Alireza Ghaffari-Hadigheh
Farzali Izadi

A First course in Mathematical English for math students, First Edition

July 21, 2020

Azarbaijan Shahid Madani University

Contents

1

Selected Topics in Calculus

1.1 Limits

[1] Central to calculus is the value of the slope of a line, $\frac{\triangle y}{\triangle x}$, but when both the numerator and denominator become almost zero. To evaluate the slope, that ratio, under those vanishing conditions, requires the idea of a limit. And central to the idea of a limit is the idea of a sequence of rational numbers.

We encounter such a sequence in geometry when we determine a value for π, which is the ratio of the circumference of a circle to the diameter. To do that, we inscribe in the circle a regular polygon. The ratio of the perimeter of the polygon to the diameter, which we can actually calculate, will be an approximation to π. And as we increase the number of sides – that is, if we consider a sequence of polygons: 60 sides, 61 sides, 62, 63, 64, and so on – then the sequence of those ratios gets closer and closer to π. Now, the circle is never equal to any polygon. But by considering a sufficiently large number of sides, the difference between the circle and that polygon, the error, will be less than any small number we name. Less even than

$$0.00000000000000000000000000000001!$$

That is the idea of a sequence approaching a limit, or a boundary. By that process, we can approximate the value of π as closely as we possibly can. For instance, the reader surely can recognize the

[1] This section has been quoted from:
"http://www.salonhogar.net/themathpage/aCalc/limits.htm".

number that is the limit of this sequence of rational numbers.

$$3, 3.1, 3.14, 3.141, 3.1415, 3.14159, 3.141592, \ldots$$

In fact, when we approximate any irrational number, it will be the limit of the sequence of its rational approximations. And again, the limit, the irrational number, will not be a member of the sequence.

> We speak of a sequence as being infinite, which is a brief way of saying that, no matter how many terms we have named already, we could always name one more. And if the sequence has a limit, then each term we add will be closer.

The limit of a variable

Consider this sequence of values of a variable x:

$$1.9, 1.99, 1.999, 1.9999, 1.99999, \ldots$$

These values are getting closer and closer to 2 – they are approaching 2 as their limit. 2 is the smallest number such that no matter which term of that sequence we consider, it will be less than 2.

We can define "closer and closer" mathematically, by saying that the differences between the terms of that sequence and 2 become and remain less than any small number we might name. That is, we can define a limit by considering the sequence of differences, $x - 2$:

$$1.9 - 2, 1.99 - 2, 1.999 - 2, 1.9999 - 2, 1.99999 - 2, \ldots$$

Now name any positive number, however small, for example, "The width of a hydrogen atom". Then if we consider a sufficient number of those differences, the absolute value of a difference (which is negative) will be less than that small number. The error, in other words, between the terms of a sequence and their limit, can be made as small as we please.

When a variable x approaches a limit l, we symbolize that as $x \to l$. Read: "The values of x approach l as a limit," or simply,"x approaches l". We also say that a sequence converges to a limit. The sequence above converges to 2.

1.2 Derivative

[2] The problem of finding the tangent line to a curve and the problem of finding the velocity of an object both involve finding the same type of limit. This special type of limit is called a derivative and we will see that it can be interpreted as a rate of change in any of the sciences or engineering.

In general, suppose an object moves along a straight line according to an equation of motion $s = f(t)$, where s is the displacement (directed distance) of the object from the origin at time t. The function f that describes the motion is called the **position function** of the object. In the time interval from $t = a$ to $t = a + h$ the change in position is $f(a + h) - f(a)$. The average velocity over this time interval is

$$\text{Average velocity} = \frac{\text{displacement}}{\text{time}} = \frac{f(a+h) - f(a)}{h},$$

which is the same as the slope of the secant line.

Now suppose we compute the average velocities over shorter and shorter time intervals $[a, a + h]$. In other words, we let h approach 0. We define the velocity (or instantaneous velocity) at time $t = a$ to be the limit of these average velocities:

$$v(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

This means that the velocity at time $t = a$ is equal to the slope of the tangent line at this point.

We have seen that the same type of limit arises in finding the slope of a tangent line or the velocity of an object. In fact, limits of the form

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

arise whenever we calculate a rate of change in any of the sciences or engineering, such as a rate of reaction in chemistry or a marginal cost in economics. Since this type of limit occurs so widely, it is given a special name and notation, namely derivative.

[2] This section is quoted from:
Stewart, James, Daniel K. Clegg, and Saleem Watson. Calculus: early transcendentals. Cengage Learning, 2020.

1.3 Definite and Indefinite Integrals

[3] A physicist who knows the velocity of a particle might wish to know its position at a given time. An engineer who can measure the variable rate at which water is leaking from a tank wants to know the amount leaked over a certain time period. A biologist who knows the rate at which a bacteria population is increasing might want to deduce what the size of the population will be at some future time. In each case, the problem is to find a function F whose derivative is a known function f. If such a function F exists, it is called an antiderivative of f.

Indefinite Integral

The Fundamental Theorem of Calculus is appropriately named because it establishes a connection between the two branches of calculus: differential calculus and integral calculus. Differential calculus arose from the tangent problem, whereas integral calculus arose from a seemingly unrelated problem, the area problem.

Newtons mentor at Cambridge, Isaac Barrow (1630 - 1677), discovered that these two problems are actually closely related. In fact, he realized that differentiation and integration are inverse processes. The Fundamental Theorem of Calculus gives the precise inverse relationship between the derivative and the integral. It was Newton and Leibnitz who exploited this relationship and used it to develop calculus into a systematic mathematical method. In particular, they saw that the Fundamental Theorem enabled them to compute areas and integrals very easily without having to compute them as limits of sums.

We end this section by bringing together the two parts of the Fundamental Theorem.

Theorem 1.1 *Suppose f is continuous on $[a, b]$.*

1. If $g(x) = \int_a^x f(t)\, dt$, then $g'(x) = f(x)$.

[3] This section is quoted from:
Stewart, James, Daniel K. Clegg, and Saleem Watson. Calculus: early transcendentals. Cengage Learning, 2020.

2. $\int_a^b f(t)\ dt = F(b) - F(a)$*,where* F *is an antiderivative of* f*, that is* $F' = f$*.*

Both parts of the Fundamental Theorem establish connections between antiderivatives and definite integrals. Part 1 says that if f is continuous, then $\int_a^x f(t)\ dt$ is an antiderivative of f. Part 2 says that $\int_a^b f(t)\ dt$ can be found by evaluating $F(b) - F(a)$, where F is an antiderivative of f. We need a convenient notation for antiderivatives that makes them easy to work with. Because of the relation given by the Fundamental Theorem between antiderivatives and integrals, the notation $\int f(x)\ dx$ is traditionally used for an antiderivative of and is called an indefinite integral.

Definite Integral

A limit of the form

$$\lim_{n\to\infty} \sum_{i=0}^{n} f(x_i^*) \triangle x = \lim_{n\to\infty} [f(x_1^*) \triangle x + f(x_2^*) \triangle x + \cdots + f(x_n^*) \triangle x]$$

arises when we compute an area. We also saw that it arises when we try to find the distance traveled by an object. It turns out that the same type of limit occurs in a wide variety of situations even when f is not necessarily a positive function. Later, we will see that limits of this form also arise in finding lengths of curves, volumes of solids, centers of mass, force due to water pressure, and work, as well as other quantities. We therefore give this type of limit a special name and notation.

Definition 1.1 If a function f is defined for $a \le x \le b$, we divide the interval $[a, b]$ into n subintervals of equal width $\triangle x = (b - a)/n$. We let $x_0 (= a), x_1, x_2, \ldots, x_n (= b)$ be the endpoints of these subintervals and we let $x_0^*, x_1^*, x_2^*, \ldots, x_n^*$ be any sample points in these subintervals, so x_i^* lies in the ith subinterval $[x_i, x_{i+1}]$. Then the definite integral of f from a to b is

$$\int_a^b f(x)\ dx = \lim_{n\to\infty} \sum_{i=0}^{n} f(x_i^*) \triangle x$$

provided that this limit exists. If it does exist, we say that f is integrable on $[a, b]$.

1.4 Exercises

1. Translate the following sentences.

1. A limit is the value that a function (or sequence) approaches as the input (or index) approaches some value. Limits are essential to calculus (and mathematical analysis in general) and are used to define continuity, derivatives, and integrals.
2. The derivative of a function of a real variable measures the sensitivity to change of the function value (output value) with respect to a change in its argument (input value).
3. The slope or gradient of a line is a number that describes both the direction and the steepness of the line.
4. The absolute value or modulus of a real number x, denoted by $|x|$, is the non-negative value of x without regard to its sign.
5. A function is a relation between sets that associates to every element of a first set exactly one element of the second set.
6. A differentiable function of one real variable is a function whose derivative exists at each point in its domain.
7. A variable is a symbol used to represent an arbitrary element of a set. In addition to numbers, variables are commonly used to represent vectors, matrices and functions.
8. A sequence is an enumerated collection of objects in which repetitions are allowed and order does matter.
9. An integral assigns numbers to functions in a way that can describe displacement, area, volume, and other concepts that arise by combining infinitesimal data.
10. There are many ways of formally defining an integral, not all of which are equivalent. The most commonly used definitions of integral are Riemann integrals and Lebesgue integrals.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
		Continuous	
Derivative			
	Converge		
	limit		
		Computational	
Calculation			
			Originally
			Approximately
		Increasing	

3. Use the correct form of the word.

1. A rigorous definition of (Continuous)..... of real functions is usually given in a first course in calculus in terms of the idea of a limit.
2. Pointwise (converge) ..... is one of various senses in which a sequence of functions can converge to a particular function.
3. The principles of (integral) ..... were formulated independently by Isaac Newton and Gottfried Wilhelm Leibniz in the late 17th century, who thought of the integral as an infinite sum of rectangles of infinitesimal width.
4. The process of finding a derivative is called (derivative) .....
5. In analytic geometry, an asymptote of a curve is a line such that the distance between the curve and the line (approach) ..... zero as one or both of the x or y coordinates tends to infinity.
6. Familiar examples of (inscribe) ..... figures include circles inscribed in triangles or regular polygons, and triangles or (regulate)..... polygons inscribed in circles.

4. Fill gaps with one of the following words

1. curvature	2. subspace	3. straight	4. denominator
5. variable	6. numerator	7. integral	8. real
9. function	10. set	11. diameter	12. ratio

1. The notion of line or ..... line was introduced by ancient mathematicians to represent straight objects (i.e., having no curvature) with negligible width and depth.
2. Intuitively, the ..... is the amount by which a curve deviates from being a straight line, or a surface deviates from being a plane.
3. ..... is a subset of a space which is a space in its own right.
4. In a formula, a dependent variable is a ..... that is implicitly a function of another (or several other) variables.
5. ..... is the part of a fraction that is above the line and signifies the number to be divided by the .....
6. ..... is a mathematical correspondence that assigns exactly one element of one set to each element of the same or another .....
7. Indefinite ..... is any function whose derivative is a given function.
8. A ..... number that is not rational is called irrational.
9. The number π is a mathematical constant. It is defined as the ..... of a circle's circumference to its ....., and it also has various equivalent definitions.

2

Selected Topics in Linear Algebra

2.1 Vector Space

[1] There are two approaches to linear algebra, each having its virtues. The first is abstract. A vector space is defined axiomatically as a collection of objects, called vectors, with a sum and a scalar-vector product. As the theory develops, matrices emerge, almost incidentally, as scalar representations of linear transformations. The advantage of this approach is generality. The disadvantage is that the hero of our story, the matrix, has to wait in the wings.

The second approach is concrete. Vectors and matrices are defined as arrays of scalars - here arrays of real or complex numbers. Operations between vectors and matrices are defined in terms of the scalars that compose them. The advantage of this approach for a treatise on matrix computations is obvious: it puts the objects we are going to manipulate to the fore. Moreover, it is truer to the history of the subject. Most decompositions we use today to solve matrix problems originated as simplifications of quadratic and bilinear forms that were defined by arrays of numbers.

Although we are going to take the concrete approach, the concepts of abstract linear algebra will not go away. It is impossible to derive and analyze matrix algorithms without a knowledge of such things as subspaces, bases, dimension, and linear transformations. Consequently, after introducing vectors and matrices and describing how they combine, we will turn to the concepts of lin-

[1] This section is quoted from:// GW Stewart, Matrix Algorithms: Volume 1: Basic Decompositions, SIAM, (1998)

ear algebra. This inversion of the traditional order of presentation allows us to use the power of matrix methods to establish the basic results of linear algebra.

The results of linear algebra apply to vector spaces over an arbitrary field . However, we will be concerned entirely with vectors and matrices composed of real and complex numbers. What distinguishes real and complex numbers from an arbitrary field of scalars is that they posses a notion of limit. This notion of limit extends in a straightforward way to finite-dimensional vector spaces over the real or complex numbers, which inherit this topology by way of a generalization of the absolute value called the norm. Moreover, these spaces have a Euclidean geometry - e.g ., we can speak of the angle between two vectors.

2.1.1 Vectors

Since we are going to define matrices as two-dimensional arrays of numbers, called scalars, we could regard a vector as a degenerate matrix with a single column, and a scalar as a matrix with one element. In fact, we will make such identifications later. However, the words "*scalar*" and "*vector*" carry their own bundles of associations, and it is therefore desirable to introduce and discuss them independently.

Scalars

Although vectors and matrices are represented on a computer by floating-point numbers - and we must ultimately account for the inaccuracies this introduces - it is convenient to regard matrices as consisting of real or complex numbers. We call these numbers, scalars.

Real and complex numbers

The set of real numbers will be denoted by $\mathbb{R}$. As usual, $|x|$ will denote the absolute value of $x \in \mathbb{R}$. The set of complex numbers will be denoted by $\mathbb{C}$. Any complex number z can be written in the form $z = x + i\ y$, where x and y are real and i is the principal square root of -1. The number x is the real part of z and is written

Re z. The number y is the imaginary part of z and is written **Im** z. The absolute value, or modulus, of z is $|z| = \sqrt{x^2 + y^2}$. The conjugate $x - iy$ of z will be written $\bar{z}$.

Vectors

In three dimensions, a directed line segment can be specified by three numbers x, y, and z. The following definition is a natural generalization of this observation.

Definition 2.1 *A vector x of dimension n or n-vector is an array of n scalars of the form $x = (x_1, x_2, \ldots, x_n)$. The scalars x_i are called the components of x. The set of n-vectors with real components will be written $\mathbb{R}^n$. The set of n-vectors with real or complex components will be written $\mathbb{C}^n$.*

Operations with vectors and scalars

Vectors can be added and multiplied by scalars. These operations are performed componentwise as specified in the following definition.

Definition 2.2 *Let x and y be n-vectors and α be a scalar. The sum of x and y is the vector*

$$x + y = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

The scalar-vector product αx is the vector

$$\alpha x = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix}.$$

The following properties are easily established from the definitions of the vector sum and scalar-vector product.

Theorem 2.1 *Let x, y, and z be n-vectors and α and β be scalars. Then*

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

The properties listed above insure that a sum of products of the form $\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$ is unambiguously defined and independent of the order of summation. Such a sum of products is called a linear combination of the vectors $x_1, x_2, \ldots, x_n$.

The properties listed in Theorem 2.1 are sufficient to define a useful mathematical object called a vector space or linear space. Specifically, a vector space consists of a field $\mathcal{F}$ of objects called scalars and a set of objects x called vectors.

2.1.2 Matrices

Matrices and the matrix-vector product arise naturally in the study of systems of equations. An $m \times n$ system of linear equations

$$\begin{array}{cccccc} a_{11}x_1 + & a_{12}x_2 + & \cdots & +a_{1n}x_n & = b_1 \\ a_{21}x_1 + & a_{22}x_2 + & \cdots & +a_{2n}x_n & = b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1}x_1 + & a_{m2}x_2 + & \cdots & +a_{mn}x_n & = b_m \end{array} \tag{2.1}$$

can be written compactly in the form

$$\sum_{j=1}^{n} a_{ij}x_{=}b_i, \qquad i = 1, 2, \ldots, m$$

However, matrices provide an even more compact representation. If we define arrays A, x, and b by

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

and define the product Ax by the left-hand side of (2.1), then (2.1) is equivalent to $Ax = b$. The scalars a_{ij} are called the elements of A. The set $m \times n$ matrices with real elements is written $\mathbb{R}^{m \times n}$. The set of $m \times n$ matrices with real or complex components is written $\mathbb{C}^{m \times n}$. The indices i and j of the elements a_{ij} of a matrix are called respectively the row index and the column index.

2.2 Eigenvalues and Eigenvectors

[2] The polynomial $p(\lambda) = \det(A - \lambda I)$ is called the characteristic polynomial of A. The roots of $p(\lambda) = 0$ are the eigenvalues of A. Since the degree of the characteristic polynomial $p(\lambda)$ equals n, the dimension of A, it has n roots, so A has n eigenvalues.

A nonzero vector x satisfying $Ax = \lambda x$ is a (right) eigenvector for the eigenvalue λ. A nonzero vector y such that $y^T A = \lambda y^T$ is a left eigenvector.

Most of algorithms will involve transforming the matrix A into simpler, or canonical forms, from which it is easy to compute its eigenvalues and eigenvectors. These transformations are called similarity transformations. The two most common canonical forms are called the Jordan form and Schur form. The Jordan form is useful theoretically but is very hard to compute in a numerically stable fashion, which is why we will aim to compute the Schur form instead.

To motivate Jordan and Schur forms, let us ask which matrices have the property that their eigenvalues are easy to compute. The easiest case would be a diagonal matrix, whose eigenvalues are simply its diagonal entries. Equally easy would be a triangular matrix, whose eigenvalues are also its diagonal entries. Below

[2] This section is quoted from:
Applied Numerical Linear Algebra - James W. Demmel Cambridge University Press, 1997

we will see that a matrix in Jordan or Schur form is triangular. But recall that a real matrix can have complex eigenvalues, since the roots of its characteristic polynomial may be real or complex. Therefore, there is not always a real triangular matrix with the same eigenvalues as a real general matrix, since a real triangular matrix can only have real eigenvalues. Therefore, we must either use complex numbers or look beyond real triangular matrices for our canonical forms for real matrices.

It will turn out to be sufficient to consider block triangular matrices, i.e. matrices of the form

$$\begin{pmatrix} A_{11} & A_{12} & \dots & A_{1b} \\ & A_{22} & \dots & A_{2b} \\ & & \ddots & \\ & & & A_{bb} \end{pmatrix}$$

where each A_{ii} is square and all entries below the A_{ii} blocks are zero. One can easily show that the characteristic polynomial $\det(A - \lambda I)$ of A is the product $\prod_{i=1}^{b} \det(A_{ii} - \lambda I)$ of the characteristic polynomials of the A_{ii} and therefore that the set $\lambda(A)$ of eigenvalues of A is the union $\bigcup_{i=1}^{b} \lambda(A_{ii})$ of the sets of eigenvalues of the diagonal blocks A_{ii}. The canonical forms that we compute will be block triangular and will proceed computationally by breaking up large diagonal blocks into smaller ones. If we start with a complex matrix A, the final diagonal blocks will be 1-by-1, so the ultimate canonical form will be triangular. If we start with a real matrix A, the ultimate canonical form will have 1-by-1 diagonal blocks (corresponding to real eigenvalues) and 2-by-2 diagonal blocks (corresponding to complex conjugate pairs of eigenvalues); such a block triangular matrix is called quasi-triangular.

2.3 Exercises

1. Translate the following sentences.

1. The first modern and more precise definition of a vector space was introduced by Peano in 1888, a theory of linear transformations of finite-dimensional vector spaces had emerged.
2. An element of a specific vector space may have various nature; for example, it could be a sequence, a function, a polynomial or a matrix. Linear algebra is concerned with those properties of such objects that are common to all vector spaces.
3. A set of vectors is linearly independent if none is in the span of the others.
4. Matrices allow explicit manipulation of finite-dimensional vector spaces and linear maps. Their theory is thus an essential part of linear algebra.
5. Systems of linear equations form a fundamental part of linear algebra. Historically, linear algebra and matrix theory has been developed for solving such systems. In the modern presentation of linear algebra through vector spaces and matrices, many problems may be interpreted in terms of linear systems.
6. A symmetric matrix is always diagonalizable. There are non-diagonalizable matrices.
7. A linear form is a linear map from a vector space V over a field F to the field of scalars F, viewed as a vector space over itself.
8. The inner product is an example of a bilinear form, and it gives the vector space a geometric structure by allowing for the definition of length and angles
9. Nearly all scientific computations involve linear algebra. Consequently, linear algebra algorithms have been highly optimized.
10. In multilinear algebra, one considers multivariable linear transformations, that is, mappings that are linear in each of a number of different variables.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
analysis			
			axiomatically
	combine		
		dependent	
	characterize		
Calculation			
			sufficiently
		stable	
representation			

3. Use the correct form of the word.

1. Linear algebra initially emerged as a method for (solve)..... systems of linear equations.
2. In a coordinate system, there is a dimension and an (associate) ..... number of coordinates.
3. Here, vector spaces are the central objects of study, and linear (transform) ..... are the mappings between them.
4. By definition of linear independence, **0** may be (express)..... as a linear combination of elements of S in only one way.
5. In order to (discussion) ..... the "dimension" of a vector space, it is important to realize that this is not a defining concept of vector space.
6. Matrices are useful in a variety of fields and form the basis for linear algebra. Their (apply) ..... include solving systems of linear equations, path-finding in graph theory, and several applications in group theory (especially representation theory).
7. Matrix multiplication is not commutative. In other words, it is not (general) ..... true that $AB = BA$.

4. Fill gaps with one of the following words

1. analysis	2. space	3. definition	4. product
5. normed	6. complete	7. finite	8. linear
9. function	10. integrable	11. metric	12. measures

Vector spaces that are not ..... dimensional often require additional structure to be tractable. A ..... vector space is a vector space along with a function called a norm, which measures the "size" of elements. The norm induces a ....., which ..... the distance between elements, and induces a topology, which allows for a ..... of continuous maps. The metric also allows for a definition of limits and completeness - a metric space that is ..... is known as a Banach space.

A complete metric space along with the additional structure of an inner ..... (a conjugate symmetric sesquilinear form) is known as a Hilbert ....., which is in some sense a particularly well-behaved Banach space. Functional analysis applies the methods of ..... algebra alongside those of mathematical analysis to study various ..... spaces; the central objects of study in functional ..... are L^p spaces, which are Banach spaces, and especially the L^p space of square ..... functions, which is the only Hilbert space among them. Functional analysis is of particular importance to quantum mechanics, the theory of partial differential equations, digital signal processing, and electrical engineering. It also provides the foundation and theoretical framework that underlies the Fourier transform and related methods.

3

Selected Topics in Statistics

3.1 Statistics

[1] Statistics is the science of sampling. How one set of measurements differs from another and what the implications of those differences might be are its primary concerns. Conceptually, the subject is rooted in the mathematics of probability, but its applications are everywhere. Statisticians are as likely to be found in a research lab or a field station as they are in a government office, an advertising firm, or a college classroom.

Properly applied, statistical techniques can be enormously effective in clarifying and quantifying natural phenomena. In general, statistical techniques are employed either to (1) describe what did happen or (2) predict what might happen. It is unarguably true that the interplay between description and prediction.

3.2 Descriptive Statistics

Descriptive statistics are used to describe the main features of a collection of data in quantitative terms. Descriptive statistics are distinguished from inferential statistics (or inductive statistics), in that descriptive statistics aim to quantitatively summarize a data set, rather than being used to support inferential statements about

[1] This section is quoted from:
An Introduction to Mathematical Statistics and Its Applications; Fourth Edition ; Richard J. Larsen, Vanderbilt University and Morris L. Marx University of West Florida ; New Jersey .(2006) pearson Education' Inc.

the population that the data are thought to represent. Even when a data analysis draws its main conclusions using inductive statistical analysis, descriptive statistics are generally presented along with more formal analysis. For example in a paper reporting on a study involving human subjects, there typically appears a table giving the overall sample size, sample sizes in important subgroups (e.g. for each treatment or exposure group), and demographic or clinical characteristics such as the average age, the proportion of subjects with each gender, and the proportion of subjects with related comorbidities.

3.2.1 Mean

In statistics, mean has two related meanings:

- the arithmetic mean (and is distinguished from the geometric mean or harmonic mean).
- the expected value of a random variable, which is also called the population mean.

For a real-valued random variable X, the mean is the expectation of X.

3.2.2 Arithmetic Mean

The arithmetic mean (or simply the mean) of a list of numbers is the sum of all of the list divided by the number of items in the list. If the list is a statistical population, then the mean of that population is called a population mean. If the list is a statistical sample, we call the resulting statistic a sample mean.

The mean is the most commonly-used type of average and is often referred to simply as the average. The term *mean* or *arithmetic mean* is preferred in mathematics and statistics to distinguish it from other averages such as the median and the mode.

3.2.3 Geometric Mean

The geometric mean is a type of mean or average, which indicates the central tendency or typical value of a set of numbers. It is similar to the arithmetic mean, which is what most people think of with the word average, except that instead of adding the set of

numbers and then dividing the sum by the count of numbers in the set, n, the numbers are multiplied and then the nth root of the resulting product is taken. For instance, the geometric mean of two numbers, say 2 and 8, is just the square root of their product which equals 4; that is $\sqrt{2 \times 8} = 4$.

The geometric mean can also be understood in terms of geometry. The geometric mean of two numbers, a and b, is the length of one side of a square whose area is equal to the area of a rectangle with sides of lengths a and b. Similarly, the geometric mean of three numbers, a, b, and c, is the length of one side of a cube whose volume is the same as that of a cuboid with sides whose lengths are equal to the three given numbers.

The geometric mean only applies to positive numbers. It is also often used for a set of numbers whose values are meant to be multiplied together or are exponential in nature, such as data on the growth of the human population or interest rates of a financial investment. The geometric mean is also one of the three classic Pythagorean means, together with the aforementioned arithmetic mean and the harmonic mean.

3.2.4 Harmonic Mean

The harmonic mean (formerly sometimes called the subcontrary mean) is one of several kinds of average. Typically, it is appropriate for situations when the average of rates is desired. The harmonic mean H of the positive real numbers $x_1, x_2, \ldots, x_n$ is defined to be

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}, \qquad x_i > 0 \text{ for all } i.$$

Equivalently, the harmonic mean is the reciprocal of the arithmetic mean of the reciprocals.

3.2.5 Median

A median is described as the numeric value separating the higher half of a sample, a population, or a probability distribution, from the lower half. The median of a finite list of numbers can be found by arranging all the observations from lowest value to highest value

and picking the middle one. If there is an even number of observations, then there is no single middle value, so one often takes the mean of the two middle values. The median can be used as a measure of location when a distribution is skewed , when end values are not known, or when one requires reduced importance to be attached to outliers, e.g. because they may be measurement errors. A disadvantage of the median is the difficulty of handling it theoretically.

3.2.6 Mode

The mode is the value that occurs the most frequently in a data set or a probability distribution. In some fields, notably education, sample data are often called scores, and the sample mode is known as the modal score. Like the statistical mean and the median, the mode is a way of capturing important information about a random variable or a population in a single quantity. The mode is in general different from the mean and median, and may be very different for strongly skewed distributions. The mode is not necessarily unique, since the same maximum frequency may be attained at different values. The most ambiguous case occurs in uniform distributions, wherein all values are equally likely.

3.2.7 Range

The range is the length of the smallest interval which contains all the data. It is calculated by subtracting the smallest observation (sample minimum) from the greatest (sample maximum) and provides an indication of statistical dispersion. It is measured in the same units as the data. Since it only depends on two of the observations, it is a poor and weak measure of dispersion except when the sample size is large.

The range, in the sense of the difference between the highest and lowest scores, is also called the crude range. When a new scale for measurement is developed, then a potential maximum or minimum will emanate from this scale. This is called the potential (crude) range. Of course this range should not be chosen too small, in order to avoid a ceiling effect. When the measurement is obtained, the resulting smallest or greatest observation, will provide the observed (crude) range.

3.2.8 Standard Deviation

The standard deviation of a statistical population, a data set, or a probability distribution is the square root of its variance. Standard deviation is a widely used measure of the variability or dispersion, being algebraically more tractable though practically less robust than the expected deviation or average absolute deviation.

It shows how much variation there is from the "*average*" (mean). It may be thought of as the average difference of the scores from the mean of distribution, how far they are away from the mean. A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values. In addition to expressing the variability of a population, standard deviation is commonly used to measure confidence in statistical conclusions.

3.2.9 Coefficient of Variation

The coefficient of variation is a normalized measure of dispersion of a probability distribution. It is defined as the ratio of the standard deviation σ to the mean $\mu : c_v = \frac{\sigma}{\mu}$

This is only defined for non-zero mean, and is most useful for variables that are always positive. It is also known as unitized risk or the variation coefficient. The coefficient of variation should only be computed for data measured on a ratio scale. As an example, if a group of temperatures are analyzed, the standard deviation does not depend on whether the Kelvin or Celsius scale is used. However the mean temperature of the data set would differ in each scale and thus the coefficient of variation would differ. So the coefficient of variation does not have any meaning for data on an interval scale. The variance-to-mean ratio, $\frac{\sigma^2}{\mu}$, is another similar ratio, but is not dimensionless, and hence not scale invariant.

A percentile is the value of a variable below which a certain percent of observations fall. So the 20th percentile is the value (or score) below which 20 percent of the observations may be found. The term percentile and the related term percentile rank are often used in descriptive statistics as well as in the reporting of scores from norm-referenced tests.

The 25th percentile is also known as the first quartile($Q1$); the 50th percentile as the median or second quartile($Q2$); the 75th percentile as the third quartile ($Q3$).

3.2.10 Variance

The variance of a random variable or distribution is the expected, or mean, value of the square of the deviation of that variable from its expected value or mean. Thus the variance is a measure of the amount of variation within the values of that variable, taking account of all possible values and their probabilities or weightings (not just the extremes which give the range).

3.3 Exercises

1. Translate the following text.

Statisticians gather information through observational studies and experiments. Observational studies observe and measure specific characteristics without modifying the subjects under study. In contrast, a statistical experiment applies a treatment to the subjects to see if a causal relationship exists. (Treatments can also be called factors, but this can be confusing because latent variables in factor analysis are also called factors.) Statistical experiments are designed to compare the outcomes of applying one or more treatments to experimental units, then comparing the results to a control group that does not receive a treatment.

Designing a statistical experiment starts with identifying the question(s) you want to answer. By carefully planning all the details of the experiment in advance, you decrease the probability of errors and increase the likelihood that the experiment will produce good data that leads to sound conclusions after analysis. The purpose of an experiment is to see if applying a treatment results in any observable differences in the experimental units. For a difference to matter, it needs to be measurable.

The independent variable is the one that you plan to change. The dependent variable is whatever you plan to measure after the treatment. Most real-life studies also have extraneous variables

that impact the results of the experiment. There are often variables that you don't even know about. A confounding variable is an extraneous variable that varies across the independent variable.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
category			
		predictable	
	normalize		
		dependent	
	indicate		
deviation			
	summarize		
			inferentially
expectation			

3. Use the correct form of the words.

1. Sample data are the result of making a series of (observe) .....s.
2. In a narrow sense, statistics is a branch of probability theory and its job is to determine stochastic parameters from (experiment) ..... data.
3. (Statistics)..... problems arise if the probability parameters of an experiment are unknown.
4. A statistical test is always based on certain model (assume) .....s about the population from which our sample comes.
5. Regardless of what we are trying to measure, the qualities that make a good measure of a scientific concept are high (reliable) ....., absence of bias, low cost, practicality, objectivity, high acceptance, and high concept (valid) .....

6. Both (category)..... and (quantity)..... variables are often recorded as numbers, so this is not a reliable guide to the major distinction between categorical and quantitative variables.

4. Fill gaps with one of the following words

1. distinguished	2. include	3. categories	4. examples
5. Continuous	6. zero	7. data	8. numbers
9. observations	10. both	11. regression	12. presenting

1. Ordinal data are also categorical, but in this case ..... have an order and can be ranked. Examples include stages of breast cancer.
2. Numeric data can be discrete or continuous. Discrete data have fixed values. ..... data can take any value, frequently within a given range.
3. Data can be broadly ..... as categorical or numeric. Categorical data may be nominal, ordinal or binary.
4. Binary, or dichotomous, data have only two possible outcomes. Common ..... are Yes/No or True/False responses, but they could also include other common epidemiological outcomes, such as “survive” and“not survive”.
5. Numeric data which may include something such as weight and length (where the range would be from ..... to, theoretically, infinity).
6. Nominal Data describes categorical data without an order. Examples ..... blood groups (O, A, B, AB), eye colour and marital status.
7. The most obvious first step in assessing a trend is to plot the ..... of interest by year.
8. Time series analysis refers to a particular collection of specialised ..... methods that illustrate trends in the data.
9. Interval data are numerical data where the differences between two ..... can be interpreted, but the ratio between two numbers is meaningless.
10. Moving averages (or rolling averages) provide a useful way of ..... time series data.

11. Ratio data are numerical and have a true zero and ..... differences and ratios are meaningful.
12. There are four data scales: nominal, ordinal, interval and ratio. Nominal and ordinal ..... have already been described.

4

Selected Topics in Probability

4.1 Probability

[1] Probability theory arose originally in connection with games of chance and then for a long time it was used primarily to investigate the credibility of testimony of witnesses in the "ethical" sciences. Nevertheless, probability has become a very powerful mathematical tool in understanding those aspects of the world that cannot be described by deterministic laws. Probability has succeeded in finding strict determinate relationships where chance seemed to reign and so terming them "laws of chance" combining such contrasting notions in the nomenclature appears to be quite justified. This introductory chapter discusses such notions as determinism, chaos and randomness, predictability and unpredictability, some initial approaches to formalizing randomness and it surveys certain problems that can be solved by probability theory. This will perhaps give one an idea to what extent the theory can answer questions arising in specific random occurrences and the character of the answers provided by the theory.

Games of chance and the analysis of testimony of witnesses were originally the basic areas of application of probability theory. Games of chance involving cards, dice and flipping coins naturally permitted the creation of appropriate random experiments (this terminology first appeared in the twentieth century) so that their

[1] This section has be quoted from:
Skorokhod, Valeriy. Basic principles and applications of probability theory. Springer Science & Business Media, 2005.

outcomes had symmetry in relation to the conditions of the experiment. These outcomes were treated as 'equally likely" and they were assigned the same probabilities. Thus, if there are s outcomes in the experiment, each elementary event was assigned a probability of $1/s$ (it is easy to see that an elementary event has that probability using the additivity of probability and the fact that the sure event has probability one). If an event is expressed as the union of r elementary events ($r \leq s$), then the probability of A is r/s by virtue of the additivity. Thus, we arrive at the definition of probability that has been in use for about two centuries.

The probability of an event A is the quotient of the number of outcomes favorable to A and the number of all possible outcomes. The outcomes favorable to A are understood to be those that imply A.

4.2 Problems of Probability Theory

Initially, probability theory was the study of ways of computing probabilities of events knowing the probabilities of other given events. The techniques developed for computing the probabilities of certain classes of events now form a constituent unit of probability but only partly and far from the main part. However, as before, probability theory only deals with the probabilities of events independently of what meaningful sense can be invested in the words "the probability of event A is p". This means that probability theory itself does interpret its results meaningfully but in so doing it does not exclude the term "probability". There is no statement like "A always occurs" but rather the statement "A occurs with probability one".

4.2.1 Independence

Independence is one of the basic concepts of probability theory. According to Kolmogorov, it is exactly this that distinguishes probability theory from measure theory. Independence will be discussed more precisely later on. For the moment, we merely point out that stochastic independence and physical independence of events (one event having no effect on another) are identical in content.

Stochastic independence is a precisely-defined mathematical concept to be given below. At this point, we note that independence was already used in latent form in the definition of random experiment. One of the requirements imposed on an experiment is the possibility of iterating it indefinitely. To iterate it assumes that the conditions of the experiment can be reconstructed after which the one just performed and all of the prior ones have no affect on the outcome of the next experiment. This means that the events occurring in different experiments must be independent.

4.2.2 Samples

A sample may be defined in general as follows. There are m finite sets $A_1, A_2, \ldots, A_m$. From each set, we choose an element $a_i \in A_i$ one by one. The collection $(a_1, \ldots, a_m)$ is then the sample. Samples are distinguished by identification rules (let us say, we are not interested in the order of the elements in a sample). Each sample is regarded as an elementary event and the elementary events are considered to be equally likely.

1. Sampling with replacement. In this instance, the A_i coincide: $A_i = A$ and the number of samples is nm, where n is the number of elements in A.
2. Sampling without replacement. A sample is constructed as follows. $A_1 = A$, $A_2 = A\backslash\{a_1\}, \ldots, A_k = A\backslash\{a_1, \ldots, a_{k-1}\}$. In other words, only samples $(a_1, \ldots, a_m)$, $a_i \in A$, are considered in which all of the elements are distinct. If A has n elements, then the number of samples without replacement is $n(n-1)\cdots(n-m+1)/m! = \binom{n}{m}$.

4.2.3 Conditional Probability

The conditional probability of an event A given event B having positive probability has occurred is the quantity

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{4.1}$$

As a function of A, $P(A|B)$ possesses all of the properties of a probability. The meaning of conditional probability may be explained as follows. Together with the original experiment, consider

a conditional probability experiment which is performed if event B has happened in the original experiment. Thus, if the original experiment has been done n times and B has happened n_B times, then this sequence contains n_B conditional experiments. The event A will have occurred in the conditional experiment if A and B occur simultaneously, i.e., if $A \cap B$ occurs. If $n_{A\cap B}$ is the number of experiments in which the event $A \cap B$ is observed (of the n carried out), then the relative frequency of occurrence in the n_B conditional experiments is $n_{A\cap B}/n_B$. If we replace the relative frequencies by the probabilities, then we have the right-hand side of (4.1).

4.3 The Poisson Distribution

The Binomial Distribution problems all had relatively small values for n, so evaluating $p_x(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ was not particularly difficult. But suppose n were 1000 and k, 500. Evaluating $p_x(500)$ would be a formidable task for many handheld calculators, even today. Two hundred years ago, the prospect of doing cumbersome binomial calculations by hand was a catalyst for mathematicians to develop some easy-to-use approximations. One of the first such approximations was the Poisson limit, which eventually gave rise to the Poisson distribution.

Simeon Denis Poisson (1781-1840) was an eminent French mathematician and physicist, an academic administrator of some note, and, according to an 1826 letter from the mathematician Abel to a friend, Poisson was a man who knew "how to behave with a great deal of dignity." One of Poisson's many interests was the application of probability to the law, and in 1837 he wrote *Recherches sur la Probabilite de Jugements*. Included in the latter is a limit for $p_x(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ that holds when n approaches ∞, p approaches 0, and np remains constant. In practice, Poisson's limit is used to approximate hard-to-calculate binomial probabilities where the values of n and p reflect the conditions of the limit-that is, when n is large and p is small.

4.3.1 The Poisson Limit

[2] Deriving an asymptotic expression for the binomial probability model is a straightforward exercise in calculus, given that np is to remain fixed as n increases.

Suppose X is a binomial random variable, where

$$p_x(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

If $n \to \infty$ and $p \to 0$ in such a way that $A = np$ remains constant, then

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=const}} P(X = k) = \lim_{\substack{n\to\infty \\ p\to 0 \\ np=const}} \binom{n}{k} p^k (1-p)^{n-k} = \frac{e^{-np}(np)^k}{k!}$$

4.4 Exercises

1. Translate the following text.

A probability distribution is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment. In more technical terms, the probability distribution is a description of a random phenomenon in terms of the probabilities of events. For instance, if the random variable X is used to denote the outcome of a coin toss ("the experiment"), then the probability distribution of X would take the value 0.5 for $X = heads$, and 0.5 for $X = tails$ (assuming the coin is fair). Examples of random phenomena can include the results of an experiment or survey.

A probability distribution is specified in terms of an underlying sample space, which is the set of all possible outcomes of the random phenomenon being observed. The sample space may be

[2] This part has been quoted from:
An Introduction to Mathematical Statistics and Its Applications; Fourth Edition ; Richard J. Larsen, Vanderbilt University and Morris L. Marx University of West Florida ; New Jersey .(2006) pearson Education' Inc.

the set of real numbers or a set of vectors, or it may be a list of non-numerical values; for example, the sample space of a coin flip would be $\{heads,\ tails\}$.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
expression			
			relatively
	coincide		
		constructive	
	explain		
frequency			
			exactly
		discussed	
condition			

3. Use the correct form of the word.

Probability distributions are generally (devision) ..... into two classes. A discrete probability (distribute) .....(applicable to the scenarios where the set of possible outcomes is discrete, such as a coin toss or a roll of dice) can be encoded by a discrete list of the probabilities of the outcomes, known as a probability mass function. On the other hand, a continuous probability distribution (applicable to the scenarios where the set of possible outcomes can take on values in a continuous range (e.g. real numbers), such as the temperature on a given day) is typically (description) ..... by probability density functions (with the probability of any individual outcome actually being 0). The normal distribution is a commonly encountered continuous probability distribution. More complex experiments, such as those (involve) ..... stochastic processes defined in continuous time, may demand the use of more general probability measures.

A probability distribution whose sample space is one-(dimension) ..... (for example real numbers, list of labels, ordered labels or binary) is called univariate, while a distribution whose sample space is a vector space of dimension 2 or more is called multivariate. A univariate distribution gives the probabilities of a single random variable taking on various alternative values; a multivariate distribution (a joint probability distribution) gives the probabilities of a random vector a list of two or more random variables taking on various (combine) ..... of values. Important and commonly encountered univariate probability distributions (inclusion) ..... the binomial distribution, the hypergeometric distribution, and the normal distribution. The multivariate normal distribution is a commonly encountered multivariate distribution.

4. Fill gaps with one of the following words

1. combination	2. sciences	3. explicit	4. curve
5. theorem	6. distributions	7. random	8. quantities
9. function	10. symmetric	11. properties	12. suitable

Normal distributions are important in statistics and are often used in the natural and social ..... to represent real-valued random variables whose distributions are not known. Their importance is partly due to the central limit ..... It states that, under some conditions, the average of many samples (observations) of a random variable with finite mean and variance is itself a ..... variable whose distribution converges to a normal distribution as the number of samples increases. Therefore, physical ..... that are expected to be the sum of many independent processes (such as measurement errors) often have distributions that are nearly normal.

Moreover, Gaussian distributions have some unique ..... that are valuable in analytic studies. For instance, any linear ..... of a fixed collection of normal deviates is a normal deviate. Many results and methods (such as propagation of uncertainty and least squares parameter fitting) can be derived analytically in ..... form when the relevant variables are normally distributed.

A normal distribution is sometimes informally called a bell ..... However, many other distributions are bell-shaped (such as the Cauchy, Student's t, and logistic .....).

The normal distribution is a subclass of the elliptical distributions. The normal distribution is ..... about its mean, and is non-zero over the entire real line. As such it may not be a ..... model for variables that are inherently positive or strongly skewed, such as the weight of a person or the price of a share. Such variables may be better described by other distributions, such as the log-normal distribution or the Pareto distribution.

5

Selected Topics in Algebra

5.1 Groups

[1] Let S be a set. A mapping

$$S \times S \to S$$

is sometimes called a law of composition (of S into itself). If x, y are elements of S, the image of the pair (x, y) under this mapping is also called their product under the law of composition, and will be denoted by xy. (Sometimes, we also write $x \cdot y$, and in many cases it is also convenient to use an additive notation, and thus to write $x + y$. In that case, we call this element the sum of x and y. It is customary to use the notation $x + y$ only when the relation $x + y = y + x$ holds.)

Let S be a set with a law of composition. If x, y, z are elements of S, then we may form their product in two ways: $(xy)z$ and $x(yz)$. If $(xy)z = x(yz)$ for all x, y, z in S then we say that the law of composition is associative.

An element e of S such that $ex = x = xe$ for all $x \in S$ is called a unit element. When the law of composition is written additively, the unit element is denoted by 0, and is called a zero element. A unit element is unique, for if e' is another unit element, we have $e = ee' = e'$ by assumption. In most cases, the unit element is written simply 1 (instead of e).

---

[1] This chapter has been quoted from:
Serge Lang, Algebra, 2002 Springer-Verlag New York, Inc.

A monoid is a set G, with a law of composition which is associative, and having a unit element (so that in particular, G is not empty).

It would be possible to define more general laws of composition, i.e. maps $S_1 \times S_2 \to S_3$ using arbitrary sets. One can then express associativity and commutativity in any setting for which they make sense. For instance, for commutativity we need a law of composition $f : S \times S \to T$ where the two sets of departure are the same. Commutativity then means $f(x, y) = f(y, x)$, or $xy = yx$ if we omit the mapping f from the notation. For associativity, we leave it to the reader to formulate the most general combination of sets under which it will work. We shall meet special cases later, for instance arising from maps

$$S \times S \to S \quad \text{and} \quad S \times T \to T$$

Then a product $(xy)z$ makes sense with $x \in S$, $y \in S$, and $z \in T$. The product $x(yz)$ also makes sense for such elements x, y, z and thus it makes sense to say that our law of composition is associative, namely to say that for all x, y, z as above we have $(xy)z = x(yz)$. If the law of composition of G is commutative, we also say that G is commutative (or Abelian).

By a submonoid of G, we shall mean a subset H of G containing the unit element e, and such that, if $x, y \in H$ then $xy \in H$ (we say that H is closed under the law of composition). It is then clear that H is itself a monoid, under the law of composition induced by that of G.

A group G is a monoid, such that for every element $x \in G$ there exists an element $y \in G$ such that $xy = yx = e$. Such an element y is called an inverse for x. Such an inverse is unique, because if y' is also an inverse for x, then $y' = y'e = y'(xy) = (y'x)y = ey = y$. We denote this inverse by x^{-1} (or by $-x$ when the law of composition is written additively).

Let G be a group. A subgroup H of G is a subset of G containing the unit element, and such that H is closed under the law of composition and inverse (i.e. it is a submonoid, such that if $x \in H$ then $x^{-1} \in H$). A subgroup is called trivial if it consists of the unit element alone. The intersection of an arbitrary non-empty family of subgroups is a subgroup (trivial verification).

Let G, G' be monoids. A monoid-homomorphism (or simply homomorphism) of G into G' is a mapping $f : G \to G'$ such that $f(xy) = f(x)f(y)$ for all $x, y \in G$, and mapping the unit element of G into that of G'. If G, G' are groups, a group-homomorphism of G into G' is simply a monoid-homomorphism.

A homomorphism $f : G \to G'$ is called an isomorphism if there exists a homomorphism $g : G' \to G$ such that $f \circ g$ and $g \circ f$ are the identity mappings (in G' and G respectively). It is trivially verified that f is an isomorphism if and only if f is bijective. The existence of an isomorphism between two groups G and G' is sometimes denoted by $G \approx G'$. If $G = G'$, we say that isomorphism is an automorphism. A homomorphism of G into itself is also called an endomorphism.

5.2 Rings and Homomorphism

A ring A is a set, together with two laws of composition called multiplication and addition respectively, and written as a product and as a sum respectively, satisfying the following conditions:

RI 1 With respect to addition, A is a commutative group.
RI 2 The multiplication is associative, and has a unit element.
RI 3 For all $x, y, Z \in A$ we have $(x+y)z = xz+yz$ (This is called distributivity.)

As usual, we denote the unit element for addition by 0, and the unit element for multiplication by 1. We do not assume that $1 \neq 0$. We observe that $0x = 0$ for all $x \in A$. Because, we have $0x + x = (0 + 1)x = 1x = x$. Hence $0x = 0$. In particular, if $1 = 0$, then A consists of 0 alone. For any $x, y \in A$ we have $(-x)y = -(xy)$. Because, we have $xy + (-x)y = (x + (-x))y = 0y = 0$, so $(-x)y$ is the additive inverse of xy.

Other standard laws relating addition and multiplication are easily proved, for instance $(-x)(-y) = xy$.

Let A be a ring, and let U be the set of elements of A which have both a right and left inverse. Then U is a multiplicative group. Indeed, if a has a right inverse b, so that $ab = 1$, and a left inverse c, so that $ca = 1$, then $cab = b$, whence $c = b$, and we see that c (or b) is a two-sided inverse, and that c itself has a two-sided inverse,

namely a. Therefore U satisfies all the axioms of a multiplicative group, and is called the group of units of A. It is sometimes denoted by A^*, and is also called the group of invertible elements of A. A ring A such that $1 \neq 0$, and such that every non-zero element is invertible is called a division ring.

A left ideal $\mathfrak{a}$ in a ring A is a subset of A which is a subgroup of the additive group of A, such that $A\mathfrak{a} \subset \mathfrak{a}$ (and hence $A\mathfrak{a} = \mathfrak{a}$ since A contains 1). To define a right ideal, we require $A\mathfrak{a} = \mathfrak{a}$, and a two-sided ideal is a subset which is both a left and a right ideal. A two-sided ideal is called simply an ideal. Note that (0) and A itself are ideals.

By a ring-homomorphism one means a mapping $f : A \to B$ where A, B are rings, and such that f is a monoid-homomorphism for the multiplicative structures on A and B, and also a monoid-homomorphism for the additive structure. In other words, f must satisfy:

$$f(a + a') = f(a) + f(a'), f(1) = 1, f(aa') = f(a)f(a'), f(0) = 0,$$

for all $a, a' \in A$. Its kernel is defined to be the kernel of f viewed as additive homomorphism.

A ring A is said to be commutative if $xy = yx$ for all $x, y \in A$. A commutative division ring is called a field. We observe that by definition, a field contains at least two elements, namely 0 and 1.

5.3 Module

Let A be a ring. A left module over A, or a left A-module M is an Abelian group, usually written additively, together with an operation of A on M (viewing A as a multiplicative monoid by **RI 2**), such that, for all $a, b \in A$ and $x, y \in M$ we have

$$(a + b)x = ax + bx \quad \text{and} \quad a(x + y) = ax + ay.$$

We leave it as an exercise to prove that $a(-x) = -(ax)$ and that $0x = 0$. By definition of an operation, we have $1x = x$.

In a similar way, one defines a right A-module. We shall deal only with left A-modules, unless otherwise specified, and hence call these simply A-modules, or even modules if the reference is clear.

Let M be an A-module. By a submodule N of M we mean an additive subgroup such that $AN \subset N$. Then N is a module (with the operation induced by that of A on M). A module over a field is called a vector space.

By a module-homomorphism one means a map $f : M \to M'$ of one module into another (over the same ring A), which is an additive group-homomorphism, and such that $f(ax) = af(x)$ for all $a \in A$ and $x \in M$. If we wish to refer to the ring A, we also say that f is an A-homomorphism, or also that it is an A-linear map.

There are some things in mathematics which satisfy all the axioms of a ring except for the existence of a unit element. Let A be a commutative ring. Let E, F be modules. By a bilinear map $g : E \times E \to F$ we mean a map such that given $x \in E$, the map $y \mapsto g(x, y)$ is A-linear, and given $y \in E$, the map $x \mapsto g(x, y)$ is A-linear. By an A-algebra we mean a module together with a bilinear map $g : E \times E \to E$. We view such a map as a law of composition on E. Aside from the examples already mentioned, we note that the group ring $A[G]$ (or monoid ring when G is a monoid) is an A-algebra, also called the group (or monoid) algebra.

5.4 Exercises

1. Translate the following sentences.

An integer n is said to be a divisor of an integer i if i is an integer multiple of n; i.e., $i = qn$ for some integer q. Thus all integers are trivially divisors of 0. The integers that have integer inverses, namely ± 1, are called the units of Z. If u is a unit and n is a divisor of i, then un is a divisor of i and n is a divisor of ui. Thus the factorization of an integer can only be unique up to a unit u, and ui has the same divisors as i. We therefore consider only factorizations of positive integers into products of positive integers. Every nonzero integer i is divisible by 1 and i; these divisors are called trivial. An integer n is said to be a factor of an integer i if n is positive and a nontrivial divisor of i. For example, 1 has no nontrivial divisors and thus no factors. A positive integer greater than 1 that has no nontrivial divisors is called a prime integer.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
multiplication			
			uniquely
	satisfy		
		invertible	
	commutate		
specification			
		divisibility	
		additive	
distribution			

3. Use the correct form of the word.

1. A ring is an Abelian group with a second binary (operate) ..... that is associative, is distributive over the Abelian group operation, and has an identity element.
2. Although ring addition is (commute) ....., ring multiplication is not required to be commutative: ab need not necessarily equal ba.
3. The study of rings (origin) ..... from the theory of polynomial rings and the theory of algebraic integers.
4. The (define) ..... of an ideal in a ring is analogous to that of normal subgroup in a group.
5. The concept of a module over a ring (general) ..... the concept of a vector space (over a field) by generalizing from multiplication of vectors with elements of a field (scalar multiplication) to multiplication with elements of a ring.
6. A field is a set on which addition, subtraction, multiplication, and division are (definition) ..... and behaved as the corresponding operations on rational and real numbers do.
7. Finite fields (also called Galois fields) are fields with (finite)..... many elements, whose number is also referred to as the order of the field.

4. Fill gaps with one of the following words

1. divides	2. rings	3. multiple	4. finite
5. degree	6. field	7. factors	8. algebraic
9. remainders	10. element	11. group	12. algebra

1. A polynomial $g(x)$ is said to be a divisor of an polynomial $f(x)$ if $f(x)$ is a polynomial ..... of $g(x)$; i.e., $f(x) = q(x)g(x)$ for some polynomial $q(x)$.
2. An important example of a finite abelian group is the set of ..... $R_n = \{0, 1, \ldots, n-1\}$ under mod-n addition, where n is any given positive integer.
3. A finite ..... G of order n is called cyclic if it is isomorphic to $\mathbb{Z}_n$.
4. By degree additivity, the ..... of a polynomial $f(x)$ is equal to the sum of the degrees of its prime factors, which are unique by unique factorization.
5. Given a positive integer i, we may factor i into a unique product of prime ..... by simply factoring out primes no greater than i until we arrive at the quotient 1.
6. If S is a subgroup of a finite group G, then $|S|$ ..... $|G|$.
7. What is the order of the smallest non-trivial ring with identity which is not a .....?
8. A finite ring is a ring that has a ..... number of elements.
9. A monoid is an algebraic structure intermediate between groups and semigroups, and is a semigroup having an identity .....
10. An interesting topic in the ring theory is the classification of finite .....
11. An associative algebra is an ..... structure with compatible operations of addition, multiplication (assumed to be associative), and a scalar multiplication by elements in some field.
12. Boolean ..... is the branch of algebra in which the values of the variables are the truth values true and false, usually denoted 1 and 0, respectively.

6

Selected Topics in Numerical Analysis

6.1 rounding Process

[1] The conventional process of rounding or "forcing" a number to n digits (or figures) consists of replacing that number by an n-digit approximation with minimum error. When this requirement leads to two permissible roundings, that one for which the nth digit of the rounded number is even is generally selected. With this rule, the associated error is never larger in magnitude than one-half unit in the place of the nth digit of the rounded number. Thus $4.05149 \simeq 4.0515, 4.051, 4.05, 4.1,$ and 4. It may be noted here that whereas 4.05149 rounds to 4.0515, which in turn rounds to 4.052; nevertheless, 4.05149 rounds directly to 4.051. Thus rounding is not necessarily transitive.

The errors introduced in the rounding of a large set of numbers, which are to be combined in a certain way, usually (but not always) tend to be equally often positive and negative, so that their effects often tend to cancel. The slight favoring of even numbers is prompted by the fact that any subsequent operations on the rounded numbers are then somewhat less likely to necessitate additional round offs.

Each digit of a number, except a zero which serves only to fix the position of the decimal point, is called a significant digit or figure of that number. Thus, the numbers 2.159, 0.04072, and 10.00 each contain four significant figures. Whether or not the last

[1] This Part has been quoted from:
Hildebrand, Francis Begnaud. Introduction to numerical analysis, Courier Corporation, 2013.

digit of 14620 is significant depends upon the context. If "a number known to be between 14615 and 14625" is intended, then that zero is not significant and the number would preferably be written in the form 1.462×10^4. Otherwise the form 1.4620×10^4 would be appropriate.

It may be seen that the concept of significant figures is related more intimately to the relative error

$$\text{Relative error} = \frac{\text{true value} - \text{approximation}}{\text{true value}}$$

than to the error (or the absolute error) itself.

6.2 Interpolation

Anyone who has had occasion to consult tables of mathematical functions is familiar with the method of linear interpolation and probably has encountered situations in which this method of "reading between the lines of the table" has appeared to be unreliable. If more reliable interpolates are desired, it is clearly necessary to make use of more information than that consisting of tabulated values (ordinates) of a function, corresponding to only two successive abscissas? Whereas that additional information could consist, for example, of known values of certain derivatives of the function at those two points, it is supposed in most of what follows that the interpolation process is to be based only on tabulated values of the function itself, with any further available information reserved for use in estimating the error involved.

There exist a number of interpolation formulas which have this property, most of which possess certain advantages in certain situations, but no one of which is preferable to all others in all respects. Whereas certain of these formulas are expressed explicitly in terms of all the ordinates on which they depend, others involve only one or two of the ordinates explicitly and express their dependence upon other ordinates only in terms of differences of ordinates and successive differences of differences.

6.3 Solving Non-Linear Equations

[2]

We are here concerned with the problem of solving equations of the form $x = f(x)$, where f is a given function. The method of iteration consists in executing the following algorithm:

Algorithm: Choose x_0 arbitrarily, and generate the sequence $\{x_n\}$ recursively from the relation

$$x_n = f(x_{n-1}), n = 1, 2, \cdots . \tag{6.1}$$

At the outset, we cannot even be sure that this algorithm is well defined. It could be that f is undefined at some point $f(x_n)$. However, let us assume that f is defined on some closed finite interval $I = [a, b]$, and that the values of f lie in the same interval. Geometrically this means that the graph of the function $y = f(x)$ is contained in the square $a \leq x \leq b$, $a \leq y \leq b$.

Under this assumption, if $x_0 \in I$, we can say that all elements of the sequence $\{x_n\}$ are in I. For if some $x_n \in I$ with $n > 0$, then also $x_{n+1} = f(x_n) \in I$ since f has its values in I. Observe that the above hypotheses are not sufficient to guarantee that equation (6.1) has a solution. However, if we assume the function f to be continuous, then the equation has at least one solution.

The above intuitive consideration can be couched in purely analytical terms, as follows. Consider the function g defined by $g(x) = x - f(x)$, $a \leq x \leq b$. This function is continuous on the interval $[a, b]$; moreover, since $f(a) > a$, $f(b) < b$, it satisfies $g(a) \leq 0$, $g(b) \geq 0$. By the intermediate value theorem of calculus it assumes all values between $g(a)$ and $g(b)$ somewhere in the interval $[a, b]$. Therefore it must assume the value zero, say at $x = s$. This implies $0 = s - f(s)$, or $s = f(s)$. Thus the number s is the desired solution.

6.4 Numerical Differentiation

[3] The mathematical definition of the derivative of $f(x)$ is

[2] This Section has been quoted from:
Henrici, Peter. Elements of numerical analysis. (1964).

[3] This section has been quoted from:
Hjorth-Jensen, Morten. Computational physics, Lecture notes (2011).

$$\frac{df(x)}{x} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

where h is the step size. If we use a Taylor expansion for $f(x)$, we can write

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \cdots$$

We can then set the computed derivative $f'_c(x)$ as

$$f'_c(x) \simeq \frac{f(x+h) - f(x)}{h} \simeq f'(x) + \frac{hf''(x)}{2} + \cdots$$

Assume now that we will employ two points to represent the function f by way of a straight line between x and $x+h$. This means that we could represent

$$f'_2(x) = \frac{f(x+h) - f(x)}{h} + \mathcal{O}(h).$$

where the suffix 2 refers to the fact that we are using two points to define the derivative and the dominating error goes like $\mathcal{O}(h)$. This is the forward derivative formula. Alternatively, we could use the backward derivative formula

$$f'_2(x) = \frac{f(x) - f(x-h)}{h} + \mathcal{O}(h).$$

If the second derivative is close to zero, this simple two point formula can be used to approximate the derivative.

Above we have used the interpolating polynomial to approximate values of a function f at points where f is not known. Another use of the interpolating polynomial, of equal or even higher importance in practice, is the imitation of the fundamental operations of calculus. In all these applications the basic idea is extremely simple: Instead of performing the operation on the function f, which may be difficult or-in cases where f is known at discrete points only-impossible, the operation is performed on a suitable interpolating polynomial.

6.5 Numerical Integration

[4] We now turn to the problem of numerical evaluation of definite integrals. The method is the same as numerical differentiation. Instead of performing the integration on the function f, which may be difficult, we perform the integration on a polynomial interpolating f at suitable points. If values of f are available on both sides of the interval of integration, it seems preferable to perform the integration on a polynomial that takes into account all these values.

It is assumed that f is continuous on $[a, b]$ and can be evaluated at arbitrary points of that interval. Several procedures offer themselves; we shall be able to dismiss two of them very briefly.

1. **Newton-Cotes formulas**. The most natural idea that offers itself seems to select a certain number of interpolating points within $[a, b]$, to interpolate f at these points, and to approximate the integral off by the integral of the interpolating polynomial. If the interpolating points divide $[a, b]$ into N equal parts, we arrive in this manner at certain integration formulas which are called the Newton-Cotes formulas. Unfortunately, these formulas have, for large values of N, some very undesirable properties. In particular, it turns out that there exist functions, even analytic ones, for which the sequence of the integrals of the interpolating polynomials does not converge towards the integral of the function f. Also, the coefficients in these formulas are large and alternate in sign, which is undesirable for the propagation of rounding error. For these reasons, the Newton-Cotes formulas are rarely used for high values of N. For $N = 2, 3, 4$ the formulas are identical with certain well-known integration formulas.
2. **Gaussian quadrature**. One may try to avoid some of the short-comings of the Newton-Cotes formulas by relinquishing the equal spacing of the interpolating points. Gauss discovered that by a proper choice of the interpolating points one can construct integration formulas which, using $N+1$ interpolating points, give the accurate value of the integral if f is a

[4] This Section has been quoted from:
Henrici, Peter. Elements of numerical analysis. (1964).

polynomial of degree $2N+1$ or less. These formulas turn out to be numerically stable, and they are in successful use at a number of computation laboratories. The formulas suffer from the disadvantage, however, that the interpolating points as well as the corresponding weights are irregular numbers that have to be stored. This practically (although not theoretically) limits the applicability of these highly interesting formulas.

6.6 Exercises

1. Translate the following sentences.

Numerical integration is the approximate computation of an integral using numerical techniques. The numerical computation of an integral is sometimes called quadrature. Ueberhuber uses the word "quadrature" to mean numerical computation of a univariate integral, and "cubature" to mean numerical computation of a multiple integral.

There are a wide range of methods available for numerical integration. The most straightforward numerical integration technique uses the Newton-Cotes formulas (also called quadrature formulas), which approximate a function tabulated at a sequence of regularly spaced intervals by various degree polynomials. If the endpoints are tabulated, then the 2- and 3-point formulas are called the trapezoidal rule and Simpson's rule, respectively. The 5-point formula is called Boole's rule. A generalization of the trapezoidal rule is Romberg integration, which can yield accurate results for many fewer function evaluations.

If the functions are known analytically instead of being tabulated at equally spaced intervals, the best numerical method of integration is called Gaussian quadrature. By picking the abscissas at which to evaluate the function, Gaussian quadrature produces the most accurate approximations possible. However, given the speed of modern computers, the additional complication of the Gaussian quadrature formalism often makes it less desirable than simply brute-force calculating twice as many points on a regular grid (which also permits the already computed values of the function to be re-used).

Modern numerical integrations methods based on information theory have been developed to simulate information systems such as computer controlled systems, communication systems, and control systems since in these cases, the classical methods (which are based on approximation theory) are not as efficient.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
algorithm			
			equally
	interpolate		
		regular	
	execute		
	enumerate		
			intuitively
instability			
	classify		

3. Use the correct form of the word.

1. spline interpolation is a form of (interpolate) ..... where the interpolant is a special type of piecewise polynomial called a spline.
2. Cubic Hermite splines are typically used for interpolation of numeric data specified at given argument values $x_1, x_2, \ldots, x_n$, to obtain a smooth (continue) ..... function.
3. The finite element method is a (numeric)..... technique for finding approximate solutions of partial differential equations as well as of integral equations.
4. Finite-difference methods are numerical methods for (approximate) ..... the solutions to differential equations using finite difference equations to approximate derivatives.

5. Spectral methods are techniques used in applied mathematics and (science) ..... computing to numerically solve certain differential equations, often involving the use of the Fast Fourier Transform.

4. Fill gaps with one of the following words

1. computation	2. approximation	3. secant	4. procedure
5. difference	6. solutions	7. functions	8. successive
9. errors	10. digits	11. finding	12. precision

1. When a number is expressed in scientific notation, the number of significant ..... (or significant figures) is the number of digits needed to express the number to within the uncertainty of calculation.
2. The Bisection Method is a ..... approximation method that narrows down an interval that contains a root of the function $f(x)$.
3. The relative error of the quotient or product of a number of quantities is less than or equal to the sum of their relative .....
4. The Newton-Raphson method (also known as Newton's method) is a way to quickly find a good .....for the root of a real-valued function $f(x) = 0f(x) = 0$.
5. The bisection method is a root-finding method that applies to any continuous ..... for which one knows two values with opposite signs.
6. The concepts of accuracy and ..... are both closely related and often confused.
7. The golden section search algorithm can be used for ..... a minimum (or maximum) of a single-variable function $f(x)$.
8. The Euler method is a first-order numerical ..... for solving ordinary differential equations with a given initial value. I
9. If rounding is performed on each of a series of numbers in a long ....., roundoff error can become important, especially if division by a small number ever occurs.

10. The secant method is a root-finding algorithm that uses a succession of roots of ..... lines to better approximate a root of a function f.
11. There are many different methods that can be used to approximate ..... to a differential equation.
12. Absolute error is the ..... between the measured or inferred value of a quantity and its actual value.

7

Selected Topics in Differential Equation

7.1 Differential Equation in a Nutshell

[1] In the calculus, you studied various methods by which you could differentiate the elementary functions. Equations which involve variables and their derivatives are called differential equations.

Let $f(x)$ define a function of x on an interval $I : a \leq x \leq b$. By an ordinary differential equation we mean an equation involving x, the function $f(x)$ and one or more of its derivatives. The order of a differential equation is the order of the highest derivative involved in the equation.

Let $y = f(x)$ define y as a function of x on an interval $I : a \leq x \leq b$. We say that the function $f(x)$ is an explicit solution or simply a solution of an ordinary differential equation involving x, $f(x)$, and its derivatives, if it satisfies the equation for every x in I, i. e., if we replace y by $f(x)$, y' by $f'(x)$, y'' by $f''(x), \cdots, y^{(n)}$ by $f^{(n)}(x)$, the differential equation reduces to an identity in x.

A relation $f(x, y) = 0$ will be called an implicit solution of the differential equation

$$F(x, y, y', \ldots, y^{(n)}) = 0. \tag{7.1}$$

on an interval $I : a \leq x \leq b$, if

[1] This section has been quoted from:
Morris. Tenenbaum, and Harry Pollard. Ordinary differential equations: an elementary textbook for students of mathematics, engineering, and the sciences. Dover Publications, 1963.

1. It defines y as an implicit function of x on I, i.e., if there exists a function $g(x)$ defined on I such that $f[x, g(x)] = 0$ for every x in I, and if
2. $g(x)$ satisfies (7.1), i.e., if

$$F[x, g(x), g'(x), \ldots, g^{(n)}(x)] = 0$$

for every x in I.

7.2 The General Solution of a Differential Equation

We assume at the outset that you have understood clearly the material of the previous lesson so that when we say "solve a differential equation" or "find a solution of a differential equation," or "the solution of a differential equation is," you will know what is meant. Or if we omit intervals for which a function or a differential equation is defined, we expect that you will be able to fill in this omission yourself.

Two conclusions seem to stem from these examples. First, if a differential equation has a solution, it has infinitely many solutions. Second, if the differential equation is of the first order, its solution contains one arbitrary constant; if of the second order, its solution contains two arbitrary constants; if of the nth order, its solution contains n arbitrary constants.

It is customary to call a solution which contains n constants $c_1, c_2, \ldots, c_n$ an n-parameter family of solutions, and to refer to the constants c_1 to c_n as parameters.

We shall now show you how to find the differential equation when its n-parameter family of solutions is known. You must bear in mind that although the family will contain the requisite number of n arbitrary constants, the nth order differential equation whose solution it is, contains no such constants. In solving problems of this type, therefore, these constants must be eliminated. Unfortunately a standard method of eliminating these constants is not always the easiest to use. There are frequently simpler methods which cannot be standardized and which will depend on your own ingenuity.

7.3 Isogonal and orthogonal Trajectories

When two curves intersect in a plane, the angle between them is defined to be the angle made by their respective tangents drawn at their point of intersection. Since these lines determine two angles, it is customary to specify the particular one desired by stating from which tangent line we are to proceed in a counterclockwise direction to reach the other.

A curve which cuts every member of a given 1-parameter family of curves in the same angle is called an isogonal trajectory of the family.

If two 1-parameter families have the property that every member of one family cuts every member of the other family in the same angle, then each family may be said to be a 1-parameter family of isogonal trajectories of the other, i. e., the curves of either family are isogonal trajectories of the other.

An interesting problem is to find a family of isogonal trajectories that makes a predetermined angle with a given 1-parameter family of curves. If we call y_1' the slope of a curve of a given 1-parameter family, y' the slope of an isogonal trajectory of the family, and α their angle of intersection measured from the tangent line with slope y' to the tangent line with slope y_1', then

$$\tan\alpha = \frac{y_l' - y'}{1 + y'y_1'}$$

A curve which cuts every member of a given 1-parameter family of curves in a 90° angle is called an orthogonal trajectory of the family. If two 1-parameter families have the property that every member of one family cuts every member of the other family in a right angle, then each family may be said to be an orthogonal trajectory of the other. Orthogonal trajectory problems are of special interest since they occur in many physical fields.

7.4 Initial Value Problem

[2] In the field of differential equations, an initial value problem is an ordinary differential equation together with specified value, called

[2] This Section has been quoted from:
http://en.wikipedia.org/wiki/Initial_value_problem

the initial condition, of the unknown function at a given point in the domain of the solution. In physics or other sciences, modeling a system frequently amounts to solving an initial value problem; in this context, the differential equation is an evolution equation specifying how, given initial conditions, the system will evolve with time.

An initial value problem is a differential equation

$$y'(t) = f(t, y(t)) \quad \text{with} \quad f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

together with a point in the domain of f

$$(t_0, y_0) \in \mathbb{R} \times \mathbb{R},$$

called the initial condition.

A solution to an initial value problem is a function y that is a solution to the differential equation and satisfies

$$y(t_0) = y_0.$$

This statement subsumes problems of higher order, by interpreting y as a vector. For derivatives of second or higher order, new variables (elements of the vector y) are introduced. More generally, the unknown function y can take values on infinite dimensional spaces, such as Banach spaces or spaces of distributions. For a large class of initial value problems, the existence and uniqueness of a solution can be demonstrated.

The Picard-Lindelöf theorem guarantees a unique solution on some interval containing t_0 if f and its partial derivative $\frac{\partial f}{\partial y}$ are continuous on a region containing t_0 and y_0. The proof of this theorem proceeds by reformulating the problem as an equivalent integral equation. The integral can be considered an operator which maps one function into another, such that the solution is a fixed point of the operator. The Banach fixed-point theorem is then invoked to show that there exists a unique fixed point, which is the solution of the initial value problem.

An older proof of the Picard-Lindelöf theorem constructs a sequence of functions which converge to the solution of the integral equation, and thus, the solution of the initial value problem. Such a construction is sometimes called “Picard’s method” or “the

method of successive approximations". This version is essentially a special case of the Banach fixed point theorem.

Hiroshi Okamura obtained a necessary and sufficient condition for the solution of an initial value problem to be unique. This condition has to do with the existence of a Lyapunov function for the system.

In some situations, the function f is not of class C^1, or even Lipschitz, so the usual result guaranteeing the local existence of a unique solution does not apply. The Peano existence theorem however proves that even for f merely continuous, solutions are guaranteed to exist locally in time; the problem is that there is no guarantee of uniqueness. An even more general result is the Carathéodory existence theorem, which proves existence for some discontinuous functions f.

7.5 Boundary Value Problem

[3] A boundary value problem is a differential equation together with a set of additional restraints, called the boundary conditions. A solution to a boundary value problem is a solution to the differential equation which also satisfies the boundary conditions.

Boundary value problems arise in several branches of physics as any physical differential equation will have them. Problems involving the wave equation, such as the determination of normal modes, are often stated as boundary value problems. A large class of important boundary value problems are the Sturm-Liouville problems. The analysis of these problems involves the eigenfunctions of a differential operator.

To be useful in applications, a boundary value problem should be well posed. This means that given the input to the problem there exists a unique solution, which depends continuously on the input. Much theoretical work in the field of partial differential equations is devoted to proving that boundary value problems arising from scientific and engineering applications are in fact well-posed.

[3] This section has been quoted from:
http://en.wikipedia.org/wiki/Boundary_value_problem

Among the earliest boundary value problems to be studied is the Dirichlet problem, of finding the harmonic functions (solutions to Laplace's equation); the solution was given by the Dirichlet's principle.

A more mathematical way to picture the difference between an initial value problem and a boundary value problem is that an initial value problem has all of the conditions specified at the same value of the independent variable in the equation (and that value is at the lower boundary of the domain, thus the term "initial" value). On the other hand, a boundary value problem has conditions specified at the extremes of the independent variable. For example, if the independent variable is time over the domain $[0, 1]$, an initial value problem would specify a value of $y(t)$ and $y'(t)$ at time $t = 0$, while a boundary value problem would specify values for $y(t)$ at both $t = 0$ and $t = 1$.

If the problem is dependent on both space and time, then instead of specifying the value of the problem at a given point for all time the data could be given at a given time for all space. For example, the temperature of an iron bar with one end kept at absolute zero and the other end at the freezing point of water would be a boundary value problem.

If the boundary gives a value to the normal derivative of the problem then it is a Neumann boundary condition. For example, if there is a heater at one end of an iron rod, then energy would be added at a constant rate but the actual temperature would not be known. If the boundary gives a value to the problem then it is a Dirichlet boundary condition. For example, if one end of an iron rod is held at absolute zero, then the value of the problem would be known at that point in space. If the boundary has the form of a curve or surface that gives a value to the normal derivative and the problem itself then it is a Cauchy boundary condition.

Aside from the boundary condition, boundary value problems are also classified according to the type of differential operator involved. For an elliptic operator, one discusses elliptic boundary value problems. For an hyperbolic operator, one discusses hyperbolic boundary value problems. These categories are further subdivided into linear and various nonlinear types.

7.6 Exercises

1. Translate the following sentences.

A partial differential equation (or briefly a PDE) is a mathematical equation that involves two or more independent variables, an unknown function (dependent on those variables), and partial derivatives of the unknown function with respect to the independent variables. The order of a partial differential equation is the order of the highest derivative involved. A solution (or a particular solution) to a partial differential equation is a function that solves the equation or, in other words, turns it into an identity when substituted into the equation. A solution is called general if it contains all particular solutions of the equation concerned.

The term exact solution is often used for second- and higher-order nonlinear PDEs to denote a particular solution Partial differential equations are used to mathematically formulate, and thus aid the solution of, physical and other problems involving functions of several variables, such as the propagation of heat or sound, fluid flow, elasticity, electrostatics, electrodynamics, etc.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
	initiate		
		general	
	involve		
			explicitly
	omit		
elimination			
		intersecting	
uniqueness			
evolution			

3. Use the correct form of the word.

1. The main idea behind the Laplace Transformation is that we can solve an equation (or system of equations) containing differential and integral terms by (transform) ..... the equation in "t-space" to one in "s-space".
2. The Laplace transform is (invert) ..... on a large class of functions.
3. The meaning of the integral depends on types of functions of interest. A necessary condition for existence of the integral is that f must be locally (integral) .....on $[0, \infty)$.
4. Two integrable functions have the same Laplace transform only if they (different) ..... on a set of Lebesgue measure zero.
5. The Laplace transform has a number of properties that make it useful for (analysis) ..... linear dynamical systems.

4. Fill gaps with one of the following words

1. string	2. gravity	3. horizontal	4. component
5. stretched	6. flexible	7. initial	8. displacement
9. perfectly	10. location	11. tangential	12. measures

Consider a vertical string of length L that has been tightly stretched between two points at $x = 0$ and $x = L$. Because the string has been tightly ..... we can assume that the slope of the displaced string at any point is small. So just what does this do for us? Let's consider a point x on the string in its equilibrium position, i.e. the .....of the point at $t = 0$. As the string vibrates this point will be displaced both vertically and horizontally, however, if we assume that at any point the slope of the string is small then the ..... displacement will be very small in relation to the vertical displacement. This means that we can now assume that at any point x on the string the displacement will be purely vertical. So, let's call this .....$u(x, t)$.

We are going to assume, at least initially, that the string is not uniform and so the mass density of the .....,$\rho(x)$ may be a function of x.

Next, we are going to assume that the string is perfectly .....This means that the string will have no resistance to bending. This in turn tells us that the force exerted by the string at

any point x on the endpoints will be .....to the string itself. This force is called the tension in the string and its magnitude will be given by $T(x,t)$.

Finally, we will let $Q(x,t)$ represent the vertical .....per unit mass of any force acting on the string. Provided we again assume that the slope of the string is small the vertical displacement of the string at any point is then given by,

$$\rho(x)\frac{\partial u^2}{\partial t^2} = \frac{\partial}{\partial x}\left(T(x,t)\frac{\partial u}{\partial x}\right) + \rho(x)Q(x,t)$$

This is a very difficult partial differential equation to solve so we need to make some further simplifications.

First, we're now going to assume that the string is .....elastic. This means that the magnitude of the tension, $T(x,t)$, will only depend upon how much the string stretches near x. Again, recalling that we're assuming that the slope of the string at any point is small this means that the tension in the string will then very nearly be the same as the tension in the string in its equilibrium position. We can then assume that the tension is a constant value, $T(x,t) = T_0$.

Further, in most cases the only external force that will act upon the string is .....and if the string light enough the effects of gravity on the vertical displacement will be small and so will also assume that $Q(x,t) = 0$. This leads to

$$\rho\frac{\partial^2 u}{\partial t^2} = T_0\frac{\partial^2 u}{\partial x^2}$$

If we now divide by the mass density and define,

$$c^2 = \frac{T_0}{\rho}$$

we arrive at the 1-D wave equation,

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

The initial conditions will also be a little different here from what we saw with the heat equation. Here we have a 2nd order time derivative and so we'll also need two initial conditions. At any

point we will specify both the .....displacement of the string as well as the initial velocity of the string. The initial conditions are then,

$$u(x,0) = f(x), \quad \frac{\partial u}{\partial t}(x,0) = g(x)$$

8

Selected Topics in Topology and Geometry

8.1 What is Topology?

[1] The concept of topological space grew out of the study of the real line and Euclidean space and the study of continuous function on these spaces. Here we define what a topological space is, and we study a number of ways of constructing a topology on a set as to make it into a topological space. We also consider some of the elementary concepts associated with topological spaces. Open and closed sets, limit points, and continuous functions are introduced as natural generalizations of the corresponding ideas for the real line and Euclidean space.

The definition of a topological space that is now standard was a long time in being formulated. Various mathematicians- Fréchet, Hausdorff, and others- proposed different definitions over a period of years during the first decades of the twentieth century, but it took quite a while before mathematicians settled on the one that seemed most suitable. They wanted, of course, a definition that was as broad as possible, so that it would include as special cases all the various examples that were useful in mathematics - Euclidean space, infinite-dimensional Euclidean space, and function spaces among them - but they also wanted the definition to be narrow enough that the standard theorems about these familiar spaces would hold for topological spaces in general. This is always the problem when one is trying to formulate a new mathemati-

[1] This section has been quoted from:
Topology (2nd Edition) James Munkresæ2000 by Prentice Hall Inc. N.J.

cal concept, to decide how general its definition should be. The definition finally settled on may seem a bit abstract, but as you work through the various ways of constructing topological spaces, you will get better feeling for what the concept means. A topology on a set X is a collection τ of subsets of X having the following properties:

1. $\varnothing$ and X are in τ.
2. The union of the elements of any subcollection of τ is in τ.
3. The intersection of the elements of any finite subcollection of τ is in τ.

A set X for which a topology τ has been specified is called a topological space.

Properly speaking, a topological space is an ordered pair (X, τ) consisting of a set X and a topology τ in X, but we often omit specific mention of τ if no confusion will arise.

If X is a topological space with topology τ, we say that a subset U of X is an open set of X if U belongs to the collection τ. Using this terminology, one can say that a topological space is a set X together with a collection of subsets of X, called open sets, such that $\varnothing$ and X are both open, and such that arbitrary union and finite intersections of open sets are open.

If X is any set, the collection of all subsets of X is a topology on X; it is called the discrete topology. The collection consisting of X and $\varnothing$ is also a topology on X; we shall call it the indiscrete topology, or the trivial topology.

Suppose that τ and τ' are two topologies on a given set X. If $\tau' \supset \tau$, we say that τ' is finer than τ; if τ' properly contains τ, we say that τ' is strictly finer than τ. We also say that τ is coarser than τ', or strictly coarser, in these two respective situations. We say τ is comparable topology with τ' if either $\tau \supset \tau'$ or $\tau' \supset \tau$.

Other terminology is sometimes used for this concept. If $\tau' \supset \tau$, some mathematicians would say that τ' is larger than τ and τ is smaller than τ'. Many mathematicians use the words "weaker" and "stronger" in this context. Unfortunately, some of them (particulary analysts) are apt to say that τ' is stronger than τ if $\tau' \supset \tau$, while others (particulary topologists) are apt to say that τ' is weaker than τ in the same situation.

If X is a set, a basis for a topology on X is a collection $\mathfrak{B}$ of subsets of X (called basis elements) such that

1. For each $x \in X$, there is at least one basis element B containing x.
2. If x belongs to the intersection of two basis elements B_1 and B_2, then there is a basis element B_3 containing x such that $B_3 \subset B_1 \cap B_2$.

if $\mathfrak{B}$ satisfies these two conditions, then we define the topology τ generated by $\mathfrak{B}$ as follows: A subset U of X is said to be open in X (that is, to be an element of τ) if for each $x \in X$, there is a basis element $B \in \mathfrak{B}$ such that $x \in B$ and $B \subset U$. Note that each basis element is itself an element of τ.

Lt X and Y be topological spaces. A function $f : X \to Y$ is said to be continuous if for each open subset V of Y, the set $f^{-1}(V)$ is an open set of X.

Recall that $f^{-1}(V)$ is the set of all points x of X for which $f(x) \in V$, it is empty if V does not intersect the image set $f(x)$ of f.

Continuity of a function depends not only upon the function f, but also on the topologies for its domain and range. if we wish to emphasize this fact, we can say that f is continuous relative to specific topologies X and Y.

8.2 Euler Characteristic

[2] The Euler characteristic is one of the most useful topological invariants. Moreover, we find the prototype of the algebraic approach to topology in it. To avoid unnecessary complication, we restrict ourselves to points, lines and surfaces in $\mathbb{R}^3$. A polyhedron is a geometrical object surrounded by faces. The boundary of two faces is an edge and two edges meet at a vertex. We extend the definition of a polyhedron a bit to include polygons and the boundaries of polygons, lines or points. We call the faces, edges and vertices of a polyhedron simplexes (or simplices). Note that the boundary of two simplexes is either empty or another simplex. (For example,

[2] The rest of this chapter has been quoted from: Nakahara, Mikio. Geometry, topology and physics. CRC Press, 2003.

the boundary of two faces is an edge.) Formal definitions of a simplex and a polyhedron in a general number of dimensions will be given later. We are now ready to define the Euler characteristic of a figure in $\mathbb{R}^3$.

Definition 8.1 *Let X be a subset of $\mathbb{R}^3$, which is homeomorphic to a polyhedron K. Then the Euler characteristic $\chi(X)$ of X is defined by*

$$\chi(X) = (\textit{number of verticies in } K) - (\textit{number of edges in } K) + (\textit{number of faces in } K). \tag{8.1}$$

The reader might wonder if $\chi(X)$ depends on the polyhedron K or not. The following theorem due to Poincaré and Alexander guarantees that it is, in fact, independent of the polyhedron K.

Theorem 8.1 *(Poincaré-Alexander) The Euler characteristic $\chi(X)$ is independent of the polyhedron K as long as K is homeomorphic to X.*

Examples are in order. The Euler characteristic of a point is $\chi(\cdot) = 1$ by definition. The Euler characteristic of a line is $\chi(-) = 2-1 = 1$, since a line has two vertices and an edge. For a triangular disc, we find $\chi(\text{triangle}) = 3 - 3 + 1 = 1$. An example which is a bit non-trivial is the Euler characteristic of S_1. The simplest polyhedron which is homeomorphic to S_1 is made of three edges of a triangle. Then $\chi(S_1) = 3 - 3 = 0$. Similarly, the sphere S_2 is homeomorphic to the surface of a tetrahedron, hence $\chi(S_2) = 4-6+4 = 2$. It is easily seen that S_2 is also homeomorphic to the surface of a cube. Using a cube to calculate the Euler characteristic of S_2, we have $\chi(S_2) = 8 - 12 + 6 = 2$, in accord with theorem 8.1. Historically this is the conclusion of Eulers theorem: if K is any polyhedron homeomorphic to S_2, with v vertices, e edges and f two-dimensional faces, then $v - e + f = 2$.

8.3 Homology Groups

The mathematical structures underlying homology groups are finitely generated Abelian groups. Let H be a subgroup of G. We say $x, y \in G$ are equivalent if $x - y \in H$, and write $x \sim y$. Clearly

$\sim$ is an equivalence relation. The equivalence class to which x belongs is denoted by $[x]$. Let G/H be the quotient space. The group operation $+$ in G naturally induces the group operation $+$ in G/H by $[x]+[y]=[x+y]$.

Definition 8.2 *If G is finitely generated by r linearly independent elements, G is called a free Abelian group of rank r.*

Let us recall how the Euler characteristic of a surface is calculated. We first construct a polyhedron homeomorphic to the given surface, then count the numbers of vertices, edges and faces. The Euler characteristic of the polyhedron, and hence of the surface, is then given by equation (8.1). We abstract this procedure so that we may represent each part of a figure by some standard object. We take triangles and their analogues in other dimensions, called simplexes, as the standard objects. By this standardization, it becomes possible to assign to each figure Abelian group structures.

8.3.1 Simplexes

Simplexes are building blocks of a polyhedron. A 0-simplex $\langle p_0\rangle$ is a point, or a vertex, and a 1-simplex $\langle p_0p_1\rangle$ is a line, or an edge. A 2-simplex $\langle p_0p_1p_2\rangle$ is defined to be a triangle with its interior included and a 3-simplex $\langle p_0p_1p_2p_3\rangle$ is a solid tetrahedron. It is common to denote a 0-simplex without the bracket; $\langle p_0\rangle$ may be also written as p_0. It is easy to continue this construction to any r-simplex $\langle p_0p_1\dots p_r\rangle$. Note that for an r-simplex to represent an r-dimensional object, the vertices p_i must be geometrically independent, that is, no $(r-1)$-dimensional hyperplane contains all the $r+1$ points. Let $p_0,\dots,p_r$ be points geometrically independent in $\mathbb{R}^m$ where $m\geq r$. The r-simplex $\langle p_0,\dots,p_r\rangle$ is expressed as

$$\sigma^r=\left\{x\in\mathbb{R}^m | x=\sum_{i=0}^{r}c_ip_i, c_i\geq 0, \sum_{i=0}^{r}c_i=1\right\}.$$

$(c_0,\dots,c_r)$ is called the barycentric coordinate of x.

Let K be a set of finite number of simplexes in $\mathbb{R}^m$. If these simplexes are nicely fitted together, K is called a simplicial complex. By "nicely" we mean:

1. an arbitrary face of a simplex of K belongs to K, that is, if $\sigma\in K$ and $\sigma'\leq\sigma$ then $\sigma'\in K$; and

2. if σ and σ' are two simplexes of K, the intersection $\sigma \cap \sigma'$ is either empty or a common face of σ and σ', that is, if $\sigma, \sigma' \in K$ then either $\sigma \cap \sigma' = \emptyset$ or $\sigma \cap \sigma' \leq \sigma$ and $\sigma \cap \sigma' \leq \sigma'$.

Let X be a topological space. If there exists a simplicial complex K and a homeomorphism $f : |K| \to X \times X$ is said to be triangulable and the pair (K, f) is called a triangulation of X. Given a topological space X, its triangulation is far from unique.

Oriented Simplexes

We may assign orientations to an r -simplex for $r \geq 1$. Instead of $\langle \ldots \rangle$ for an unoriented simplex, we will use $(\ldots)$ to denote an oriented simplex. The symbol σ_r is used to denote both types of simplex. An oriented 1-simplex $\sigma_1 = (p_0 p_1)$ is a directed line segment traversed in the direction $p_0 \to p_1$. Now $(p_0 p_1)$ should be distinguished from $(p_1 p_0)$.

Definition 8.3 *The r-**chain** group $C_r(K)$ of a simplicial complex K is a free Abelian group generated by the oriented r-simplexes of K. If $r > \dim K$, $C_r(K)$ is defined to be 0. An element of $C_r(K)$ is called an r-chain.*

Before we define the cycle group and the boundary group, we need to introduce the boundary operator. Let us denote the boundary of an r-simplex σ_r by $\partial_r \sigma_r$. ∂_r should be understood as an operator acting on σ_r to produce its boundary.

Definition 8.4 *If $c \in C_r(K)$ satisfies $\partial_r c = 0$, c is called an r-cycle. The set of r-cycles $Z_r(K)$ is a subgroup of $C_r(K)$ and is called the r-cycle group.*

Let K be an n-dimensional simplicial complex. The rth homology group $H_r(K)$,$0 \leq r \leq n$, associated with K is defined by $H_r(K) \equiv Z_r(K)/B_r(K)$.

If necessary, we define $H_r(K) = 0$ for $r > n$ or $r < 0$. If we want to stress that the group structure is defined with integer coefficients, we write $H_r(K; \mathbb{Z})$.

8.4 Exercises

1. Translate the following text.

Euclidean geometry is a mathematical well-known system attributed to the Greek mathematician Euclid of Alexandria. Euclid's text Elements was the first systematic discussion of geometry. It has been one of the most influential books in history, as much for its method as for its mathematical content. The method consists of assuming a small set of intuitively appealing axioms, and then proving many other propositions (theorems) from those axioms. Although many of Euclid's results had been stated by earlier Greek mathematicians, Euclid was the first to show how these propositions could be fitted together into a comprehensive deductive and logical system.

The Elements begin with plane geometry, still often taught in secondary school as the first axiomatic system and the first examples of formal proof. The Elements goes on to the solid geometry of three dimensions, and Euclidean geometry was subsequently extended to any finite number of dimensions. Much of the Elements states results of what is now called number theory, proved using geometrical methods.

For over two thousand years, the adjective "Euclidean" was unnecessary because no other sort of geometry had been conceived. Euclid's axioms seemed so intuitively obvious that any theorem proved from them was deemed true in an absolute sense. Many other consistent formal geometries are now known, the first ones being discovered in the early 19th century. It also is no longer taken for granted that Euclidean geometry describes physical space. An implication of Einstein's theory of general relativity is that Euclidean geometry is only a good approximation to the properties of physical space if the gravitational field is not too strong.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
orientation			
			cyclically
	structure		
		simplicial	
	simulate		
variation			
			infinitely
		triangulable	
verification			

3. Use the correct form of the word.

To the ancients, the parallel postulate seemed less obvious than the others; (verify) ..... it physically would require us to inspect two lines to check that they never intersected, even at some very (distance) .....point, and this inspection could potentially take an infinite amount of time. Euclid himself seems to have considered it as being (quality) ..... different from the others, as evidenced by the organization of the Elements: the first 28 propositions he presents are those that can be proved without it.

Many geometers tried in vain to prove the fifth postulate from the first four. By 1763 at least 28 (differ) ..... proofs had been published, but all were found to be incorrect. In fact the parallel postulate cannot be proved from the other four: this was shown in the 19th century by the (construct) ..... of alternative (non-Euclidean) systems of geometry where the other axioms are still true but the parallel postulate is replaced by a (conflict) .....axiom.

One distinguishing aspect of these systems is that the three angles of a triangle do not add to 180: in hyperbolic geometry the sum of the three angles is always less than 180 and can approach

zero, while in elliptic geometry it is greater than 180. If the parallel postulate is dropped from the list of axioms without (replace) .....the result is the more general geometry called absolute geometry.

4. Fill gaps with one of the following words

1. another	2. inevitably	3. following	4. right
5. subtracted	6. centre	7. greater	8. segment
9. sphere	10. straight	11. concepts	12. derived

Following a precedent set in the Elements, Euclidean geometry has been exposited as an axiomatic system, in which all theorems ("true statements") are ..... from a finite number of axioms. Near the beginning of the first book of the Elements, Euclid gives five postulates (axioms):

1. Any two points can be joined by a ..... line.
2. Any straight line can be extended indefinitely in a straight line.
3. Given any straight line segment, a circle can be drawn having the segment as radius and one endpoint as .....
4. All ..... angles are congruent.
5. Parallel postulate. If two lines intersect a third in such a way that the sum of the inner angles on one side is less than two right angles, then the two lines ..... must intersect each other on that side if extended far enough.

These axioms invoke the following .....: point, straight line segment and line, side of a line, circle with radius and centre, right angle, congruence, inner and right angles, sum. The following verbs appear: join, extend, draw, intersect. The circle described in postulate 3 is tacitly unique. Postulates 3 and 5 hold only for plane geometry; in three dimensions, postulate 3 defines a .....

The Elements also include the ..... five "common notions":

1. Things that equal the same thing also equal one .....
2. If equals are added to equals, then the wholes are equal.
3. If equals are ..... from equals, then the remainders are equal.
4. Things that coincide with one another equal one another.
5. The whole is ..... than the part.

9

Selected Topics in Discrete Mathematics

9.1 The Language of Logic

[1] Logic is the study of the principles and techniques of reasoning. Logic plays a central role in the development of every area of learning, especially in mathematics and computer science. Computer scientists, for example, employ logic to develop programming languages and to establish the correctness of programs. Electronics engineers apply logic in the design of computer chips.

A declarative sentence that is either true or false, but not both, is a proposition (or a statement), which we will denote by the lowercase letter p, q, r, s, or t. The variables p, q, r, s, or t are boolean variables (or logic variables).

Consider the sentence, This sentence is false. It is certainly a valid declarative sentence, but is it a proposition? To answer this, assume the sentence is true. But the sentence says it is false. This contradicts our assumption. On the other hand, suppose the sentence is false. This implies the sentence is true, which again contradicts our assumption. Thus, if we assume that the sentence is true, it is false; and if we assume that it is false, it is true. It is a meaningless and self-contradictory sentence, so it is not a proposition, but a paradox.

A compound proposition is formed by combining two or more simple propositions called components. Compound propositions can be formed in several ways. The conjunction of two arbitrary

[1] This chapter has been quoted from:
Koshy, Thomas. Discrete mathematics with applications. Elsevier, 2004.

propositions p and q, denoted by $p \wedge q$, is the proposition p and q. It is formed by combining the propositions using the word `and`, called a connective. If both p and q are true, then $p \wedge q$ is true; otherwise it is false. A second way of combining two propositions p and q is by using the connective `or`. The resulting proposition p *or* q is the disjunction of p and q and is denoted by $p \vee q$. The disjunction of two propositions is true if at least one component is true; it is false only if both components are false.

The negation of a proposition p is It is not the case that p, denoted by $\neg p$. You may read $\neg p$ as the negation of p or simply not p. If a proposition p is true, then $\neg p$ is false; if p is false, then $\neg p$ is true.

Two propositions p and q can be combined to form statements of the form: If p, then q. Such a statement is an implication, denoted by $p \Rightarrow q$. Since it involves a condition, it is also called a conditional statement. The component p is the hypothesis (or premise) of the implication and q the conclusion. To construct the truth table for an implication *If* p, *then* q, we shall think of it as a conditional promise. If you do p, then I promise to do q. If the promise is kept, we consider the implication true; if the promise is not kept, we consider it false. We can use this analogy to construct the truth table, as shown below. Consider the following implication:

$p \Rightarrow q$: If you wax my car, then I will pay you $25.

If you wax my car (p true) and if I pay you $25 ($q$ true), then the implication is true. If you wax my car (p true) and if I do not pay you $25 ($q$ false), then the promise is violated; hence the implication is false. What if you do not wax my car (p false)? Then I may give you $25 (being generous!) or not. (So q may be true or false). In either case, my promise has not been tested and hence has not been violated. Consequently, the implication has not been proved false. If it is not false, it must be true. In other words, if p is false, the implication $p \Rightarrow q$ is true by default. (If p is false, the implication is said to be vacuously true.)

9.2 Combinatorics

At the beginning of the 18th century, the following problem was proposed:

A secretary had written n different letters and addressed n different envelopes for them. Unfortunately, a wind storm mixed up the letters and the envelopes. After the storm was over, each letter was placed in an envelope. In how many ways can the letters be placed in the envelopes, so that every letter is in a wrong envelope?

This problem has several variations. A variation involves n guests checking in their coats at the coat room of a fancy restaurant. In how many ways can the attendant return their coats, so no person gets the right coat?

Before answering these problems, we make the following definition. A permutation of n distinct items $a_1, a_2, \cdots, a_n$ in which no item a_i appears in its original position i for any i, $1 \leq i \leq n$, is called a derangement. We would like to find the number of possible derangements of n items.

Let D_n denote the number of derangements of n distinct items. To find an explicit formula for D_n, first we derive a recurrence relation satisfied by D_n as:

$$D_n = (n-1)(D_{n-1} + D_{n-2}), n \geq 2$$

where $D_0 = 1$ and $D_1 = 0$.

It can be proven that the number of derangements of n distinct elements is

$$D_n = n! \left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!}\right], n \geq 0$$

It is shown in calculus that

$$e^{-1} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}$$

so the expression inside the brackets in the formula is the sum of the first $(n+1)$ terms in the expansion of e^{-1}.

9.3 Graphs and Trees

Graph theory, a fascinating branch of mathematics, has numerous applications to such diverse areas as computer science, engineering, linguistics, and management science, as well as the natural

and social sciences. Like many important discoveries, graph theory grew out of an interesting physical problem, the celebrated Königsberg Bridge Puzzle. The outstanding Swiss mathematician Leonhard Euler solved the puzzle in 1736, thus laying the foundation for graph theory and earning his title as the father of graph theory.

A graph (or undirected graph) G consists of a nonempty finite set V of points (called vertices or nodes) and a set E of unordered pairs of elements in V (called edges). The graph G is the ordered pair $(V, E) : G = (V, E)$. An edge connecting the vertices u and v is denoted by $\{u, v\}$,$u - v$, or some label. Geometrically, edges are denoted by arcs or line segments.

An edge emanating from and terminating at the same vertex is a loop. Parallel edges have the same vertices. A simple graph contains no loops or parallel edges.

Two vertices v and w in a graph are adjacent, if an edge runs between them; if a loop occurs at v, v is adjacent to itself. An isolated vertex is not adjacent to any vertex. Adjacent edges have a common vertex. An edge is incident with a vertex v if v is an endpoint of the edge.

The degree of a vertex v in a graph is the number of edges meeting at v; it is denoted by $\deg(v)$. Clearly, a vertex v is isolated if $\deg(v) = 0$. In addition, a loop at v contributes two to its degree.

The adjacency matrix of a graph with n vertices $v_1, v_2, \ldots, v_n$ is an $n \times n$ matrix $A = (a_{ij})$, where a_{ij} = number of edges from v_i to v_j. Because every edge in a graph is undirected, $a_{ij} = a_{ji}$ for every i and j, so the adjacency matrix of every graph is symmetric.

A subgraph of a graph $G = (V, E)$ is a graph $G_1 = (V_1, E_1)$ where $V_1 \subseteq V$ and $E_1 \subseteq E$. A simple graph with an edge between every two distinct vertices is a complete graph. A complete graph with n vertices is denoted by K_n. The cycle graph C_n of length $n(\geq 3)$ consists of n vertices $v_1, \ldots, v_n$ and edges $\{v_i, v_{i+1}\}$, where $1 \leq i \leq n$ and $v_{n+1} = v_0$

If the vertex set V of a simple graph $G = (V, E)$ can be partitioned into two disjoint (nonempty) sets V_1 and V_2, so every edge in G is incident with a vertex in V_1 and a vertex V_2, then G is bipartite. Let G be a bipartite graph with $|V_1| = m$ and $|V_2| = n$.

If an edge runs between every vertex in V_1 and V_2, G is a complete bipartite graph, denoted by $K_{m,n}$.

A simple graph in which each edge e is assigned a positive real number w is a weighted graph.

A graph is planar if it can be drawn in the plane, so its edges meet only at the vertices. Such a drawing is a planar representation of the graph. Let G be a connected planar graph with e edges and v vertices. Let r be the number of regions formed by a planar representation of G. Then $r = e - u + 2$ (Euler's formula).

Trees are the most important class of graphs and they make fine modeling tools. A connected, acyclic graph is a tree.

- A connected graph is a tree if and only if there is a unique, simple path between any two vertices.
- A connected graph with n vertices is a tree if and only if it has exactly $n - 1$ edges.

A subgraph H of a connected graph G is a spanning tree of G if H is a tree containing every vertex of G. Every connected graph has a spanning tree.

Let G be a connected weighted graph. The weight of a spanning tree of G is the sum of the weights of its edges. A minimal spanning tree of G weighs the least. Several algorithms can find a minimal spanning tree T of a connected weighted graph G. Two of them are Kruskal's and Prim's.

A tree with a root is a rooted tree. Rooted trees are drawn with the root at the top, especially in computer science; they grow downward.

Let T be a rooted tree with root v_0. Let $v_0 - v_1 - \cdots - v_{n-1} - v_n$ be the path from v_0 to v_n. Then:

- v_{n-1} is the parent of v_n.
- v_n is a child of v_{n-1}.
- Vertices with the same parent are siblings.
- The vertices $v_0, v_1, \ldots, v_{n-1}$ are ancestors of v_n.
- The descendants of a vertex v are those vertices for which v is an ancestor.
- A vertex with no children is a leaf or a terminal vertex.
- A vertex that is not a leaf is an internal vertex.

- The subtree of T rooted at v consists of v, its descendants, and all edges incident with them.

9.4 Recursion

Recursion is an elegant and powerful problem-solving technique, used extensively in both discrete mathematics and computer science.

Leonardo Fibonacci, the most outstanding Italian mathematician of the Middle Ages, proposed the following problem around 1202:

> Suppose there are two newborn rabbits, one male and the other female. Find the number of rabbits produced in a year if:
> - Each pair takes one month to become mature.
> - Each pair produces a mixed pair every month, from the second month.
> - No rabbits die.

Suppose, for convenience, that the original pair of rabbits was born on January 1. They take a month to become mature. So there is still only one pair on February 1. On March 1, they are 2 months old and produce a new mixed pair, a total of two pairs. Continuing like this, there will be three pairs on April 1, five pairs on May 1, and so on.

The numbers $1, 1, 2, 3, 5, 8, \cdots$ are Fibonacci numbers. They have a fascinating property: Any Fibonacci number, except the first two, is the sum of the two immediately preceding Fibonacci numbers. (At the given rate, there will be 144 pairs of rabbits on December 1.)

This yields the following recursive definition of the nth Fibonacci number F_n:

$$\begin{aligned} F_1 = F_2 &= 1 && \longleftarrow \text{ initial conditions} \\ F_n &= F_{n-1} + F_{n-2} && \longleftarrow \text{ recurrence relation} \end{aligned}$$

Solving the recurrence relation for a function f means finding an explicit formula for F_n. The iterative method of solving it involves two steps:

- Apply the recurrence formula iteratively and look for a pattern to predict an explicit formula.
- Use induction to prove that the formula does indeed hold for every possible value of the integer n.

Unfortunately, the iterative method illustrated above can be applied to only a small and simple class of recurrence relations. The recurrence relation $F_n = F_{n-1} + F_{n-2}$ is linear and homogeneous. If it has a nonzero solution of the form $c\alpha^n$, then $c\alpha^n = c\alpha^{n-1} + c\alpha^{n-2}$. Since $c\alpha \neq 0$, this yields $\alpha^2 = \alpha + 1$; that is, $\alpha^2 - \alpha - 1 = 0$, so α must be a solution of the characteristic equation $x^2 - x - 1 = 0$, and its solutions are a $\alpha = \frac{1+\sqrt{5}}{2}$ and $\beta = \frac{1-\sqrt{5}}{2}$. You may verify $\alpha + \beta = 1$ and $\alpha\beta = -1$. The general solution is $F_n = A\alpha^n + B\beta^n$. To find A and B, we have:

$$F_1 = A\alpha + B\beta = 1$$
$$F_2 = A\alpha^2 + B\beta^2 = 1$$

Solving these two equations, we get (Verify):

$$A = \frac{1}{\sqrt{5}} \quad \text{and} \quad B = -\frac{1}{\sqrt{5}}.$$

9.5 Exercises

1. Translate the following sentences.

Certain graph problems deal with finding a path between two vertices such that each edge is traversed exactly once, or finding a path between two vertices while visiting each vertex exactly once. These paths are better known as Euler path and Hamiltonian path respectively. The Euler path problem was first proposed in the 1700s.

- An Euler path is a path that uses every edge of a graph exactly once. It starts and ends at different vertices.
- An Euler circuit is a circuit that uses every edge of a graph exactly once. It starts and ends at the same vertex.

There are simple criteria for determining whether a multigraph has a Euler path or a Euler circuit. For any multigraph to have a Euler circuit, all the degrees of the vertices must be even.

Theorem: A connected multigraph (and simple graph) with at least two vertices has a Euler circuit if and only if each of its vertices has an even degree.
Proof of the above statement is that every time a circuit passes through a vertex, it adds twice to its degree. Since it is a circuit, it starts and ends at the same vertex, which makes it contribute one degree when the circuit starts and one when it ends. In this way, every vertex has an even degree.

Theorem: A connected multigraph (and simple graph) has an Euler path but not an Euler circuit if and only if it has exactly two vertices of odd degree.
The proof is an extension of the proof given above. Since a path may start and end at different vertices, the vertices where the path starts and ends are allowed to have odd degrees.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
logic			
			iteratively
	contradict		
		homogeneous	
	weigh		
paradox			
			correctly
		inductive	
recurrence			

3. Use the correct form of the word.

1. Mathematical proof is an argument we give logically to (valid) .....a mathematical statement.

2. When we want to (proof) ..... a conditional statement p implies q, we assume that p is true, and follow (imply) .....to get to show that q is then true.
3. Symmetric and anti-symmetric (relate)..... are not opposite because a relation can contain both the properties or may not.
4. A relation is an (equal) ..... relation if it is reflexive, symmetric, and transitive.
5. A relation R on a set A is called (refelect) ..... if $(a, a) \in R$ holds for every element $a \in A$.

4. Fill gaps with one of the following words

1. exactly	2. route	3. path	4. diverse
5. electronic	6. octahedron	7. district	8. nodes
9. determine	10. unique	11. complete	12. visiting

A Hamiltonian cycle is a closed loop on a graph where every node (vertex) is visited exactly once. A loop is just an edge that joins a node to itself; so a Hamiltonian cycle is a path traveling from a point back to itself, ... every node en route. If a graph with more than one node (i.e. a non-singleton graph) has a Hamiltonian cycle, we call it a Hamiltonian graph. There isnt any equation or general trick to finding out whether a graph has a Hamiltonian cycle; the only way to ..... this is to do a complete and exhaustive search, going through all the options.

Every ..... graph with more than two vertices is a Hamiltonian graph. This follows from the definition of a complete graph: an undirected, simple graph such that every pair of nodes is connected by a ..... edge. The graph of every platonic solid is a Hamiltonian graph. So the graph of a cube, a tetrahedron, an ....., or an icosahedron are all Hamiltonian graphs with Hamiltonian cycles. A graph with n vertices (where $n > 3$) is Hamiltonian if the sum of the degrees of every pair of non-adjacent vertices is n or greater. This is known as Ores theorem.

A search for Hamiltonian cycles is not just a fun game for the afternoon off. It has real applications in such ..... fields as computer graphics, ... circuit design, mapping genomes, and operations research. For instance, when mapping genomes scientists

must combine many tiny fragments of genetic code ("read", they are called), into one single genomic sequence (a "superstring"). This can be done by finding a Hamiltonian cycle or Hamiltonian ....., where each of the reads are considered nodes in a graph and each overlap (place where the end of one read matches the beginning of another) is considered to be an edge. In a much less complex application of exactly the same math, school districts use Hamiltonian cycles to plan the best ..... to pick up students from across the ..... Here, students may be considered ....., the paths between them edges, and the bus wishes to travel a route that will pass each students house ..... once.

10

Selected Topics in Optimization

10.1 The Origins of Operations Research

[1] The roots of operations research (OR) can be traced back many decades, when early attempts were made to use a scientific approach in the management of organizations. However, the beginning of the activity called operations research has generally been attributed to the military services early in World War II. Because of the war effort, there was an urgent need to allocate scarce resources to the various military operations and to the activities within each operation in an effective manner. Therefore, the British and then the U.S. military management called upon a large number of scientists to apply a scientific approach to dealing with this and other strategic and tactical problems. In effect, they were asked to do research on (military) operations. These teams of scientists were the first OR teams. By developing effective methods of using the new tool of radar, these teams were instrumental in winning the Air Battle of Britain. Through their research on how to better manage convoy and antisubmarine operations, they also played a major role in winning the Battle of the North Atlantic. Similar efforts assisted the Island Campaign in the Pacific.

When the war ended, the success of OR in the war effort spurred interest in applying OR outside the military as well. As the industrial boom following the war was running its course, the problems

[1] This chapter has been quoted from:
Introduction to Operations Research Seventh Edition, Hillier and Lieberman 2001
The McGraw-Hill Companies.

caused by the increasing complexity and specialization in organizations were again coming to the forefront. It was becoming apparent to a growing number of people, including business consultants who had served on or with the OR teams during the war, that these were basically the same problems that had been faced by the military but in a different context. By the early 1950s, these individuals had introduced the use of OR to a variety of organizations in business, industry, and government. The rapid spread of OR soon followed.

10.2 Linear Programming

The development of linear programming has been ranked among the most important scientific advances of the mid-20th century, and we must agree with this assessment. Its impact since just 1950 has been extraordinary. Today it is a standard tool that has saved many thousands or millions of dollars for most companies or businesses of even moderate size in the various industrialized countries of the world; and its use in other sectors of society has been spreading rapidly. A major proportion of all scientific computation on computers is devoted to the use of linear programming. Dozens of textbooks have been written about linear programming, and published articles describing important applications now number in the hundreds.

Linear programming uses a mathematical model to describe the problem of concern. The adjective linear means that all the mathematical functions in this model are required to be linear functions. The word programming does not refer here to computer programming; rather, it is essentially a synonym for planning. Thus, linear programming involves the planning of activities to obtain an optimal result, i.e., a result that reaches the specified goal best (according to the mathematical model) among all feasible alternatives.

Although allocating resources to activities is the most common type of application, linear programming has numerous other important applications as well. In fact, any problem whose mathematical model fits the very general format for the linear programming model is a linear programming problem. Furthermore, a re-

markably efficient solution procedure, called the simplex method, is available for solving linear programming problems of even enormous size. These are some of the reasons for the tremendous impact of linear programming in recent decades.

Let us begin with some basic terminology and notation. The key terms are resources and activities, where m denotes the number of different kinds of resources that can be used and n denotes the number of activities being considered. Some typical resources are money and particular kinds of machines, equipment, vehicles, and personnel. Examples of activities include investing in particular projects, advertising in particular media, and shipping goods from a particular source to a particular destination. In any application of linear programming, all the activities may be of one general kind (such as any one of these three examples), and then the individual activities would be particular alternatives within this general category.

The model poses the problem in terms of making decisions about the levels of the activities, so $x_1, x_2, \ldots, x_n$ are called the decision variables. The values of c_j, b_i, and a_{ij} (for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$) are the input constants for the model. The c_j, b_i, and a_{ij} are also referred to as the parameters of the model.

we can now formulate the mathematical model for this general problem of allocating resources to activities. In particular, this model is to select the values for $x_1, x_2, \ldots, x_n$ so as to

$$\text{Maximize} \quad Z = c_1x_1 + c_2x_2 + \cdots + c_nx_n,$$

subject to the restrictions

$$\begin{array}{ccccccc} a_{11}x_1 & +a_{12}x_2 & + \cdots & +a_{1n}x_n & \leq b_1 \\ a_{21}x_1 & +a_{22}x_2 & + \cdots & +a_{2n}x_n & \leq b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1}x_1 & +a_{m2}x_2 & + \cdots & +a_{mn}x_n & \leq b_m \end{array}$$

and

$$x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0.$$

Common terminology for the linear programming model can now be summarized. The function being maximized, $c_1x_1 + c_2x_2 + \cdots +$

$c_n x_n$, is called the objective function. The restrictions normally are referred to as constraints. The first m constraints (those with a function of all the variables $a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i$ on the left-hand side) are sometimes called functional constraints (or structural constraints). Similarly, the $x_j \geq 0$ restrictions are called nonnegativity constraints (or nonnegativity conditions).

You may be used to having the term solution mean the final answer to a problem, but the convention in linear programming (and its extensions) is quite different. Here, any specification of values for the decision variables $(x_1, x_2, \ldots, x_n)$ is called a solution, regardless of whether it is a desirable or even an allowable choice. Different types of solutions are then identified by using an appropriate adjective.

A feasible solution is a solution for which all the constraints are satisfied. An infeasible solution is a solution for which at least one constraint is violated. The feasible region is the collection of all feasible solutions. An optimal solution is a feasible solution that has the most favorable value of the objective function. The most favorable value is the largest value if the objective function is to be maximized, whereas it is the smallest value if the objective function is to be minimized.

10.3 The Transportation and Assignment Problems

We continue to broaden our horizons here by discussing two particularly important (and related) types of linear programming problems. One type, called the transportation problem, received this name because many of its applications involve determining how to optimally transport goods. However, some of its important applications (e.g., production scheduling) actually have nothing to do with transportation. The second type, called the assignment problem, involves such applications as assigning people to tasks. Although its applications appear to be quite different from those for the transportation problem, we shall see that the assignment problem can be viewed as a special type of transportation problem.

10.3.1 Transportation Problem

To describe the general model for the transportation problem, we need to use terms that are considerably less specific than those for the components of the prototype example. In particular, the general transportation problem is concerned (literally or figuratively) with distributing any commodity from any group of supply centers, called sources, to any group of receiving centers, called destinations, in such a way as to minimize the total distribution cost.

> **The requirements assumption:** Each source has a fixed supply of units, where this entire supply must be distributed to the destinations. (We let s_i denote the number of units being supplied by source i, for $i = 1, 2, \ldots, m$.) Similarly, each destination has a fixed demand for units, where this entire demand must be received from the sources. (We let d_j denote the number of units being received by destination j, for $j = 1, 2, \ldots, n$.)

This assumption that there is no leeway in the amounts to be sent or received means that there needs to be a balance between the total supply from all sources and the total demand at all destinations.

> **The feasible solutions property:** A transportation problem will have feasible solutions if and only if
>
> $$\sum_{i=1}^{m} s_i = \sum_{j=1}^{m} d_j$$

In some real problems, the supplies actually represent maximum amounts (rather than fixed amounts) to be distributed. Similarly, in other cases, the demands represent maximum amounts (rather than fixed amounts) to be received. Such problems do not quite fit the model for a transportation problem because they violate the requirements assumption. However, it is possible to reformulate the problem so that they then fit this model by introducing a dummy destination or a dummy source to take up the slack between the actual amounts and maximum amounts being distributed. We will

illustrate how this is done with two examples at the end of this section.

This reference to a *unit* cost implies the following basic assumption for any transportation problem.

> **The cost assumption:** The cost of distributing units from any particular source to any particular destination is directly proportional to the number of units distributed. Therefore, this cost is just the unit cost of distribution times the number of units distributed. (We let c_{ij} denote this unit cost for source i and destination j.)

The only data needed for a transportation problem model are the supplies, demands, and unit costs. These are the parameters of the model.

> **The model:** Any problem (whether involving transportation or not) fits the model for a transportation problem if it can be described completely in terms of a parameter and it satisfies both the requirements assumption and the cost assumption. The objective is to minimize the total cost of distributing the units.

10.3.2 Assignment Problem

The assignment problem is a special type of linear programming problem where assignees are being assigned to perform tasks. For example, the assignees might be employees who need to be given work assignments. Assigning people to jobs is a common application of the assignment problem. However, the assignees need not be people. They also could be machines, or vehicles, or plants, or even time slots to be assigned tasks. The first example below involves machines being assigned to locations, so the tasks in this case simply involve holding a machine. A subsequent example involves plants being assigned products to be produced.

To fit the definition of an assignment problem, these kinds of applications need to be formulated in a way that satisfies the following assumptions.

1. The number of assignees and the number of tasks are the same. (This number is denoted by n.)

2. Each assignee is to be assigned to exactly one task.
3. Each task is to be performed by exactly one assignee.
4. There is a cost c_{ij} associated with assignee i ($i = 1, 2, \ldots, n$) performing task j ($j = 1, 2, \ldots, n$).
5. The objective is to determine how all n assignments should be made to minimize the total cost.

Any problem satisfying all these assumptions can be solved extremely efficiently by algorithms designed specifically for assignment problems.

The first three assumptions are fairly restrictive . Many potential applications do not quite satisfy these assumptions. However, it often is possible to reformulate the problem to make it fit. For example, dummy assignees or dummy tasks frequently can be used for this purpose.

10.4 Exercises

1. Translate the following text.

According to the number of time periods considered in the model, optimization models can be classified as static (single time period) or multistage (multiple time periods). Even when all relationships are linear, if several time periods are incorporated in the model the resulting linear program could become prohibitively large for solution by standard computational methods. Fortunately, in most of these cases, the problem exhibits some form of special structure that can be adequately exploited by the application of special types of mathematical programming methods. Dynamic programming is one approach for solving multistage problems. Further, there is a considerable research effort underway today, in the field of large-scale linear programming, to develop special algorithms to deal with multistage problems.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
proportion			
			specially
	assume		
		restrictive	
	allocate		
complexity			
			optimally
		scientific	
solution			

3. Use the correct form of the word.

Another important way of (classify)..... optimization models refers to the behavior of the parameters of the model. If the parameters are known constants, the optimization model is said to be deterministic. If the parameters are specified as uncertain quantities, whose values are (characterization) .....by probability distributions, the optimization model is said to be stochastic. If some of the parameters are allowed to (variation) .....systematically, and the changes in the optimum solution corresponding to changes in those parameters are determined, the optimization model is said to be parametric. (general)....., stochastic and parametric mathematical programming give rise to much more difficult problems than deterministic mathematical programming. Although important theoretical and practical contributions have been made in the areas of stochastic and parametric programming, there are still no (effect).....general procedures that cope with these problems. Deterministic linear programming, however, can be (efficient).....applied to very large problems of up to 5000 rows and an almost unlimited number of variables. Moreover, in linear programming, sensitivity analysis and parametric

programming can be conducted effectively after (obtain) ..... the deterministic optimum solution.

4. Fill gaps with one of the following words

1. amount	2. program-ming	3. class	4. models
5. efficiently	6. structure	7. variables	8. satisfies
9. progress	10. optimal	11. discrete	12. continuous

A way of classifying optimization ..... deals with the behavior of the variables in the ..... solution. If the variables are allowed to take any value that ..... the constraints, the optimization model is said to be continuous. If the variables are allowed to take on only ..... values, the optimization model is called integer or discrete. Finally, when there are some integer variables and some continuous ..... in the problem, the optimization model is said to be mixed. In general, problems with integer variables are significantly more difficult to solve than those with ..... variables. Network models are a ..... of linear programming models that are an exception to this rule, as their special ..... results in integer optimal solutions. Although significant ..... has been made in the general area of mixed and integer linear ....., there is still no algorithm that can ..... solve all general medium-size integer linear programs in a reasonable ..... of time though, for special problems, adequate computational techniques have been developed.

11

Selected Topics in Analysis

11.1 Compact space

[1] Intuitively speaking, a space is said to be compact if whenever one takes an infinite number of "steps" in the space, eventually one must get arbitrarily close to some other point of the space. Thus, while disks and spheres are compact, infinite lines and planes are not, nor is a disk or a sphere with a missing point. In the case of an infinite line or plane, one can set off making equal steps in any direction without approaching any point, so that neither space is compact. In the case of a disk or sphere with a missing point, one can move towards the missing point without approaching any point within the space. This demonstrates that the disk or sphere with a missing point are not compact either.

Compactness generalizes many important properties of closed and bounded intervals in the real line; that is, intervals of the form $[a, b]$ for real numbers a and b. For instance, any continuous function defined on a compact space into an ordered set (with the order topology) such as the real line is bounded. Thus, what is known as the extreme value theorem in calculus generalizes to compact spaces. In this fashion, one can prove many important theorems in the class of compact spaces, that do not hold in the context of non-compact ones.

Various definitions of compactness may apply, depending on the level of generality. A subset of Euclidean space in particular is

[1] This chapter has been quoted from:
A Course in Functional Analysis, John B. Conway , Springer-Verlag New York Berlin Heidelberg Tokyo, 1985.

called compact if it is closed and bounded. This implies, by the Bolzano-Weierstrass Theorem, that any infinite sequence from the set has a subsequence that converges to a point in the set. This puts a fine point on the idea of taking "steps" in a space. Various equivalent notions of compactness, such as sequential compactness and limit point compactness, can be developed in general metric spaces.

In general topological spaces, however, the different notions of compactness are not equivalent, and the most useful notion of compactness originally called bicompactness involves families of open sets that "cover" the space in the sense that each point of the space must lie in some set contained in the family. Specifically, a topological space is compact if, whenever a collection of open sets covers the space, some subcollection consisting only of finitely many open sets also covers the space. That this form of compactness holds for closed and bounded subsets of Euclidean space is known as the Heine-Borel Theorem. Compactness, when defined in this manner, often allows one to take information that is known locally - in a neighborhood of each point of the space - and to extend it to information that holds globally throughout the space. An example of this phenomenon is Dirichlet's theorem, to which it was originally applied by Heine, that a continuous function on a compact interval is uniformly continuous: here continuity is a local property of the function, and uniform continuity the corresponding global property.

11.2 Hilbert space

A Hilbert space is the abstraction of the finite-dimensional Euclidean spaces of geometry. Its properties are very regular and contain few surprises, though the presence of an infinity of dimensions guarantees a certain amount of surprise. Historically, it was the properties of Hilbert spaces that guided mathematicians when they began to generalize.

Definition 11.1 *If X is a vector space over $\mathbb{F}$, a semi-inner product on X is a function $u : X \times X \to \mathbb{F}$ such that for all α, β in $\mathbb{F}$ and x, y, z in $\mathcal{H}$, the following are satisfied:*

(a) $u(\alpha x + \beta y, z) = \alpha u(x, z) + \beta u(y, z)$,
(b) $u(x, \alpha y + \beta z) = \alpha u(x, y) + \beta u(x, z)$,
(c) $u(x, x) \geq 0$,
(d) $u(x, y) = u(y, x)$.

An inner product on $\mathcal{H}$ is a semi-inner product that also satisfies the following: If $u(x, x) = 0$, then $x = 0$. An inner product here will be denoted by $\langle x, y\rangle = u(x, y)$.

A Hilbert space is a vector space $\mathcal{H}$ over $\mathbb{F}$ together with an inner product $\langle ., .\rangle$ such that relative to the metric $d(x, y) = ||x - y||$ induced by the norm, $\mathcal{H}$ is a complete metric space.

The greatest advantage of a Hilbert space is its underlying concept of orthogonality .

Definition 11.2 *If $\mathcal{H}$ is a Hilbert space and $f, g \in \mathcal{H}$, then f and g are orthogonal if $\langle f, g\rangle = 0$. In symbols, $f \perp g$. If $A, B \subseteq \mathcal{H}$, then $A \perp B$ if $f \perp g$ for every f in A and g in B.*

Definition 11.3 *If $\mathcal{H}$ is any vector space over $\mathbb{F}$ and $A \subseteq \mathcal{H}$, then A is a convex set if for any x and y in A and $0 \leq t \leq 1$, $tx + (1 - t)y \in A$.*

Note that $\{tx + (1 - t)y : 0 \leq t \leq 1\}$ is the straight-line segment joining x and y. So a convex set is a set A such that if x and y in A, the entire line segment joining x and y is contained in A.

11.3 Orthogonal Sets of Vectors and Basis

It cab be shown that, as in Euclidean space, each Hilbert space can be coordinatized. The vehicle for introducing the coordinates is an orthogonal basis.

Definition 11.4 *An orthogonal subset of a Hilbert space $\mathcal{H}$ is a subset $\mathcal{E}$ having the properties: (a) for e in $\mathcal{E}$, $||e|| = 1$; (b) if $e_l, e_2 \in \mathcal{E}$ and $e_1 \neq e_2$, then $e_1 \perp e_2$.*

A basis for $\mathcal{H}$ is a maximal orthogonal set.

Actually, the sum that appears in the following can be given a better interpretation- a mathematically precise one that will be useful later. The question is, what is meant by $\sum\{h_i : i \in I\}$ if $h_i \in \mathcal{H}$ and I is an infinite, possibly uncountable set? Let $\mathcal{F}$ be

the collection of all finite subsets of I and order $\mathcal{F}$ by inclusion , so $\mathcal{F}$ becomes a directed set. For each F in $\mathcal{F}$, define

$$h_F = \sum\{h_i : i \in F\}.$$

Since this is a finite sum, h_F is a well-defined element of $\mathcal{H}$. Now $\{h_F : F \in \mathcal{F}\}$is a net in $\mathcal{H}$.

With the notation above, the sum $\sum\{h_i : i \in I\}$ converges if the net $\{h_F : F \in \mathcal{F}\}$ converges; the value of the sum is the limit of the net. If $\mathcal{H} = \mathbb{F}$, the definition above gives meaning to an uncountable sum of scalars. If the set I is countable, then this definition of convergent sum is not the usual one. That is, if $\{h_n\}$ is a sequence in $\mathcal{H}$ then the convergence of $\sum\{h_n : n \in \mathbb{N}\}$ is not equivalent to the convergence of $\displaystyle\sum_{n=1}^{\infty} h_n$.

11.4 Isomorphic Hilbert Spaces

Every mathematical theory has its concept of isomorphism. In topology there is homeomorphism and homotopy equivalence; algebra calls them isomorphisms. The basic idea is to define a map which preserves the basic structure of the spaces in the category.

Definition 11.5 *If $\mathcal{H}$ and $\mathcal{K}$ are Hilbert spaces, an isomorphism between $\mathcal{H}$ and $\mathcal{K}$ is a linear surjection $U : \mathcal{H} \to \mathcal{K}$ such that*

$$\langle Uh, Ug\rangle = \langle h, g\rangle.$$

for all h, g in $\mathcal{H}$. In this case $\mathcal{H}$ and $\mathcal{K}$ are said to be isomorphic.

It is easy to see that if $U : \mathcal{H} \to \mathcal{K}$ is an isomorphism, then so is $U^{-1} : \mathcal{K} \to \mathcal{H}$. Similar such arguments show that the concept of isomorphic is an equivalence relation on Hilbert spaces. It is also certain that this is the correct equivalence relation since an inner product is the essential ingredient for a Hilbert space and isomorphic Hilbert spaces have the same inner product. One might object that completeness is another essential ingredient in the definition of a Hilbert space. So it is! However, this too is preserved by an isomorphism. An isometry between metric spaces is a map that preserves distance.

Note that an isometry between metric spaces maps Cauchy sequences into Cauchy sequences. Thus an isomorphism also preserves completeness. That is, if an inner product space is isomorphic to a Hilbert space, then it must be complete.

A word about terminology. Many call what we call an isomorphism a unitary operator. We shall define a unitary operator as a linear transformation $U : \mathcal{H} \to \mathcal{K}$ that is a surjective isometry. That is, a unitary operator is an isomorphism whose range coincides with its domain. This may seem to be a minor distinction, and in many ways it is. But experience has taught me that there is some benefit in making such a distinction, or at least in being aware of it.

Theorem 11.1 *Two Hilbert spaces are isomorphic if and only if they have the same dimension.*

11.5 Operators on Hilbert Space

There is a marked contrast here between Hilbert spaces and the Banach spaces. Essentially all of the information about the geometry of Hilbert space is contained in the preceding chapter. The geometry of Banach space lies in darkness and has attracted the attention of many talented research mathematicians. However, the theory of linear operators (linear transformations) on a Banach space has very few general results, whereas Hilbert space operators have an elegant and well-developed general theory. Indeed, the reason for this dichotomy is related to the opposite status of the geometric considerations. Questions concerning operators on Hilbert space don't necessitate or imply any geometric difficulties.

Proposition 11.1 *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces and $A : \mathcal{H} \to \mathcal{K}$ a linear transformation. The following statements' are equivalent.*

(a) A is continuous.
(b) A is continuous at 0.
(c) A is continuous at some point.
(d) There is a constant $c > 0$ such that $||Ah|| \leq c||h||$ for all h in $\mathcal{H}$.

By virtue of the preceding proposition, $d(A, B) = ||A - B||$ defines a metric on $\mathcal{B}(\mathcal{H}, \mathcal{K})$. So it makes sense to consider $\mathcal{B}(\mathcal{H}, \mathcal{K})$ as a metric space.

Definition 11.6 *If $\mathcal{H}$ and $\mathcal{K}$ are Hilbert spaces, a function $u : \mathcal{H} \times \mathcal{K} :\to \mathbb{F}$ is a sesquilinear form if for h, g in $\mathcal{H}$, k, f in $\mathcal{K}$, and α, β in $\mathbb{F}$,*

(a) $u(\alpha h + \beta g, k) = \alpha u(h, k) + \beta u(g, k)$;
(b) $u(h, \alpha k + \beta f) = \alpha u(h, k) + \beta u(h, f)$.

The prefix sesqui is used because the function is linear in one variable but (for $\mathbb{F} = \mathcal{C}$) only conjugate linear in the other. (Sesqui means One-and-a-half.)

a sesquilinear form is bounded, if there is a constant M such that $|u(h, k)| \leq M||h||||k||$ for all h in $\mathcal{H}$ and k in $\mathcal{K}$. The constant M is called a bound for u.

If $K : L^2(\mu) \to L^2(\mu)$ is defined by

$$(Kf)(x) = \int k(x, y)f(y)dt\mu(y),$$

then K is a bounded linear operator. The operator described above is called an integral operator and the function k is called its kernel.

Definition 11.7 *If $A \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, then the unique operator B in $\mathcal{B}(\mathcal{K}, \mathcal{H})$ satisfying $u(h, k) = (Ah, k) = (h, Bk)$ is called the adjoint of A and is denoted by $B = A^*$.*

11.6 Exercises

1. Translate the following text.

The complex numbers are the field $\mathbb{C}$ of numbers of the form $x + iy$, where x and y are real numbers and i is the imaginary unit equal to the square root of -1, $\sqrt{-1}$. When a single letter $z = x + iy$ is used to denote a complex number, it is sometimes called an "affix". In component notation, $z = x+iy$ can be written (x, y). The field of complex numbers includes the field of real numbers as a subfield.

Complex numbers are useful abstract quantities that can be used in calculations and result in physically meaningful solutions.

However, recognition of this fact is one that took a long time for mathematicians to accept. For example, John Wallis wrote, "These Imaginary Quantities (as they are commonly called) arising from the Supposed Root of a Negative Square (when they happen) are reputed to imply that the Case proposed is impossible".

Through the Euler formula, a complex number $z = x + iy$ may be written in "phasor" form

$$z = |z|(\cos\theta + i\sin\theta) = |z|e^{i\theta}.$$

Here, $|z|$ is known as the complex modulus (or sometimes the complex norm) and theta is known as the complex argument or phase. Historically, the geometric representation of a complex number as simply a point in the plane was important because it made the whole idea of a complex number more acceptable. In particular,"imaginary" numbers became accepted partly through their visualization.

Unlike real numbers, complex numbers do not have a natural ordering, so there is no analog of complex-valued inequalities. This property is not so surprising however when they are viewed as being elements in the complex plane, since points in a plane also lack a natural ordering.

2. Write down other forms of the following words.

Noun	**Verb**	**adjective**	**adverb**
coordinates			
			equivalently
	linearize		
		sequential	
	necessitate		
operation			
			uniformly
		compact	
distinction			

3. Use the correct form of the word.

1. Show that if we (calculation) .....with symbols $x+iy$, where x and y are real numbers, according to the usual rules for adding and multiplying numbers and in addition use $i^2 = -1$, then all the (require) ..... for a field are satisfied.
2. After (introduction) ..... complex numbers we can, for any given real number, find a real or complex number whose square is the given number.
3. A map is called conformal if it preserves angles and their (orient) .....
4. The angle between two smooth curves in an oriented surface at a point of (intersect) .....of the curves is the angle between the tangents at the point.
5. So far we have said nothing about (converge) ..... on the boundary of the circle of convergence. There is a good reason for this; nothing much can be said in general. One can have divergence at every point of the circle, convergence at some points and (diverge)..... at others or one can have absolute convergence at every point of the circle.

4. Fill gaps with one of the following words

1. properties	2. integrals	3. fundamental	4. complex
5. numbers	6. analytic	7. applications	8. derivative
9. regions	10. nice	11. derivative	12. extremely

Complex analysis is the study of complex ..... together with their derivatives, manipulation, and other ..... Complex analysis is an ..... powerful tool with an unexpectedly large number of practical ..... to the solution of physical problems. Contour integration, for example, provides a method of computing difficult ..... by investigating the singularities of the function in ..... of the complex plane near and between the limits of integration.

The key result in complex analysis is the Cauchy integral theorem, which is the reason that single-variable complex analysis has so many ..... results. A single example of the unexpected power of ..... analysis is Picard's great theorem, which states that an ..... function assumes every complex number, with possibly one

exception, infinitely often in any neighborhood of an essential singularity!

A ..... result of complex analysis is the Cauchy-Riemann equations, which give the conditions a function must satisfy in order for a complex generalization of the ....., the so-called complex derivative, to exist. When the complex ..... is defined "everywhere" the function is said to be analytic.

12

Selected Topics in Number Theory

12.1 What is number theory?

[1] Number theory (or arithmetic or higher arithmetic in older usage) is a branch of pure mathematics devoted primarily to the study of the integers and integer-valued functions. German mathematician Carl Friedrich Gauss (1777-1855) said,"Mathematics is the queen of the sciences-and number theory is the queen of mathematics" Number theorists study prime numbers as well as the properties of objects made out of integers (for example, rational numbers) or defined as generalizations of the integers (for example, algebraic integers). Number theory seeks to understand the properties of integer systems in spite of their apparent complexity.

12.2 Main subdivisions

12.2.1 Elementary tools

The term elementary generally denotes a method that does not use complex analysis. For example, the prime number theorem was first proven using complex analysis in 1896, but an elementary proof was found only in 1949 by Erdős and Selberg.[76] The term is somewhat ambiguous: for example, proofs based on complex Tauberian theorems (for example, Wiener-Ikehara) are often seen as quite enlightening but not elementary, in spite of using Fourier

[1] This chapter has been quoted from:
https://en.wikipedia.org/wiki/Number_theory#Elementary_tools

analysis, rather than complex analysis as such. Here as elsewhere, an elementary proof may be longer and more difficult for most readers than a non-elementary one.

12.2.2 Analytic number theory

Some subjects generally considered to be part of analytic number theory, for example, sieve theory, are better covered by the second rather than the first definition: some of sieve theory, for instance, uses little analysis, yet it does belong to analytic number theory.

The following are examples of problems in analytic number theory: the prime number theorem, the Goldbach conjecture (or the twin prime conjecture, or the Hardy-Littlewood conjectures), the Waring problem and the Riemann hypothesis. Some of the most important tools of analytic number theory are the circle method, sieve methods and L-functions (or, rather, the study of their properties). The theory of modular forms (and, more generally, automorphic forms) also occupies an increasingly central place in the toolbox of analytic number theory.

12.2.3 Algebraic number theory

An algebraic number is any complex number that is a solution to some polynomial equation $f(x) = 0$ with rational coefficients; for example, every solution x of $x^5 + \frac{11}{2}x^3 - 7x^2 + 9 = 0$ (say) is an algebraic number. Fields of algebraic numbers are also called algebraic number fields, or shortly number fields. Algebraic number theory studies algebraic number fields. Thus, analytic and algebraic number theory can and do overlap: the former is defined by its methods, the latter by its objects of study.

It could be argued that the simplest kind of number fields (viz., quadratic fields) were already studied by Gauss, as the discussion of quadratic forms in Disquisitiones arithmeticae can be restated in terms of ideals and norms in quadratic fields. (A quadratic field consists of all numbers of the form $a + b\sqrt{d}$, where a and b are rational numbers and d is a fixed rational number whose square root is not rational.) For that matter, the 11th-century chakravala method amounts-in modern terms-to an algorithm for finding the units of a real quadratic number field. However, neither Bhāskara nor Gauss knew of number fields as such.

12.2.4 Diophantine geometry

The central problem of Diophantine geometry is to determine when a Diophantine equation has solutions, and if it does, how many. The approach taken is to think of the solutions of an equation as a geometric object.

For example, an equation in two variables defines a curve in the plane. More generally, an equation, or system of equations, in two or more variables defines a curve, a surface or some other such object in n-dimensional space. In Diophantine geometry, one asks whether there are any rational points (points all of whose coordinates are rationals) or integral points (points all of whose coordinates are integers) on the curve or surface. If there are any such points, the next step is to ask how many there are and how they are distributed. A basic question in this direction is if there are finitely or infinitely many rational points on a given curve (or surface).

12.2.5 Probabilistic number theory

Much of probabilistic number theory can be seen as an important special case of the study of variables that are almost, but not quite, mutually independent. For example, the event that a random integer between one and a million be divisible by two and the event that it be divisible by three are almost independent, but not quite.

It is sometimes said that probabilistic combinatorics uses the fact that whatever happens with probability greater than 0 must happen sometimes; one may say with equal justice that many applications of probabilistic number theory hinge on the fact that whatever is unusual must be rare. If certain algebraic objects (say, rational or integer solutions to certain equations) can be shown to be in the tail of certain sensibly defined distributions, it follows that there must be few of them; this is a very concrete non-probabilistic statement following from a probabilistic one.

12.2.6 Computational number theory

While the word algorithm goes back only to certain readers of al-Khwārizmī, careful descriptions of methods of solution are older than proofs: such methods (that is, algorithms) are as old as any

recognisable mathematics-ancient Egyptian, Babylonian, Vedic, Chinese-whereas proofs appeared only with the Greeks of the classical period.

An interesting early case is that of what we now call the Euclidean algorithm. In its basic form (namely, as an algorithm for computing the greatest common divisor) it appears as Proposition 2 of Book VII in Elements, together with a proof of correctness. However, in the form that is often used in number theory (namely, as an algorithm for finding integer solutions to an equation $ax + by = c$, or, what is the same, for finding the quantities whose existence is assured by the Chinese remainder theorem) it first appears in the works of Aryabhata (5th-6th century CE) as an algorithm called kuttaka ("pulverise"), without a proof of correctness.

12.3 Modular arithmetic

[2] In mathematics, modular arithmetic is a system of arithmetic for integers, where numbers "wrap around" when reaching a certain value, called the modulus. The modern approach to modular arithmetic was developed by Carl Friedrich Gauss in his book Disquisitiones Arithmeticae, published in 1801.

A familiar use of modular arithmetic is in the 12-hour clock, in which the day is divided into two 12-hour periods. If the time is 7:00 now, then 8 hours later it will be 3:00. Usual addition would suggest that the later time should be $7 + 8 = 15$, but this is not the answer because clock time "wraps around" every 12 hours. Because the hour number starts over after it reaches 12, this is arithmetic modulo 12. In terms of the definition below, 15 is congruent to 3 modulo 12, so the (military) time called "15:00" has the equivalent clock form"3:00".

12.3.1 Definition of congruence relation

Modular arithmetic can be handled mathematically by introducing a congruence relation on the integers that is compatible with

[2] This section has been quoted from:
https://en.wikipedia.org/wiki/Modular_arithmetic

the operations on integers: addition, subtraction, and multiplication. For a positive integer n, two numbers a and b are said to be congruent modulo n, if their difference $a - b$ is an integer multiple of n (that is, if there is an integer k such that $a - b = kn$). This congruence relation is typically considered when a and b are integers, and is denoted:

$$a \equiv b \pmod{n}$$

The parentheses mean that (mod n) applies to the entire equation, not just to the right-hand side (here b). Sometimes, $=$ is used instead of $\equiv$; in this case, if the parentheses are omitted, this generally means that "mod" denotes the modulo operation applied to the righthand side, and the equality implies thus that $0 \leq a < n$. The number n is called the modulus of the congruence.

12.3.2 Applications

In theoretical mathematics, modular arithmetic is one of the foundations of number theory, touching on almost every aspect of its study, and it is also used extensively in group theory, ring theory, knot theory, and abstract algebra. In applied mathematics, it is used in computer algebra, cryptography, computer science, chemistry and the visual and musical arts.

A very practical application is to calculate checksums within serial number identifiers. For example, International Standard Book Number (ISBN) uses modulo 11 (for 10 digit ISBN) or modulo 10 (for 13 digit ISBN) arithmetic for error detection. Likewise, International Bank Account Numbers (IBANs), for example, make use of modulo 97 arithmetic to spot user input errors in bank account numbers. In chemistry, the last digit of the CAS registry number (a unique identifying number for each chemical compound) is a check digit, which is calculated by taking the last digit of the first two parts of the CAS registry number times 1, the previous digit times 2, the previous digit times 3 etc., adding all these up and computing the sum modulo 10.

In cryptography, modular arithmetic directly underpins public key systems such as RSA and Diffie-Hellman, and provides finite fields which underlie elliptic curves, and is used in a variety of

symmetric key algorithms including Advanced Encryption Standard (AES), International Data Encryption Algorithm (IDEA), and RC4. RSA and Diffie-Hellman use modular exponentiation.

12.4 Exercises

1. Translate the following text.

A continued fraction is an expression obtained through an iterative process of representing a number as the sum of its integer part and the reciprocal of another number, then writing this other number as the sum of its integer part and another reciprocal, and so on. In a finite continued fraction (or terminated continued fraction), the iteration/recursion is terminated after finitely many steps by using an integer in lieu of another continued fraction. In contrast, an infinite continued fraction is an infinite expression. In either case, all integers in the sequence, other than the first, must be positive. The integers a_i are called the coefficients or terms of the continued fraction.

Continued fractions have a number of remarkable properties related to the Euclidean algorithm for integers or real numbers. Every rational number $\frac{p}{q}$ has two closely related expressions as a finite continued fraction, whose coefficients a_i can be determined by applying the Euclidean algorithm to (p, q). The numerical value of an infinite continued fraction is irrational; it is defined from its infinite sequence of integers as the limit of a sequence of values for finite continued fractions. Each finite continued fraction of the sequence is obtained by using a finite prefix of the infinite continued fraction's defining sequence of integers. Moreover, every irrational number α is the value of a unique infinite continued fraction, whose coefficients can be found using the non-terminating version of the Euclidean algorithm applied to the incommensurable values α and 1. This way of expressing real numbers (rational and irrational) is called their continued fraction representation.

2. Use the correct form of the word.

1. Number theory is a vast and fascinating field of mathematics, sometimes called "higher arithmetic," (consist) .....of the study of the properties of whole numbers.

2. The Euclidean algorithm is an (effect) ..... method for computing the greatest common divisor of two integers, without explicitly factoring the two integers.
3. Sum of squares theorems are theorems in additive number theory concerning the (express) ..... of integers as sums of squares of other integers.
4. Sum of squares theorems have found various (apply) .....s in applied number theory, such as cryptography and integer factoring algorithms.
5. Arithmetic functions are real- or complex-valued functions defined on the set $\mathbb{Z}^+$ of positive integers. They (description) ..... arithmetic properties of numbers and are widely used in the field of number theory.

3. Fill gaps with one of the following words

1.geometrical	2. states	3. figurate	4. equation
5. congruences	6. possess	7. factored	8. square
9. dealing	10. residue	11.finding	12. ranges

1. A Diophantine equation is an ..... in which only integer solutions are allowed.
2. Arithmetic is the branch of mathematics ..... with integers or, more generally, numerical computation.
3. The fundamental theorem of arithmetic .....that every positive integer (except the number 1) can be represented in exactly one way apart from rearrangement as a product of one or more primes.
4. An abnormal number is a hypothetical number which can be ..... into primes in more than one way.
5. Interval arithmetic is the arithmetic of quantities that lie within specified ..... (i.e., intervals) instead of having definite known values.
6. Modular arithmetic is the arithmetic of ....., sometimes known informally as "clock arithmetic."
7. The word .....is used in a number of different contexts in mathematics. Two of the most common uses are the complex residue of a pole, and the remainder of a congruence.

8. A cubic number is a ..... number of the form n^3 with n a positive integer.
9. A square number, also called a perfect ....., is a figurate number of the form $S_n = n^2$, where n is an integer.
10. A figurate number, also known as a figural number, is a number that can be represented by a regular ..... arrangement of equally spaced points.
11. A set of real numbers $x_1, \ldots, x_n$ is said to ..... an integer relation if there exist integers a_i such that $a_1x_1 + a_2x_2 + ... + a_nx_n = 0$, with not all $a_i = 0$.
12. The subset sum problem is the problem of ..... what subset of a list of integers has a given sum, which is an integer relation problem where the relation coefficients a_i are 0 or 1.

13

Famous problems in Math history

13.1 Fermat's Last Theorem

[1]Pierre de Fermat died in 1665. Today we think of Fermat as a number theorist, in fact as perhaps the most famous number theorist who ever lived. It is therefore surprising to find that Fermat was in fact a lawyer and only an amateur mathematician. Also surprising is the fact that he published only one mathematical paper in his life, and that was an anonymous article written as an appendix to a colleague's book.

Because Fermat refused to publish his work, his friends feared that it would soon be forgotten unless something was done about it. His son, Samuel undertook the task of collecting Fermat's letters and other mathematical papers, comments written in books, etc. with the object of publishing his father's mathematical ideas. In this way the famous 'Last theorem' came to be published. It was found by Samuel written as a marginal note in his father's copy of Diophantus's Arithmetica.

Fermat's Last Theorem states that

$$x^n + y^n = z^n$$

has no non-zero integer solutions for x, y and z when $n > 2$. Fermat wrote

> I have discovered a truly remarkable proof which this margin is too small to contain.

[1] This section has been quoted from:
http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Fermat's_last_theorem.html

Fermat almost certainly wrote the marginal note around 1630, when he first studied Diophantus's Arithmetica. It may well be that Fermat realized that his remarkable proof was wrong, however, since all his other theorems were stated and restated in challenge problems that Fermat sent to other mathematicians. Although the special cases of $n = 3$ and $n = 4$ were issued as challenges (and Fermat did know how to prove these) the general theorem was never mentioned again by Fermat.

In fact in all the mathematical work left by Fermat there is only one proof. Fermat proves that the area of a right triangle cannot be a square. Clearly this means that a rational triangle cannot be a rational square. In symbols, there do not exist integers x, y, z with $x^2 + y^2 = z^2$ such that $xy/2$ is a square. From this it is easy to deduce the $n = 4$ case of Fermat's theorem.

It is worth noting that at this stage it remained to prove Fermat's Last Theorem for odd primes n only. For if there were integers x, y, z with $x^n + y^n = z^n$ then if $n = pq$,

$$(x^q)^p + (y^q)^p = (z^q)^p.$$

Euler wrote to Goldbach on 4 August 1753 claiming he had a proof of Fermat's Theorem when $n = 3$. However his proof in Algebra (1770) contains a fallacy and it is far from easy to give an alternative proof of the statement which has the fallacious proof. There is an indirect way of mending the whole proof using arguments which appear in other proofs of Euler so perhaps it is not too unreasonable to attribute the $n = 3$ case to Euler.

Euler's mistake is an interesting one, one which was to have a bearing on later developments. He needed to find cubes of the form $p^2 + 3q^2$ and Euler shows that, for any a, b if we put $p = a^3 - 9ab^2$, $q = 3(a^2b - b^3)$ then $p^2 + 3q^2 = (a^2 + 3b^2)^3$.

This is true but he then tries to show that, if $p^2 + 3q^2$ is a cube then an a and b exist such that p and q are as above. His method is imaginative, calculating with numbers of the form $a+b\sqrt{3}$. However numbers of this form do not behave in the same way as the integers, which Euler did not seem to appreciate.

The next major step forward was due to Sophie Germain. A special case says that if n and $2n+1$ are primes then $x^n + y^n = z^n$ implies that one of x, y, z is divisible by n. Hence Fermat's Last Theorem splits into two cases.

Case 1: None of x, y, z is divisible by n.
Case 2: One and only one of x, y, z is divisible by n.

Sophie Germain proved Case 1 of Fermat's Last Theorem for all n less than 100 and Legendre extended her methods to all numbers less than 197. At this stage, Case 2 had not been proved for even $n = 5$ so it became clear that Case 2 was the one on which to concentrate. Now Case 2 for $n = 5$ itself splits into two. One of x, y, z is even and one is divisible by 5. Case 2(i) is when the number divisible by 5 is even; Case 2(ii) is when the even number and the one divisible by 5 are distinct.

Case 2(i) was proved by Dirichlet and presented to the Paris Académie des Sciences in July 1825. Legendre was able to prove Case 2(ii) and the complete proof for $n = 5$ was published in September 1825. In fact Dirichlet was able to complete his own proof of the $n = 5$ case with an argument for Case 2(ii) which was an extension of his own argument for Case 2(i).

In 1832 Dirichlet published a proof of Fermat's Last Theorem for $n = 14$. Of course he had been attempting to prove the $n = 7$ case but had proved a weaker result. The $n = 7$ case was finally solved by Lamé in 1839. It showed why Dirichlet had so much difficulty, for although Dirichlet's $n = 14$ proof used similar (but computationally much harder) arguments to the earlier cases, Lamé had to introduce some completely new methods. Lamé's proof is exceedingly hard and makes it look as though progress with Fermat's Last Theorem to larger n would be almost impossible without some radically new thinking.

The year 1847 is of major significance in the study of Fermat's Last Theorem. On 1 March of that year Lamé announced to the Paris Académie that he had proved Fermat's Last Theorem. He sketched a proof which involved factorizing $x^n + y^n = z^n$ into linear factors over the complex numbers. Lamé acknowledged that the idea was suggested to him by Liouville. However, Liouville addressed the meeting after Lamé and suggested that the problem of this approach was that uniqueness of factorization into primes was needed for these complex numbers and he doubted if it were true. Cauchy supported Lamé but, in rather typical fashion, pointed out that he had reported to the October 1847 meeting of the Académie an idea which he believed might prove Fermat's Last Theorem.

Much work was done in the following weeks in attempting to prove the uniqueness of factorization. Wantzel claimed to have proved it on 15 March but his argument

> It is true for $n = 2$, $n = 3$ and $n = 4$ and one easily sees that the same argument applies for $n > 4$

was somewhat hopeful. [Wantzel is correct about $n = 2$ (ordinary integers), $n = 3$ (the argument Euler got wrong) and $n = 4$ (which was proved by Gauss).]

On 24 May Liouville read a letter to the Académie which settled the arguments. The letter was from Kummer, enclosing an offprint of a 1844 paper which proved that uniqueness of factorization failed but could be 'recovered' by the introduction of ideal complex numbers which he had done in 1846. Kummer had used his new theory to find conditions under which a prime is regular and had proved Fermat's Last Theorem for regular primes. Kummer also said in his letter that he believed 37 failed his conditions.

By September 1847 Kummer sent to Dirichlet and the Berlin Academy a paper proving that a prime p is regular (and so Fermat's Last Theorem is true for that prime) if p does not divide the numerators of any of the Bernoulli numbers $B_2, B_4, \ldots, B_{p-3}$. The Bernoulli number B_i is defined by

$$\frac{x}{e^x - 1} = \sum_{i-0}^{\infty} \frac{B_i x^i}{i!}.$$

Kummer shows that all primes up to 37 are regular but 37 is not regular as 37 divides the numerator of B_{32}.

The only primes less than 100 which are not regular are 37, 59 and 67. More powerful techniques were used to prove Fermat's Last Theorem for these numbers. This work was done and continued to larger numbers by Kummer, Mirimanoff, Wieferich, Furtwängler, Vandiver and others. Although it was expected that the number of regular primes would be infinite even this defied proof. In 1915 Jensen proved that the number of irregular primes is infinite.

Despite large prizes being offered for a solution, Fermat's Last Theorem remained unsolved. It has the dubious distinction of being the theorem with the largest number of published false proofs. For example over 1000 false proofs were published between 1908

and 1912. The only positive progress seemed to be computing results which merely showed that any counter-example would be very large. Using techniques based on Kummer's work, Fermat's Last Theorem was proved true, with the help of computers, for n up to $4,000,000$ by 1993.

In 1983 a major contribution was made by Gerd Faltings who proved that for every $n > 2$ there are at most a finite number of coprime integers x, y, z with $x^n + y^n = z^n$. This was a major step but a proof that the finite number was 0 in all cases did not seem likely to follow by extending Faltings' arguments.

The final chapter in the story began in 1955, although at this stage the work was not thought of as connected with Fermat's Last Theorem. Yutaka Taniyama asked some questions about elliptic curves, i.e. curves of the form $y^2 = x^3 + ax + b$ for constants a and b. Further work by Weil and Shimura produced a conjecture, now known as the Shimura-Taniyama-Weil Conjecture. In 1986 the connection was made between the Shimura-Taniyama-Weil Conjecture and Fermat's Last Theorem by Frey at Saarbrücken showing that Fermat's Last Theorem was far from being some unimportant curiosity in number theory but was in fact related to fundamental properties of space.

Further work by other mathematicians showed that a counter-example to Fermat's Last Theorem would provide a counter-example to the Shimura-Taniyama-Weil Conjecture. The proof of Fermat's Last Theorem was completed in 1993 by Andrew Wiles, a British mathematician working at Princeton in the USA. Wiles gave a series of three lectures at the Isaac Newton Institute in Cambridge, England the first on Monday 21 June, the second on Tuesday 22 June. In the final lecture on Wednesday 23 June 1993 at around 10.30 in the morning Wiles announced his proof of Fermat's Last Theorem as a corollary to his main results. Having written the theorem on the blackboard he said I will stop here and sat down. In fact Wiles had proved the Shimura-Taniyama-Weil Conjecture for a class of examples, including those necessary to prove Fermat's Last Theorem.

This, however, is not the end of the story. On 4 December 1993 Andrew Wiles made a statement in view of the speculation. He said that during the reviewing process a number of problems had

emerged, most of which had been resolved. However one problem remains and Wiles essentially withdrew his claim to have a proof. He states

> The key reduction of (most cases of) the Taniyama-Shimura conjecture to the calculation of the Selmer group is correct. However the final calculation of a precise upper bound for the Selmer group in the semisquare case (of the symmetric square representation associated to a modular form) is not yet complete as it stands. I believe that I will be able to finish this in the near future using the ideas explained in my Cambridge lectures.

In March 1994 Faltings, writing in Scientific American, said

> If it were easy, he would have solved it by now. Strictly speaking, it was not a proof when it was announced.

Weil, also in Scientific American, wrote

> I believe he has had some good ideas in trying to construct the proof but the proof is not there. To some extent, proving Fermat's Theorem is like climbing Everest. If a man wants to climb Everest and falls short of it by 100 yards, he has not climbed Everest.

In fact, from the beginning of 1994, Wiles began to collaborate with Richard Taylor in an attempt to fill the holes in the proof. However they decided that one of the key steps in the proof, using methods due to Flach, could not be made to work. They tried a new approach with a similar lack of success. In August 1994 Wiles addressed the International Congress of Mathematicians but was no nearer to solving the difficulties.

Taylor suggested a last attempt to extend Flach's method in the way necessary and Wiles, although convinced it would not work, agreed mainly to enable him to convince Taylor that it could never work. Wiles worked on it for about two weeks, then suddenly inspiration struck.

> In a flash, I saw that the thing that stopped it [the extension of Flach's method] working was something that would make another method I had tried previously work.

On 6 October Wiles sent the new proof to three colleagues including Faltings. All liked the new proof which was essentially simpler than the earlier one. Faltings sent a simplification of part of the proof.

No proof of the complexity of this can easily be guaranteed to be correct, so a very small doubt will remain for some time. However when Taylor lectured at the British Mathematical Colloquium in Edinburgh in April 1995 he gave the impression that no real doubts remained over Fermat's Last Theorem.

13.2 The Four Color Problem

[2] The Four Colour Conjecture first seems to have been made by Francis Guthrie. He was a student at University College London where he studied under De Morgan. After graduating from London he studied law but by this time his brother Frederick Guthrie had become a student of De Morgan. Francis Guthrie showed his brother some results he had been trying to prove about the coloring of maps and asked Frederick to ask De Morgan about them.

De Morgan was unable to give an answer but, on 23 October 1852, the same day he was asked the question, he wrote to Hamilton in Dublin. De Morgan wrote:

> A student of mine asked me today to give him a reason for a fact which I did not know was a fact - and do not yet. He says that if a figure be anyhow divided and the compartments differently colored so that figures with any portion of common boundary line are differently coloured - four colours may be wanted, but not more - the following is the case in which four colors are wanted. Query cannot a necessity for five or more be invented. ..... If you retort with some very simple case which makes me out a stupid animal, I think I must do as the Sphynx did....

Hamilton replied on 26 October 1852 (showing the efficiency of both himself and the postal service):

[2] This section has been quoted from:
http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/The_four_colour_theorem.html

> I am not likely to attempt your quaternion of color very soon.

Before continuing with the history of the Four Color Conjecture we will complete details of Francis Guthrie. After practising as a barrister he went to South Africa in 1861 as a Professor of Mathematics. He published a few mathematical papers and became interested in botany. A heather (Erica Guthriei) is named after him.

De Morgan kept asking if anyone could find a solution to Guthrie's problem and several mathematicians worked on it. Charles Peirce in the USA attempted to prove the Conjecture in the 1860's and he was to retain a lifelong interest in the problem. Cayley also learnt of the problem from De Morgan and on 13 June 1878 he posed a question to the London Mathematical Society asking if the Four Color Conjecture had been solved. Shortly afterwards Cayley sent a paper "On the coloring of maps" to the Royal Geographical Society and in was published in 1879. The paper explains where the difficulties lie in attempting to prove the conjecture.

On 17 July 1879 Alfred Bray Kempe announced in Nature that he had a proof of the Four Color Conjecture. Kempe was a London barrister who had studied mathematics under Cayley at Cambridge and devoted some of his time to mathematics throughout his life. At Cayley's suggestion Kempe submitted the Theorem to the American Journal of Mathematics where it was published in 1879. Story read the paper before publication and made some simplifications. Story reported the proof to the Scientific Association of Johns Hopkins University in November 1879 and Charles Peirce, who was at the November meeting, spoke at the December meeting of the Association of his own work on the Four Color Conjecture.

Kempe used an argument known as the method of Kempe chains. If we have a map in which every region is colored red, green, blue or yellow except one, say X. If this final region X is not surrounded by regions of all four colors there is a colour left for X. Hence suppose that regions of all four colours surround X. If X is surrounded by regions A, B, C, D in order, colored red, yellow, green and blue then there are two cases to consider.

(i) There is no chain of adjacent regions from A to C alternately colored red and green.
(ii) There is a chain of adjacent regions from A to C alternately colored red and green.

If (i) holds there is no problem. Change A to green, and then interchange the colour of the red/green regions in the chain joining A. Since C is not in the chain it remains green and there is now no red region adjacent to X. Colour X red.

If (ii) holds then there can be no chain of yellow/blue adjacent regions from B to D. [It could not cross the chain of red/green regions.] Hence property (i) holds for B and D and we change colours as above.

Kempe received great acclaim for his proof. He was elected a Fellow of the Royal Society and served as its treasurer for many years. He was knighted in 1912. He published two improved versions of his proof, the second in 1880 aroused the interest of P. G. Tait, the Professor of Natural Philosophy at Edinburgh. Tait addressed the Royal Society of Edinburgh on the subject and published two papers on the (what we should now call) Four Colour Theorem. They contain some clever ideas and a number of basic errors.

The Four Colour Theorem returned to being the Four Colour Conjecture in 1890. Percy John Heawood, a lecturer at Durham England, published a paper called Map Colouring Theorem. In it he states that his aim is

> rather destructive than constructive, for it will be shown that there is a defect in the now apparently recognized proof.

Although Heawood showed that Kempe's proof was wrong he did prove that every map can be 5-coloured in this paper. Kempe reported the error to the London Mathematical Society himself and said he could not correct the mistake in his proof. In 1896 de la Vallée Poussin also pointed out the error in Kempe's paper, apparently unaware of Heawood's work.

Heawood was to work throughout his life on map colouring, work which spanned nearly 60 years. He successfully investigated the number of colours needed for maps on other surfaces and gave

what is known as the Heawood estimate for the necessary number in terms of the Euler characteristic of the surface.

Heawood's other claim to fame is raising money to restore Durham Castle as Secretary of the Durham Castle Restoration Fund. For his perseverance in raising the money to save the Castle from sliding down the hill on which it stands Heawood received the O.B.E.

Heawood was to make further contributions to the Four Colour Conjecture. In 1898 he proved that if the number of edges around each region is divisible by 3 then the regions are 4-colourable. He then wrote many papers generalizing this result.

To understand the later work we need to define some concepts.

Clearly a graph can be constructed from any map the regions being represented by the vertices and two vertices being joined by an edge if the regions corresponding to the vertices are adjacent. The resulting graph is planar, that is can be drawn in the plane without any edges crossing. The Four Colour Conjecture now asks if the vertices of the graph can be coloured with 4 colours so that no two adjacent vertices are the same colour.

From the graph, a triangulation can be obtained by adding edges to divide any non-triangular face into triangles. A configuration is part of a triangulation contained within a circuit. An unavoidable set is a set of configurations with the property that any triangulation must contain one of the configurations in the set. A configuration is reducible if it cannot be contained in a triangulation of the smallest graph which cannot be 4-coloured.

The search for avoidable sets began in 1904 with work of Weinicke. Renewed interest in the USA was due to Veblen who published a paper in 1912 on the Four Colour Conjecture generalising Heawood's work. Further work by G. D. Birkhoff introduced the concept of reducibility (defined above) on which most later work rested.

Franklin in 1922 published further examples of unavoidable sets and used Birkhoff's idea of reducibility to prove, among other results, that any map with ≤ 25 regions can be 4-coloured. The number of regions which resulted in a 4-colourable map was slowly increased. Reynolds increased it to 27 in 1926, Winn to 35 in 1940, Ore and Stemple to 39 in 1970 and Mayer to 95 in 1976.

However the final ideas necessary for the solution of the Four Colour Conjecture had been introduced before these last two results. Heesch in 1969 introduced the method of discharging. This consists of assigning to a vertex of degree i the charge $6 - i$. Now from Euler's formula we can deduce that the sum of the charges over all the vertices must be 12. A given set S of configurations can be proved unavoidable if for a triangulation T which does not contain a configuration in S we can redistribute the charges (without changing the total charge) so that no vertex ends up with a positive charge.

Heesch thought that the Four Colour Conjecture could be solved by considering a set of around 8900 configurations. There were difficulties with his approach since some of his configurations had a boundary of up to 18 edges and could not be tested for reducibility. The tests for reducibility used Kempe chain arguments but some configurations had obstacles to prevent reduction.

The year 1976 saw a complete solution to the Four Colour Conjecture when it was to become the Four Colour Theorem for the second, and last, time. The proof was achieved by Appel and Haken, basing their methods on reducibility using Kempe chains. They carried through the ideas of Heesch and eventually they constructed an unavoidable set with around 1500 configurations. They managed to keep the boundary ring size down to 14, making computations easier that for the Heesch case. There was a long period where they essentially used trial and error together with unbelievable intuition to modify their unavoidable set and their discharging procedure. Appel and Haken used 1200 hours of computer time to work through the details of the final proof. Koch assisted Appel and Haken with the computer calculations.

The Four Colour Theorem was the first major theorem to be proved using a computer, having a proof that could not be verified directly by other mathematicians. Despite some worries about this initially, independent verification soon convinced everyone that the Four Colour Theorem had finally been proved. Details of the proof appeared in two articles in 1977. Recent work has led to improvements in the algorithm.

Index