CALCULUS

Dynamic Mathematics

For Multivariable

Fourth Edition

Chuck Garner, Ph.D.
Rockdale Magnet School for Science and Technology

Lulu
Rahleigh London Toronto Bangalore

RMMT PUBLISHING
Conyers

Calculus: Dynamic Mathematics
For Multivariable

Fourth Edition

Printed by LULU, published by RMMT PUBLISHING.

This book was produced directly from the author's LaTeX files.
Figures were drawn by the author using the TeXdraw package and PSTricks.
Set in *STIX2*, designed by Ross Mills and John Hudson in 2016, and Optima, designed by Hermann Zapf in 1958.

Cover and titlepage photograph:
"Light Travel"
by Alex Grichenko. Photo courtesy of `www.publicdomainpictures.net`.

Printed and bound in the United States of America.

Contents

Preface

Think for yourself.
— The Beatles

Although this is the first edition of this book, the material has appeared before in my other books. This is the first time the material for multivariable calculus has appeared on its own as a stand-alone volume.

When I first wrote my set of calculus books, I wrote a two-volume set for single-variable and multivariable calculus to be used over a two-year course sequence at my school. However, the second of the two volumes included the multivariable calculus and also some single-variable topics. This meant that the two-volume set – while useful for me – was not that useful for those teaching a traditional AP Calculus BC course and following it up with a course in multivariable calculus. What this means is the book assumes the student has completed AP Calculus BC, or its equivalent.

"Advanced Placement" and "AP" are registered trademarks of the College Board, which is not associated or affiliated with the production of this book.

As with my other calculus books, proofs are an important aspect of this book. The study of calculus begins the transition between computation-driven mathematics to proof-driven mathematics. Yet proofs are often neglected in the teaching of high school calculus. The book is some attempt at rectifying that situation. I have tried to present rigorous proofs without delving into the higher analysis of the real numbers. Where rigor is not possible, or where such a proof would deliver more confusion than insight, I present an inuitive argument based on assumptions of certain properties of the real numbers. As the student progresses through the book, the rigor increases slightly. But in every case, the proofs are written in an accessible way in order to give high school students practice at thinking about proofs and reading proofs, as well as understanding the logic behind a statement of a theorem.

Organization. One may take a quick look at the contents page, but that does not reveal the reasons why the book is organized the way it is. What follows is an attempt at explaining those reasons.

Chapter 1: Linear Algebra. This is a short introduction to the concepts of linear algebra. The theory of three-dimensional vectors is fully developed, as are matrices and determinants. All of this is needed for applications of partial derivatives, but many calculus books give this a quick and incomplete treatment. Linear algebra is useful and important in its own right and not just a tool for calculus in three dimensions. Continuing that theme, in the remainder of the chapter are topics traditional in linear algebra courses, but not usually included in calculus books. It is a shame, as eigenvectors and eigenvalues fit nicely with differential equations that model predator-prey problems (which are included in Chapter 4).

CHAPTER 2: PARTIAL DERIVATIVES. A more traditional approach to partial derivatives is found here, but with a few twists. Mirroring the development of single-variable derivatives in Volume 1, the partial derivative is introduced numerically before the algebraic aspect takes over. Applications appear throughout this chapter: biophysics, business, and chemistry, among others. The Jacobian and its purpose is introduced as a differential matrix and used extensively in the next chapter. Differentials of composite functions and the general chain rule lead into the Implicit Function Theorem. We also discuss extrema, including use of the Lagrange multiplier. Eigenvalues and eigenvectors also make a return when discussing extrema on quadratic surfaces.

CHAPTER 3: MULTIPLE INTEGRALS. The multiple integrals only run the first five sections, then line and surface integrals take over for the remainder of the chapter. Line and surface integrals are included here since they may be expressed as multiple integrals under certain conditions. Many applications are included throughout. All theorems in this chapter are stated in terms of rectangular equations as well as in terms of vectors. Transforming regions and integrals from rectangular to polar is used throughout; however, spherical and cylindrical coordinates are not used. The reason for the omission of these coordinate systems is that I believe both systems can be easily learned by the student if the need arises, given a firm background in polar coordinates. The three major theorems of multivariable integrals are developed and explained (Green, Gauss, and Stokes), as well as the concept of line integrals which are independent of path.

CHAPTER 4: MORE DIFFERENTIAL EQUATIONS. This short chapter includes simple techniques that rely on partial derivatives and infinite series. In a sense this chapter is an "application" of multivariable ideas to single variable equations.

Problems and Exercises. Most sections have at least one "Exercise" in the text. These exercises are designed to be completed by the student at that point in the lesson. They extend recent concepts, foreshadow upcoming ideas, or are for practicing new skills. I use these as classwork assignments, completed individually or in small groups.

One section common to each chapter is the section entitled "Preparation and Extension." The Preparation problems are designed to prepare students for a chapter test. The problems here are similar to test questions that I write and are a good measure of how well prepared the student is for the test. The Extension problems are meant to challenge the students by extending the concepts and skills learned in that chapter.

CHUCK GARNER
Conyers, Georgia
June 2021

Chapter 1

LINEAR ALGEBRA

MANY STUDENTS WHO learn linear algebra consider it to be the mathematics of matrices. While matrices are a central component of linear algebra, matrices are not the fundamental idea, only the representation for that idea. The idea is a *linear transformation*. By a linear transformation, we mean any set of linear equations that describes how to transform one set of points to another. The theory and concept behind linear transformations gives the study its name: *linear algebra*.

Because we are interested in linear transformations in general (that is, we will investigate more than simply transformation in two-dimensions), we will begin with a treatment of three-dimensional vectors. This will result in a natural justification for the matrix representation of linear transformations. We will also develop some of the theory and applications of linear algebra.

1.1 The Geometry of Vectors

Mathematical reality lies outside us,
. . . our function is to discover or observe it.
— G. H. Hardy

Vectors are well-suited to describe geometrical concepts. We use this section to re-interpret some familiar ideas in two dimensions with vectors, and then to extend these ideas to three dimensions. We begin with lines in the plane.

Suppose we have a point $P_0(x_0, y_0)$ on a line and a vector $\mathbf{n} = \langle a, b \rangle$ that is perpendicular to the line. Then another point $P(x, y)$ is on the line precisely when the vectors $\mathbf{n}$ and $\overrightarrow{P_0P}$ are perpendicular. (See Figure 1.1.) We have the following definition.

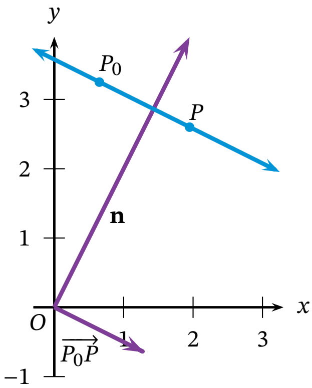

Figure 1.1 – The equation of the line $\overline{P_0P}$ is derived by taking the dot product of the vector from P_0 to P and a vector normal to the line

Suppose $P_0(x_0, y_0)$ is a point on a line and $\mathbf{n} = \langle a, b\rangle$ is a vector perpendicular to the line. Then the equation of the line is

$$\mathbf{n} \cdot \overrightarrow{P_0P} = 0$$

where P is any point on the line.

Notice that this dot product results in the familiar equation of a line. We have

$$\begin{aligned}\mathbf{n} \cdot \overrightarrow{P_0P} &= \langle a, b\rangle \cdot \langle x - x_0, y - y_0\rangle \\ &= a(x - x_0) + b(y - y_0) = 0\end{aligned}$$

which simplifies to $ax + by = d$, where $d = ax_0 + by_0$.

Example 1.1.1

Suppose the point $P_0(4, 5)$ is on a line and the vector $\mathbf{n} = \langle -3, 1\rangle$ is perpendicular to the line. Then, for any point $P(x, y)$ on the line, we have

$$\langle -3, 1\rangle \cdot \langle x - 4, y - 5\rangle = -3(x - 4) + (y - 5) = -3x + 12 + y - 5 = 0,$$

or $3x - y = 7$. ♦

This new way to write a line is very useful, as we will now show. Given an equation of a line $ax + by = d$, we see that the vector $\langle a, b\rangle$ is perpendicular to the line. This is simply what we learned in algebra: the slopes of perpendicular lines are negative reciprocals. Note that the slope of $ax + by = d$ is $-\frac{a}{b}$ and the slope of the vector $\langle a, b\rangle$ is $\frac{b}{a}$.

The representation of a line by an equation $\mathbf{n} \cdot \overrightarrow{P_0P} = 0$ is not unique. Any point on the line can replace P_0, and a nonzero multiple of $\mathbf{n}$ can replace $\mathbf{n}$ without changing the set of points P that satisfy the equation. We can use some of this freedom by requiring the vector $\mathbf{n}$ to be a unit vector. The result is called the **normalized equation**, which then becomes

$$\mathbf{u} \cdot \overrightarrow{P_0P} = 0 \qquad \text{where} \quad \mathbf{u} = \frac{\mathbf{n}}{\|\mathbf{n}\|}.$$

Example 1.1.2

Suppose the point $P_0(-1, -1)$ lies on a line and $\mathbf{n} = \langle 3, 4\rangle$ is orthogonal to the line. Then the normalized equation of the line is

$$\frac{\langle 3, 4\rangle}{\|\langle 3, 4\rangle\|} \cdot \langle x + 1, y + 1\rangle = \tfrac{3}{5}(x + 1) + \tfrac{4}{5}(y + 1) = 0,$$

or $\frac{3}{5}x + \frac{4}{5}y = -\frac{7}{5}$. ♦

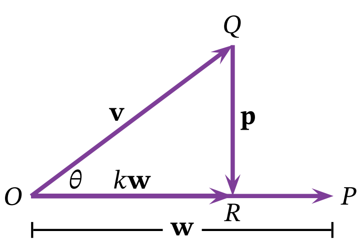

Figure 1.2 – The projection of **v** on **w** is $k\mathbf{w}$.

At this point, it is prudent to give another interpretation of the dot product of two vectors. Consider a vector $\mathbf{w} = \overrightarrow{OP}$ and a vector $\mathbf{v} = \overrightarrow{OQ}$. If we drop a perpendicular from Q to $\overrightarrow{OP}$, as in Figure 1.2, we create a vector $\mathbf{p} = \overrightarrow{QR}$ where R is the foot of the perpendicular. The vector $\overrightarrow{OR}$ is called the **projection** of $\mathbf{v}$ on $\mathbf{w}$. Note that the vector $\overrightarrow{OR}$, which lies on $\mathbf{w}$, is a scalar multiple of $\mathbf{w}$; i.e., $\overrightarrow{OR} = k\mathbf{w}$.

Since $\cos\theta = \pm\|k\mathbf{w}\|/\|\mathbf{v}\|$, we have that

$$\mathbf{v}\cdot\mathbf{w} = \|\mathbf{v}\|\,\|\mathbf{w}\|\cos\theta = \pm\,\|\mathbf{v}\|\,\|\mathbf{w}\|\,\frac{\|k\mathbf{w}\|}{\|\mathbf{v}\|} = \pm\,\|\mathbf{w}\|\,\|k\mathbf{w}\|$$

where the sign is chosen as positive if θ is an acute angle or negative if θ is an obtuse angle. Hence, we have the following alternate definition of the dot product.

The dot product of $\mathbf{v}$ and $\mathbf{w}$ is

$$\mathbf{v}\cdot\mathbf{w} = \pm\,\|\mathbf{w}\|\,\|\text{projection of } \mathbf{v} \text{ on } \mathbf{w}\| = \pm\,\|\mathbf{w}\|\,\|k\mathbf{w}\| \;= \pm k\,\|\mathbf{w}\|^2$$

This interpretation of the dot product helps to prove the following.

LEMMA 1.A (Distance from a Point to a Line) *Let $\mathbf{u}\cdot\overrightarrow{P_0P} = 0$ be the normalized equation of a line and let P_1 be any point in the plane. Then the distance from P_1 to the line is given by $D = |\mathbf{u}\cdot\overrightarrow{P_0P_1}|$.*

Proof. From point P_1 we drop a perpendicular to $\overrightarrow{P_0P}$, meeting the vector at point Q. Then the distance we seek is given by $D = \|\overrightarrow{P_1Q}\|$. Recall that $\mathbf{u}$ is a unit vector perpendicular to the line. Hence, the projection of $\overrightarrow{P_0P_1}$ on $\mathbf{u}$ is exactly $\overrightarrow{P_1Q}$. Thus,

$$\mathbf{u}\cdot\overrightarrow{P_0P_1} = \pm\,\|\mathbf{u}\|\,\|\overrightarrow{P_1Q}\| = \pm\,\|\mathbf{u}\|\,D.$$

Therefore, since $\mathbf{u}$ is a unit vector and distance is always positive, we have

$$|\mathbf{u}\cdot\overrightarrow{P_0P_1}| = D.$$

■

Example 1.1.3
To find the distance from the point $(2,-1)$ to the line $3x + 4y = 9$, we first write the normalized equation of the line. The vector perpendicular to the line is $\mathbf{n} = \langle 3,4\rangle$, so the unit vector is $\left\langle \frac{3}{5}, \frac{4}{5}\right\rangle$. Then the normalized equation is $\frac{3}{5}x + \frac{4}{5}y - \frac{9}{5} = 0$. Now we substitute the point $(2,-1)$ into the normalized expression to get

$$\tfrac{3}{5}(2) + \tfrac{4}{5}(-1) - \tfrac{9}{5} = \tfrac{6}{5} - \tfrac{4}{5} - \tfrac{9}{5} = -\tfrac{7}{5},$$

which indicates that $\left|-\frac{7}{5}\right| = \frac{7}{5}$ is the distance from the point to the line. ♦

Examining distances with vectors naturally leads to measuring other geometric properties. One such property is area.

THEOREM 1.B (Area of a Parallelogram) *Let vectors $\mathbf{v}$ and $\mathbf{w}$ lie in the plane such that they form adjacent sides of a parallelogram. Then the area of the parallelogram is given by*

$$A = \|\mathbf{v}\|\,\|\mathbf{w}\|\sin\theta$$

where θ is the angle between the two vectors.

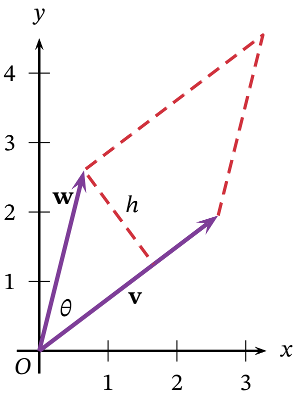

Figure 1.3 – The area of a parallelogram.

Proof. Suppose $\mathbf{w} = \langle w_1, w_2 \rangle$ and $\mathbf{v} = \langle v_1, v_2 \rangle$ both emanate from the origin with an angle of θ between them. Without loss of generality, assume $\mathbf{v}$ is long enough so that when a perpendicular is dropped from the end of $\mathbf{w}$, the perpendicular intersects $\mathbf{v}$, as in Figure 1.3. Call the length of the perpendicular h. Then, since $\sin\theta = h/\|\mathbf{w}\|$, we have $h = \|\mathbf{w}\| \sin(\theta)$. Thus, because the base is $\|\mathbf{v}\|$, the area of the parallelogram is

$$A = bh = \|\mathbf{v}\|\,\|\mathbf{w}\| \sin(\theta),$$

and the theorem is proved. ■

Exercise 1.1.4 Find the distance from the point $(4, -7)$ to the line through the point $(-3, 1)$ and perpendicular to the vector $\langle 2, 3 \rangle$.

Exercise 1.1.5 Find the area of the parallelogram formed by $\left\langle 3, \sqrt{3} \right\rangle$ and $\langle 0, 9 \rangle$.

Problems for Section 1.1

1 Find the equation of the line through $(3, -1)$ and perpendicular to $\langle -1, 2 \rangle$.

2 Find the unit vector that is perpendicular to the line $2x - y + 4 = 0$.

3 Find the distance from the point $(-1, 7)$ to the line $4x - y - 11 = 0$.

4 Find the area of the parallelogram formed by the vectors $\langle -2, 4 \rangle$ and $\langle 3, 3 \rangle$. What is the area of the triangle formed by these vectors?

5 Given that $\mathbf{v} = \langle x, y \rangle$, explain why $\|\mathbf{v}\| = r$ is the equation of a circle of radius r. What is the circle's center?

6 Given three points $R(2, 5)$, $S(-4, 2)$, and $T(1, -1)$ in the plane, find:

(a) the equation of the line through S if $\overrightarrow{TR}$ is the normal vector;

(b) the distance from R to the line found in part (a);

(c) the unit vector perpendicular to the line through R and S;

(d) the area of the parallelogram formed by $\overrightarrow{SR}$ and $\overrightarrow{ST}$;

(e) the equations of the line through R that is perpendicular to the line through S and T;

(f) the angles of the triangle formed by vertices R, S, and T.

1.2 Vectors in Three Dimensions

There exists, if I am not mistaken, an entire world which is the totality of mathematical truths, to which we have access only with our mind, just as a world of physical reality exists, the one like the other independent of ourselves, both of divine creation.

— Charles Hermite

Vectors in three-dimensional space are analogous to vectors in the plane but with a third component to represent the third dimension. A vector in space is $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$ where the third component can be considered “height” (if the first two are “length” and “width”). This leads to a third standard unit vector, and modifications to the familiar $\mathbf{i}$ and $\mathbf{j}$:

$$\mathbf{i} = \langle 1, 0, 0 \rangle, \quad \mathbf{j} = \langle 0, 1, 0 \rangle, \quad \mathbf{k} = \langle 0, 0, 1 \rangle.$$

Problems for Section 1.8

Find the eigenvalues and eigenvectors of the following matrices.

1 $\begin{bmatrix} -3 & 2 \\ 8 & 3 \end{bmatrix}$

2 $\begin{bmatrix} 0 & 1 & -2 \\ 2 & 1 & 0 \\ 4 & -2 & 5 \end{bmatrix}$

3 $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

4 $\begin{bmatrix} \pi & 2 \\ 0 & 2\pi \end{bmatrix}$

5 $\begin{bmatrix} 5 & -2 & 8 \\ -4 & 0 & -5 \\ -4 & 2 & -7 \end{bmatrix}$

6 $\begin{bmatrix} 0 & 1 & -2 \\ -6 & 5 & -4 \\ 0 & 0 & 3 \end{bmatrix}$

7 Prove that the product of the eigenvalues of a square matrix is equal to its determinant.

8 Prove that every square matrix is similar to itself.

9 Prove that if A is similar to B and B is similar to C, then A is similar to C.

1.9 Orthogonal Matrices

In many cases, mathematics is an escape from reality. The mathematician finds his own monastic niche and happiness in pursuits that are disconnected from external affairs. Some practice it as if using a drug. Chess sometimes plays a similar role. In their unhappiness over the events of this world, some immerse themselves in a kind of self-sufficiency in mathematics. (Some have engaged in it for this reason alone.)

— Stanislaw Ulam

We begin this section with a new matrix operation in which we interchange the rows and columns of a matrix.

The **transpose** of a matrix A, denoted A^T, is obtained by swapping rows of A with the columns of A. That is, if $A = [a_{ij}]$, then $A^T = [a_{ji}]$.

Example 1.9.1

We have that if $A = \begin{bmatrix} 1 & 0 \\ -4 & 5 \\ 4 & 3 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & -4 & 4 \\ 0 & 5 & 3 \end{bmatrix}$.

Clearly, if the matrix is square, so is its transpose. ♦

The transpose obeys several rules, as the theorem below indicates.

THEOREM 1.P (Properties of the Transpose) *Suppose A and B are matrices and k is a constant. Then*

i. $(A^T)^T = A$.

ii. $(kA)^T = kA^T$.

iii. $(A+B)^T = A^T + B^T$.

iv. $(AB)^T = B^T A^T$.

v. If A is nonsingular, then $(A^T)^{-1} = (A^{-1})^T$.

Proof. Proofs of statements (i) and (ii) are left as Problem 15.

To prove statement (iii), we set $C = A + B$ and $D = A^T + B^T$. Then $c_{ij} = a_{ij} + b_{ij}$ and $d_{ij} = a_{ji} + b_{ji}$ for all i and j. Then $C^T = [c_{ji}] = [a_{ji} + b_{ji}] = [d_{ij}] = D$, which proves the statement.

To prove statement (iv), we let A be an $m \times p$ matrix and B a $p \times n$ matrix so that the multiplication of A and B is defined. Now set $E = AB = [e_{ij}]$ and $F = B^T A^T = [f_{ij}]$, where

$$f_{ij} = b_{1i}a_{j1} + \cdots + b_{pi}a_{jp} = a_{j1}b_{1i} + \cdots + a_{jp}b_{pi} = e_{ji}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Thus, $F = E^T$, and the statement follows.

Finally, to prove statement (v), write $AA^{-1} = I$. Then by statement (iv), $(AA^{-1})^T = (A^{-1})^T A^T = I^T = I$, so that A^T has an inverse $(A^T)^{-1}$, which, since inverses are unique, must be equal to $(A^{-1})^T$. ■

With the transpose, we may succintly write a column vector as the transpose of the corresponding row vector:

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}^T$$

Note that this gives us a way to express the dot product of $\mathbf{x}$ and $\mathbf{y}$ as the product of the "matrix" $\mathbf{x}$ and the "matrix" $\mathbf{y}^T$. Hence, $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}\mathbf{y}^T$.

A square matrix M for which $M = M^T$ is called a **symmetric matrix**.

Example 1.9.2
Both the matrices

$$\begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 1 & 0 \\ 1 & -1 & -3 \\ 0 & -3 & 5 \end{bmatrix}$$

are symmetric. ♦

Concerning symmetric matrices, we have the following interesting result linking symmetry with eigenvalues.

THEOREM 1.Q *Suppose A is an $n \times n$ symmetric matrix. Then any two eigenvectors that are associated with distinct eigenvalues are orthogonal.*

Proof. For symmetric matrix A, suppose λ_1 has eigenvector $\mathbf{x}_1$ and λ_2 has eigenvector $\mathbf{x}_2$. Then

$$\begin{aligned} \lambda_2(\mathbf{x}_1 \cdot \mathbf{x}_2) &= (\lambda_2\mathbf{x}_1)\mathbf{x}_2^T \\ &= \mathbf{x}_1(\lambda_2\mathbf{x}_2^T) \\ &= \mathbf{x}_1 A\mathbf{x}_2^T, \end{aligned}$$

and since A is symmetric, $A = A^T$, so that we have

$$
\begin{aligned}
&= \mathbf{x}_1 A^T \mathbf{x}_2^T \\
&= (\mathbf{x}_1 A^T)\mathbf{x}_2^T \\
&= (A\mathbf{x}_1^T)^T \mathbf{x}_2^T \\
&= (\lambda_1 \mathbf{x}_1^T)^T \mathbf{x}_2^T \\
&= \lambda_1 (\mathbf{x}_1^T)^T \mathbf{x}_2^T \\
&= \lambda_1 (\mathbf{x}_1 \mathbf{x}_2^T) \\
&= \lambda_1 (\mathbf{x}_1 \cdot \mathbf{x}_2).
\end{aligned}
$$

Thus, $\lambda_1(\mathbf{x}_1 \cdot \mathbf{x}_2) = \lambda_2(\mathbf{x}_1 \cdot \mathbf{x}_2)$, or $(\lambda_1 - \lambda_2)(\mathbf{x}_1 \cdot \mathbf{x}_2) = 0$. Since $\lambda_1 \neq \lambda_2$, we must have that $\mathbf{x}_1 \cdot \mathbf{x}_2 = 0$. Therefore, since the dot product is zero, the eigenvectors are orthogonal. ■

We introduce some notation. The symbol $\text{diag}(a_{11}, a_{22}, \dots, a_{nn})$ denotes the matrix whose every entry is zero except for the entries along the main diagonal; such a matrix is called a **diagonal matrix**. For example,

$$
\text{diag}(2, 7, -1) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -1 \end{bmatrix}.
$$

Note that every diagonal matrix $\text{diag}(a_{11}, a_{22}, \dots, a_{nn})$ is also symmetric.

The next theorem relates the concept of similarity to the eigenvalues of a matrix.

COROLLARY 1.R *Let the $n \times n$ matrix A have n distinct eigenvalues λ_i for $i = 1, 2, \dots, n$. Then A is similar to $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.*

Proof. Let A have n distinct eigenvalues λ_i. Then A also has n distinct eigenvectors $\mathbf{x}_i$ so that $A\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for every i. Let C be the matrix whose column vectors are the eigenvectors of A; in other words,

$$
C = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}.
$$

Then we have

$$
\begin{aligned}
AC &= A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \\
&= A \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 x_{11} & \lambda_2 x_{12} & \cdots & \lambda_n x_{1n} \\ \lambda_1 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_n x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{n1} & \lambda_2 x_{n2} & \cdots & \lambda_n x_{nn} \end{bmatrix} \\
&= \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \\
&= C\,\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).
\end{aligned}
$$

Since C is the matrix of eigenvectors arising from n distinct eigenvalues, by Theorem 1.Q, we know that C consists of mutually orthogonal columns; hence, $\det C$ cannot be zero, which means that C is nonsingular. Therefore, we may multiply both sides of $AC = C\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ by C^{-1} to get $\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = C^{-1}AC$. We conclude that $A \sim \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. ■

Since a matrix is always similar to its diagonal matrix of its eigenvalues, we have a computationally easier way to study the effect of large matrices. For example, let $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and suppose we wish to calculate A^2. Since we have $A = CDC^{-1}$, then

$$A^2 = (CDC^{-1})(CDC^{-1}) = CD(C^{-1}C)DC^{-1} = CD^2C^{-1}.$$

Similarly,

$$A^3 = (CD^2C^{-1})(CDC^{-1}) = CD^2(CC^{-1})DC^{-1} = CD^3C^{-1}.$$

In this manner, we see that

$$A^k = CD^kC^{-1}.$$

Powers of diagonal matrices are very easy to compute:

$$D^k = [\operatorname{diag}(\lambda_1, \ldots, \lambda_n)]^k = \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k).$$

Example 1.9.3

From Example 1.8.1, we know the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}.$$

Thus, instead of performing three complicated mutiplications, we compute

$$\begin{aligned} A^4 &= \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix}^4 \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} (-1)^4 & 0 \\ 0 & 4^4 \end{bmatrix} \begin{bmatrix} \frac{3}{5} & -\frac{2}{5} \\ -\frac{1}{5} & -\frac{1}{5} \end{bmatrix} \\ &= \begin{bmatrix} 103 & 102 \\ 153 & 154 \end{bmatrix}. \end{aligned}$$

Computing A to any other large power is much simpler than multiplying A to itself over and over again. ♦

Questions of orthogonality bring us to the following definition.

> A square matrix M is called an **orthogonal matrix** if $MM^T = I$. In other words, M is orthogonal if and only if $M^T = M^{-1}$.

There are two consequences of this definition. The first is that every orthogonal matrix must also be nonsingular. The second is that, due to the commutativity of a matrix and its inverse, $MM^T = M^TM = I$.

Let us give a reason for this definition. We may call a matrix M such that $MM^T = I$ orthogonal because each row vector is orthogonal to every other row vector, and each column vector is orthogonal to every other column vector. In particular, we have the following theorem.

THEOREM 1.S *A square matrix M is orthogonal if and only if each column and row vector is a unit vector and each column and row vector is orthogonal to every other column and row vector.*

Proof. Let us denote the columns of M by $\mathbf{m}_i$ for $i = 1, 2, \ldots, n$. Then

$$M = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \cdots & \mathbf{m}_n \end{bmatrix} \quad \text{and} \quad M^T = \begin{bmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^T \\ \vdots \\ \mathbf{m}_n^T \end{bmatrix}$$

so that

$$M^T M = \begin{bmatrix} \mathbf{m}_1^T\mathbf{m}_1 & \mathbf{m}_1^T\mathbf{m}_2 & \cdots & \mathbf{m}_1^T\mathbf{m}_n \\ \mathbf{m}_2^T\mathbf{m}_1 & \mathbf{m}_2^T\mathbf{m}_2 & \cdots & \mathbf{m}_2^T\mathbf{m}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{m}_n^T\mathbf{m}_1 & \mathbf{m}_n^T\mathbf{m}_2 & \cdots & \mathbf{m}_n^T\mathbf{m}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I.$$

Thus, since $\mathbf{m}_i^T\mathbf{m}_i = 1$ for each i, each column vector is a unit vector; since $\mathbf{m}_i^T\mathbf{m}_j = 0$ when $i \neq j$, the column vectors are orthogonal.

The proof for the row vectors is similar and is omitted. ■

Exercise 1.9.4 Is the matrix $B = \frac{1}{7}\begin{bmatrix} 2 & 3 & 6 \\ 3 & -6 & 2 \\ -6 & -2 & 3 \end{bmatrix}$ orthogonal? Explain.

Exercise 1.9.5 What is the determinant of B?

Example 1.9.6

The matrix

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

is orthogonal for every θ. Note that

$$R^T R = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

$$= \begin{bmatrix} \cos^2(\theta) + \sin^2(\theta) & 0 \\ 0 & \cos^2(\theta) + \sin^2(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Also the magnitude of each column is 1, and the columns are orthogonal for every θ. ♦

The matrix $R(\theta)$ in the above example is called a **rotational matrix**. This matrix represents a counterclockwise rotation of θ about the origin.

Example 1.9.7

The point $(2, 0)$ is rotated to the point $(0, 2)$ when the matrix $R(\frac{\pi}{2})$ is applied. To wit:

$$\begin{bmatrix} \cos\left(\frac{\pi}{2}\right) & -\sin\left(\frac{\pi}{2}\right) \\ \sin\left(\frac{\pi}{2}\right) & \cos\left(\frac{\pi}{2}\right) \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \end{bmatrix}.$$

In fact we may rotate any point (x, y).

$$\begin{bmatrix} \cos\left(\frac{\pi}{2}\right) & -\sin\left(\frac{\pi}{2}\right) \\ \sin\left(\frac{\pi}{2}\right) & \cos\left(\frac{\pi}{2}\right) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y & x \end{bmatrix}.$$

Hence, the rotation sends the point (x, y) to the point $(-y, x)$.

If a locus of points are described by some equation, say $y = f(x)$, then this allows us to rotate all such points through an angle of $\frac{\pi}{2}$. If $y = x^2$, then the point (x, x^2) becomes $(-x^2, x)$. In other words, the upward positive parabola $y = x^2$ rotates to become $x = -y^2$, which describes a parabola that opens to the left. ♦

Problems for Section 1.9

1 What is $\det R(\theta)$?

Find the transpose of each matrix.

2 $\begin{bmatrix} 1 & 2 & 3 \\ 7 & 0 & 5 \end{bmatrix}$

3 $\begin{bmatrix} 2 & 8 & -4 & 1 \end{bmatrix}$

4 $\begin{bmatrix} 2 \\ 1 \\ -6 \end{bmatrix}$

Find the values of a and b so that each matrix is symmetric.

5 $\begin{bmatrix} -3 & 3a-1 \\ 2a & 7 \end{bmatrix}$

6 $\begin{bmatrix} 1 & a & 8 \\ b-a & 9 & 4+a \\ 8 & b & 3 \end{bmatrix}$

Are the following matrices orthogonal? Justify your answers.

7 $\begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$

8 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

9 $\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

10 $\begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$

11 $\begin{bmatrix} 2 & 1 & 2 \\ 1 & 2 & -2 \\ -2 & 2 & 1 \end{bmatrix}$

12 Suppose $A = \begin{bmatrix} 3 & 1 \\ 4 & 3 \end{bmatrix}$.

(a) Find a matrix B and a nonsingular matrix C so that $A = CBC^{-1}$.

(b) Compute A^5.

Each matrix is a rotational matrix. Determine the angle of rotation, then determine the new, rotated equations of the lines $x = 0$ and $y = 0$.

13 $\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$

14 $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

15 Prove statements (i) and (ii) of Theorem 1.P.

16 Suppose A is an orthogonal matrix.

(a) Prove that $\det A = \pm 1$.

(b) Prove that A^T and A^{-1} are also orthogonal.

1.10 Preparation and Extension

One of the penalties for refusing to participate in politics is that you end up being governed by your inferiors.

— Plato

Preparation Problems for Chapter 1

Find the velocity and acceleration vectors, given the position vector $\mathbf{p}(t)$.

1 $\mathbf{p}(t) = \left\langle 6te^{-2t}, 8e^{-2t}, -16t^2 \right\rangle$

2 $\mathbf{p}(t) = \left\langle 6e^{-3t}, \sin(3t), t^3 - 6t \right\rangle$

Find the position vector, given either the velocity vector $\mathbf{v}(t)$ or the acceleration vector $\mathbf{a}(t)$.

3 $\mathbf{v}(t) = \left\langle 10, 3e^t, -32t + 8 \right\rangle, \quad \mathbf{p}(0) = \langle 0, 6, 20 \rangle$

4 $\mathbf{v}(t) = \left\langle t+1, t^2, e^{-t/3} \right\rangle, \quad \mathbf{p}(0) = \langle 4, 0, -3 \rangle$

(Note that at the critical point, the price of butter becomes 279 cents per pound and cheese becomes 470 cents per pound. According to the webpage *Understanding Dairy Markets*, by Brian W. Gould of the University of Wisconsin, in 2009 the average national cost of butter was $2.80 per pound and that of cheddar cheese was $4.70 per pound. ♦

Pricing information from `future.aae.wisc.edu/tab/prices.html`*, retrieved January 2011.*

Problems for Section 2.10

For each of the following surfaces, find critical points and test those points for extrema.

1 $z = 2x^3 - xy^2 + 5x^2 - y^2$

2 $z = x^3 - 12xy + 8y^3$

3 $z = 2x^2 - xy - 3y^2 - 3x + 7y$

4 $z = x\cosh(y)$

5 $z = x^2 - 2x(\sin(y) + \cos(y)) + 1$

6 $z = x^2 + y^2 + 2x^{-1}y^{-2}$

7 $z = x^3 + y^2 - 3x^2 - 3y - 9x$

8 *(Calculator)* Products A and B are jointly produced. The price of product A is $P_A(x) = 300 - x$ and that of product B is $P_B(y) = 150 - 4y$. The joint cost is

$$C(x, y) = 2x^2 + \tfrac{1}{2}y^2 + xy + 30.$$

What is the maximum profit of this production model, and what prices should be established for A and B to achieve the maximum profit?

2.11 The Lagrange Multiplier

The reader will find no figures in this work. The methods which I set forth do not require either constructions or geometrical or mechanical reasonings: but only algebraic operations, subject to a regular and uniform rule of procedure.

— Joseph-Louis Lagrange

Often we are required to maximize or minimize a function $f(x, y, z)$ subject to a condition $g(x, y, z)$. For instance we may want to maximize profit, but under the constraint of the availability of supplies. So in this section we develop methods to find extrema of functions subject to additional **constraints**, or **side conditions**.

The viability of such methods is a direct result of the existence of the following type of extrema. Given a function and a constraint, the effect of the constraint on the function is to retrict the function's domain to a closed subset E of the domain D.

Let $z = f(x, y)$ be a continuous function defined on a closed region E formed from a domain D of the xy-plane. If P_0 is a point of E such that $f(P_0) \geq f(x, y)$ for all points (x, y) in E, then P_0 is an **absolute maximum**. If P_0 is a point of E such that $f(P_0) \leq f(x, y)$ for all points (x, y) in E, then P_0 is an **absolute minimum**.

This, combined with the following theorem, is exploited to find extrema of functions with constraints.

THEOREM 2.M *Suppose D is a bounded domain of the xy-plane. If $z = f(x, y)$ is defined and continuous in the closed region E formed of*

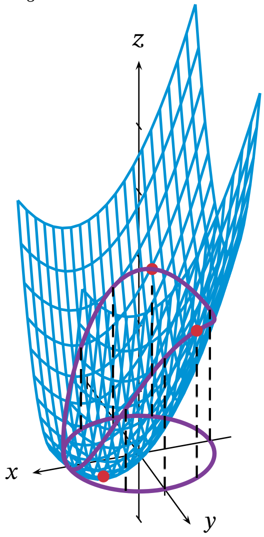

Figure 2.10 – Figure for Example 2.11.1. The points of absolute extrema of the surface under the constraint are marked.

D and its boundary, then f has an absolute maximum and an absolute minimum in E.

This theorem is the three-dimensional analogue of the Extreme Value Theorem. Since every continuous function has absolute extrema on a closed region, we are guaranteed to find extrema. However, in the course of our search we may find (as in the last section) saddle points or other local extrema as well. One method to find absolute extrema, which is illustrated in the example below, relies on being able to write the constraint in terms of one variable. This method for finding absolute extrema is known as the *direct method*.

Example 2.11.1

Let us find the absolute maximum and absolute minimum of $z = x^2 + 2y^2 - x$ on the unit disk $x^2 + y^2 \le 1$.

Since $\frac{\partial z}{\partial x} = 2x - 1$ and $\frac{\partial z}{\partial y} = 4y$, we have a critical point at $(\frac{1}{2}, 0)$. Note that this critical point is within the unit disk, so it is a candidate for absolute extrema.

Now we investigate the function values on the boundary of the unit disk. Figure 2.10 shows the surface z and the projection of the constraint onto the surface. It is along this projected curve that we find extrema. To do this, we substitute the boundary equation $x^2 + y^2 = 1$ into z to get $z = 2 - x - x^2 = (2 + x)(1 - x)$, which is valid for $-1 \le x \le 1$. This function (which is now a single-variable function) has the critical points $x = -1$, $x = -\frac{1}{2}$, and $x = 1$ on the boundary.

Finally, to determine absolute extrema, we evaluate z at these points and see which is the smallest and which is the largest. At the point $(\frac{1}{2}, 0)$, we compute $z = -\frac{1}{4}$. On the boundary, we may compute values of z using the equation $z = (2 + x)(1 - x)$. Hence, $x = -1$ gives $z = 2$; $x = -\frac{1}{2}$ gives $z = \frac{9}{4}$; and $x = 1$ gives $z = 0$. Thus we conclude that the absolute maximum is $\frac{9}{4}$ which occurs at the two points $(-\frac{1}{2}, \pm\frac{\sqrt{3}}{2})$ and the absolute minimum is $-\frac{1}{4}$ which occurs at $(\frac{1}{2}, 0)$. (Since the minimum occurs inside the unit disk, it is also a relative minimum.) ♦

This direct method is very useful in the simple cases, but is implausible for more complicated functions and constraints: it may be impossible to solve the constraint equation for one of the variables. So although the direct method is fine, we need a more general method. This method is called the **Lagrange multiplier** method and its basis is the following theorem.

Joseph-Louis Lagrange, 18th century Italian-born French mathematician, advanced all fields of analysis, number theory, and celestial mechanics. He first wrote about the multiplier 1764, and fully realized the method in his groundbreaking book Mechanical Analysis *of 1788.*

THEOREM 2.N (Lagrange's Theorem) *Let functions $f(\mathbf{x})$ and $g(\mathbf{x})$ have continuous partial derivatives in a neighborhood of $\mathbf{x}_0$. If $\mathbf{x}_0$ gives an extreme of f subject to the constraint $g(\mathbf{x}) = 0$ where $\nabla g(\mathbf{x}_0) \ne \mathbf{0}$, then $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$ for some constant λ.*

Proof. Suppose the point $\mathbf{x}_0$ gives an extreme of $f(\mathbf{x})$ where $g(\mathbf{x}) = 0$ is given as a constraint. Let $C(t)$ be any differentiable curve lying on the surface f that passes through the point $f(\mathbf{x}_0)$. Then there is a value t_0 such that $C(t_0) = \mathbf{x}_0$. Then the function $f(C(t))$ has a an extreme for $t = t_0$. The derivative of $f(C(t))$ at the point t_0 is therefore zero. In other words, by the Chain Rule,

$$\frac{d}{dt} f(C(t))\bigg|_{t_0} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}\bigg|_{t_0} = \nabla f(\mathbf{x}_0) \cdot C'(t_0) = 0.$$

This implies that $\nabla f(\mathbf{x}_0)$ is normal to the curve C at $f(\mathbf{x}_0)$. Since the curve C was an arbitrary curve passing through $f(\mathbf{x}_0)$, we have that $\nabla f(\mathbf{x}_0)$ is normal to *every* curve passing through $f(\mathbf{x}_0)$.

Since $\mathbf{x}_0$ maximizes f subject to $g = 0$, it must be that g passes through the point $f(\mathbf{x}_0)$ also. By definition, $\nabla g(\mathbf{x}_0)$ is normal to the surface $g = 0$; but so is $\nabla f(\mathbf{x}_0)$. Thus, $\nabla g(\mathbf{x}_0)$ is parallel to $\nabla f(\mathbf{x}_0)$. Therefore there is a constant λ such that $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$. ■

So we are using λ for the Lagrange multiplier and we used λ for eigenvalues. . . is there some connection?

The constant λ in Theorem 2.N is called the **Lagrange multiplier**. This method consists of solving the system of equations

$$\begin{cases} \nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0) \\ g(\mathbf{x}_0) = 0 \end{cases}$$

where λ is treated as an extra variable to ensure the number of equations and variables is identical. Like with other methods of finding extrema, not all solutions to the system will give absolute extrema. Just as with critical points, we must determine which of all the solutions will give extrema. Fortunately, this is simply done by evaluating f at our possible candidates. The following example illustrates the method.

Example 2.11.2

To find the extreme values of $z = x + 2y$ on the circle $x^2 + y^2 = 1$, we compute the gradients

$$\nabla f = \langle 1, 2 \rangle \qquad \text{and} \qquad \nabla g = \langle 2x, 2y \rangle$$

and then form the system

$$\begin{aligned} 1 &= 2\lambda x \\ 2 &= 2\lambda y \\ x^2 + y^2 &= 1. \end{aligned}$$

We begin with the first equation to get that $\lambda = \frac{1}{2x}$. Substituting this into the second, we obtain $2x = y$. Then we use this relation in the last equation to write $x^2 + (2x)^2 = 1$. Solving this equation gives $x = \pm 1/\sqrt{5}$; then $y = \pm 2/\sqrt{5}$. The maximum is then $z = 1/\sqrt{5} + 4/\sqrt{5} = \sqrt{5}$ and the minimum is $z = -1/\sqrt{5} - 4/\sqrt{5} = -\sqrt{5}$. ♦

Exercise 2.11.3 Find the extrema of $z = x + y$ given the constraint $x^2 + y^2 \leq 1$.

Example 2.11.4

To find the point on the surface $x^2 + xy - z^2 + 4 = 0$ that is closest to the origin requires us to minimize the distance function $D = \sqrt{x^2 + y^2 + z^2}$. However, to minimize this function is the same as minimizing $D^2 = x^2 + y^2 + z^2$, which is simpler to do!

For this problem, our function to minimize is distance, which makes the surface the constraint. So we have the system

$$\begin{aligned} 2x &= \lambda(2x + y) \\ 2y &= \lambda x \\ 2z &= -2\lambda z \\ x^2 + xy - z^2 + 4 &= 0 \end{aligned}$$

The third equation simplifies to $\lambda = -1$, and we use this value in the first equation to get that $y = -4x$. Using this relation and the value of λ in the second equation, we find that $x = 0$. Thus, $y = 0$ and $z = \pm 2$. Hence, there are two points on the surface closest to the origin: $(0, 0, 2)$ and $(0, 0, -2)$. ♦

Some Applications. The method of Lagrange multipliers allows us to find extrema of a wide variety of problems. The next few examples illustrate business and physics applications.

Example 2.11.5

A rocket is launched with a constant acceleration of a feet per second per second. The rocket's height after t seconds is given by $f(t, a) = \frac{1}{2}(a - 32)t^2$ feet. Fuel usage for t seconds is proportional to $a^2 t$ and the limited fuel capacity of the rocket satisfies $a^2 t = 8000$. We are to find the value of a that maximizes the height that the rocket attains when the fuel runs out.

In order to find this value, we maximize $f(t, a)$ given the constraint $a^2 t - 8000 = 0$. Hence, we have the system

$$\begin{aligned} (a - 32)t &= \lambda a^2 \\ \tfrac{1}{2}t^2 &= 2\lambda a t \\ a^2 t - 8000 &= 0. \end{aligned}$$

The first equation gives us $\lambda = (a - 32)t/a^2$, which, when substituted into the second equation, gives us

$$\begin{aligned} \frac{1}{2}t^2 &= \frac{2a(a - 32)t^2}{a^2} \\ a^2 t^2 &= 4a(a - 32)t^2 \\ 3a^2 t^2 &= 128 a t^2 \\ a &= \tfrac{128}{3}. \end{aligned}$$

(Note that the solutions $a = 0$ and $t = 0$ are rejected because they do not satisfy the constraint.) Hence, with an acceleration of $\frac{128}{3}$, we have

$$t = \frac{8000}{a^2} = \frac{8000}{(128/3)^2} = \frac{1125}{256} \approx 4.395 \text{ seconds.}$$

The height the rocket attains when the fuel runs out is $f(4.395, \frac{128}{3}) = \frac{1}{2}(\frac{128}{3} - 32)(\frac{1125}{256})^2 \approx 102.997$ feet. ♦

Next we demonstrate an application to business and economics. In 1928 Charles Cobb and Paul Douglas published a study in which they modeled the growth of the American economy during the period 1899–1922. They considered a simplified view of the economy in which production output is determined by the amount of labor involved and the amount of capital invested. Their model is the function

$$P(K, L) = aK^m L^n$$

where K is the capitol investment (value of all facilities and equipment in a year), L is the labor investment (person-hours per year), P is the total production (value of goods produced per year), and a, m, and n are constants reflecting the types of capitol, labor, and available technology. This is known as the **Cobb-Douglas production function**.

This production function is now known as being too simplified for complex economies like the United States, but can still describe some economies in emerging markets.

Example 2.11.6
Suppose the production of a company follows the Cobb-Douglas production model $P = 300K^{1/3}L^{2/3}$. However, cost constraints on the business force $5K + 2L \leq 180$. How do we maximize production?

The answer is easy: Lagrange multipliers! Although the constraint is an inequality, we may approach this problem in the same manner as the others. The only thing we must take into consideration is that the Lagrange multiplier method only gives us extrema *on the boundary* of the constraint. So the first partials of the function must be searched for critical points also. We begin with that. The partials of P are

$$\frac{\partial P}{\partial K} = 100K^{-2/3}L^{2/3} \qquad \text{and} \qquad \frac{\partial P}{\partial L} = 200K^{1/3}L^{-1/3}$$

and we must reject the only solutions of $K = 0$ and $L = 0$. So we continue with the system

$$\begin{aligned} 100K^{-2/3}L^{2/3} &= 5\lambda \\ 200K^{1/3}L^{-1/3} &= 2\lambda \\ 5K + 2L - 180 &= 0 \end{aligned}$$

The first equation tells us $\lambda = 20L^{2/3}/K^{2/3}$. We substitute this into the second equation to get $200K^{1/3}/L^{1/3} = 40L^{2/3}/K^{2/3}$, or $5K = L$. Using this relation in the third equation, we get $L = 60$; whence $K = 12$. Therefore the production is maximized at $(12, 60)$ giving $P = 300(12^{1/3})(60^{2/3}) \approx \$10,526.46$ per year. ♦

Quadratic Forms. An important theoretical application of Lagrange multipliers is in extrema of quadratic forms on the unit circle. A quadratic form is a surface of the form

$$f(x, y) = ax^2 + 2bxy + cy^2$$

for constants a, b, and c. (Ellipsoids, paraboloids, and hyperboloids have such equations.) To maximize such a function, we proceed as before, with the constraint $g(x, y) = x^2 + y^2 - 1$. We have the system

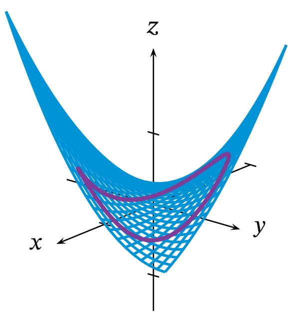

Figure 2.11 – A quadratic form (a hyperboloid of one sheet) with the unit circle projected on its surface.

$$\begin{aligned} 2ax + 2by &= 2\lambda x \\ 2bx + 2cy &= 2\lambda y \\ x^2 + y^2 - 1 &= 0. \end{aligned}$$

But notice that the first two equations can be represented as the matrix equation $A\mathbf{x} = \lambda\mathbf{x}$ where

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \qquad \text{and} \qquad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Moreover, $\mathbf{x}$ must meet the constraint $x^2 + y^2 = 1$; hence $\|\mathbf{x}\| = 1$. Therefore the Lagrange multiplier is actually an eigenvalue of A, and $\mathbf{x}$ is a unit eigenvector of A.

So here's the connection!

Hence, the eigenvalues are the critical points of f on the unit circle: particularly, the absolute maximum of f on the unit circle is the largest eigenvalue; the absolute minimum of f is the smallest eigenvalue. The corresponding unit eigenvectors are the values that give the critical points.

Example 2.11.7
To find the extrema of $x^2 + 4xy - 2y^2$ on the unit circle, we compute the eigeinvalues and eigenvectors of the associated matrix.

The matrix is $\left[\begin{smallmatrix}1 & 2\\ 2 & -2\end{smallmatrix}\right]$ with eigenvalues -3 and 2 and unit eigenvectors $\frac{1}{\sqrt{5}}\langle 1,-2\rangle$ and $\frac{1}{\sqrt{5}}\langle 2,1\rangle$. Hence, the maximum of f is 2, given by $(\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$; and the minimum is -3, given by $(\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}})$. ♦

Problems for Section 2.11

Find the absolute extrema of the function f subject to the given constraint.

1 $f(x,y) = 3x^2 + 3y^2 + 5$, subject to $x - y = 1$

2 $f(x,y) = x^3y$, subject to $2x + y = 5$

3 $f(x,y) = x^3y$, subject to $\sqrt{x} + \sqrt{y} = 1$

4 $f(x,y) = x^3 - y^3$, subject to $x - y = 2$

5 $f(x,y) = e^{x+y}$, subject to $x^2 + y^2 = 2$

6 $f(x,y,z) = x^2+2y^2+4z^2$, subject to $x^2+y^2+z^2 = 1$

7 Find the point on the surface $z = (y+1)^2-(x-2)^2+1$ nearest to the point $(2,-1,-2)$.

8 A closed rectangular box is to be made so that its volume is 60 cubic feet. The costs of the material for the top and bottom are 10 cents per square foot and 20 cents per square foot, respectively. The cost of the sides is 2 cents per square foot.

(a) Determine the cost function $C(x,y)$, where x and y are the length and width of the box.

(b) Find the dimensions of the box that will give the minimum cost and then compute the minimum cost. (Zill, 1985, p. 794)

9 A clothing company makes two types of overcoats, and the cost of manufacturing these overcoats is $C(x,y) = 2x^2 + 6y^2 + 4xy + 10$. If a total of 20 overcoats can be made daily, how many of each type should be made to minimize the cost? What is the minimum cost?

10 *(Calculator)* Maximize the production given by $P(K,L) = 10K^{0.6}L^{0.4}$ with constraint $2K + 3L \le 60$.

11 Find the extrema of the quadratic form $13x^2-8xy-2y^2$ on the unit circle.

2.12 Two Differential Operators

A modern branch of mathematics, having achieved the art of dealing with the infinitely small, can now yield solutions in other more complex problems of motion, which used to appear insoluble. This modern branch of mathematics, unknown to the ancients, when dealing with problems of motion, admits the conception of the infinitely small, and so conforms to the chief condition of motion (absolute continuity) and thereby corrects the inevitable error which the human mind cannot avoid when dealing with separate elements of motion instead of examining continuous motion. In seeking the laws of historical movement just the same thing happens. The movement of humanity, arising as it does from innumerable human wills, is continuous. To understand the laws of this continuous movement is the aim of history. Only by taking an infinitesimally small unit for observation (the differential of history, that is, the individual tendencies of man) and attaining to the art of integrating them (that is, finding the sum of these infinitesimals) can we hope to arrive at the laws of history.

— Leo Tolstoy

In this final section of the present chapter, we introduce two differential operators. We have already encountered one differential operator, the gradient. Recall that the gradient of a function $f(x,y,z)$ is the vector

$$\nabla f = \left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle.$$

The gradient ∇ by itself is meaningless; it is an operator, and as such, we take the gradient *of* something. The two new differential operators

3.4 Integration in the Physical World

We speak of invention: it would be more correct to speak of discovery. The distinction between these two words is well known: discovery concerns a phenomenon, a law, a being which already existed, but had not been perceived. Columbus discovered America: it existed before him; on the contrary, Franklin invented the lightning rod: before him there had never been any lightning rod.

Such a distinction has proved less evident than it appears at first glance. Torricelli has observed that when one inverts a closed tube on a mercury trough, the mercury ascends to a certain determinate height: this is a discovery; but in doing this, he has invented the barometer; and there are plenty of examples of scientific results which are just as much discoveries as inventions.

— Jacques Hadamard

In this section, we present six applications of double integrals. Some of these applications we have already encountered, and some are three-dimen- sional analogues to familiar concepts. All are important to the physical sciences. We begin with two straightforward ideas.

I. Volume. If $f(x, y)$ is the equation of a surface, then $V = \iint_R f\, dx\, dy$ is the volume between the surface and the xy-plane.

II. Area. For $f(x, y) = 1$, we have $A = \iint_R dx\, dy$ as the area of R.

These definitions of volume and area arise from the definition of double integrals. Note, however, that we may use single-variable integrals to compute the same things.

For instance, consider the region R bounded by the x-axis, the line $x = 3$, and the curve $y = x^2$. Using a single-variable integral, we may compute

$$\int_0^3 x^2\, dx = \frac{x^3}{3}\bigg|_0^3 = 9.$$

On the other hand, we may also compute the same area using a double integral:

$$\int_0^3 \int_0^{x^2} dy\, dx = \int_0^3 y\bigg|_0^{x^2} dx = \int_0^3 x^2\, dx = 9.$$

The same goes for volume. In Volume 1, we determined the volume of a solid with defined cross-sections. The following is one such problem.

Example 3.4.1

Suppose the region R in the xy-plane, bounded by the x-axis, the line $x = 3$, and the curve $y = x^2$, is the base of a solid whose cross-sections are isosceles right triangles with a leg of each triangle in R. We may find the volume by first finding an expression for the area of a cross section. Since the cross sections are triangles of height x^2 and base x^2, we have $A(x) = \frac{1}{2}x^4$. Then we compute

$$\int_0^3 \frac{x^4}{2}\, dx = \frac{x^5}{10}\bigg|_0^3 = 24.3.$$

On the other hand, we could use a double integral. The region R for our double integral is the same, but in this interpretation, we want the volume of the solid above R and below the plane $y = z$. Hence, we compute

$$\int_0^3 \int_0^{x^2} y\, dy\, dx = \int_0^3 \left. \frac{y^2}{2} \right|_0^{x^2} dx = \int_0^3 \frac{x^4}{2}\, dx = 24.3.$$

Of course, using a double integral in place of a single integral may complicate things if one cannot easily determine the surface under which we wish to find the volume. ♦

The next application concerns the mass of a **lamina**, or thin plate, sheet, or layer of a larger composite structure. A lamina will usually be defined by a region R.

III. Mass. If $\rho(x, y)$ is the equation of the density of a lamina (in mass per unit area), then $M = \iint_R \rho\, dx\, dy$ is the mass of R.

IV. Center of Mass. If ρ is density, then the center of mass $(\bar{x}, \bar{y})$ of the lamina represented by R is given by

$$\bar{x} = \frac{1}{M} \iint_R x\rho\, dx\, dy, \qquad \bar{y} = \frac{1}{M} \iint_R y\rho\, dx\, dy$$

where M is the mass of R.

The use of the Greek letter rho *(ρ) is traditional notation for density.*

Example 3.4.2

We compute the center of mass of the lamina covering the triangular region with vertices $(0, 0)$, $(1, 0)$, and $(1, 2)$, given that the plate's density is $\rho(x, y) = 12x + 12y + 6$.

First, we find M, the mass of the plate:

$$\int_0^1 \int_0^{2x} (12x + 12y + 6)\, dy\, dx = \int_0^1 \left. \left(12xy + 6y^2 + 6y\right) \right|_0^{2x} dx$$
$$= \int_0^1 \left(48x^2 + 12x\right)\, dx = \left. 16x^3 + 6x^2 \right|_0^1 = 22.$$

Now we have

$$\bar{x} = \frac{1}{22} \int_0^1 \int_0^{2x} \left(12x^2 + 12xy + 6x\right)\, dy\, dx$$
$$= \frac{1}{22} \int_0^1 \left. \left(12x^2 y + 6xy^2 + 6xy\right) \right|_0^{2x} dx = \frac{1}{22} \int_0^1 \left(48x^3 + 12x^2\right)\, dx$$
$$= \frac{1}{22} \left. \left(12x^4 + 4x^3\right) \right|_0^1 = \frac{1}{22}(16) = \frac{8}{11}.$$

Similarly, we find $\bar{y} = \frac{9}{11}$; hence, the center of mass is at the point $\left(\frac{8}{11}, \frac{9}{11}\right)$. ♦

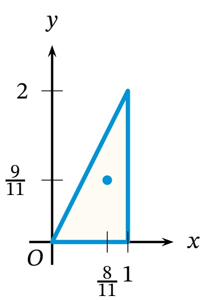

Figure 3.8 – The center of mass of the lamina from Example 3.4.2.

V. Moment of Inertia. If ρ is density, and R represents a lamina, the moments of inertia about the x-axis and y-axis are

$$I_x = \iint_R y^2 \rho\, dx\, dy, \qquad I_y = \iint_R x^2 \rho\, dx\, dy,$$

and the moment of inertia about the origin is $I_O = I_x + I_y$.

The moment of inertia quantifies the resistance of a physical object to angular acceleration. The moment of inertia is to rotational motion as mass is to linear motion. An object's moment of inertia depends on its shape and the distribution of mass within that shape: the greater the concentration of material away from the object's center, the larger the moment of inertia.

This is calculated in the above manner because the definition of the moment of inertia is the product of the mass times the square of the distance from an axis. The mass is $\iint_R \rho\, dx\, dy$ and the distance from the x-axis to a point of mass is y; hence the moment aound the x-axis is $I_x = \iint_R y^2 \rho\, dx\, dy$. Likewise for the moment around the y-axis.

Example 3.4.3
We compute the moments of inertia of the lamina covering the triangular region with vertices $(0,0)$, $(1,0)$, and $(1,2)$, given that the plate's density is $\rho(x,y) = 12x + 12y + 6$.

The moment of inertia about the x-axis is given by

$$\begin{aligned} I_x &= \int_0^1 \int_0^{2x} \left(12xy^2 + 12y^3 + 6y^2\right) dy\, dx \\ &= \int_0^1 \left(4xy^3 + 3y^4 + 2y^3\right)\Big|_0^{2x} = \int_0^1 \left(80x^4 + 16x^3\right) dx \\ &= 16x^5 + 4x^4\Big|_0^1 = 20. \end{aligned}$$

Similarly, the moment of inertia about the y-axis is $\frac{63}{5}$; hence the moment of inertia about the origin is $\frac{63}{5} + 20 = \frac{163}{5}$. ♦

VI. Surface Area. The surface area of a surface in space over a region R in the xy-plane is proved as the following theorem.

THEOREM 3.C (Area of a Surface) *If $z = f(x,y)$ is defined and has continuous partial derivatives in $R \subseteq D$, then the surface area of a surface in space is*

$$S = \iint_R \sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\, dx\, dy.$$

Proof. Let (x_i, y_i) be a point of the ith rectangle of the subdivision of $R \subseteq D$. (See Figure 3.9.) Then the tangent plane at (x_i, y_i) is

$$z - z_i = \frac{\partial z}{\partial x}(x - x_i) + \frac{\partial z}{\partial y}(y - y_i)$$

where the partials are evaluated at (x_i, y_i). Let S_i be the area of the part of tangent plane above the ith rectangle. Then S_i is the area of a parallelogram whose projection is a rectangle on R with area A_i. Let $\mathbf{n}_i$ be the normal to z at (x_i, y_i); hence,

$$\mathbf{n}_i = -\frac{\partial z}{\partial x}\mathbf{i} - \frac{\partial z}{\partial y}\mathbf{j} + \mathbf{k}.$$

Then $S_i = A_i \sec(\gamma_i)$, where γ_i is the angle between $\mathbf{n}_i$ and $\mathbf{k}$. Note that

$$\cos(\gamma_i) = \frac{\mathbf{n}_i \cdot \mathbf{k}}{||\mathbf{n}_i||||\mathbf{k}||} = \frac{1}{\sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}}$$

so that

$$\sec(\gamma_i) = \sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}.$$

Hence, if we let d_i denote the diagonal of the ith rectangle,

$$\begin{aligned}\lim_{\substack{n\to\infty \\ d\to 0}} \sum_{i=1}^{n} S_i &= \lim_{\substack{n\to\infty \\ d\to 0}} \sum_{i=1}^{n} A_i \sec(\gamma_i) \\ &= \lim_{\substack{n\to\infty \\ d\to 0}} \sum_{i=1}^{n} A_i \sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2} \\ &= \iint_R \sqrt{1 + \left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\, dx\, dy\end{aligned}$$

■

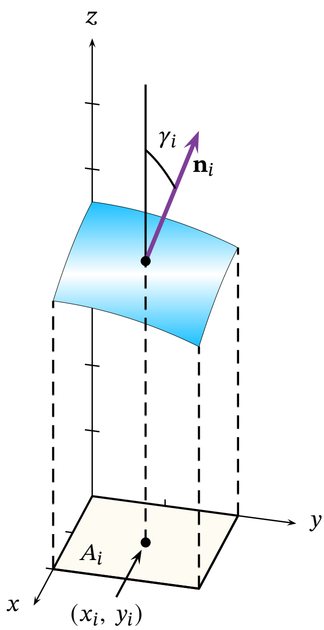

Figure 3.9 – The derivation of the surface area.

Example 3.4.4

We use the previous theorem to compute the surface area of the paraboloid $z = x^2 + y^2$ bounded by $x^2 + y^2 = 4$.

We have that $-\sqrt{4 - x^2} \le y \le \sqrt{4 - x^2}$ and $-2 \le x \le 2$ Thus,

$$A = \iint_R \sqrt{1 + (2x)^2 + (2y)^2}\, dx\, dy = \iint_R \sqrt{1 + 4x^2 + 4y^2}\, dx\, dy.$$

This integral is tedious; so we change coordinates to polar. Then the region is the complete circular disk of radius 2. Hence, $0 \le r \le 2$ and $0 \le \theta \le 2\pi$, and, since the Jacobian is r, we have

$$\begin{aligned}A &= \int_0^{2\pi} \int_0^2 r\sqrt{4r^2 + 1}\, dr\, d\theta = \int_0^{2\pi} \frac{1}{12}\left(4r^2 + 1\right)^{3/2}\Big|_0^2 d\theta \\ &= \int_0^{2\pi} \frac{1}{12}\left(17^{3/2} - 1\right) d\theta = \frac{\pi}{6}(17^{3/2} - 1).\end{aligned}$$

This is approximately 36.177. ♦

Problems for Section 3.4

1 Show the computations that verify the values of $\bar{y}$ from Example 3.4.2 and I_y from Example 3.4.3.

2 Compute the moment of inertia about the origin for the triangular lamina with vertices $(0,0)$, $(1,0)$, and $(0,3)$ and density $\rho(x,y) = 6$.

3 Compute the center of mass of the unit square in the first quadrant with density $\rho(x,y) = y \arctan(x)$.

4 Using a double integral, compute the volume of the solid whose base is the region R in the xy-plane bounded by the x-axis, the line $x = \frac{\pi}{2}$, the line $x = 0$, and the curve $y = \cos(x)$, and whose cross-sections are rectangles with width $\cos(x)$ and height 5.

5 Compute the surface area of the hemisphere $z = \sqrt{1 - x^2 - y^2}$ using a double integral.

6 Compute the surface area of the cone $2x^2 + 2y^2 = 5z^2$, $z \ge 0$, bounded by $x^2 + y^2 = 4$.