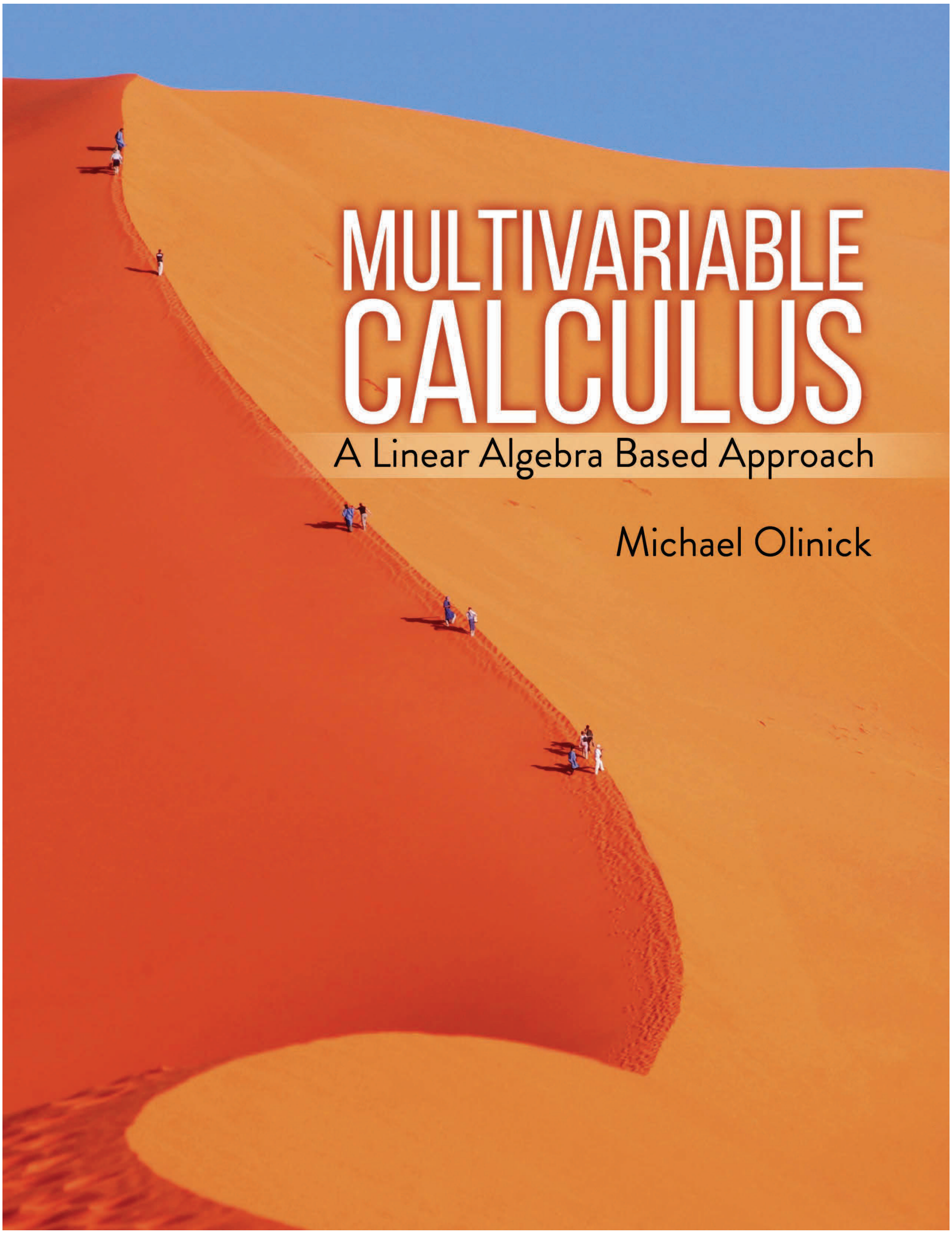
MULTIVARIABLE
CALCULUS
A Linear Algebra Based Approach
Michael Olinick

Multivariable Calculus
A Linear Algebra Based Approach

Michael Olinick[1]

August 16, 2021

[1] www.middlebury.edu/~molinick

Dedicated to Joshua, Zoey, Sydney and Raya.

Contents

Preface

> *Is mathematical analysis . . . only a vain play of the mind? It can give to the physicist only a convenient language; is this not a mediocre service, which, strictly speaking, could be done without; and even is it not to be feared that this artificial language may be a veil interposed between reality and the eye of the physicist? Far from it; without this language most of the intimate analogies of things would have remained forever unknown to us; and we should forever have been ignorant of the internal harmony of the world, which is, we shall see, the only true objective reality.*
>
> – Henri Poincaré, 1854 – 1912

A few years ago, a Vermont gym teacher took a year's leave to walk the perimeter of the continental United States. He recorded, for each day of his trek, how many miles he walked along with the latitude, longitude, and elevation of his resting place for the night.

During the recent COVID-19 pandemic, epidemiologists considered for each week of the crisis the numbers of new cases, deaths from the virus, hospitalizations, and recoveries.

Building a laptop device requires dozens of different materials including metal, glass, plastics, and a range of other products from aluminum to zinc sulfide. A computer manufacturer is contemplating producing various numbers of laptop models and wants to know how much of each of these materials she would need.

These are three instances which we can study through a function which takes a single number as input (day n of the walk, week n of the pandemic, n laptops) and outputs a collection of numbers (miles walked, latitude, longitude, elevation, for example).

There are also a myriad of examples where we use functions which take a collection of numbers as input and produce a single number as output. The *volume* of a circular cylinder, for example, is a function of the *radius* r of the *base* b and its *height* h. Physicists need both the mass and velocity of a moving object to compute its momentum or kinetic energy.

Economists regularly calculate the Gross National Product (GNP) of a given country. It is a single numerical estimate of the total value of all the final products and services turned out in a given period by the means of production owned by the country's residents. Calculating the GNP involves knowing a host of numbers including personal consumption expenditures, private domestic investment, government expenditure, net exports, income earned by residents from overseas investments, and income earned within the domestic economy by foreign residents.

Political scientists use the *Gini Index*, a number between 0 and 1, as a measure of inequality within a society. To compute a Gini Index for income inequality, for example, they might first divide up the range of annual incomes into 20 or 25 different levels. Then they would use the number of people at each income level to calculate the index.

The world also abounds with processes that take in multiple numbers as input and also produce many numbers as output. As a simple example, consider our circular cylinder of radius r and height h and think about taking these two numbers and calculating both the volume and surface area of the cylinder.

As a more complicated situation where multiple numbers may be fed into a function that processes them in some way and produces multiple quantities as output, imagine going for dinner at your favorite fast food restaurant. You've decided in advance that you want a beverage, burger, salad, dessert and a container of French fries. The menu displays 20 beverages, 9 different types of burgers, 5 salad choices, a dozen desserts, and 3 sizes of fries. For each of the 49 items, you know the cost, calorie count, and number of grams of fat, carbohydrates, and protein. You have a limited amount of money to spend and while you need to keep the total calories, fat and carbohydrates below some specific levels, you also need a dinner that meets or exceeds your need for protein. You'd like a function that outputs some possible dinner choices along with their costs, total calories and other nutritional data all of which meet your constraints.

Multivariable Calculus deals with situations like these where there are functions whose inputs and outputs may be single numbers or collections (vectors) of numbers. The few examples we've described hopefully begin to give you a sense of the importance of Multivariable Calculus in gaining a better understanding of complex real world problems.

We may classify these functions into four groups which we will consider in this order:

Input	Output	Notation	Description
Real Number	Real Number	$f : \mathbb{R}^1 \to \mathbb{R}^1$	Real-Valued Function of a Real Variable
Real Number	Vector	$f : \mathbb{R}^1 \to \mathbb{R}^n$	Vector-Valued Function of a Real Variable
Vector	Real Number	$f : \mathbb{R}^m \to \mathbb{R}^1$	Real-Valued Function of a Vector
Vector	Vector	$f : \mathbb{R}^m \to \mathbb{R}^n$	Vector-Valued Function of a Vector

Prerequisites

Our first category, $f : \mathbb{R}^1 \to \mathbb{R}^1$, real-valued functions of a real variable, is the subject matter of courses typically labeled something like *Calculus I* and *Calculus II.* We assume you have completed a year's study of this material, which we will refer to as *Elementary Calculus* or *Single Variable Calculus.*

Since the other three categories all involve functions dealing with vectors, it shouldn't be surprising that many concepts in *Linear Algebra* will help you understand them better. The *Chain Rule*, for instance, really amounts to matrix multiplication. Hidden in the *Change of Variable* or *Method of Substitution* technique for dealing with integrals is the determinant of a matrix. Eigenvalues play an important role in identifying maximum and minimum values of functions.

A one semester introduction to Linear Algebra will provide a sufficient background for you. Chapter 1 provides a review of the pertinent ideas from Single Variable Calculus and Linear Algebra

Our Approach

I hope to show you how the ideas of limit, continuity, derivative and integral you studied in single variable calculus extend naturally to a multivariable setting. You'll also learn about several different generalizations of the Fundamental Theorem of Calculus.

Isaac Newton forged some of the first tools of multivariable calculus and used them with spectacular success to derive Kepler's laws of planetary motion from a handful of basic assumptions about gravity, forces and movement (See Chapter 2). Newton's work led to the exploration, intense in the 19th Century, of how mathematics could help us understand mechanics, fluid dynamics, and electricity among other phenomena in the physical world. In more recent times, the range of applications of multivariable calculus has expanded across physics, chemistry, biology, geology, neuroscience, engineering, economics and the social sciences.

Useful and powerful in its applications, multivariable calculus is also beautiful mathematics. The subject weaves together geometry, analysis and linear algebra, Our treatment assumes the reader has completed a one term linear algebra course. Linear algebra provides both the language and, more importantly, an appropriate perspective for the material you are going to study here. We stress, for example, how the derivative is really a matrix and the chain rule amounts to matrix multiplication.

The central concept of calculus, the **limit** is deep. Mathematicians needed many centuries to develop a rigorous understanding of this idea. At first glance it seems intuitively clear and one can quickly master using its properties to do computations. It is a subtle concept , however, when we actually need to *prove* a mathematical result. The initial goals of exploring multivariable calculus may be to master the tools of differentiating and integrating functions with multiple inputs and multiple outputs and to see how they help us understand the physical and social world around us. Studying multivariable calculus also provides us with opportunities to revisit and deepen our understanding of single variable calculus.

Throughout our development of multivariable calculus, we will employ a *spiral* approach: you will encounter the same concept at various stages with an increasing level of rigor, abstraction, complexity and power. We will emphasize how the multidimensional approach to the derivative and the integral grow naturally out of the one-dimensional versions and where they require fresh perspectives.

About the companion website

The website[1] for this file contains:

- Additional Exercises and Projects.
- More Extended Historical and Biographical Notes.
- Miscellaneous material (e.g. suggested readings etc).

Acknowledgements

- I am grateful for the feedback from all my students who engaged with various versions of this approach to multivariable calculus. Their responses have been very helpful in developing the material.
- A special word of thanks goes to my faculty colleagues for many insightful conversations on making learning and teaching more effective.
- Funding from the Middlebury College Faculty Leave Program was essential in providing the time and resources to prepare this text.
- I would also like to thank Daniel Pinto, Elizabeth Palmer, Deepthi Mohan, and the entire staff of Kendall Hunt for their work on this project.

[1] `www.middlebury.edu/~molinick`

- Duncan McCabe, an undergraduate mathematics major at Middlebury, did splendid work preparing solutions manuals, which should helpful to students and instructors.

- I'm much indebted to my family and friends for their support and encouragement.

- I want to express deep gratitude to my wife Judy who not only was superb proof reader, syntax checker and wordsmith, but unselfishly gave up use of our living and dining room for 15 months as I filled them with books, papers, legal pads, pencils, pens, USB drives, and all the other clutter that goes with my writing projects.

All remaining mistakes and errors are, of course, my responsibility. Please report them to me as well any questions, comments, and suggestions that might be useful for subsequent editions.

Michael Olinick
molinick@middlebury.edu

1

Remembrance of Things Past

> *"If I have seen further than others, it is by standing upon the shoulders of giants."*
>
> – Isaac Newton, 1643 – 1727

In this introductory chapter, we review the major concepts of single variable calculus and linear algebra that are crucial for our study of the multivariable situation. The two great pillars of calculus, derivatives and integrals, are initially studied separately before we see how they are linked together through the Fundamental Theorem of Calculus. Linear Algebra provides vectors and matrices, tools which enable us to treat a collection of numbers as a single entity.

1.1 Caclulus

There are many situations whose analysis leads to the discovery or creation of the principal tools of calculus: differentiation and integration. We begin with two problems, one with a physical background and one with a geometric one, that both lead to the derivative.

Question 1: How fast is an object moving at a particular moment?

You may well have begun your study of calculus with considering the situation of an object moving along a straight line in such a way that its position on the line at any time t was described by a function f, the value $f(t)$ giving the object's location at that time. The question is then posed: How fast is the object moving precisely at the time a? In other words, what is the *instantaneous velocity* at a?

The answer lies in thinking about the *average velocity* during a brief time interval beginning at $t = a$ and ending at $t = a + h$ and asking what happens to that average velocity as the length of the time interval shrinks. If the average approaches a limiting value, we'll call that limit the instantaneous velocity. Using the idea that the average velocity is the ratio of the change of position to the change of time, we have

$$\text{Average Velocity on } [a, a+h] = \frac{f(a+h) - f(a)}{h}$$

and

$$\text{Instantaneous Velocity at } t = a = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

if this limit exists.

If the limit exists, then we say f is *differentiable at a* and use the notation $f'(a)$ for the limit. The number $f'(a)$ is called the *derivative of f at a.*

Question 2: If we understand the behavior of a curve at a particular point P on that curve, what can we say about its behavior at a point close to P?

Single Variable Calculus considers Question 1 when the curve is the graph of a function and we can construct a tangent line to the curve at the point P. We can build the tangent line if we know the location of the point and how rapidly the curve's height is changing there. Consider the picture below in Figure 1.1, the most important diagram from beginning calculus.

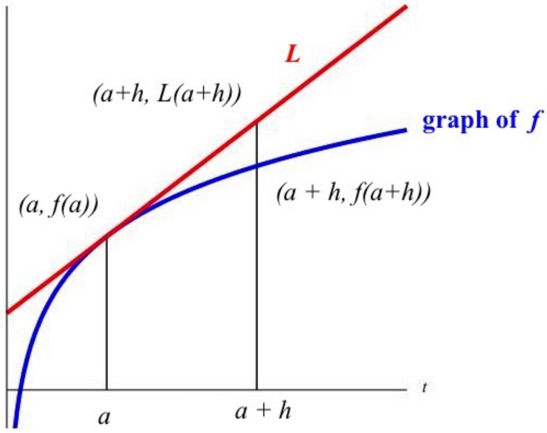

Figure 1.1

Here f is a real-valued function defined on some strip of the real line, typically a ray or an interval which might be open or closed. The function f is differentiable at the number a and $a + h$ is a close-by neighbor. The line L is the tangent line to the graph of f at the point $P = (a, f(a))$.

What exactly do we mean by the "tangent line" to a curve at a point? We learn in calculus that it is the line passing through the point that best approximates the curve near that point. There is an important alternate, but equivalent, perspective. If we place a sequence of lenses of ever increasing magnification over the point, the piece of the curve we will see will appear more and more like a straight line segment. Roughly speaking, the tangent line is what the curve would look like if we could blow it up indefinitely. We call this property *local linearity*

The tangent line has slope $f'(a)$ and passes through P so we can represent it by a simple equation $L(t) = f(a) + f'(a)(t - a)$. It is very simple to determine values of $L(t)$ for any value of t. The computation only involves the simplest arithmetic operations of subtraction, multiplication and addition. Calculation of values of $f(t)$ may be quite a bit more challenging as a formula for $f(t)$ could involve abstruse combinations of logarithmic, exponential, trigonometric and inverse trigonometric functions. It's a lot easier to compute $L(a+h)$ than $f(a+h)$ so it's tempting to use $L(a+h)$ to *estimate* or *approximate* $f(a + h)$. Our definition of the tangent line gives us license to succumb a bit to this temptation. If h is very small in magnitude, then $L(a+h)$ will approximate $f(a+h)$ fairly well. Roughly speaking, the smaller

h is, the better the approximation will be. More exactly,

$$f(a+h) \approx L(a+h) = f(a) + hf'(a)$$

with the relative difference

$$\frac{f(a+h) - L(a+h)}{h} = \frac{f(a+h) - f(a) - hf'(a)}{h} = \frac{f(a+h) - f(a)}{h} - f'(a)$$

having a limit of 0 as h approaches 0.

You learn quickly that, among other things,

- We can use the same definition for an arbitrary real-valued function of the real numbers, not just those describing positions of moving objects. The *Difference Quotient* $\frac{f(a+h)-f(a)}{h}$ measures an average rate of change. The derivative then tells us how quickly a function is changing at a given input so it has a myriad of interesting applications. **A fundamental concern of Calculus is understanding *change*, how rapidly or slowly things are fluctuating.**
- The limit of the Difference Quotient does not always exist. The classic example is $f(t) = |t|$ at $a = 0$.
- Sometimes the limit is "obvious" as for example, $\lim_{h\to 0}(5+h) = 5$
- Often, however, it is difficult to see what the limit might be or even whether it exists at all. Consider the examples

$$\lim_{h\to 0} \frac{\sin h}{h} \text{ or } \lim_{h\to 0} \sin \frac{1}{h}$$

- Such examples require us to have a careful and rigorous definition of *limit* if we are to make substantial progress in understanding the behavior of functions.
- The *limit* concept is foundational in the study of derivatives, integrals, sequences and series.

1.1.1 Limits, Continuity and the Derivative

The core concept of calculus is *limit*. The limit idea is explicit in the definition of the derivative and the definite integral. It is a subtle concept that mathematicians wrestled with for several centuries before reaching a full understanding and a rigorous definition. The challenge is developing a clear and unambiguous meaning of the intuitive notion of "nearness."

We often use terms like *nearness* and *closeness* interchangably. Standard discussions of limits, for example, involve terms such as "arbitrarily close" and "sufficiently close." We can operationalize *closeness* through the notion of *distance*: two objects are close if the distance between them is small.

For a pair of numbers, we can measure the distance between them by the size of their difference. The distance between a and b is given by $|a-b|$. For example, the distance between 3 and 5 or between -3 and -5 is 2 because $|3-5| = |5-3| = |-3-(-5)| = |-5-(-3)| = 2$. Note also that $|a-b| = \sqrt{(a-b)^2}$. If we visualize the numbers a and b as points on the real number line, then the distance between them is simply the length of the interval separating them.

It's very convenient to consider the set of all points that lie within a certain distance of some given point. If r is positive number and x_0 is a point, then the set of all points x that lie within distance r of x_0 is the set of points x satisfying

$$|x - x_0| < r$$

and is called a **r-neighborhood** of x_0. Geometrically, the r-neighborhood of x_0 is the open interval $(x_0 - r, x_0 + r)$.

By a **deleted neighborhood** or **punctured neighborhood** of x_0 we mean a neighborhood of x_0 which excludes x_0. A deleted neighborhood would be the set of all points x such that $0 < |x - x_0| < r$ for some positive radius r. Geometrically it is the open interval $(x_0 - r, x_0 + r)$ with the center deleted.

The formal definition of the limit of a function uses the ideas of neighborhood and deleted neighborhood:

Definition: Limit of a Function: If f is a real-valued function defined on a deleted neighborhood of a and L is a real number, then

$$\lim_{x \to a} f(x) = L$$

means that for every neighborhood $\mathcal{N}$ of L, there is a *deleted neighborhood* $\mathcal{M}$ of a such that $f(x)$ is in $\mathcal{N}$ whenever x is in $\mathcal{M}$.

We can also write this last condition as an "ϵ, δ" statement: "for each positive number ϵ, there is some positive number δ such that $|f(x)-L| < \epsilon$ whenever $0 < |x-a| < \delta$".

Note that the function f may or may not be defined at a itself. If f does have a value at a and that value is equal to L, we say that f is continuous at a:

Definition: Continuity of a Function at a Point: The function f is continuous at a means that for every neighborhood $\mathcal{N}$ of L, there is a *neighborhood* M of a such that $f(x)$ is in $\mathcal{N}$ whenever x is in $\mathcal{M}$. A function is *continuous on a set* if it is continuous at each point of the set.

Let's observe that if $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ are points in the plane, then the Pythagorean Theorem conveniently gives the distance between them, which we will denote as $|\mathbf{a} - \mathbf{b}|$, as

$$|\mathbf{a} - \mathbf{b}| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

.

The arithmetic and algebraic properties of limits and continuity are nice. The limit of a sum is the sum of the limits, the limit of a product is the product of the limits, and so on. More precisely, we have
Theorem 1.1.1.

$$\textit{If } \lim_{x\to a} f(x) \textit{ and } \lim_{x\to a} g(x) \textit{ exist, then}$$

$$(1) : \lim_{x\to a}[f(x) + g(x)] = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$$

$$(2) : \lim_{x\to a}[f(x) - g(x)] = \lim_{x\to a} f(x) - \lim_{x\to a} g(x)$$

$$(3) : \lim_{x\to a}[cf(x)] = c\lim_{x\to a} f(x) \textit{ for any constant } c$$

$$(4) : \lim_{x\to a}[f(x) \times g(x)] = \lim_{x\to a} f(x) \times \lim_{x\to a} g(x)$$

$$(5) : \lim_{x\to a}\left[\frac{f(x}{g(x)}\right] = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}, \textit{ if } [\lim_{x\to a} g(x) \neq 0$$

$$(6) \lim_{x\to a}[f(x)]^n = [\lim_{x\to a}[f(x)]^n, \textit{ if } n \textit{ is a positive integer}$$

$$(7)\lim_{x\to a} c = c \textit{ for al constants } c$$

$$(8)\lim_{x\to a} x = a$$

As a consequence of this theorem, it is easy to show that the sum, difference and product of continuous functions are continuous, as is the quotient wherever the denominator is nonzero. Polynomial functions are continuous everywhere and rational functions (quotients of polynomials) are continuous except possibly where the denominator is 0. Powers, roots, exponential and logarithm functions are similarly behaved as are the standard trigonometric functions wherever they are defined.

Continuous functions also have intermediate value and extreme value properties:

Theorem 1.1.2. ***Intermediate Value Theorem*** *Suppose f is continuous on the closed interval [a,b] and K is any number between f(a) and f(b), then there is at least number c in the interval with f(c) = K.*

Theorem 1.1.3. ***Extreme Value Theorem*** *If f is continuous on the closed interval I= [a,b] then there are points c and d in I such that* $f(c) \geq f(x)$ *for all x in I and* $f(d) \leq f(x)$ *for all x in I.*

Finally, recall that that differentiability implies continuity:

Theorem 1.1.4. *If f is differentiable at a, then f is continuous at a.*

The converse of this last theorem is false in general. An important example is $f(x) = |x|$ which is continuous, but not differentiable at 0.

1.1.2 Properties and Uses of the Derivative

Let's recall the basic properties of the derivative. if f and g are each differentiable at a, then so are their sum, difference, product and, if $g(a) \neq 0$, quotient $\frac{f}{g}$. Moreover, we have the following elementary formulas:

1. If f is a constant function, then $f'(a) = 0$ for all values of a.
2. For any constant c, $(cf)'(a) = c\,f'(a)$.
3. The derivative of a sum is the sum of the derivatives: $(f + g)'(a) = f'(a) + g'(a)$.

4. Product Rule: $(f \times g)'(a) = f'(a) \times g(a) + f(a) \times g'(a)$.
5. Reciprocal Rule: $(\frac{1}{g})'(a) = \frac{-g'(a)}{g^2(a)}$ if $g(a) \neq 0$.
6. Quoitent Rule: $(\frac{f}{g})'(a)\frac{f'(a)g(a)-f(a)g'(a)}{g^2(a)}$ if $g(a) \neq 0$.

The other major result about differentiation is the *Chain Rule* for finding the derivative of a composition of functions. If g is defined on the range of f, then the composition $f \circ g$ is defined by $(f \circ g)(x) = f(g(x))$. if g is differentiable at a and f is differentiable at $g(a)$, then we have

$$\text{Chain Rule: } (f \circ g)'(a) = f'(g(a)) \times g'(a)$$

We collect the derivatives of commonly used functions in the table below:

Table 1.1: Table of Derivatives

Function	Derivative	Function	Derivative
x^n	nx^{n-1}	$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
e^x	e^x	$\ln x$	$1/x$
$\sin x$	$\cos x$	$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$
$\cos x$	- $\sin x$	$\arccos x$	$\frac{-1}{\sqrt{1-x^2}}$
$\tan x$	$\sec^2 x$	$\arctan x$	$\frac{1}{1+x^2}$
$\cot x$	$-\csc^2 x$	$\text{arccot}\, x$	$\frac{-1}{1+x^2}$
. [1 ex] $\sec x$	$\sec x \tan x$	$\text{arcsec}\, x$	$\frac{1}{x\sqrt{x^2-1}}$
$\csc x$	$-\csc x \cot x$	$\text{arccsc}\, x$	$\frac{-1}{x\sqrt{x^2-1}}$

The derivative is useful for sketching the graph of a function and for solving optimization problems by locating extreme values. If f is differentiable on an open interval a with $f'(a)$ positive, then f is increasing at a while it is decreasing there if $f(a)$ is negative. If the first derivative f' changes sign from negative to positive at a, then f has a *relative minimum* at a. There is a *relative maximum* if the sign changes from positive to negative. This is the gist of what is often called *The First Derivative Test.*

There is also a *Second Derivative Test* that can often tell you the nature of a point where the first derivative is 0. If $f'(a) = 0$ and $f''(a) > 0$, then f has a local minimum at a but a local maximum if $f''(a) > 0$. Unfortunately, the test is silent in the case that both the first and second derivatives are 0 at a.

The *Mean Value Theorem* is one of the most important results about differentiation of real-valued functions of real variables. Recall its formal statement:
Theorem 1.1.5. *(Mean Value Theorem) If f is continuous on the closed interval [a,b] and differentiable on the open interval (a, b), then there is at least one number c in (a, b) such that* $f'(c) = \frac{f(b)-f(a)}{b-a}$

One of the consequences of the Mean Value Theorem is that two functions which have the same derivative can only differ by a constant; that is, if $f'(t) = g'(t)$ for all t in some interval, then there is a constant C such that $f(t) = g(t) + C$ for all t in that interval. This result is critical in addressing the *Antidifferentiation Problem*: Given a function f, find an **antiderivative**, a function whose derivative is f. The Mean Value Theorem tells us that there can not be two *essentially* different antiderivatives. The graphs of two antiderivatives of f would always be the same fixed distance apart.

1.1.3 The Definite Integral

A major motivation for studying the derivative is a geometric one: can we approximate the behavior of a curve near a particular point by a straight line? This is a question about the local of behavior of the function whose graph is the curve of concern, The study of the integral often also starts with a geometric question about a curve, a question that addresses global behavior.

Consider a positive valued function f defined on a closed interval $I = [a, b]$ as shown in Figure 1.2.

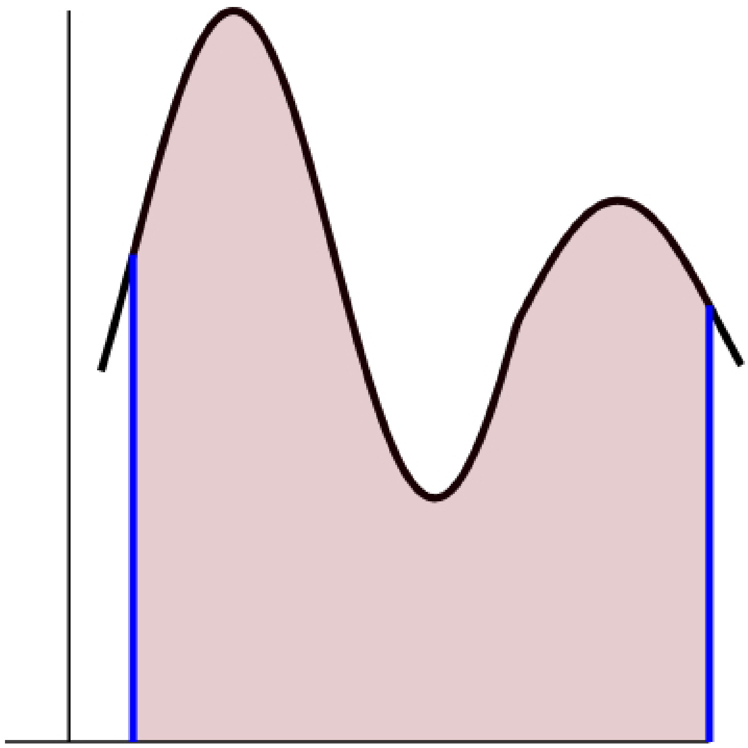

Figure 1.2

The graph of the function together with the horizontal axis and the vertical lines at a and b form the boundaries of a region $\mathcal{R}$ in the plane. What is the area of this region? The definite integral was devised to answer this question and related ones. The underlying idea is a geometric one: split the region into thin strips and approximate the area of each strip by the area of a rectangle. Adding up these areas gives an approximation to the area of $\mathcal{R}$. We can obtain better approximations by choosing more and more thinner and thinner strips. The region's area is then defined to be the limiting value of such a process.

Revisiting this process, recall that we start with a *partition* P of the interval $[a, b]$ which is a collection of points $t_0, t_1, ..., t_{j-1}, t_j, ..., t_n$ in the interval with $a = t_0 < t_1 < t_2 < ... < t_{j-1} < t_j < ... < t_{n-1} < t_n = b$. These partition points carve up $[a, b]$ into n subintervals with a typical subinterval denoted by $[t_{j-1}, t_j]$ whose width is Δt_j is $t_j - t_{j-1}$. We then select a point t_j^* in each subinterval and form the *Riemann Sum*

$$\sum_{j=1}^{n} f(t_j^*)\Delta t_j$$

. If $f(t)$ is positive throughout the interval $[a, b]$, then the product $f(t_j^*)\Delta t_j$ represents the area of a rectangle of height $f(t_j^*)$ and width Δt_j as in Figure 1.3.

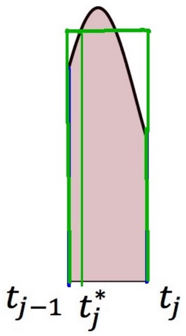

Figure 1.3

Note that a small piece of the region $\mathcal{R}$ sticks outside the rectangle and that some part of the rectangle sits outside $\mathcal{R}$, but if the width of the rectangle is very small and the graph of f does not "wiggle" very much over the interval $[t_{j-1}, t_j]$, then the area of the rectangle should be a good approximation of the area of the shaded region.

The *mesh* of a partition is the length of the longest subinterval. If the number of subintervals increases in a manner which makes the mesh shrink toward 0, then we expect that the associated Riemann sums approach a limit that would make sense to call the actual area of $\mathcal{R}$.

Definition The **definite integral** of a function f defined on the closed interval $[a, b]$ we denote by $\int_a^b f(t)\, dt$ is

$$\int_a^b f(t)\, dt = \lim_{mesh\ P \to 0} \sum_{j=1}^{n} f(t_j^*)\, \Delta t_j$$

if this limit exists.

There are certainly many functions for which the definite integral does not exist. We are fortunate that most of the functions we will encounter in our study of calculus and its applications are sufficiently well-behaved to insure that they have integrals. The principal theorem we will use is

Theorem 1.1.6. *If f is continuous on $[a, b]$, then $\int_a^b f(t)\, dt$ will exist.*

There are also functions on closed intervals which fail to be continuous but for which the definite integral does exist. Such functions are appropriately called *integrable* functions .

We summarize some familiar properties of the definite integral. If f and g are integrable on the interval $[a, b]$ and α and β are any real numbers, then $\alpha f + \beta$ is integrable on $[a, b]$. Moreover, each of the following is true:

- $\int_a^b \alpha\, dt = \alpha(b - a)$.
- $\int_a^b (\alpha f + \beta g)(t)\, dt = \alpha \int_a^b f(t)\, dt + \beta \int_a^b g(t)\, dt$.
- if f is integrable on some closed interval I containing the numbers a, b and c, then $\int_a^b f(t)\, dt = \int_a^c f(t)\, dt + \int_c^b f(t)\, dt$.
- Suppose we have a triple of numbers a, b, c with $a < b < c$ and f is integrable on both $[a, b]$ and $[b, c]$. The f is integrable on $[a, c]$ with $\int_a^c f(t)\, dt = \int_a^b f(t)\, dt + \int_b^c f(t)\, dt$.
- If $f(t) \geq g(t)$ for all t in $[a, b]$, then $\int_a^b f(t)\, dt \geq \int_a^b g(t)\, dt$.

In the rectangular $xy-$coordinate system, if $y = f(x)$ is a nonnegative valued function on the closed interval $[a, b]$, then the area bounded by the horizontal axis, the vertical lines at $x = a$ and $x = b$ and the graph of f is given by $\int_a^b f(x)\, dx$.. Recall that in the polar $(r, \theta)-$ coordinate system, if $r = f(\theta)$ is a nonnegative valued function on the closed θ interval $[\theta_1, \theta_2]$, then the area bounded by rays $\theta = \theta_1$ and $\theta = \theta_2$ and the graph of f is given by $\int_{\theta_1}^{\theta_2} \frac{r^2}{2}\, d\theta = \int_{\theta_1}^{\theta_2} \frac{f(\theta)^2}{2}\, d\theta$.

There is also a useful Intermediate Value Theorem for definite integrals.
Theorem 1.1.7. *If f is continuous on $[a, b]$ then there is at least one number K in the interval for which $\int_a^b f(t)\, dt = f(K)(b - a)$.*

We call the number $\frac{1}{b-a} \int_a^b f(t)\, dt$ the **average value** of f on $[a, b]$.

1.1.4 Fundamental Theorem of Calculus

The paths we followed in getting to the derivative and the integral began in different places with different questions. The derivative developed in response to the problem of determining how fast an object is moving at a particular instant if we know its position at every moment. The definite integral was the result of asking what is the area of a region in the plane whose boundary contains the graph of a function.

The discovery and investigation of the intimate link between the derivative and the integral is generally credited to independent work by Isaac Newton and Gottfried Leibniz. They formulated the *Fundamental Theorem of Calculus.* In today's notation, it takes the form

Theorem 1.1.8. *Fundamental Theorem of Calculus. Suppose f is a continuous function on the closed interval [a,b].*

- *Part I: One antiderivative of f on [a,b] is the function G defined by*

$$G(x) = \int_a^x f(t)\, dt$$

  for all x in [a,b]

- *Part II: If F is any antiderivative of f on [a,b], then*

$$\int_a^b f(t)\, dt = F(b) - F(a)$$

Our most familiar application of the Fundamental Theorem is to avoid computing a definite integral via the limit of Riemann sums, but finding an antiderivative first and evaluating it at the endpoints.

One of the major goals of multivariable calculus is to find generalizations of the Fundamental Theorem of Calculus for vector-valued functions of vectors. We will see that there are many situations where there is an equality between an operation on a function along the boundary of a set and another operation on a derivative of that function over the entire set. The theorems of Green, Gauss and Stokes we will study in Chapter 8 are major examples.

1.1.5 Taylor's Theorem

Taylor's Theorem (Brook Taylor, 1685 – 1731) and its close relative, Maclaurin's Theorem (Colin Maclaurin, 1698 – 1746) concern power series representations of functions which allow the calculation of very close approximation to the values of complicated functions by evaluating polynomials, a computation that involves only multiplication and addition. If f is a real-valued function of one variable which has derivatives of all orders in an open interval centered at x_0, then

$$f(x_0+h) = f(x_0)+f'(x_0)h+\frac{f''(x_0)}{2}h^2+\frac{f^{(3)}(x_0)}{3!}h^3+...+\frac{f^{(k)}(x_0)}{k!}h^k+R_k(x_0,h)$$

where the remainder

$$R_k(x_0, h) = \int_{x_0}^{x_0+h} \frac{(x_0 + h - s)^k}{k!} f^{(k+1)}(s)\, ds.$$

For h close to 0, the remainder is very small in the sense that

$$\lim_{h \to 0} \frac{R_k(x_0, h)}{h^k} = 0.$$

The remainder term is also called the **error** term because it measures the difference between the true value of the function and the approximation you get by stopping with the h^k term. We can also express the remainder term in the *Lagrange* form

$$R_k(x_0, h) = \frac{f^{(k+1)}(c)}{(k+1)!} h^{k+1} \text{ for some } c \text{ between } x_0 \text{ and } x_0 + h.$$

Proofs of Taylor's Theorem may be found in many single variable calculus texts [1]

Let's examine briefly the First Derivative and Second Derivative Tests for Local Maxima and Minima through the lens of Taylor's Theorem. Setting $k = 1$ in the theorem yields

$$f(x_0 + h) = f(x_0) + f'(x_0)h + R_1(x_0, h) \text{ where } \lim_{h \to 0} \frac{R_1(x_0, h)}{h} = 0.$$

Now if $f'(x_0)$ is positive, then $f'(x_0)h > 0$ when $h > 0$ and $f'(x_0)h < 0$ when $h < 0$. Choosing h so small that we can neglect the remainder term, we see that $f(x_0 + h)$ is smaller than $f(x_0)$ when $h < 0$ and greater than $f(x_0)$ when $h > 0$ so f is increasing at x_0. A similar argument shows that f is decreasing at x_0 when $f'(x_0) < 0$.

The Second Derivative Test allows us to determine the nature of a critical point if the value of the second derivative is nonzero. To be more precise, the test says that if $f'(x_0) = 0$ and $f''(x_0) >$, then there is local minimum of f at x_0 and a local maximum if $f''(x_0) < 0$. Letting $k = 2$ in Taylor's Theorem when $f'(x_0) = 0$ gives us

$$f(x_0 + h) = f(x_0) + 0 + \frac{f''(x_0)}{2} h^2 + R_2(x_0, h)$$

[1] My favorite is the 6th edition of *Calculus* by Ear; Swokowski, Michael Olinick, and Dennis Pence.

where the remainder term contributes a negligible amount when h is tiny. Thus, for very small values of h (positive or negative)

$$f(x_0 + h) - f(x_0) \approx \frac{f''(x_0)}{2}h^2$$

where the sign of the difference is the same as the sign of $f''(x_0)$. If $f''(x_0) > 0$, then $f(x_0 + h) > f(x_0)$ so f has a local minimum at x_0. You will see an analogous result for real-valued functions of several variables in Chapter 5.

1.2 Linear Algebra

1.2.1 Vectors

John Harris (1666 – 1719) was an English clergyman with a strong interest in mathematics and science. His *Lexicon Technicum*, published in 1704, gives the earliest usage in English of the word *vector*. In the context of astronomy, Harris writes, a vector is a line drawn between a planet and the Sun "because 'tis that Line by which the Planet seems to be carried round its Centre, and with which it describes proportional Areas in proportional Time." As the planet orbits about the Sun, the line varies in length and direction. *Vector* comes from Latin, meaning "one who carries or conveys."

In 1846, the Irish mathematician Sir William Rowan Hamilton (1805?1865) introduced *vector* into the vocabulary of mathematics as a "quantity having magnitude and direction" in a paper published in *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. In pure mathematics, Hamilton is best known for his creation of quaternions, but he made notable contributions to physics. His work on Newtonian mechanics paved the way for classical field theories and quantum mechanics.

Walking outside on a blustery day, you can feel the wind blowing in a particular direction with a particular velocity at each step. Imagine a jet plane climbing into the sky at a certain angle and airspeed. Think about a river otter scurrying across the water's edge, zigzagging directions and changing the pace of its motion. These are all instances which can be described with vectors.

We can specify direction and magnitude with a set of numbers, typically three describing a direction in space and one for the magnitude. Mathematicians soon realized that a vector was a convenient way to deal with a multitude of numbers as a single object. They developed a language and rules for manipulating and using vectors in a wide variety of contexts.

Definition A **vector** $\mathbf{x}$ in $\mathbb{R}^n$ is an ordered n-tuple of real numbers, $\mathbf{x} = (x_1, x_2, ..., x_n)$ We call the individual numbers the **components** of the vector. In most of our treatment of vectors as inputs or outputs of functions, a vector $\mathbf{x}$ is really a **column** vector (or $n \times 1$ matrix) and should be written vertically

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}.$$

To conserve on space, we'll write these vectors horizontally.

For $n = 2$, we identify $\mathbb{R}^2$ as the points in the plane. Here a vector would have the form of an ordered pair $\mathbf{x} = (x_1, x_2)$. By *ordered*, we mean for example that (2,3) is different from (3,2). Similarly, we use $\mathbb{R}^3$ as the mathematical structure representing our usual idea of 3-dimensional space. The symbol $\mathbb{R}^1 = \mathbb{R}$ represents the real numbers.

We will use bold face letters such as $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{a}$, $\mathbf{b}$ to denote vectors. An alternative notation you are likely to see displays arrows above the letters; for example, $\vec{x}, \vec{y}, \vec{z}, \vec{a}, \vec{b}$. Lightface letters, such as t, α, β, denote **scalars** which are real numbers.

There are four operations involving vectors we will use throughout our study of multivariable calculus: sum, scalar multiple, dot product, and cross product. Recall first that we add two vectors by adding their corresponding components and we form the product of a scalar and a vector by multiplying each component by that scalar.

Definition If $\mathbf{x} = (x_1, x_2, ..., x_n)$ and $\mathbf{y} = (y_1, y_2, ., , y_n)$ are vectors in $\mathbb{R}^n$, then their sum $\mathbf{x} + \mathbf{y}$ is the vector

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, ..., x_j + y_j, ..., x_n + y_n)$$

For a scalar α and a vector $\mathbf{x} = (x_1, x_2, .., x_j, ..., x_n)$, the scalar multiple is

$$\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, ..., \alpha x_j, ..., \alpha x_n$$

By a **zero vector 0**, we mean a vector each of whose components is the scalar zero. We will usually write $-(x)$ for $(-1)\mathbf{x}$

In your Linear Algebra course, you verified that these operations satisfy many nice algebraic and arithmetical properties. For all vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ and scalars α, β,we have

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
3. $\mathbf{x} + \mathbf{0} = \mathbf{x}$
4. $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
5. $1\mathbf{x} = \mathbf{x}$
6. $0\mathbf{x} = \mathbf{0}$
7. $\alpha(\mathbf{x} + \mathbf{y} = \alpha\mathbf{x} + \alpha\mathbf{y}$
8. $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$
9. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$

Definition The **length** $|\mathbf{x}\ |$ of a vector $\mathbf{x}$ in $\mathbb{R}^n$ is the number

$$|\mathbf{x}| = \sqrt{x_1^2 + x_2^2 + ... + x_j^2 + ... + x_n^2} = \sqrt{\sum_{j=1}^{n} x_j^2}$$

A vector of length 1 is called a **unit** vector. There are three special unit vectors in $\mathbb{R}^3$: $\mathbf{i} = (1, 0, 0), (\mathbf{j} = (0, 1, 0), \mathbf{k} = (0, 0, 1)$. The linear algebra perspective is that $\mathbb{R}^3$ is a vector space with $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ as a basis. In a similar fashion, the **standard basis** for $\mathbb{R}^n$ is the set $\{\mathbf{e_1}, \mathbf{e_2}, .., \mathbf{e_j}, ..., \mathbf{e_n}, \}$ where for each j, the vector $\mathbf{e_j}$ is all 0's except for the jth coordinate which is 1.

Our notation for the length of a vector,$|\mathbf{x}|$ looks like the absolute function notation for a real number. In fact, if $n = 1$, then the vector $\mathbf{x}$ really coincides with the real number $x = x_1$; its length and absolute value are the same. We can think of the length of a vector in terms of a function L from $\mathbb{R}^n$ to $\mathbb{R}^1$ with $L(\mathbf{x}) = |\mathbf{x}|$. This function has several nice properties:

Theorem 1.2.1. *For all vectors* $\mathbf{x}$ *and all scalars* α*, we have*

1. $|\mathbf{x}| \geq 0$ *for all* $\mathbf{x}$
2. $|\mathbf{x}| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$
3. $|\alpha\mathbf{x}| = |\alpha||\mathbf{x}|$

Once we have defined the *length* or *magnitude* of a vector, it is easy to measure the distance between two vectors as the length of their difference:

Definition. If $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mathbb{R}^n$, then the **distance** $D(\mathbf{x}, \mathbf{y})$ between them is $D(x, y) = |\mathbf{x} - \mathbf{y}|$.

Here are some properties of the distance function:

Theorem 1.2.2. *If* $\mathbf{x}$*,* $\mathbf{y}$*, and* $\mathbf{z}$ *are any vectors in* $\mathbb{R}^n$*, then*

1. $D(\mathbf{x}, \mathbf{y}) \geq 0$
2. $D(\mathbf{x}, \mathbf{y}) = 0$ *if and only if* $\mathbf{x} = \mathbf{y}$
3. $D(\mathbf{x}, \mathbf{y}) = D(\mathbf{y}, \mathbf{x})$
4. *(Triangle Inequality)* $D(\mathbf{x}, \mathbf{z}) \leq D(\mathbf{x}, \mathbf{y}) + D(\mathbf{y}, \mathbf{z})$

Besides multiplying a vector by a scalar and adding two vectors together, there are in various contexts, operations involving multiplication of two vectors.

Definition The **Dot Product** $\mathbf{x} \cdot \mathbf{y}$ in $\mathbb{R}^n$ is the scalar defined by

$$\mathbf{x} \cdot \mathbf{y} = \sum_{j=1}^{n} x_j y_j = x_1 y_1 + x_2 y_2 + ... + x_j y_j + ... + x_n y_n$$

Here are some of the important properties of the dot product:

Theorem 1.2.3. *For any vector* $\mathbf{x}$ *in* $\mathbb{R}^n$ *and any scalar* α*, we have*

1. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
2. $\mathbf{x} \cdot \mathbf{x} = |\mathbf{x}|^2$
3. $(\alpha\mathbf{x}) \cdot \mathbf{y} = \alpha(\mathbf{x} \cdot \mathbf{y})$
4. $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$
5. $\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos\theta$ *where* θ *is the angle between* $\mathbf{x}$ *and* $\mathbf{y}$
6. $\mathbf{x} \cdot \mathbf{y} = 0$ *if and only if the vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *are orthogonal (perpendicular) to each other.*

We can also state one of the most important inequalities in mathematics, the **Cauchy – Schwarz Inequality.**

Theorem 1.2.4. ***Cauchy – Schwarz Inequality***: *For any pair of vectors* $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^n$, *we have*

$$|\mathbf{u}\cdot\mathbf{v}| \le |\mathbf{u}||\mathbf{v}|$$

Proof: The inequality does follow from part (5) of Theorem 1.2.3 since $|\cos\theta| \le 1$ for all θ. We'll present a purely algebraic proof, however, based on a simple result about quadratic polynomials.

Suppose $Q(x)$ is the quadratic ax^2+bx+c. If $Q(x)\ge 0$ for all values of x, then it cannot have two distinct real roots. From the quadratic formula, it must be the case that the discriminant $b^2-4ac\le 0$ so $b^2\le 4ac$.

Given the vectors $\mathbf{u}=(u_1,u_2,...,u_j,...,u_n)$ and $\mathbf{v}=(v_1,v_2,...,v_j,...,v_n)$, we form the polynomial $Q(x)=\sum_{j=1}^n(u_jx+v_j)^2$. Since each term in the sum is non-negative, we have $Q(x)\ge 0$ for all x. On the other hand, expanding the sum, we have

$$\begin{aligned}
Q(x) &= \sum_{j=1}^n (u_jx+v_j)^2 \\
&= \sum_{j=1}^n \left(u_j^2x^2+2u_jv_jx+v_j^2\right) \\
&= \sum_{j=1}^n (u_j^2x^2) + \sum_{j=1}^n (2u_jv_jx) + \sum_{j=1}^n v_j^2 \\
&= \left(\sum_{j=1}^n u_j^2\right)x^2 + \left(\sum_{j=1}^n 2u_jv_j\right)x + \sum_{j=1}^n v_j^2 \\
&= ax^2+bx+c
\end{aligned}$$

Thus

$$\left(2\sum_{j=1}^n u_jv_j\right)^2 = b^2 \le 4ac = 4\left(\sum_{j=1}^n u_j^2\right)\left(\sum_{j=1}^n v_j^2\right)$$

so

$$|\mathbf{u}\cdot\mathbf{v}|^2 = \left(\sum_{j=1}^n u_jv_j\right)^2 \le \left(\sum_{j=1}^n u_j^2\right)\left(\sum_{j=1}^n v_j^2\right) = |\mathbf{u}|^2|\mathbf{v}|^2$$

Taking the square roots of these non-negative quantities yields

$$|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|$$

□

1.2.2 Linear Independence

A **linear combination** of vectors $\{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}\}$ is a vector of the form $\mathbf{v} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_k\mathbf{v_k}$ where the coefficients $c_1, c_2, ..., c_k$ are constant scalars. An **interesting linear combination** is a linear combination where at least one coefficient is nonzero.

Example Let $\mathbf{v_1} = (1, 5), \mathbf{v_2} = (7, -1), \mathbf{v_3} = (23, 7)$ be three vectors in $\mathbb{R}^2$. Then $2\mathbf{v_1} + 3\mathbf{v_2} - 1\mathbf{v_3}$ is an interesting linear combination. Note that this linear combination is

$$\begin{aligned} 2(1,5) + 3(7,-1) - 1(23,7) &= (2,10) + (21,-3) + (-23,-7) \\ &= (2 + 21 - 23, 10 - 3 - 7) = (0,0) = \mathbf{0}. \end{aligned}$$

We call a set of vectors **linearly dependent** if we can write the zero vector as an interesting linear combination of the vectors in the set. Thus $\{\mathbf{v_1} = (1,5), \mathbf{v_2} = (7,-1), \mathbf{v_3} = (23,7)\}$ is a linearly dependent set. Note that we can always express the zero vector as an *uninteresting* linear combination of any nonempty set of vectors by taking all of the c_i coefficients equal to 0.

A set of vectors is **linearly independent** if there is no interesting linear combination of those vectors equal to the zero vector. An equivalent way to state this condition is $S = \{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_k}\}$ is linearly independent if and only if

$$\mathbf{0} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_k\mathbf{v_k} \text{ implies each } c_i = 0$$

The **Span** of a set S of vectors is the collection of all vectors which can be written as linear combinations of members of S. A **Basis** for $\mathbb{R}^n$ is a linearly independent set of vectors whose span is all of $\mathbb{R}^n$.

1.2.3 Matrices

A vector is just a string of symbols or numbers that can be written horizontally as a row (to save space) or vertically as a column. A **matrix** is a rectangular array of objects, treated as a single entity. J. J. Sylvester(1814

–1897), who introduced the word into mathematics in 1850, described it as "an oblong arrangement of terms."

The **rows** of a matrix are its horizontal lines of numbers; the **columns** are the vertical lines. The **dimensions** of a matrix are the number of rows followed by the number of columns. When the number of rows and columns are the same, we say the matrix is **square**. Here are five matrices with their dimensions:

Matrix A	Matrix B	Matrix C	Matrix D	Matrix E
$\begin{pmatrix} 12 & 13 & 07 \\ 11 & 4 & 09 \end{pmatrix}$	$\begin{pmatrix} 10 & \pi \\ 23 & -9 \\ 18 & \sqrt{2} \end{pmatrix}$	$\begin{pmatrix} 0 & 5 \\ 75 & 3 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 13 \\ 07 \end{pmatrix}$	$\begin{pmatrix} 1 & 2 & \ln 3 & e^{-1} \end{pmatrix}$
2 by 3	3 by 2	2 by 2 square	3 by 1 column vector	1 by 4 row vector

A generic $m \times n$ ("m by n") matrix has the form

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}$$

where the entry in row i and column j may be written as $A_{ij} = a_{i,j} = a_{ij}$

We multiply a matrix by a scalar by simply multiplying each entry of the matrix by that scalar:

$$(\alpha A)_{ij} = \alpha A_{ij}$$

We can form the sum of two matrices A and B of the same dimensions by simply adding the corresponding entries:

$$(A + B)_{ij} = A_{ij} + B_{ij}$$

Recall that **matrix multiplication** is a bit more complicated. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then we can form the product AB as an $m \times p$ matrix whose ijth entry is the dot product of the ith row of A and the jth column of B. To form a product AB, the number of columns of

A must equal the number of rows of B.

For our examples above AB is 2 by 2, BA is 3 by 3, and BC is 3 by 2, but CB is not defined. Even if A and B are two square n by n matrices such that AB and BA are both defined and n by n, AB need not equal BA; that is, matrix multiplication is not alway commutative. It is always associative, however: if A is $m \times n$, B is $n \times p$ and C is $p \times r$, then $A(BC) = (AB)C$.

The **diagonal entries** of a square matrix A are those entries A_{ij} where $i = j$; that is $a_{1,1}, a_{2,2}, ..., a_{n,n}$. An **Identity Matrix I** is a square matrix whose diagonal entries are each equal to 1 and whose nondiagonal entries are 0. The 2×2 and 3×3 identity matrices are

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Invertibility: if A and B are square matrices of the same size with the property that

$$AB = I = BA$$

then we say that A is **invertible** with **inverse** B. If C is also an inverse of A, then

$$B = IB = (CA)B = C(AB) = CI = C$$

. We see then that a matrix can have at most one inverse and we can denote it by A^{-1}.

Invertibility of square matrices is such an important topic in Linear Algebra that you learn many different equivalent conditions for invertibility; for example, the rows are linearly independent, the columns are linearly independent, the reduced row echelon form is the identity, and so on. A handy criterion for us will be the nature of the **determinant**: A square matrix A is invertible if and only if the determinant det A is nonzero.

The determinant of a 2×2 matrix A of the form $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $ad - bc$. For a 3×3 matrix, we can find the determinant by expansion along the first row:

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} = a_1 \det \begin{pmatrix} b_2 & b_3 \\ c_2 & c_3 \end{pmatrix} - a_2 \det \begin{pmatrix} b_1 & b_3 \\ c_1 & c_3 \end{pmatrix} + a_3 \det \begin{pmatrix} b_1 & b_2 \\ c_1 & c_2 \end{pmatrix}$$

We can compute determinants of larger size square matrices in a similar fashion. Expand along the first row. If A is an $n \times n$ matrix, then for each element $a_{1,j}$ in the first row, form $c_j = a_{1,j} \det A^{[i,j]}$ where $A^{[i,j]}$ is the $(n-1) \times (n-1)$ submatrix obtained by deleting row 1 and column j from A. Then the determinant of A is the alternating sum of the c_j's:

$$\det A = c_1 - c_2 + c_3 + ... + (-1)^{n-1} c_n$$

There is a nice geometric interpretation of the determinant which we will make use of in Chapter 6. If $X = \begin{pmatrix} x \\ y \end{pmatrix}$ is a vector in $\mathbb{R}^2$ and A is a 2×2 matrix, then AX is also a vector in $\mathbb{R}^2$. We can then view multiplication by A as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ with $f_A(X) = AX$. Let $\mathcal{S}$ be the unit square determined by the unit vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$; that is, the square with vertices (0,0), (1,0), (0,1) and (1,1). With $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ Then the function f_A takes the following action:

$$A \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}, A \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}, A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} a+b \\ c+d \end{pmatrix}$$

Thus f_A maps the unit square $\mathcal{S}$ onto a parallelogram $\mathcal{P}$ with vertices $(0,0)$, (a,b), (c,d), and $(a+b, c+d)$. The area of this parallelogram turns out to be $|\det A|$.

In a similar fashion, let $\mathcal{C}$ be the unit cube in $\mathbb{R}^3$ determined by the unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Its vertices are (0,0,0), (0,0,1), (1,0,0), (1,0,1), (0,1,0), (0,1,1), (1,1,0), and (1,1,1). If A is a 3×3 matrix, then the function f_A which maps each vector $\mathbf{X}$ to $A\mathbf{X}$ sends the unit cube to a parallelopiped $\mathcal{P}$ in $\mathbb{R}^3$. The volume of $\mathcal{P}$ is $|\det A|$.

There is a corresponding result in higher dimensions. Let $\mathcal{C}$ be the unit hypercube in $\mathbb{R}^n$ determined by the unit vectors $\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_j}, ..., \mathbf{e_n}$. If A is a

$n \times n$ matrix, then the function f_A which maps each vector $\mathbf{X}$ to $A\mathbf{X}$ sends the unit hypercube to a parallelopiped $\mathcal{P}$ in $\mathbb{R}^n$. The volume of $\mathcal{P}$ is $|\det A|$.

There is another type of product for a pair of 3-dimensional vectors. **Definition** The **cross product** $\mathbf{x} \times \mathbf{y}$ of two vectors in $\mathbb{R}^3$ is defined as the 3-dimensional vector

$$\mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1)$$

It's a bit of a challenge to remember this formula. There's an easy way to generate the cross product by using determinants:

$$\mathbf{x} \times \mathbf{y} = \det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix} = \det \begin{pmatrix} x_2 & x_3 \\ y_2 & y_3 \end{pmatrix} \mathbf{i} - \det \begin{pmatrix} x_1 & x_3 \\ y_1 & y_3 \end{pmatrix} \mathbf{j} + \det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \mathbf{k}$$

Note that the cross product is only defined for 3-dimensional vectors, unlike the dot product. Moreover, the cross product is itself a vector, while the dot product is a scalar.

We summarize some of the important properties of the cross product:

Theorem 1.2.5. *Properties of the Cross Product If* $\mathbf{x}, \mathbf{y},$ *and* $\mathbf{z}$ *are any three vectors in* $\mathbb{R}^3$ *and* α *is any scalar, then*

1. $\mathbf{x} \times \mathbf{y}$ *is orthogonal to both* $\mathbf{x}$ *and* $\mathbf{y}$
2. $\mathbf{y} \times \mathbf{x}$ = - $\mathbf{x} \times \mathbf{y}$
3. $\alpha\mathbf{x} \times \mathbf{y} = (\alpha\mathbf{x}) \times \mathbf{y} = \mathbf{x} \times (\alpha\mathbf{y})$
4. $\mathbf{x} \times (\mathbf{y} + \mathbf{z}) = \mathbf{x} \times \mathbf{y} + \mathbf{x} \times \mathbf{z}$
5. $(\mathbf{x} + \mathbf{y}) \times \mathbf{z} = \mathbf{x} \times \mathbf{z} + \mathbf{y} \times \mathbf{z}$

1.2.4 Eigenvalues and Eigenvectors

Suppose A is a square matrix, λ is a scalar and $\mathbf{x}$ is a nonzero vector such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

Then λ is an **eigenvalue** of A with associated **eigenvector x**.

The eigenvalues of A are roots of the **characteristic polynomial** $det(A - \lambda I)$; that is, solutions to the polynomial equation $\det(A - \lambda I) = 0$.

Example: Let $A = \begin{pmatrix} -8 & -10 & 5 \\ 20 & 22 & -10 \\ 30 & 30 & -13 \end{pmatrix}$. Then the characteristic polynomial of A is $\lambda^3 - \lambda^2 - 8\lambda + 12 = (\lambda - 2)^2(\lambda + 3)$ so the eigenvalues are 2 and -3. The vectors $\{\mathbf{v_1}, \mathbf{v_2}\}$ form a linearly independent set of eigenvectors associated with $\lambda = 2$ where

$$\mathbf{v_1} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \text{ and } \mathbf{v_2} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

The vector $\mathbf{w} = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}$ is an eigenvector associated with $\lambda = -3$.

1.3 Exercises and Projects

Exercises

Calculus

1. Let f be the function defined by $f(x) = \begin{cases} \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$. Show that any neighborhood of 0 contains points x_0 and x_1 with $f(x_0) = 0$ and $f(x_1) = 1$. Explain why this property shows that f is not continuous at 0.

2. Let f be the function defined by $f(x) = \begin{cases} x \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$. Show that f is continuous at 0 but not differentiable at 0.

3. Suppose f is a function which is continuous at $x = a$ and $f(a) > 0$. Explain why there must be an open interval U containing a such that $f(x) > 0$ for all x in U.

4. A number a is a **fixed point** for a function f if $f(a) = a$. For each of the following functions mapping the closed interval [0,1] into [0,1], find at least one fixed point:

(a) $f(x) = x^2$

(b) $f(x) = \sqrt{x}$

(c) $f(x) = \frac{x+1}{3}$

5. Find a continuous function from the open interval (0,1) onto the open interval (0,1) which has no fixed point

6. Sketch the graphs of $g(x) = x$ and $f(x) = cos\left(\frac{\pi}{2}x\right)$ defined on [0,1]. Explain geometrically why f must have a fixed point.

7. Let f be any continuous function from [0,1] to [0,1]. Use the Intermediate Value Theorem to prove that f must have at least one fixed point. [Hint: Consider $g(x) = f(x) - x$]

8. Darboux's Theorem asserts that even though the derivative of a function need not be continuous, it will always have an Intermediate Value property. More precisely, suppose I is a closed interval and $f : I \to \mathbb{R}^1$ is differentiable at all points of I. If a and b are distinct points of I and k is any number between $f'(a)$ and $f'(b)$, then there is at least one number c between a and b such that $f'(c) = k$.

   (a) Give several examples of functions f differentiable on closed intervals I where f' fails to be continuous on all points of the interval.

   (b) Prove Darboux's Theorem. [The Extreme Value Theorem may be usedful].

Linear Algebra

9. Find a linearly dependent set of three vectors in $\mathbb{R}^3$ such that every pair of the vectors forms a linearly independent set.

10. Show that if S is a linearly dependent set of vectors, then at least one of them may be written as a linear combination of the others.

11. Show that no linearly independent set of vectors contains the zero vector.

12. Show that every set of m vectors in $\mathbb{R}^n$ is linearly dependent if $m > n$.

13. Let $\mathbf{v}$ and $\mathbf{w}$ be any pair of vectors in $\mathbb{R}^3$ such that $\{\mathbf{v}, \mathbf{w}\}$ is a linearly independent set and suppose $\mathbf{N} = \mathbf{v} \times \mathbf{w}$. Show that $\mathbf{N}$ is normal to any plane which contains both a line parallel to $\mathbf{v}$ and a line parallel to $\mathbf{w}$.

14. Suppose $S = \{\mathbf{v_1}, \mathbf{v_2}, ..., , \mathbf{v_n}\}$ is a set of n vectors in $\mathbb{R}^n$. Let M be the $n \times n$ matrix whose jth column is the vector $\mathbf{v_j}$ for each $j = 1, 2, ..., n$.. Show that S is linearly independent if and only if the matrix M is invertible.

15. Prove that each of the asserted properties of the cross product is true.

16. Show that the cross product is **bilinear** ; that is,

    (a) $(\alpha\mathbf{x} + \beta\mathbf{y}) \times \mathbf{z} = \alpha(\mathbf{x} \times \mathbf{z}) + \beta(\mathbf{y} \times \mathbf{z})$, and

    (b) $\mathbf{x} \times (\alpha\mathbf{y} + \beta\mathbf{z}) = \alpha(\mathbf{x} \times \mathbf{y}) + \beta(\mathbf{x} \times \mathbf{z})$

17. Let M be an $n \times n$ invertible matrix. Show the rows of the matrix form a linearly independent set of vectors in $\mathbb{R}^n$.

18. Show that any set whose span is $\mathbb{R}^n$ must contain at least n distinct vectors.

Projects

1. Investigate the continuity and differentiability at 0 of the function $f_n(x) = \begin{cases} x^n \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$ for different values of the integer n.

2. The expressions $x^x, x^{\sqrt{x}}, (\sin x)^x, x^{\ln(1-x)}$ all take on the *indeterminate form* 0^0 at $x = 0$. Why is 0^0 called an indeterminate form? Sketch the graphs of these functions on the open interval (0,1). What appears to happen as x approaches 0? Suppose f and g are two functions chosen from the list $x, x^2, x^3, \sqrt{x}, \sin x, \ln 1 - x, \tan x$. Use l'Hôpital's Rule, or other approaches, to show that

$$(*) \qquad \lim_{x \to 0^+} f(x)^{g(x)} = 1$$

   Before you leap to the conclusion that 0^0 should always be taken to be 1, consider $x^{1/\ln x}$ and $(e^{-1/x})^x$ for $x > 0$.
   Can you find necessary and sufficient conditions on arbitrary functions so that the limit statement (*) is true?
   As a partial result, prove the following: Let g satisfy $f(0) = 0$ and possess a derivative in an open interval centered at the origin. If $\lim_{x\to 0} \frac{g(x)}{g'(x)}$ exists, show that it must have the value 0 and hence

$$\lim_{x \to 0} x^{g(x)} = 1.$$

3. We compute the sum of two vectors or two matrices *componentwise*: the ith component of $\mathbf{x} + \mathbf{y}$ is $x_i + y_i$; similarly, if A and B are both $m \times n$ matrices, then $(A + B)_{ij} = A_{ij} + B_{ij}$. Multiplications (dot, cross or matrix product) are done in a different way. For this project, investigate the properties of componentwise multiplication. Denote this new multiplication operation by the symbol $\odot$:

$$(A \odot B)_{ij} = A_{ij} B_{ij}, \text{ for all } i, j$$

where A and B are both $m \times n$ matrices. We call the matrix $A \odot B$ the *Hadamard Product* [2] of A and B. Determine if the Hadamard product

   (a) is commutative: does $A \odot B = B \odot A$?

   (b) is associative: does $A \odot (B \odot C) = (A \odot B) \odot C$?

   (c) satisifies the distributive law: does $A \odot (B + C) = A \odot B + A \odot C$?

   Find matrices I and O such that $I \odot A = A \odot I = A$ and $O \odot A = A \odot O = O$.
   Find two nonzero matrices A and B so that $A \odot B$ is a matrix of all zeros.
   Photographic images undergo *lossy compression* to reduce their original size for storage and transmission as digital files without significantly sacrificing quality. JPEG is the most commonly employed algorithm for this purpose. Investigate how the Hadamard product plays a role in this compression.

4. An *orthogonal set* in $\mathbb{R}^n$ is a collection of nonzero vectors $\{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_k}\}$ such that every pair of vectors is orthogonal; i.e., $\mathbf{x_i} \cdot \mathbf{x_j} = 0$. If, in addition, the norm of each vector is 1, then the orthogonal set is called *orthonormal* .

   (a) Show that every orthogonal set is linearly independent.

   (b) Discuss how to convert an orthogonal set to an orthonormal one.

   (c) Show that the standard basis for $\mathbb{R}^n$ is an orthonormal set.

   (d) Show that $\{(1, 2, 2), (-4/3, -2/3, 4/3), (2/9, -2/9, 1/9)\}$ is an orthogonal basis for $\mathbb{R}^3$.

[2] Jacques Salomon Hadamard 1865 –1963) was a French mathematician who made major contributions in many fields of mathematics including analysis, number theory, differential geometry, and partial differential equations

The *Gram – Schmidt Process* is an algorithm for converting a linearly independent set S of vectors to an orthonormal one with the same span as S. [3] Investigate how this process works, prove that it does yield an orthonormal one, and discuss its complexity as an algorithm.

[3] Jørgen Pedersen Gram (1850 ? 1916) was a Danish actuary and mathematician who published a number of papers in pure and applied mathematics, including the distribution of prime numbers and mathematical models of forest management. Erhard Schmidt (1876 ?-1959), a German mathematician, worked in the fields of integral equations and functional analysis. He held positions at the University of Berlin during the Nazi regime and expressed support for Hitler. Laplace had developed a similar procedure more than half a century before Gram and Schmidt separately described theirs.