.

8.5 Generalized Stokes' Theorem

We saw the exterior derivative unifies the gradient, curl and divergence of three-dimensional vector calculus. Now we see the mirror result for integration. I'll begin by stating the theorem in more generality than we require:

Theorem 8.5.1. *Generalized Stokes' Theorem:*

Let α be a differential k-form in $\mathbb{R}^n$ and suppose M is a $(k+1)$-dimensional space embedded in $\mathbb{R}^n$ with the k-dimensional positively oriented boundary ∂M then

$$\int_M d\alpha = \int_{\partial M} \alpha.$$

The concept of a positively oriented boundary can be made rigorous, but I forego details here. Rest assured the concept lines up nicely with the usual discussions of the boundary of a surface or volume.

Example 8.5.2. *Given a curve C from p to q we define*

$$\partial C = q - p.$$

If f is a smooth function and C is an oriented curve from p to q then

$$\int_C df = \int_{\partial C} f \quad \Rightarrow \quad \int_C \omega_{\nabla f} = \int_{q-p} f \quad \Rightarrow \quad \int_C \nabla f \bullet d\vec{r} = f(q) - f(p).$$

We see the Generalized Stokes' Theorem reproduces the Fundamental Theorem for Line Integrals.

Example 8.5.3. *Let S be a surface with positively oriented boundary ∂S. Suppose $\vec{F}$ is a vector field with work form $\omega_{\vec{F}}$ then*

$$\int_S d\omega_{\vec{F}} = \int_{\partial S} \omega_{\vec{F}} \quad \Rightarrow \quad \int_S \Phi_{\nabla\times\vec{F}} = \int_{\partial S} \omega_{\vec{F}} \quad \Rightarrow \quad \int_S (\nabla \times \vec{F}) \bullet d\vec{S} = \int_{\partial S} \vec{F} \bullet d\vec{r}.$$

Indeed, the Generalized Stokes' Theorem reproduces Stokes' Theorem.

Example 8.5.4. *Let E be a volume with positively oriented boundary ∂E. Suppose $\vec{F}$ is a vector field with flux form $\Phi_{\vec{F}}$ then*

$$\int_E d\Phi_{\vec{F}} = \int_{\partial E} \Phi_{\vec{F}} \quad \Rightarrow \quad \int_E (\nabla \bullet \vec{F}) dx \wedge dy \wedge dz = \int_{\partial E} \Phi_{\vec{F}} \quad \Rightarrow \quad \int_E (\nabla \bullet \vec{F}) dx\, dy\, dz = \int_{\partial E} \vec{F} \bullet d\vec{S}.$$

The Generalized Stokes' Theorem reproduces Gauss' Theorem.

Let me attempt to make further use of Example 8.4.2.

Example 8.5.5. *Notice $\alpha = (t^2 + x^2) dx \wedge dy \wedge dz$ has*

$$\alpha = d\left(xt^2 + \frac{1}{3}x^3\right) \wedge dy \wedge dz = d\left[\left(xt^2 + \frac{1}{3}x^3\right) \wedge dy \wedge dz\right]$$

Define $\beta = \left(xt^2 + \frac{1}{3}x^3\right) \wedge dy \wedge dz$. Then, using M given in Example 8.4.2,

$$\int_M \alpha = \int_M d\beta = \int_{\partial M} \beta = 16.$$

The direct calculation of ∂M would be somewhat tedious. Recall $M = \vec{X}(D)$ where $D = [0,1]^3$

$$\vec{X}(u,v,w) = \langle 1, 2u, 3v, 4w^2 \rangle$$

then ∂M includes six two-dimensional surfaces obtained by setting $u = 0, 1$ while varying v, w, or setting $v = 0, 1$ while varying u, w or by setting $w = 0, 1$ and varying u, v. In each case the parameters must be ordered such that the resulting surface is part of the positively oriented boundary of E. Please forgive me for not detailing the mechanics of this here.

8.6 Maxwell's Equations in differential forms

We studied Maxwell's Equations in vector calculus. These are coupled partial differential equations which detail how the components of the electric and magnetic fields must behave locally in response to charge, current and each other:

$$\nabla \bullet \vec{E} = \frac{\rho}{\epsilon_o}, \qquad \nabla \times \vec{B} = \mu_o\left(\vec{J} + \epsilon_o \frac{\partial \vec{E}}{\partial t}\right), \qquad \nabla \bullet \vec{B} = 0, \qquad \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

These are Gauss' Law, Ampere's Law with Maxwell's correction, no magnetic monopoles and Faraday's Law all written as local field equations. We also saw that both the electric and magnetic fields can be derived from potentials. In particular: for a given $\vec{E}$ and $\vec{B}$ we insist the potentials satisfy:

$$\vec{E} = -\nabla V + \frac{\partial A}{\partial t} \qquad \& \qquad \vec{B} = \nabla \times \vec{A}.$$

We call V the **scalar potential** or **voltage** function whereas $\vec{A}$ is known as the **vector potential**. If we use calculus in spacetime, or perhaps I should say time-space since we use $\mathbb{R}^4_{txyz}$ then we can write Maxwell's Equations as a pair of differential form equations and we can unify the scalar and vector potentials into a single 4-potential. These equations are best understood with an introduction of Hodge duality contructed with respect to the Minkowski metric. I will simply quote these equations since my intention here is just to give you a snapshot of how the exterior calculus and algebra unifies the notation of basic physical law[3].

The **electromagnetic field tensor** or **Faraday form** is a two-form given by

$$F = \omega_{\vec{E}} \wedge dt + \Phi_{\vec{B}}$$

It may be helpful to look at the components of F as a matrix:

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}$$

After a bit of calculation we can show:

$$dF = \Phi_{\nabla \times \vec{E} + \partial_t \vec{B}} \wedge dt + (\nabla \cdot \vec{B}) dx \wedge dy \wedge dz$$

Therefore, $\nabla \bullet \vec{B} = 0$ and $\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$ can be expressed simply as $\boxed{dF = 0}$.

[3] I can recommend reading if you desire further detail

The Hodge dual of the Faraday tensor is given by

$$\star F = \Phi_E - \omega_B \wedge dt$$

By the nearly the same calculus as for dF, making the substitutions $\vec{E} \mapsto -\vec{B}$ and $\vec{B} \mapsto \vec{E}$ we deduce

$$d \star F = \Phi_{-\nabla\times\vec{B}+\partial_t\vec{E}} \wedge dt + (\nabla \cdot \vec{E})dx \wedge dy \wedge dz$$

The current one-form is given by $J = -\rho dt + J_x dx + J_y dy + J_z dz$ and the Hodge dual of J gives the three-form

$$\star J = -\Phi_{\vec{J}} \wedge dt + \rho dx \wedge dy \wedge dz.$$

Now, I'm being a bit careless with units since some of the equations implicit $c = 1$ whereas in truth $c = \frac{1}{\sqrt{\mu_o \epsilon_o}} = 2.98 \times 10^8 \ m/s$. But setting aside minor numerical details, we find Gauss' Law $\nabla \bullet \vec{E} = \frac{\rho}{\epsilon_o}$ and Ampere's Law with Maxwell's correction $\nabla \times \vec{B} = \mu_o \bigg(\vec{J} + \epsilon_o \frac{\partial \vec{E}}{\partial t} \bigg)$ can be compactly expressed as the differential form equation $\boxed{d \star F = \mu_o \star J}$.

Next, consider the 4-potential given by $A = -V dt + \omega_{\vec{A}}$. A little four dimensional calculus shows

$$dA = \omega_{(-\nabla V - \partial_t \vec{A})} \wedge dt + \Phi_{\nabla \times \vec{A}}$$

However, since $\vec{E} = -\nabla V - \frac{\partial A}{\partial t}$ and $\vec{B} = \nabla \times \vec{A}$ we have that

$$dA = \omega_{\vec{E}} \wedge dt + \Phi_{\vec{B}} \quad \Rightarrow \quad \boxed{F = dA}.$$

Appreciate that $dF = 0$ arises naturally from the fact $d^2 = 0$; $dF = d(dA) = 0$. Another way to look at this is since F is closed, we know A exists for which $F = dA$ for some potential one-form A on any simply connected region on which F is defined. Well, in fact, up to the gauge freedom, A is uniquely defined if we are given a set of source charges and currents. Griffith's *Introduction to Electrodynamics* explains why the solutions are unique in the context of the standard theory of electricity and magnetism where $\nabla \bullet \vec{B} = 0$.

However, if we entertain problems which involve magnetic monopoles (where $\nabla \bullet \vec{B} \neq 0$ at at least one point) then the story gets a bit more interesting. Much like the necessity of the angle jump around the origin in the plane, the vector potential for a magnetic monopole must suffer a line of divergence which is known as the **Dirac string**. This has nearly nothing to do with string theory, but it has everything to do with the origin story of algebraic topology. This string is not a physical divergence, it can point in any direction, but it must point somewhere because of the nontrivial topology involved.

8.6.1 Electrostatics in Five dimensions

We will endeavor to determine the electric field of a point charge in 5 dimensions where we are thinking of adding an extra spatial dimension. Lets call the fourth spatial dimension the w-direction so that a typical point in space time will be (t, x, y, z, w). First we note that the electromagnetic field tensor can still be derived from a one-form potential,

$$A = -\rho dt + A_1 dx + A_2 dy + A_3 dz + A_4 dw$$

we will find it convenient to make our convention for this section that $\mu, \nu, ... = 0, 1, 2, 3, 4$ whereas $m, n, ... = 1, 2, 3, 4$ so we can rewrite the potential one-form as,

$$A = -\rho dt + A_m dx^m$$

This is derived from the vector potential $A^\mu = (\rho, A^m)$ under the assumption we use the natural generalization of the Minkowski metric, namely the 5 by 5 matrix,

$$(\eta_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = (\eta^{\mu\nu}) \tag{8.10}$$

we could study the linear isometries of this metric, they would form the group $O(1, 4)$. Now we form the field tensor by taking the exterior derivative of the one-form potential,

$$F = dA = \frac{1}{2}(\partial_\mu \partial_\nu - \partial_\nu \partial_\mu) dx^\mu \wedge dx^\nu$$

now we would like to find the electric and magnetic "fields" in 4 dimensions. Perhaps we should say 4+1 dimensions, just understand that I take there to be 4 spatial directions throughout this discussion if in doubt. Note that we are faced with a dilemma of interpretation. There are 10 independent components of a 5 by 5 antisymmetric tensor, naively we wold expect that the electric and magnetic fields each would have 4 components, but that is not possible, we'd be missing two components. The solution is this, the time components of the field tensor are understood to correspond to the electric part of the fields whereas the remaining 6 components are said to be magnetic. This aligns with what we found in 3 dimensions, its just in 3 dimensions we had the fortunate quirk that the number of linearly independent one and two forms were equal at any point. This definition means that the magnetic field will in general not be a vector field but rather a "flux" encoded by a 2-form.

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & -E_x & -E_y & -E_z & -E_w \\ E_x & 0 & B_z & -B_y & H_1 \\ E_y & -B_z & 0 & B_x & H_2 \\ E_z & B_y & -B_x & 0 & H_3 \\ E_w & -H_1 & -H_2 & -H_3 & 0 \end{pmatrix} \tag{8.11}$$

Now we can write this compactly via the following equation,

$$F = E \wedge dt + B$$

I admit there are subtle points about how exactly we should interpret the magnetic field, however I'm going to leave that to your imagination and instead focus on the electric sector. What is the generalized Maxwell's equation that E must satisfy?

$$d^*F = \mu_o {}^*\mathcal{J} \implies d^*(E \wedge dt + B) = \mu_o {}^*\mathcal{J}$$

where $\mathcal{J} = -\rho dt + J_m dx^m$ so the 5 dimensional Hodge dual will give us a $5 - 1 = 4$ form, in particular we will be interested in just the term stemming from the dual of dt,

$$^*(-\rho dt) = \rho dx \wedge dy \wedge dz \wedge dw$$

the corresponding term in d^*F is $d^*(E \wedge dt)$ thus, using $\mu_o = \frac{1}{\epsilon_o}$,

$$d^*(E \wedge dt) = \frac{1}{\epsilon_o}\rho dx \wedge dy \wedge dz \wedge dw \tag{8.12}$$

is the 4-dimensional Gauss's equation. Now consider the case we have an isolated point charge which has somehow always existed at the origin. Moreover consider a 3-sphere that surrounds the charge. We wish to determine the generalized Coulomb field due to the point charge. First we note that the solid 3-sphere is a 4-dimensional object, it the set of all $(x, y, z, w) \in \mathbb{R}^4$ such that

$$x^2 + y^2 + z^2 + w^2 \leq r^2$$

We may parametrize a three-sphere of radius r via generalized spherical coordinates,

$$\begin{aligned} x &= rsin(\theta)cos(\phi)sin(\psi) \\ y &= rsin(\theta)sin(\phi)sin(\psi) \\ z &= rcos(\theta)sin(\psi) \\ w &= rcos(\psi) \end{aligned} \tag{8.13}$$

Now it can be shown that the volume[4] and surface area of the radius r three-sphere are as follows,

$$vol(S^3) = \frac{\pi^2}{2}r^4 \qquad\qquad area(S^3) = 2\pi^2 r^3$$

We may write the charge density of a smeared out point charge q as,

$$\rho = \begin{cases} 2q/\pi^2 a^4, & 0 \leq r \leq a \\ 0, & r > a \end{cases}. \tag{8.14}$$

Notice that if we integrate ρ over any four-dimensional region which contains the solid three sphere of radius a will give the enclosed charge to be q. Then integrate over the Gaussian 3-sphere S^3 with radius r call it M,

$$\int_M d^*(E \wedge dt) = \frac{1}{\epsilon_o}\int_M \rho dx \wedge dy \wedge dz \wedge dw$$

now use the Generalized Stokes Theorem to deduce,

$$\int_{\partial M} {}^*(E \wedge dt) = \frac{q}{\epsilon_o}$$

but by the "spherical" symmetry of the problem we find that E must be independent of the direction it points, this means that it can only have a radial component. Thus we may calculate the integral with respect to generalized spherical coordinates and we will find that it is the product of $E_r \equiv E$ and the surface volume of the four dimensional solid three sphere. That is,

$$\int_{\partial M} {}^*(E \wedge dt) = 2\pi^2 r^3 E = \frac{q}{\epsilon_o}$$

Thus,

$$\boxed{E = \frac{q}{2\pi^2 \epsilon_o r^3}}$$

[4] see Example 6.7.1

the Coulomb field is weaker if it were to propogate in 4 spatial dimensions. Qualitatively what has happened is that the have taken the same net flux and spread it out over an additional dimension, this means it thins out quicker. A very similar idea is used in some *brane world* scenarios. String theorists posit that the gravitational field spreads out in more than four dimensions while in contrast the standard model fields of electromagnetism, and the strong and weak forces are confined to a four-dimensional brane. That sort of model attempts an explanation as to why gravity is so weak in comparison to the other forces. Also it gives large scale corrections to gravity that some hope will match observations which at present don't seem to fit the standard gravitational models.

8.7 search for potentials in differential form

We argued in a previous chapter that if $\vec{E}$ is the electric field due to charge density ρ then the voltage function V for which $\vec{E} = -\nabla V$ can be calculated from the following integral of the density:

$$V(\vec{r}) = \iiint_M \frac{\rho(\vec{r}\,')}{4\pi\epsilon_o\|\vec{r}-\vec{r}\,'\|}dV'.$$

In the case of a point charge at the origin we have $\rho(\vec{r}) = Q\delta(\vec{r})$ (see §7.8.1). To calculate $V(\vec{r})$ for the point charge at the origin we simple do the integral.

$$V(\vec{r}) = \iiint_M \frac{Q\delta(\vec{r}\,')}{4\pi\epsilon_o\|\vec{r}-\vec{r}\,'\|}dV' = \frac{Q}{4\pi\epsilon_o\|\vec{r}\|} = \frac{Q}{4\pi\epsilon_o r}.$$

where $r = \sqrt{x^2+y^2+z^2}$ in this context. Moreover, partial differentiation such as $\frac{\partial r}{\partial x} = \frac{x}{r}$ and $\frac{\partial r}{\partial y} = \frac{y}{r}$ and $\frac{\partial r}{\partial z} = \frac{z}{r}$ yield

$$\vec{E} = -\nabla V = -\frac{Q}{4\pi\epsilon_o}\nabla\left[\frac{1}{r}\right] = \frac{Q}{4\pi\epsilon_o r^2}\left\langle \frac{x}{r}, \frac{y}{r}, \frac{z}{r}\right\rangle = \frac{Q}{4\pi\epsilon_o r^2}\widehat{r}.$$

The field above is known as the **Coulomb field**. In the usual math spherical coordinate system we have

$$\vec{E} = \frac{Q}{4\pi\epsilon_o\rho^2}\widehat{\rho}$$

Suppose for a moment that we didn't know the formula for V of the Coulomb field, yet we knew the form of the Coulomb field from experimental data. Can we calculate V from $\vec{E}$? Can we solve $\vec{E} = -\nabla V$ for V ? In Cartesian coordinates this is somewhat challenging:

$$\frac{Q}{4\pi\epsilon_o(x^2+y^2+z^2)^{3/2}}\langle x,y,z\rangle = \left\langle -\frac{\partial V}{\partial x}, -\frac{\partial V}{\partial y}, -\frac{\partial V}{\partial z}\right\rangle.$$

Some students will be able to solve this given sufficient time. They will derive:

$$V(x,y,z) = \frac{Q}{4\pi\epsilon_o(x^2+y^2+z^2)^{1/2}}.$$

On the other hand, if we face the problem in spherical coordinates then the spherical symmetry of V is implied from the spherical symmetry of $\vec{E} = \frac{Q}{4\pi\epsilon_o\rho^2}\widehat{\rho}$. In short, V can only depend on ρ and as such[5]:

$$\vec{E} = \frac{Q}{4\pi\epsilon_o\rho^2}\widehat{\rho} = -\nabla V = -\frac{\partial V}{\partial\rho}\widehat{\rho}$$

hence we simply need to solve $\frac{Q}{4\pi\epsilon_o\rho^2} = -\frac{\partial V}{\partial \rho}$. Integration reveals $V = \frac{Q}{4\pi\epsilon_o\rho}$. This is a typical story, using coordinates which mirror the symmetry of a given problem greatly simplify calculation.

[5]note generally $\nabla f = \widehat{\rho}\frac{\partial f}{\partial\rho} + \widehat{\phi}\frac{1}{\rho}\frac{\partial f}{\partial\phi} + \widehat{\theta}\frac{1}{\rho\sin(\phi)}\frac{\partial f}{\partial\theta}$

8.7.1 electrostatic scalar potential from work form formalism

Let's examine the electrostatic Coulomb field problem using differential forms. Scalar potential V has $\vec{E} = -\nabla V$. Recall $df = \omega_{\nabla f}$ so

$$\omega_{\vec{E}} = \omega_{-\nabla V} = -\omega_{\nabla V} = -dV.$$

Our aim is to find V for which $-dV = \omega_{\vec{E}}$.

Let us pause to note how the work form maps translate to spherical coordinates. Of course $\vec{F} = a\,\widehat{x} + b\,\widehat{y} + c\,\widehat{z}$ gives $\omega_{\vec{F}} = adx + bdy + cdz$. However, if we use the coframe notation $\theta_1 = d\rho$, $\theta_2 = \rho\, d\phi$ and $\theta_3 = \rho \sin\phi\, d\theta$ which is dual to $\{\widehat{\rho}, \widehat{\phi}, \widehat{\theta}\}$ then the vector field $\vec{F} = a\widehat{\rho} + b\widehat{\phi} + c\widehat{\theta}$ has

$$\omega_{\vec{F}} = a\theta_1 + b\theta_2 + c\theta_3 = ad\rho + b\rho d\phi + c\rho \sin\theta\, d\theta$$

Let us turn to the explicit Coulomb field problem, but this time we'll drop the $\frac{Q}{4\pi\epsilon_o}$ to reduce clutter. Given $\vec{E} = \frac{1}{\rho^2}\widehat{\rho}$ we construct work form

$$\omega_{\vec{E}} = \frac{1}{\rho^2} d\rho = -d\left(\frac{1}{\rho}\right) \quad \Rightarrow \quad V = \frac{1}{\rho}.$$

8.7.2 magnetic monopoles and the flux form formalism

Ok, now we have some background on the electric monopole field which is called the Coulomb field. What if the same field pattern was possible in magnetostatics ? In such a context, $\vec{B} = \nabla \times \vec{A}$ where $\vec{A}$ is the vector potential. Can we calculate $\vec{A}$ given $\vec{B} = \frac{1}{\rho^2}\widehat{\rho}$? That is, can you find $\vec{A}$ for which $\nabla \times \vec{A} = \frac{1}{\rho^2}\widehat{\rho}$? Even if you happen to know the formula for the curl in spherical coordinates this calculation is far from trivial. In my experience, if you give this on a test with very intelligent students, it will stop them in their tracks. Yet, a solution is known from Dirac. For context, Dirac was one of the greatest physicists of the twentieth century, he played a large role in the maturation of quantum mechanics as well as relativistic quantum mechanics.

We should translate the problem of finding a vector potential into the differential form context: we desire to find $\vec{A}$ for which $\vec{B} = \nabla \times \vec{A}$. Consider:

$$\Phi_{\vec{B}} = \Phi_{\nabla \times \vec{A}} = d\omega_{\vec{A}}.$$

Therefore, we reduce the problem of finding a vector potential to the problem of finding a one-form whose exterior derivative corresponds to the flux-form of the magnetic field. This is actually easier than the curl formula, if we play our cards just the right way.

The flux form in Cartesian coordinates is given by $\Phi_{\vec{F}} = ady \wedge dz + bdz \wedge dx + cdx \wedge dy$ for $\vec{F} = a\,\widehat{x} + b\,\widehat{y} + c\,\widehat{z}$. Likewise, in the spherical frame with coframe $\theta_1 = d\rho$, $\theta_2 = \rho\, d\phi$ and $\theta_3 = \rho \sin\phi\, d\theta$ which is dual to $\{\widehat{\rho}, \widehat{\phi}, \widehat{\theta}\}$ then the vector field $\vec{F} = a\widehat{\rho} + b\widehat{\phi} + c\widehat{\theta}$ has

$$\Phi_{\vec{F}} = a\theta_2 \wedge \theta_3 + b\theta_3 \wedge \theta_1 + c\theta_1 \wedge \theta_2 = a\rho^2 \sin\phi d\phi \wedge d\theta + b\rho \sin\phi d\phi \wedge d\rho + c\rho d\rho \wedge d\theta.$$

Thus consider $\vec{B} = \frac{1}{\rho^2}\widehat{\rho}$ we face

$$\Phi_{\vec{B}} = \frac{1}{\rho^2}\rho^2 \sin\phi d\phi \wedge d\theta = \sin\phi d\phi \wedge d\theta = d\left(-\cos\phi d\theta\right)$$

We need to find $\vec{A}$ for which $\omega_{\vec{A}} = -\cos\phi d\theta$. Notice, $\theta_3 = \rho\sin\phi\, d\theta$ implies $d\theta = \frac{1}{\rho\sin\phi}\theta_3$ hence

$$\omega_{\vec{A}} = -\cos\phi\frac{1}{\rho\sin\phi}\theta_3 = \frac{-\cos\phi}{\rho\sin\phi}\theta_3 \quad\Rightarrow\quad \boxed{\vec{A} = \frac{-\cos\phi}{\rho\sin\phi}\widehat{\theta}}$$

Now, to appreciate this formula we should return to Cartesian coordinates, recall $x = \rho\cos\theta\sin\phi$, $y = \rho\sin\theta\sin\phi$ and $z = \rho\cos\phi$. Also, recall $\rho\sin\phi = r = \sqrt{x^2+y^2}$. Furthermore, $\cos\theta = \frac{x}{r}$ and $\sin\theta = \frac{y}{r}$. Thus,

$$\widehat{\theta} = \langle -\sin\theta, \cos\theta, 0\rangle = \left\langle -\frac{y}{r}, \frac{x}{r}, 0\right\rangle$$

Note, $\cos\phi = \frac{z}{\rho}$ and $r = \rho\sin\phi$. Putting it all together,

$$\vec{A} = \frac{-\cos\phi}{\rho\sin\phi}\widehat{\theta} = \frac{-\frac{z}{\rho}}{r}\left\langle -\frac{y}{r}, \frac{x}{r}, 0\right\rangle \quad\Rightarrow\quad \boxed{\vec{A} = \frac{z}{(x^2+y^2)\sqrt{x^2+y^2+z^2}}\langle y, -x, 0\rangle.}$$

This vector potential is **singular** where $x^2+y^2=0$ and this particular type of singularity is known as a **Dirac string**. In fact, the formula due to Dirac is a bit more beautiful than our formula, somehow Dirac found

$$\vec{A} = \frac{1}{\rho(z-\rho)}\langle -y, x, 0\rangle$$

where $r = \sqrt{x^2+y^2+z^2}$ which has its Dirac string on the positive z-axis where $z=\rho$. Dirac's formula can be seen by a shift of $\omega_{\vec{A}} = -\cos\phi d\theta$ to $(1+\cos\phi)d\theta$[6].

The Dirac string is very different than what we encounter in standard electricity and magnetism. In the absense of magnetic monopoles it turns out the singularities in the potential indicate the presence of localized charge or current. Notice that the Coulomb potential $V = 1/\rho$ is singular at the origin where $\rho = 0$, but that is hardly surprising since the Coulomb field is the electric field of a point charge placed at the origin. In contrast, the magnetic monopole's Dirac string is mostly removed from the magnetic charge at the origin. If we study the mathematics here then we'll soon learn that the Dirac string is not an accident. It is a necessary feature for the vector potential of the magnetic monopole. Why ?

In Section 7.5.6 we saw that the flux in two dimensions of the reciprocal field $\vec{F} = \frac{1}{\sqrt{x^2+y^2}}\langle x, y\rangle$ through a loop C which encloses the origin is given by

$$\int_C \vec{F}\bullet\hat{n}\,ds = 2\pi.$$

This calculation is related to the circulation of the vector field $\vec{G} = \langle\frac{-y}{\sqrt{x^2+y^2}}, \frac{x}{\sqrt{x^2+y^2}}\rangle$ in the punctured plane since

$$\int_C \vec{F}\bullet\hat{n}\,ds = \int_C \frac{-ydx+xdy}{\sqrt{x^2+y^2}} = \int_C \vec{G}\bullet d\vec{r}.$$

Hence, if $\vec{G} = \nabla g$ then it is impossible for g to be smooth on the whole curve which loops around the origin since that would imply that $\int_C \vec{G}\bullet d\vec{r} = 0$. For example, $g = \tan^{-1}(y/x)$ suffices as a

[6]see www.supermath.info/monopoles.pdf equation 84, but beware, have to translate from physics sphericals to math sphericals

potential for $\vec{G}$ in the half-plane with $x > 0$ since $\nabla g = \vec{G}$. You could also use $g = \cot^{-1}(x/y)$ for $y > 0$. Indeed, choose any choice of angle, this will serve as a potential for $\vec{G}$ on all of the plane except where the angle jumps in general. There is a connection between the curl and divergence as they reduce to the plane with Green's Theorem: suppose S is subset of the plane then for $\vec{F} = \langle P, Q\rangle$

$$\int_S \nabla \bullet \langle P, Q\rangle dA = \int_{\partial S} \langle P, Q\rangle \bullet \hat{n}\, ds = \int_{\partial S} Pdy - Qdx = \int_{\partial S} \langle -Q, P\rangle \bullet d\vec{r} = \int_S \nabla \times \langle -Q, P\rangle \bullet d\vec{S}$$

where $d\vec{S} = \langle 0, 0, 1\rangle dxdy$ so we only consider the z-component of the curl and naturally we use $\langle -Q, P, 0\rangle$ if we think of the curl in the three dimensional sense.

Actually, the two options above are naturally understood in terms of the two ways to formulate a vector field. Given $\vec{F} = \langle P, Q\rangle$ in $\mathbb{R}^2$ we write $\omega_{\vec{F}} = Pdx + Qdy$ for the **work form** corresponding to $\vec{F}$. Notice Hodge duality can be defined in the plane,

$$\star dx = dy \qquad \& \qquad \star\, dy = -dx$$

since $dx \wedge \star dx = dx \wedge dy$ and $dy \wedge \star dy = dx \wedge dy$ must hold true. Then,

$$\star\omega_{\vec{F}} = Pdy - Qdx$$

hence $\Phi_{\vec{F}} = -Qdx + Pdy$ defines the **flux form** of $\vec{F}$ in the plane.

In polar coordinates, $\widehat{r}, \widehat{\theta}$ define the polar frame with coframe $\theta_1 = dr, \theta_2 = rd\theta$. Hodge duality for the polar frame is given by $\star\theta_1 = \theta_2$ whereas $\star\theta_2 = -\theta_1$. That is, $\star dr = rd\theta$ and $\star d\theta = -\frac{1}{r}dr$. Given vector field $\vec{F} = a\widehat{r} + b\widehat{\theta}$ we find

$$\omega_{\vec{F}} = adr + brd\theta \qquad \& \qquad \Phi_{\vec{F}} = ard\theta - \frac{b}{r}dr.$$

The vector field $\vec{F} = \frac{1}{x^2+y^2}\langle x, y\rangle = \frac{1}{r}\widehat{r}$ has $a = 1/r$ and $b = 0$ for components with respect to the polar frame. Thus,

$$\omega_{\vec{F}} = \frac{1}{r}dr \qquad \& \qquad \Phi_{\vec{F}} = \frac{1}{r}rd\theta = d\theta$$

Apparently $\Phi_{\vec{F}}$ is exact with potential zero-form θ and since $\omega_{\vec{F}} = d(\ln r)$ we find potential zero-form $\ln(r)$ for $\omega_{\vec{F}}$. Notice θ can only be continuous on at most a slit-plane whereas $\ln(r)$ serves as a potential on the whole punctured plane. This story plays out for the Coulomb field and whether we view it as an electric or magnetic field. In the electric sense, $\vec{E} = -\nabla V$ has solution on punctured three space whereas in the magnetic sense $\vec{B} = \nabla \times \vec{A}$ fails to have a global solution on punctured three dimensional space.

Ok, what does the punctured plane have to do with the magnetic monopole ? Notice that if $\vec{B} = \nabla \times \vec{A}$ on an entire sphere S which encloses the origin then $\int_S \vec{B} \bullet d\vec{S} = \int_S (\nabla \times \vec{A}) \bullet d\vec{S} = \int_{\partial S} \vec{A} \bullet d\vec{r} = 0$ since $\partial S = \emptyset$. However, direct calculation of the flux of the monopole yields $\int_S \vec{B} \bullet d\vec{S} = 4\pi$. Thus it is impossible to find $\vec{A}$ globally defined on $\mathbb{R}^3 - \{(0, 0, 0)\}$ since we obtain a contradiction when we consider any sphere enclosing the origin. Curiously, we are able to define $\vec{A}$ on almost the whole space, but there must be some subset on which $\vec{A}$ fails to be smooth.

8.8 generalized Pythagorean identity and flux forms in $\mathbb{R}^n$

Finally, we note the Hodge dual in $\mathbb{R}^n$ is likewise defined:

Definition 8.8.1. *Hodge duality*

Given a p-form σ there exists a unique $n-p$ form $\star\sigma$ such that

$$\sigma \wedge \star\sigma = \|\sigma\|^2 dx_1 \wedge \cdots \wedge dx_n$$

where $\|\sigma\|$ is the Euclidean length of σ.

Given the definition above it is also natural to define the **flux form** corresponding to a one-form.