First Course in Operations Research

Content:

- Problem formulation of operation research
- Modeling Life and scientific problems
- One-dimensional minimization methods
- Linear programming problem
- Geometry of linear programming problems
- Classical optimization techniques
- Multi-variable optimization with no constrains
- Multi-variable optimization with equality constraints
- Unconstrained multi-variable optimization methods
    - Direct methods
    - Indirect methods

Recommended references

1. Wenyu Sun and Ya-Xiang Yuan, Optimization theory and methods: nonlinear programming, 2006 Springer Science & Business Media, LLC.
2. Hamdy Taha, Operations research: An introduction (Eight Edition), 2006, Prentice Hall.
3. Frederick S. Hillier and Gerald J. Lieberman, Introduction to operations research, Seventh edition, 2000, McGraw-Hill.

Chapter 1

Introduction

Operation research, for short OR, is the act of obtaining the best result under given circumstances. Thus, we may have several solutions for a certain problem and our aim is to find the best solution among those solutions which leads to the presentation of the optimization problem.

Problem Formulation

1. Define the quantity to be maximize or minimize.
   This quantity is called objective function.
2. Define the constraints
   Those are the restriction under which we have to solve our problem.
3. Define the non-negative constraints
   We have to be sure that all the variables are of non-negative type. If this is not the case, then we have to modify them as we will see later on in our study.

Examples

Example1:

The Haty shop makes its sandwiches from a combination beef and goat meat. The beef contains 80% meat and 20% fat, and it costs 24 pounds per kilo. The goat meat contains 68% meat and 32% fat, and it costs 18 pounds per kilo. What is the amount of meat from each type must be used in each kilo of meat if it wants to minimize its costs and keep the ratio of fat so that no more than 25%?

Solution 1:

Let $x_1$weight of beef meat and $x_2$weight of goat meat

Objective function is

minimize $z = 24x_1 + 18x_2$

The constrains

(1) Rate of fat

$$0.20x_1 + 0.32x_2 \leq 0.25$$

(2) Per kilo

$$x_1 + x_2 = 1$$

Non-negative condition

$$x_1 \geq 0, x_2 \geq 0$$

Thus, the final formula for the linear programming problem is

Minimize $z = 24x_1 + 18x_2$

Subject to

$$0.20x_1 + 0.32x_2 \leq 0.25$$

$$x_1 + x_2 = 1$$

$$x_1 \geq 0. x_2 \geq 0$$

Example2:

A factory wants in the production of 2 models. The first one needs 3 units of wood; and 3 units of iron; 5 units of aluminum, models II needs a single unit of wood; 8 units of iron; 4 units of aluminum. If you know that the maximum available of wood is 53 units, Steel 127 and 100 for aluminum. Form the mathematical model in the following cases

A - If the first model is given a profit of unit and the second 2 units.

B – If the first model gives a profit of two units and the second gives a single unit.

Solution 2:

Let the factory produce x unit of 1st one and y from the 2nd.

Objective function

(a) $Max\ Z = x + 2y$ (b) $Max\ Z = 2x + y$

and the constraints are

For wood;

$$3x + y \leq 53$$

For iron;

$$3x + 8y \leq 127$$

For Aluminum;

$$5x + 4y \leq 100$$

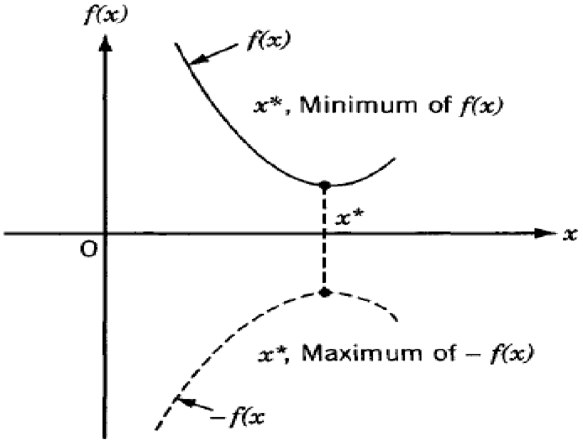
Non-negative condition

$$x \geq 0\ ,\quad y \geq 0$$

Nonlinear Programming I:
One-Dimensional Minimization Methods

1. Introduction

It can be seen from the blow figure that if a point $x*$ corresponds to the minimum value of a function $f(x)$, the same point also corresponds to the maximum value of the negative of the function, - $f(x)$.



Optimization can be taken to mean minimization since the maximum of a function can be found by seeking the minimum of the negative of the same function.

2. Statement of an optimization problem

An optimization or a mathematical programming problem can be stated as follows.

$$\text{Find } \mathbf{X} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix} \text{ which minimizes } f(\mathbf{X})$$

subject to the constraints

$$g_j(\mathbf{X}) \leq 0, \quad j = 1,2,\ldots,m$$
$$l_j(\mathbf{X}) = 0, \quad j = 1,2,\ldots,p \tag{1.1}$$

Where, X is an n-dimensional vector called the design vector, $f(X)$ is termed the objective function, and $g_j(X)$ and $l_j(X)$ are known as inequality and equality constraints, respectively.

- The problem stated in Eq. (1.1) is called a constrained optimization problem.

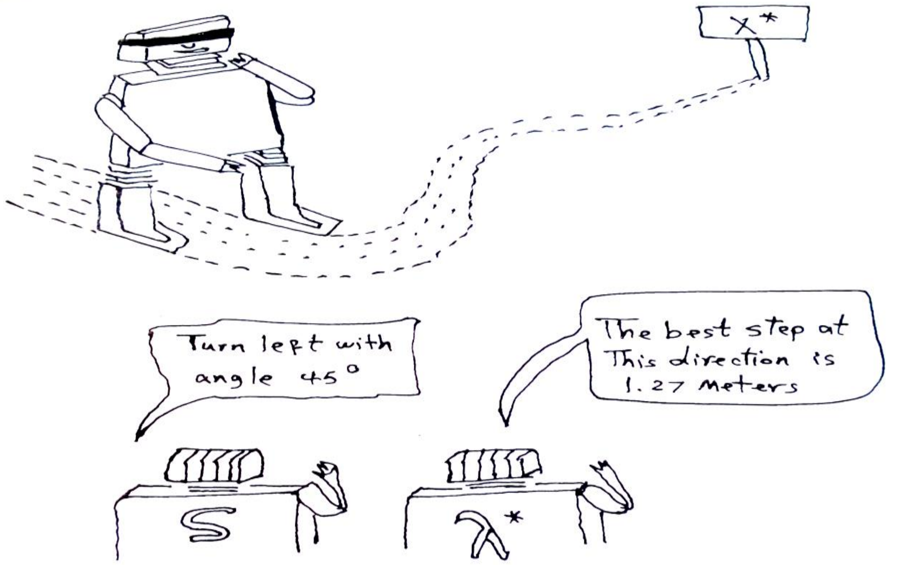
<u>The algorithm that treats a nonlinear programming problem.</u>

1. Start with an initial trial point $\mathbf{X}_1$.
2. Find a suitable direction $\mathbf{S}_i$ ($i = 1$ to start with) which points in the general direction of the optimum.
3. Find an appropriate step length λ_i^* for movement along the direction $\mathbf{S}_i$.
4. Obtain the new approximation $\mathbf{X}_{i+1}$ as

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \lambda_i^* \mathbf{S}_i$$

5. Test whether $\mathbf{X}_{i+1}$ is optimum. If $\mathbf{X}_{i+1}$ is optimum, stop the procedure. Otherwise, set a new $i = i + 1$ and repeat step (2) onward.

- From this algorithm, we conclude that finding a minimum of single variable objective function is an important step (step3) in solving unconstrained multivariable optimization problem. So we start with studying unconstrained single optimization problem



Theorem 1: Necessary Condition

If a function f(x) is defined in the interval *a* <*x* <*b* and have a relative minimum at $x = x^*$, where *a* < x^* <*b,* and if the derivative *df(x)/dx* = $f'(x^*)$ exists as a finite number at $x = x^*$, then $f'(x^*)$ = 0.

Theorem 2: Sufficient Condition:

Let $f\,'(x^*) = f\,''(x^*) = \ldots = f^{(n-1)}(x^*) = 0$, but $f^{(n)}(x^*) \neq 0$. Then $f(x^*)$ is

(i) A minimum value of $f(x^*)$ if $f^{(n)}(x^*) > 0$ and *n* is even;

(ii) A maximum value of $f(x^*)$ if $f^{(n)}(x^*) < 0$ and *n* is even;

(iii) Neither a maximum nor a minimum if *n* is odd.

Example:

Use theorems 1 and 2 to find the optimum values of

$$f(x) = 12x^5 - 45x^4 + 40x^3 + 5$$

Answer:

$$f'(x) = 60x^4 - 3 * 60x^3 + 60 * 2 * x^2$$
$$= 60x^2(x^2 - 3x + 2)$$
$$= 60x^2(x - 1)(x - 2) = 0$$

The extreme points are $\boldsymbol{x = 0, x = 1}$ **and** $\boldsymbol{x = 2}$

$x = 0$	$x = 1$	$x = 2$
$f''(x) = 240x^3 - 540x^2 + 240x$ $f''(0) = 0$ We evaluate the nex derivative $f'''(x) = 3 * 240x^2 - 2 * 540x + 240$ $f'''(0) = +240$, Order of derivative i odd. So this point is neithe maximum nor minimun	$f''(1) = -60$ this point is relative maximum $f_{Max} = 12(1) - 45(1) + 40(1) + 5$ $=12$	$f''(2) = 240$ this point is relative minimum $f_{Min} = -11$

<u>Excercises 3:</u>

Find the maxima and minima, if any, of the functions

(a) $f(x) = \frac{x^4}{(x-1)(x-3)^3}$

(b) $f(x) = 4x^3 - 18x^2 + 27x - 7$

(c) $f(x) = 10x^6 - 48x^5 + 15x^4 + 200x^3 - 120x^2 - 480x + 100$

Answer:

A unimodal function is one that has only one peak in a given interval	T	F
A unimodal function is one that has several peaks in a given interval	T	F
In the **Interval halving method**, the function value at the middle point of the interval will be available in the stage except the first stage	T	F
The interval of uncertainty of the **Interval halving method** remaining at the end of n experiments) $n \geq 3$ and odd) is given by $$L_n = \left(\frac{1}{2}\right)^{\frac{(n-1)}{2}} L_0$$	T	F
the best value for the eliminating part of the interval in Fibonacci method assuming we conduct a large number of iterations is 0.38٢	T	F

Unconstrained single optimization problem

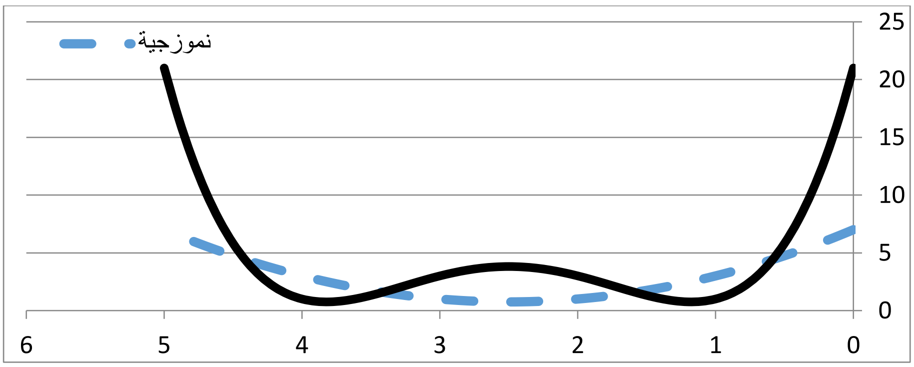
Analytical methods (differential calculus methods)	Numerical methods		
	Elimination methods	Interpolation methods	
		Requiring no derivatives	Requiring derivatives
	Unrestricted search	(quadratic)	Cubic
	Exhaustive search		Direct root
	Dichotomous search		Newton
	Fibonacci method		Quasi-Newton
	Golden section method		Secant

Unimodal function

A unimodal function is one that has only one peak in a given interval

A *unimodal function* is one that has only one peak (maximum) or valley (minimum) in a given interval. Thus a function of one variable is said to be *unimodal* if, given that two values of the variable are on the same side of the optimum, the one nearer the optimum gives the better functional value (i.e., the smaller value in the case of a minimization problem). This can be stated mathematically as follows:

A function $f(x)$ is unimodal if (i) $x_1 < x_2 < x^*$ implies that $f(x_2) < f(x_1)$, and (ii) $x_2 > x_1 > x^*$ implies that $f(x_1) < f(x_2)$, where x^* is the minimum point.



Elimination methods

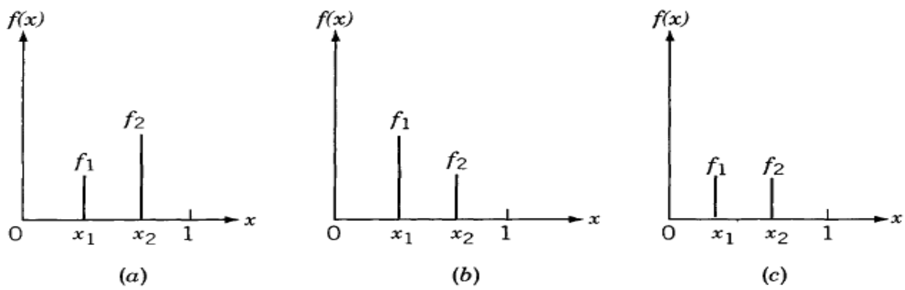
Unrestricted search

In the most practical problems, the optimum solution is known to lie within restricted ranges of the design variables. In some cases this range is not known, and hence the reach has to be made with no restrictions on the values of the variables.

Search with fixed step size

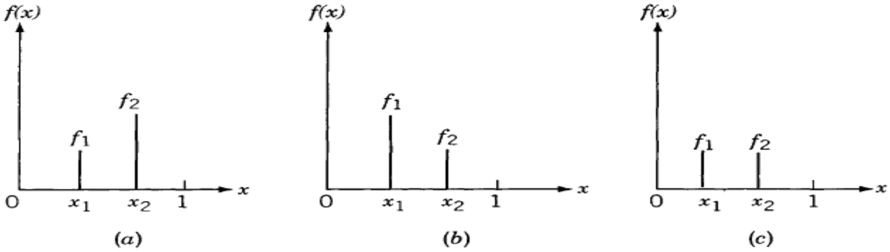
The most elementary approach for such a problem is to use a fixed step size and move from an initial guess points in a favorable direction (positive or negative). The step size used must be small in the relation to the final accuracy desired. Although this method is very simple to implement, it is not efficient in many cases. This method is described in the following steps:

1. Start with an initial guess point, say, x_1.
2. Find $f_1 = f(x_1)$.
3. Assuming a step size s, find $x_2 = x_1 + s$.
4. Find $f_2 = f(x_2)$.
5. If $f_2 < f_1$, and if the problem is one of minimization, the assumption of unimodality indicates that the desired minimum cannot lie at $x < x_1$.



Hence the search can be continued further along points $x_3, x_4, \ldots$ using the unimodality assumption while testing each pair of experiments. This procedure is continued until a point, $x_i = x_1 + (i - 1)s$, shows an increase in the function value.

6. The search is terminated at x_i, and either x_{i-1} or x_i can be taken as the optimum point.
7. Originally, if $f_2 > f_1$, the search should be carried in the reverse direction at points $x_{-2}, x_{-3}, \ldots$, where $x_{-j} = x_1 - (j - 1)s$.
8. If $f_2 = f_1$, the desired minimum lies in between x_1 and x_2, and the minimum point can be taken as either x_1 or x_2.
9. If it happens that both f_2 and f_{-2} are greater than f_1, it implies that the desired minimum will lie in the double interval $x_{-2} < x < x_2$.



Example:

Use unrestricted search with Fixed Step Size to Find the maximum of

$$f(x) = \begin{cases} \frac{1}{2}x, & x \le 2 \\ 3 - x, & x > 2 \end{cases}$$

by starting from $x_1 = 0$ **with an initial step size of 0.4.**

Solution: this problem corresponds to Find the minimum of

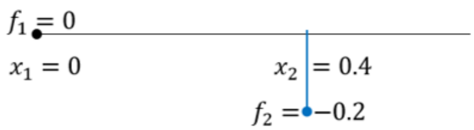
$$f(x) = \begin{cases} -0.5x; & x \le 2 \\ x - 3; & x > 2 \end{cases}$$

$x_1 = 0$, $f(x_1) = f(0) = 0 \; S = 0.4$

$$x_2 = x_1 + S = 0.4$$

$$f(x) = \begin{cases} -\frac{1}{2}x, & x \le 2 \\ 3 - x, & x > 2 \end{cases}$$

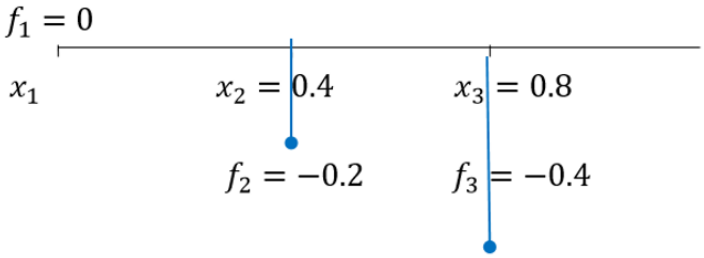
$$f(x_2) = f(0.4) = -\frac{1}{2}(0.4) = -0.2$$



$$x_3 = x_2 + S = 0.4 + 0.4 = 0.8$$

$$f(x_3) = f(0.8) = -\frac{1}{2}(0.8) = -0.4$$

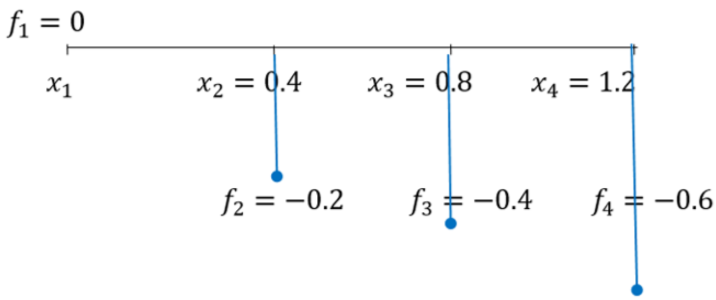
$$f(x) = \begin{cases} -\frac{1}{2}x, & x \le 2 \\ 3 - x, & x > 2 \end{cases}$$



$$x_4 = x_3 + S = 0.8 + 0.4 = 1.2$$

$$f(x_4) = f(1.2) = -\frac{1}{2}(1.2) = -0.6$$

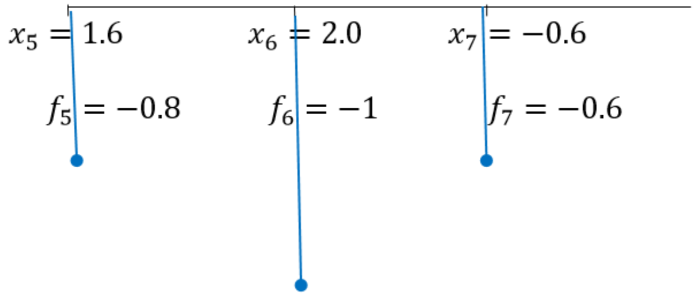
$$f(x) = \begin{cases} -\frac{1}{2}x, & x \le 2 \\ 3 - x, & x > 2 \end{cases}$$



$x_5 = x_4 + S = 1.2 + 0.4 = 1.6 \quad f(x_5) = f(1.6) = -\frac{1}{2}(1.6) = -0.8$

$x_6 = x_5 + S = 1.6 + 0.4 = 2.0 \quad f(x_6) = f(2.0) = -\frac{1}{2}(2.0) = -1$

$x_7 = x_6 + S = 2.0 + 0.4 = 2.4 \quad f(x_7) = f(2.4) = 2.4 - 3 = -0.6$



$\therefore\ x_6 = 2.0$ *is the minimum point and* $f(2.0) = -1$

Fibonacci method

As stated earlier, the *Fibonacci method* can be used to find the minimum of a function of one variable even if the function is not continuous. This method, like many other elimination methods, has the following limitations:

1. The initial interval of uncertainty, in which the optimum lies, has to be known.
2. The function being optimized has to be unimodal in the initial interval of uncertainty.
3. The exact optimum cannot be located in this method. Only an interval known as the *final interval of uncertainty* will be known. The final interval of uncertainty can be made as small as desired by using more computations.
4. The number of function evaluations to be used in the search or the resolution required has to be specified beforehand.

This method makes use of the sequence of Fibonacci numbers, $\{F_n\}$, for placing the experiments. These numbers are defined as

$$F_0 = F_1 = 1$$

$$F_n = F_{n-1} + F_{n-2}, \qquad n = 2,3,4,\ldots$$

which yield the sequence 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ...

Procedure. Let L_0 be the initial interval of uncertainty defined by $a \le x \le b$ and n be the total number of experiments to be conducted. Define

$$L_2^* = \frac{F_{n-2}}{F_n} L_0$$

and place the first two experiments at points x_1 and x_2, which are located at a distance of L_2^* from each end of L_0.[†] This gives[‡]

$$x_1 = a + L_2^* = a + \frac{F_{n-2}}{F_n} L_0$$

$$x_2 = b - L_2^* = b - \frac{F_{n-2}}{F_n} L_0 = a + \frac{F_{n-1}}{F_n} L_0$$

Discard part of the interval by using the unimodality assumption.

Example:

Use Fibonacci method to find the maximum of

$$f(x) = \begin{cases} \frac{1}{2}x, & x \le 2 \\ 3 - x, & x > 2 \end{cases}$$

by starting from [0,3] **with n=6**

Solution:

This problem corresponds to

Find the minimum of

$$f(x) = \begin{cases} -0.5x; & x \le 2 \\ x - 3; & x > 2 \end{cases}$$

the sequence of Fibonacci numbers, is

$$f_0 = f_1 = 1, \quad 1,1,2,3,5,8,13,21$$

$n = 6$ $\qquad f_0 = f_1 = 1, \qquad 1,1,2,3,5,8,13,21$

$$L^* = \frac{f_{n-2}}{f_n} L_o = \frac{f_4}{f_6}(3-0) = \frac{5}{13}(3) = 1.15$$

$$x_1 = a + L^* = 0 + 1.15 = 1.15$$

$$x_2 = b - L^* = 3 - 1.15 = 1.85$$

$a = 0 \qquad x_1 = 1.15 \qquad x_2 = 1.85 \qquad b = 3$

$f_a = 0 \qquad f_1 = -0.57 \qquad f_2 = -0.925 \qquad f_b = 0$

Discard part of the interval by using the unimodality assumption.

$$[a = 1.15, b = 3]$$

$a = 0 \qquad x_1 = 1.15 \qquad x_2 = 1.85 \qquad b = 3$

$f_a = 0 \qquad f_1 = -0.57 \qquad f_2 = -0.925 \qquad f_b = 0$

$f_0 = f_1 = 1, \quad 1,1,2,3,5,8,13,21 \qquad n = 5$

$$L^* = \frac{f_{n-2}}{f_n} L_o = \frac{3}{8}(3 - 1.15) = \frac{3}{8}(1.85) = 0.694$$

$$x_1 = a + L^* = 1.15 + 0.694 = 1.84,$$

$$x_2 = b - L^* = 3 - 0.694 = 2.31$$

$a = 1.15 \qquad x_1 = 1.84 \qquad x_2 = 2.31 \qquad b = 3$

$f_a = -058 \qquad f_1 = -0.92 \qquad f_2 = -0.69 \qquad f_b = 0$

Discard part of the interval by using the unimodality assumption.

$$[a = 1.15, b = x_2 = 2.31]$$

$f_0 = f_1 = 1,\ \ 1,1,2,3,5,8,13,21 / \quad n = 4$

$$L^* = \frac{f_{n-2}}{f_n} L_o = \frac{2}{5}(2.31 - 1.15) = \frac{2}{5}(1.16) = 0.464$$

$x_1 = a + L^* = 1.15 + 0.464 = 1.614\,,$

$x_2 = b - L^* = 2.31 - 0.464 = 1.846$

$a = 1.15 \qquad x_1 = 1.614 \qquad x_2 = 1.846 \qquad b = 2.31$

$f_a = -0.57 \qquad f_1 = -0.807 \qquad f_2 = -0.923 \qquad f_b = -0.69$

Discard part of the interval by using the unimodality assumption.

$$[a = x_1 = 1.614, b = 2.31]$$

$n = 3$

$$L^* = \frac{f_{n-2}}{f_n} L_o = \frac{1}{3}(2.31 - 1.614) = \frac{1}{3}(0.696) = 0.232$$

$x_1 = a + L^* = 1.614 + 0.232 = 1.846\,,$

$x_2 = b - L^* = 2.31 - 0.232 = 2.078$

$a = 1.614 \qquad x_1 = 1.846 \qquad x_2 = 2.078 \qquad b = 2.31$

$f_a = -0.807 \qquad f_1 = -0.923 \qquad f_2 = -0.922 \qquad f_b = -0.69$

n	3	4	5	6	7

$a = 1.614 \quad x_1 = 1.846 \quad x_2 = 2.078 \quad b = 2.31$

$f_a = -0.807 \quad f_1 = -0.923 \quad f_2 = -0.922 \quad f_b = -0.69$

The Last interval is [1.846, 2.31]

Thus the minimum must located at the middle

Hence the minimum is x*=2.078

Golden Section Method

The *golden section method* is same as the Fibonacci method except that in the Fibonacci method the total number of experiments to be conducted has to be specified before beginning the calculation, whereas this is not required in the golden section method. In the Fibonacci method, the location of the first two experiments is determined by the total number of experiments, n. In the golden section method we start with the assumption that we are going to conduct a large number of experiments. Of course, the total number of experiments can be decided during the computation.

Example:

Deduce the best value for the eliminating part of the interval in Fibonacci method assuming we conduct a large number of iterations.

$f_0 = f_1 = 1, \quad 1,1,2,3,5,8,13,21$

$$L^* = \frac{f_{n-2}}{f_n} L_o$$

$\frac{f_{n-2}}{f_n}$	$\frac{f_1}{f_3}=\frac{1}{3}=0.33$	$\frac{f_2}{f_4}=\frac{2}{5}=0.4$	$\frac{3}{8}=0.37$	$\frac{5}{13}=0.382$	$\frac{8}{21}=0.382$

$$\lim_{n\to\infty}\frac{f_{n-٢}}{f_n}=٠.٣٨٢$$

The algorithm of method of Golden Section Method

Procedure. The procedure is same as the Fibonacci method except that the location of the first two experiments is defined by

$L^* = 0.382\text{L}_\text{o}$

1. Let L_o be the initial interval : $L_o = [a, b]$

2. Define $L^* = 0.382\text{L}_\text{o}$

3. Put points of test to be $x_1 = a + L^*, x_2 = b - L^*$

4. Eliminate the non-desired part of the interval depending on the unimodality property

5. define the new interval $L_o = [a, b]$, repeat steps 2-5 until a desired accuracy is obtained.

In step 5, we can use one of the following accuracy formula:

$$|f(x_1) - f(x_2)| \leq \varepsilon$$

Or

$$|L_o| \leq \varepsilon$$

Where ε is small chosen value (such as 0.1).

Exercise:

Use Golden Section method to find the maximum of

$$f(x) = \begin{cases} \frac{1}{2}x, & x \leq 2 \\ 3 - x, & x > 2 \end{cases}$$

By starting from [0,3] with n=6

Chapter 4: Linear Programming Problem

Standard form of a linear programming problem

The general linear programming problem can be stated in the following standard form:

1. *Scalar form*

$$\text{Minimize } f(x_1, x_2, \ldots, x_n) = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

subject to the constraints

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \\
x_1 &\geq 0 \\
x_2 &\geq 0 \\
&\vdots \\
x_n &\geq 0
\end{aligned}$$

where c_j, b_j, and a_{ij} ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) are known constants, and x_j are the decision variables.

2. Matrix form

$$\text{Minimize } f(\mathbf{X}) = \mathbf{c}^T\mathbf{X}$$

subject to the constraints

$$\mathbf{aX} = \mathbf{b}$$

$$\mathbf{X} \geq \mathbf{0}$$

where

$$\mathbf{X} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix}, \quad \mathbf{b} = \begin{Bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{Bmatrix}, \quad \mathbf{c} = \begin{Bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{Bmatrix},$$

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

The characteristics of a linear programming problem, stated in the standard form, are:

1. The objective function is of the minimization type.
2. All the constraints are of the equality type.
3. All the decision variables are nonnegative.

It is now shown that any linear programming problem can be expressed in the standard form by using the following transformations.

The maximization of a function $f(x_1, x_2, \ldots, x_n)$ is equivalent to the minimization of the negative of the same function. For example, the objective function

$$\text{minimize } f = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

is equivalent to

$$\text{maximize } f' = -f = -c_1x_1 - c_2x_2 - \cdots - c_nx_n$$

Consequently, the objective function can be stated in the minimization form in any linear programming problem.

2. In most engineering optimization problems, the decision variables represent some physical dimensions, and hence the variables x_j will be nonnegative. However, a variable may be unrestricted in sign in some problems. In such cases, an unrestricted variable (which can take a positive, negative, or zero value) can be written as the difference of two nonnegative variables. Thus if x_j is unrestricted in sign, it can be written as

$$x_j = x_j' - x_j'', \text{ where } x_j' \geq 0 \text{ and } x_j'' \geq 0$$

Slack variable

A non-negative variable which must be added to an inequality constraint of the form ≤ to be in an equality form

If a constraint appears in the form of a ''less than or equal to'' type of inequality as

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n \leq b_k$$

it can be converted into the equality form by adding a nonnegative slack variable x_{n+1} as follows:

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n + x_{n+1} = b_k$$

surplus variable

A non-negative variable which must be subtracted from an inequality constraint of the form ≥ to be in an equality form

if the constraint is in the form of a ''greater than or equal to'' type of inequality as

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n \geq b_k$$

it can be converted into the equality form by subtracting a variable as

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n - x_{n+1} = b_k$$

where x_{n+1} is a nonnegative variable known as a *surplus variable*.

Exercises

Detect which of the following Mathematical statements is true and which is false.

1	The Matrix form of the standard form of linear programming problem is Minimize $f(X) = c^T X$, where c is unknown constant.	□	□
2	The decision variables in the standard form of linear programming problem must be positive or zero	□	□
3	The maximization of a function is equivalent to the minimization of the negative of the same function	□	□
4	An unrestricted variable can be written as the difference of two nonnegative variables to agree the standard form of linear programming problem	□	□
5	A non negative variable which must be added to an inequality constraint of the form ≤ to be in an equality form is called slack variable	□	□
6	A non negative variable which must be subtracted from an inequality constraint of the form ≥ to be in an equality form is called surplus variable	□	□
7	A non negative variable which must be added to an inequality constraint of the form ≤ to be in an equality form is called surplus variable	□	□
8	A non negative variable which must be subtracted from an inequality constraint of the form ≥ to be in an equality form is called slack variable	□	□

<u>Geometry of linear programming problems</u>

The following general geometrical characteristics can be noted from the graphical solution.

1. The feasible region is a convex polygon.
2. The optimum value occurs at an extreme point or vertex of the feasible region.

Example:

Find the set of points that satisfies the following set of inequalities:

$$4x + 5y \leq 33, x + 4y \geq 11, 2x - 3y \geq -11$$

Answer:

We consider the line

$$4x + 5y = 33$$

$$x = 0 \rightarrow 5y = 33 \rightarrow y = \frac{33}{5} = 6\frac{3}{5} = 6.6$$

$$4x + 5y = 33$$

$$y = 0 \rightarrow 4x = 33 \rightarrow x = \frac{33}{4} = 8\frac{1}{4} = 8.25$$

$$\left(0, \frac{33}{5}\right), \left(\frac{33}{4}, 0\right)$$

(0,0) satisfies $4x + 5y \leq 33$ then this inequality is satisfied by the set of point down and left the line passes through $\left(0, \frac{33}{5}\right), \left(\frac{33}{4}, 0\right)$

<u>The line $x + 4y = 11$</u>

$$x = 0 \rightarrow 4y = 11 \rightarrow y = \frac{11}{4} = 2\frac{3}{4}$$

$$y = 0 \rightarrow x = 11$$

The point to that satisfies the inequality are over and on the line

<u>The line $2x - 3y = -11$</u>

$$x = 0 \rightarrow -3y = -11 \rightarrow y = \frac{11}{3} = 3\frac{2}{3} = 3.67$$

$$y = 0 \rightarrow 2x = -11 \rightarrow x = \frac{-11}{2} = -5\frac{1}{2} = -5.5$$

(0,0) satisfies $2x - 3y \geq -11$ then this inequality is satisfied by the set of point down the line

Solving $4x + 5y = 33$, $x + 4y = 11$,

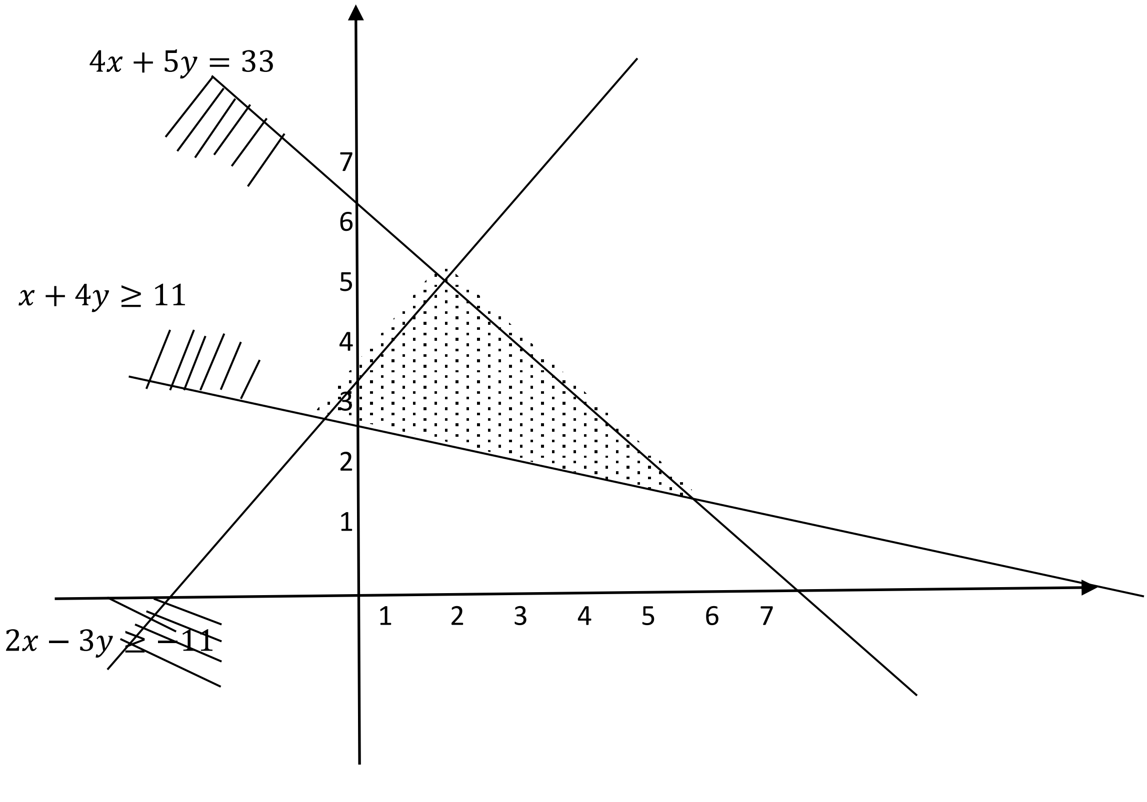
we obtain $\mathrm{P}(7,1)$

Solving $4x + 5y = 33$, $2x - 3y = -11$,

we obtain $\mathrm{Q}(2,5)$

Solving $x + 4y = 11$, $2x - 3y = -11$

We obtain $\mathrm{R}(-1,3)$



The set of points that satisfy the three inequalities are those inside and at the triangle described at the figure ABC.

The optimum value occurs at an extreme point or vertex of the feasible region.

Exercises

Select the correct word

(1) The inequality $4x + 5y \leq 33$ is satisfied by the set of points down and left the line passes through

□	$\left(\mathbf{0},\frac{\mathbf{33}}{\mathbf{5}}\right),\left(\frac{\mathbf{33}}{\mathbf{4}},\mathbf{0}\right)$	□	$\left(\frac{\mathbf{33}}{\mathbf{5}},\mathbf{0}\right),\left(\frac{\mathbf{33}}{\mathbf{4}},\mathbf{0}\right)$	□	$\left(\frac{\mathbf{33}}{\mathbf{5}},\mathbf{0}\right),\left(\mathbf{0},\frac{\mathbf{33}}{\mathbf{4}}\right)$	□	$(\mathbf{1},\mathbf{0}),(\mathbf{0},\mathbf{1})$

(2)The inequality $2x - 3y \geq -11$ is satisfied by the set of point down and left the line passes through

□	$\left(\frac{11}{3},\mathbf{0}\right),\left(\frac{-11}{2},\mathbf{0}\right)$	□	$\left(\mathbf{0},\frac{11}{3}\right),\left(\frac{-11}{2},\mathbf{0}\right)$	□	$\left(\frac{11}{3},\mathbf{0}\right),\left(\mathbf{0},\frac{-11}{2}\right)$	□	$(\mathbf{9},\mathbf{0}),(\mathbf{0},\mathbf{6})$

Example

Find the solution of the following LP problem graphically:
Maximize$f(x, y) = 3x + y + 2$,
Subject to $2x + y + 9 \geq 0$, $3y - x + 6 \geq 0$, $x + 2y \leq 3$, $y \leq x + 3$
Answer:

$$2x + y + 9 = 0$$

$$2x + y = -9 \Rightarrow x = 0, y = -9$$

$$y = 0, x = -4.5$$

(0,0) Satisfies it, so the proposed area is up right the line

----

$$3y - x + 6 \geq 0$$

$$3y - x = -6 \Rightarrow x = 0, y = -2$$

$$y = 0, x = 6$$

(0,0) Satisfies it, so the proposed area is up Left the line

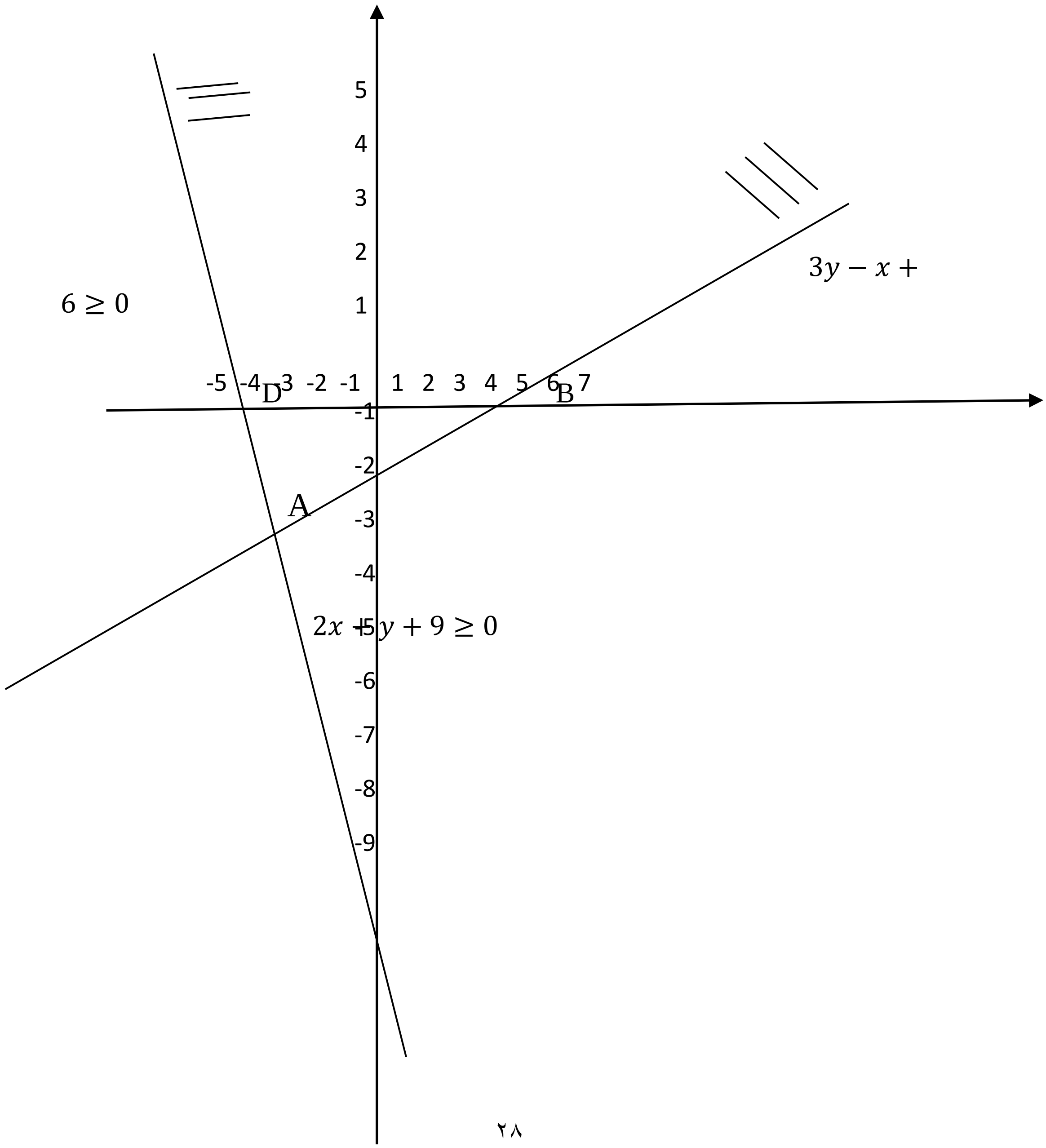
The intersection of

$2x + y = -9,\ 3y - x = -6$

is obtained by solving these two eqs. To obtain

$$x = -3, y = -3$$

$$f(at\ A) = 3(-3) + (-3) + 2 = -10$$



$x + 2y \leq 3$

=============

$x + 2y = 3$ ➔ $x = 0, y = 1.5$

$y = 0, x = 3$

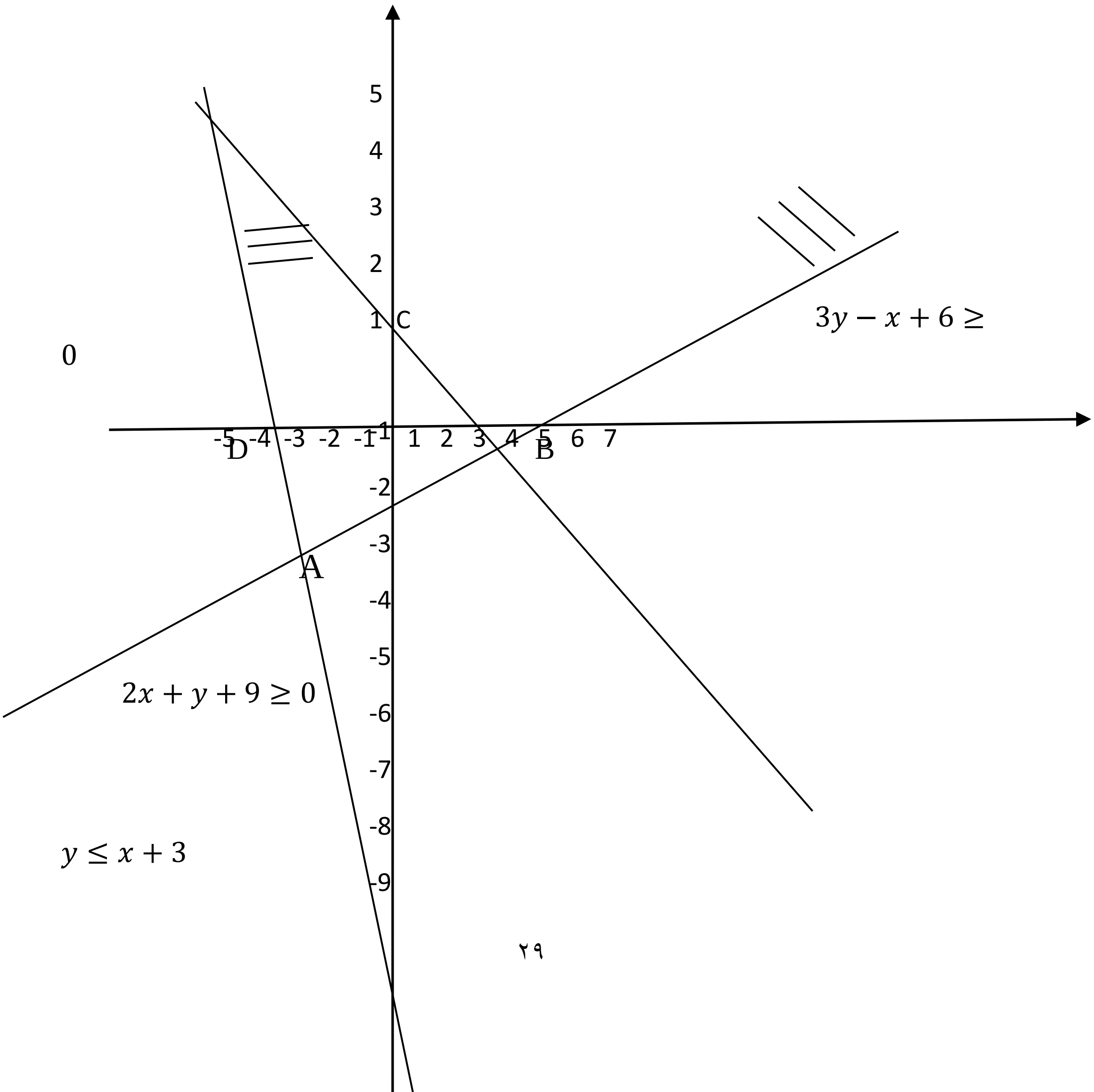
(0,0) satisfies it, so the proposed area is Down Left the line

===========

The intersection of $3y - x = -6, x + 2y = 3$

is obtained by solving these two eqs. To obtain $x = 4.5, y = -0.6$

$f(at\ B) = 3(4.5) + (-0.6) + 2 = 14$



$y - x = 3 \Rightarrow x = 0, y = 3$

$$y = 0, x = -3$$

(0,0) satisfies it, so the proposed area is Down Wright the line

===========

The intersection of $y - x = 3, x + 2y = 3$

Is obtained by solving these two eqs. To obtain $x = -1, y = 2$

$$f(at\ C) = 3(-1) + (2) + 2 = 1$$

===========

The intersection of $y + x = 3, 2x + y = -9$

Is obtained by solving these two eqs. To obtain $x = -4, y = -12$

$$f(x, y) = 3x + y + 2$$

Thus,

$$f_A = -10 \text{ at } A(-3, -3)$$

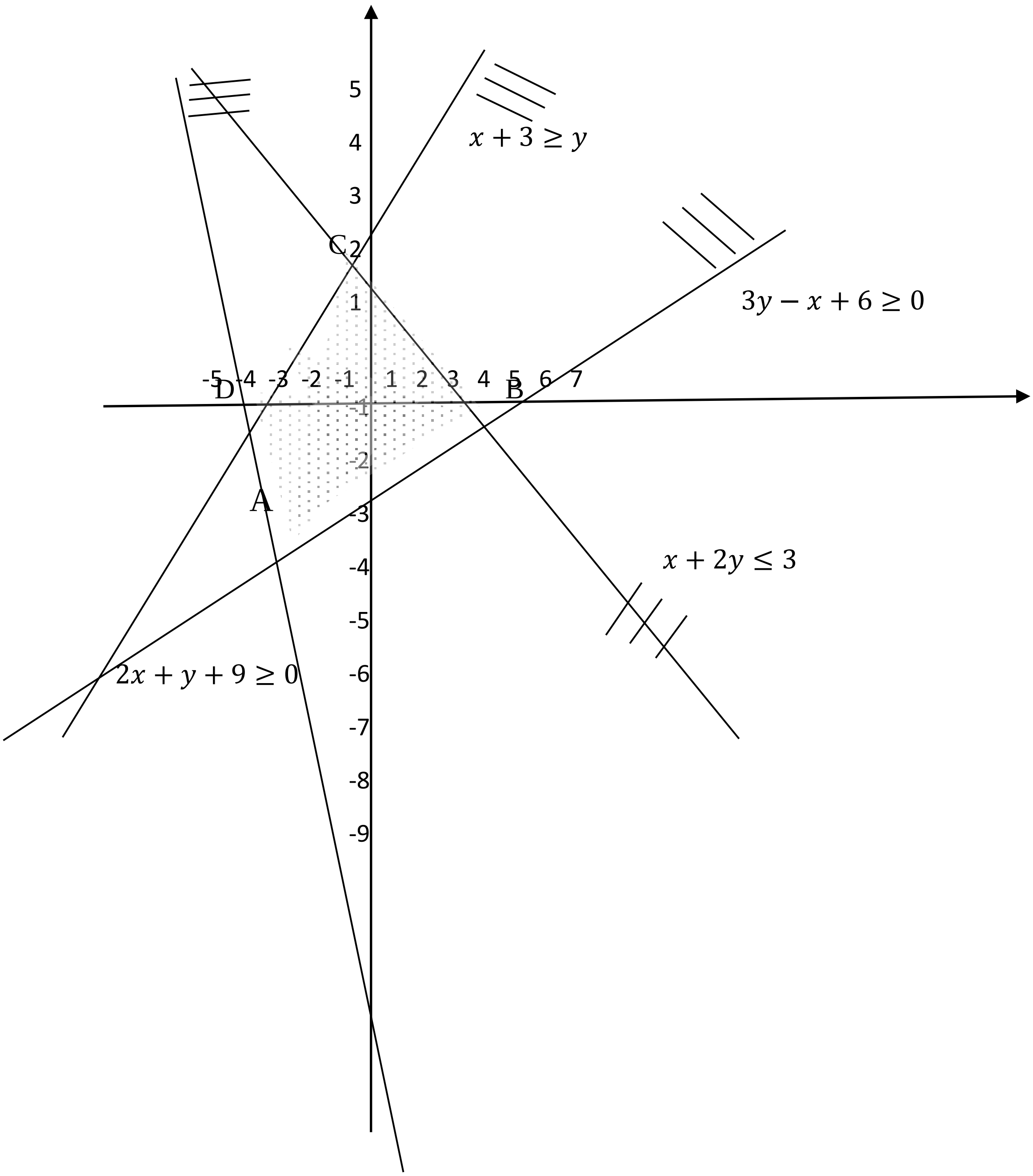
$$f_C = 1 \text{ at } C(-1,2)$$

$$f_B = 14 \text{ at } B(4.2, -0.6)$$

$$f_D = -11 \text{ at } D(-4, -1)$$

Hence the Maximum value is $f_B = 14$ at $B(4.2, -0.6)$

And the Minimum value is $f_D = -11$ at $D(-4, -1)$


$x + 3 \geq y$
$3y - x + 6 \geq 0$
$x + 2y \leq 3$
$2x + y + 9 \geq 0$
A
B
C
D
5
4
3
2
1
-1
-2
-3
-4
-5
-6
-7
-8
-9
-5 -4 -3 -2 -1 1 2 3 4 5 6 7

Exercises

Consider the following graph that represent four inequalities constraints of linear programming problem with the objective function $f(x,y) = 3x + y + 2$. Answer the following:

(1) The Point of intersection A is

☐	$(4.2, -0.6)$	☐	$(-3, -3)$	☐	$(-1,2)$	☐	$(\mathbf{1}, \mathbf{0}), (\mathbf{0}, \mathbf{1})$

(2) The Point of intersection B is

☐	$(4.2, -0.6)$	☐	$(-3, -3)$	☐	$(-1,2)$	☐	$(\mathbf{1}, \mathbf{0}), (\mathbf{0}, \mathbf{1})$

(3) The Point of intersection C is

☐	$(-3, -3)$	☐	$(4.2, -0.6)$	☐	$(-1,2)$	☐	$(-\mathbf{1}, \mathbf{0}), (\mathbf{0}, -\mathbf{1})$

(4) The Point of intersection D is

☐	$(-4.2,0.6)$	☐	$(-3, -3)$	☐	$(4.2, -0.6)$	☐	$(-4, -1)$

(5) The Maximum value occurs at

☐	$B(4.2, -0.6)$	☐	$D(-4, -1)$	☐	$C(-1,2)$	☐	$(\mathbf{1}, \mathbf{0}), (\mathbf{0}, \mathbf{1})$

(9) the Minimum value occurs at

☐	$B(4.2, -0.6)$	☐		☐		☐	$(-\mathbf{1}, \mathbf{0}), (\mathbf{0}, -\mathbf{1})$

Example:

A manufacturing firm produces two machine parts using lathes, milling machines, and grinding machines. The different machining times required for each part, the machining times available on different machines, and the profit on each machine part are given in the following table.

	Machining Time Required (min)		Maximum Time Available
Type of Machine	Machine Part I	Machine Part II	per Week (min)
Lathes	10	5	2500
Milling machines	4	10	2000
Grinding machines	1	1.5	450
Profit per unit	$50	$100	

Determine the number of parts I and II to be manufactured per week to maximize the profit.

Solution

Let the number of machine parts I and II manufactured per week be denoted by x and y, respectively.

The constraints due to the maximum time limitations on the various machines are given by

$$10x + 5y \le 2500 \quad (E_1)$$

$$4x + 10y \le 2000 \quad (E_2)$$

$$x + 1.5y \le 450 \quad (E_3)$$

Since the variables x and y cannot take negative values, we have

$$x \ge 0, \quad y \ge 0 \quad (E_4)$$

The total profit is given by

$$f(x,y) = 50x + 100y \qquad (E_5)$$

Exercise:
Find the solution of the following LP problem graphically:

Maximize

$$f(x,y) = 50x + 100y$$

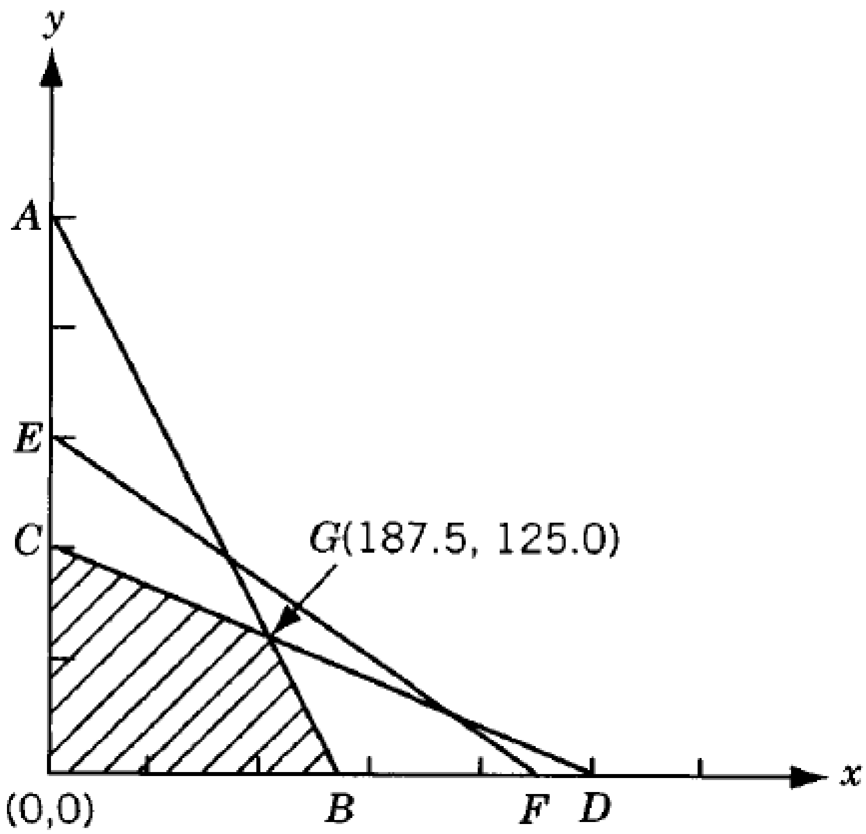
subject to

$$10x + 5y \leq 2500$$
$$4x + 10y \leq 2000$$
$$x + 1.5y \leq 450$$

$$x \geq 0, y \geq 0$$

Thus the problem is to determine the nonnegative values of x and y that satisfy the constraints stated in Eqs. (E_1) to (E_3) and maximize the objective function given by Eq. (E_5). The inequalities (E_1) to (E_4) can be plotted in the xy plane and the feasible region identified as shown in Fig. 3.3. Our objective is to find

at least one point out of the infinite points in the shaded region of Fig. which maximizes the profit function (E_5).



Definitions and Theorems

Definitions

1. Point in n-Dimensional Space

$$(x_1, x_2, \ldots, x_n).$$

2. Line Segment in **n-*Dimensions (L)*** If the coordinates of two points A and B are given by $x_j^{(1)}$ and $x_j^{(2)}$ ($j = 1,2,\ldots,n$), the line segment (L) joining these points is the collection of points $\mathbf{X}$ (λ) whose coordinates are given by $x_j = \lambda x_j^{(1)} + (1 - \lambda)x_j^{(2)}$, $j = 1,2,\ldots,n$, with $0 \le \lambda \le 1$.

Thus

$$L = \{\mathbf{X} | \mathbf{X} = \lambda \mathbf{X}^{(1)} + (1 - \lambda)\mathbf{X}^{(2)}\}$$

In one dimension, for example, it is easy to see that the definition is in accordance with our experience (Fig.):

$$x^{(2)} - x(\lambda) = \lambda[x^{(2)} - x^{(1)}], \qquad 0 \le \lambda \le 1$$

Figure Line segment.

whence

$$x(\lambda) = \lambda x^{(1)} + (1 - \lambda)x^{(2)}, \qquad 0 \le \lambda \le 1$$

Convex Set

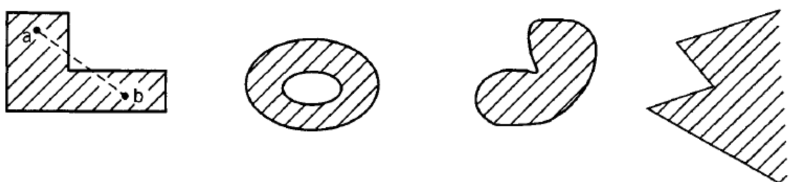
4. Convex Set A convex set is a collection of points such that if $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are any two points in the collection, the line segment joining them is also in the collection. A convex set, *S*, can be defined mathematically as follows:

If $X^{(1)}, X^{(2)} \in S$, implies

$$X = \lambda X^{(1)} + (1 - \lambda)X^{(2)} \in S, \quad 0 \le \lambda \le 1.$$

Convex Set

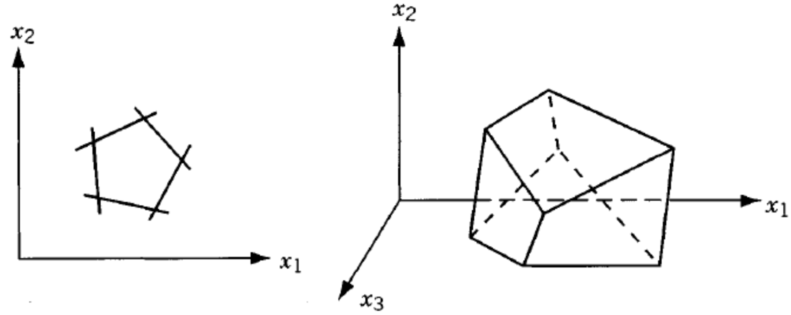
A set containing only one point is always considered to be convex



Non Convex Set

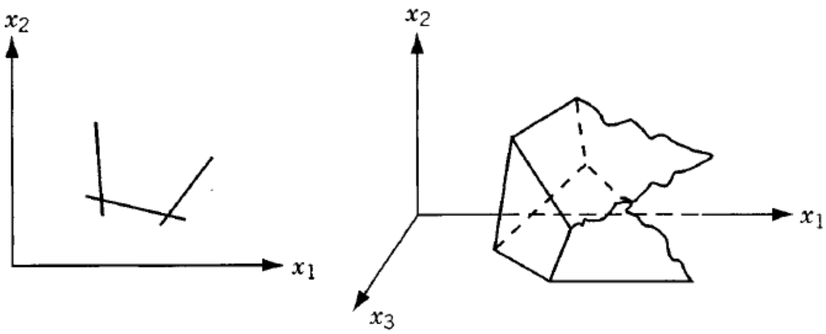
Convex Polyhedron and Convex Polytope

A convex polyhedron is a set of points common to one or more half-spaces. A convex polyhedron that is bounded is called a convex polytope.



(*a*) (*b*)

Figure *a* and *b* represent convex polytopes in two and three dimensions,



(*c*) (*d*)

and Fig. c and *d* denote convex polyhedra in two and three dimensions.
It can be seen that a convex polygon, shown in Fig. a and c, can be considered as the intersection of one or more half-planes.

Vertex or Extreme Point

This is a point in the convex set that **does not lie** on a line segment joining two other points of the set. For example,

every point on the circumference of a circle and each **corner** point of a polygon can
be called a vertex or extreme point.

7. ***Feasible Solution*** In a linear programming problem, any solution that satisfies the constraints

$$\mathbf{aX} = \mathbf{b}$$
$$\mathbf{X} \geq \mathbf{0}$$

is called a *feasible solution.*

8. Basic Solution A basic solution is one in which *n* - *m* variables are set equal to zero. A basic solution can be obtained by setting *n* - *m* variables to zero and solving the constraint Eqs. (3.2) simultaneously.

9. ***Basis*** The collection of variables not set equal to zero to obtain the basic solution is called the basis.

10. Basic Feasible Solution This is a basic solution that satisfies the non negativity conditions of the problem

$$\mathbf{aX} = \mathbf{b} \qquad \mathbf{X} \geq \mathbf{0}$$

11. Non degenerate Basic Feasible Solution This is a basic feasible solution that has got exactly *m* positive x_i.

12. Optimal Solution A feasible solution that optimizes the objective function is called an optimal solution.

13. Optimal Basic Solution This is a basic feasible solution for which the
objective function is optimal.

Theorems

The basic theorems of linear programming can now be stated and proved.

Theorem 1 The intersection of any number of convex sets is also convex.

Proof: Let the given convex sets be represented as R_i $(i = 1,2,\ldots,K)$ and their intersection as R, so that‡

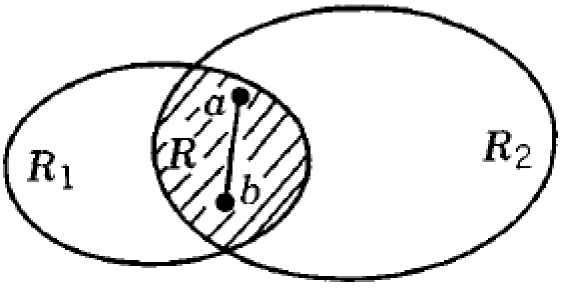
$$R = \bigcap_{i=1}^{K} R_i$$

If the points $\mathbf{X}^{(1)}, \mathbf{X}^{(2)} \in R$, then from the definition of intersection,

$$\mathbf{X} = \lambda \mathbf{X}^{(1)} + (1 - \lambda)\,\mathbf{X}^{(2)} \in R_i \qquad (i = 1,2,\ldots,K)$$
$$0 \le \lambda \le 1$$

Thus

$$\mathbf{X} \in R = \bigcap_{i=1}^{K} R_i$$



and the theorem is proved.

Theorem 2 The feasible region of a linear programming problem is convex.

Proof: The feasible region S of a standard linear programming problem is defined as

$$S = \{\mathbf{X} | \mathbf{aX} = \mathbf{b}, \mathbf{X} \geq 0\}$$

Let the points $\mathbf{X}_1$ and $\mathbf{X}_2$ belong to the feasible set S so that

$$\mathbf{aX}_1 = \mathbf{b}, \quad \mathbf{X}_1 \geq 0$$

$$\mathbf{aX}_2 = \mathbf{b}, \quad \mathbf{X}_2 \geq 0$$

Multiply the 1st eq by λ and the second by $1-\lambda$, and adding them, we obtain:

$$\mathbf{a}[\lambda\mathbf{X}_1 + (1-\lambda)\mathbf{X}_2] = \lambda\mathbf{b} + (1-\lambda)\mathbf{b} = \mathbf{b}$$

that is,

$$\mathbf{aX}_\lambda = \mathbf{b}$$

where

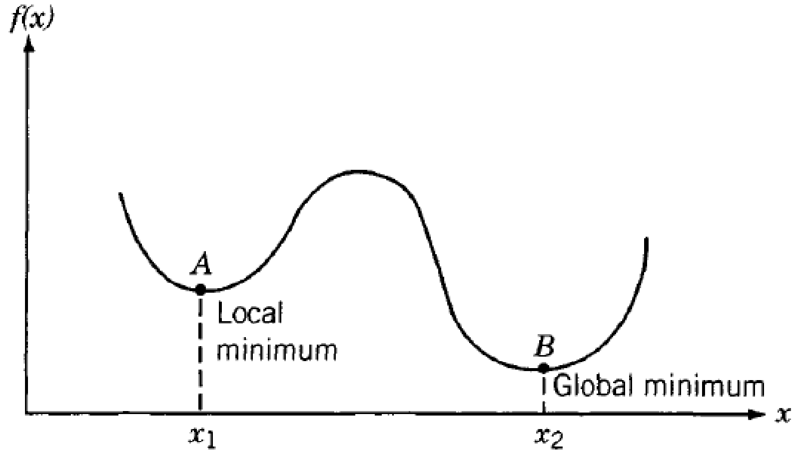
$$\mathbf{X}_\lambda = \lambda\mathbf{X}_1 + (1-\lambda)\mathbf{X}_2$$

Thus the point $\mathbf{X}_\lambda$ satisfies the constraints and if

$$0 \leq \lambda \leq 1, \quad \mathbf{X}_\lambda \geq 0$$

Hence the theorem is proved.

Theorem 3 Any local minimum solution is global for a linear programming problem.



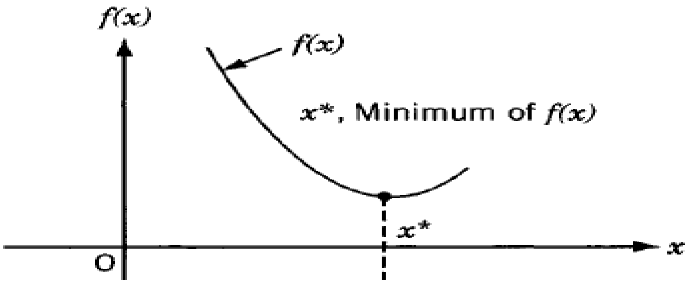
Theorem 4: Every basic feasible solution is an extreme point of the convex set of feasible solutions.

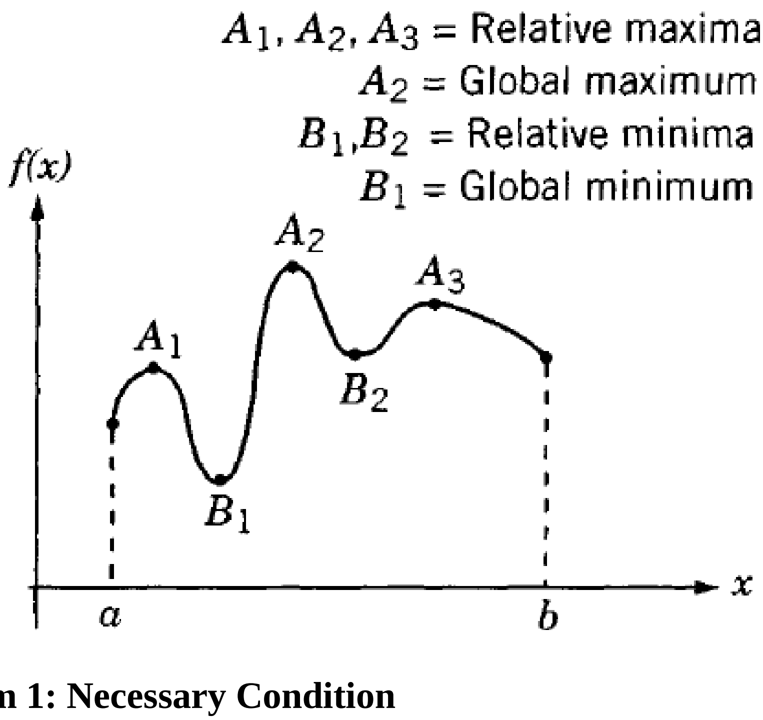
Theorem 5 Let S be a closed convex polyhedron. Then the minimum of a linear function over *S* is attained at an extreme point of *S*.

Classical Optimization Techniques

Single-variable optimization

A function of one variable $f(x)$ is said to have a relative or local minimum at $x = x^*$ if $f(x^*) < f(x^* + h)$ for all sufficiently small positive and negative values of h.





Theorem 1: Necessary Condition

If a function $f(x)$ is defined in the interval $a < x < b$ and have a relative minimum at $x = x^*$, where $a < x^* < b$, and if the derivative *df(x)/dx* = $f'(x^*)$ exists as a finite number at $x = x^*$, then $f'(x^*)$ = 0.

Theorem 2: Sufficient Condition:

Let f '(x^*) = f "(x^*) = …. =$f^{(n-1)}(x^*)$ = 0, but$f^{(n)}$ (x^*) ≠ 0. Then $f(x^*)$ is

(i) a minimum value of $f(x^*)$ if $f^{(n)}$ (x^*) > 0 and *n* is even;

(ii) a maximum value of $f(x^*)$ if $f^{(n)}(x^*)$< 0 and *n* is even;

(iii) neither a maximum nor a minimum if *n* is odd.

(b) Use theorems in(a) to find the optimum values of

$$f(x) = 12x^5 - 45x^4 + 40x^3 + 5$$

Answer:

$$f'(x) = 60x^4 - 3 * 60x^3 + 60 * 2 * x^2$$
$$= 60x^2(x^2 - 3x + 2)$$
$$= 60x^2(x-1)(x-2) = 0$$

The extreme points are

$x = 0, x = 1$ and $x = 2$

$x = 0$	$x = 1$	$x = 2$
$f''(x)$= $240x^3 - 540x^2 + 240x$ $f''(0) = 0$ We evaluate the next derivative $f'''(x)$= $3 * 240x^2 - 2 * 540x + 240$ $f'''(0)$= $+240$, order of derivative is odd. So this point is neither maximum nor minimum	$f''(1) = -60$ this point is relative maximum $f_{Max} = 12(1) - 45\ (1) + 40(1) + 5$ =12	$f''(2) = 240$ this point is relative minimum $f_{Min} = -11$

<u>**Excercises:**</u>

(1)Find the maxima and minima, if any, of the functions

$$f(x) = \frac{x^4}{(x-1)(x-3)^3}$$

$$f(x) = 4x^3 - 18x^2 + 27x - 7$$

$$f(x) = 10x^6 - 48x^5 + 15x^4 + 200x^3 - 120x^2 - 480x + 100$$

[2] Detect which of the following Mathematical statements is true and which is false. Write the false one(s) in the correct case.

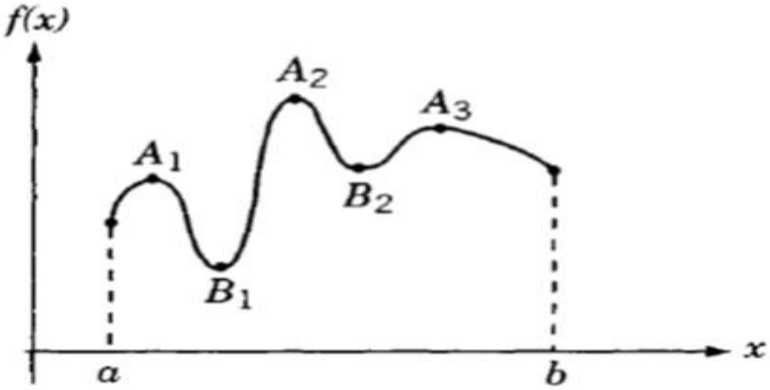


Figure I

In Figure I,

1	A_1 is relative minimum	□	□
2	A_2 is Global Maximum	□	□
	A_3 is relative Maximum	□	□
	B_1 is Global minimum	□	□
	B_2 is Global minimum	□	□
	The necessary condition for a function $f(x)$ to have a relative minimum at $x = x^*$, is $f'(x^*) = 0$.	□	□
	The sufficient condition for a function $f(x)$ to have a relative minimum at $x = x^*$ depends on the order (even- or odd) of the first non zero derivative of $f(x)$.	□	□

[3] Select the correct word

(1) A function of one variable $f(x)$ is said to have a relative or local minimum at $x = x^*$ if $f(x^*) \dots f(x^* + h)$ for all sufficiently small positive and negative values of h.

□	$<$	□	$\geq$	□	$\leq$	□	Else

Consider using the necessary and sufficient condition to find the optimum values of
$f(x) = 12x^5 - 45x^4 + 40x^3 + 5$. Answer the following questions:

$$f'(x) = ax^2(x - b)(x - c)$$

(1) $a = \cdots$

□	40	□	80	□	60	□	1

(1) $b = \cdots$

□	40	□	1	□	60	□	10

(1) $c = \cdots$

□	40	□	1	□	60	□	2

(1)The extreme point $x = \cdots$ is neither maximum nor minimum

□	0	□	1	□	60	□	2

The extreme point $x = \cdots$ is relative maximum

□	40	□	1	□	60	□	10

The extreme point $x = \cdots$ is relative minimum

□	40	□	1	□	60	□	2

Multivariable optimization with no constraints

Definition: ***r_th*** *Differential of* f: If all partial derivatives of the function f through order $r \geq 1$ exist and are continuous at a point X*, the polynomial

$$d^r f(\mathbf{X}^*) = \underbrace{\sum_{i=1}^{n} \sum_{j=1}^{n} \cdots \sum_{k=1}^{n}}_{r \text{ summations}} h_i h_j \cdots h_k \frac{\partial^r f(\mathbf{X}^*)}{\partial x_i \, \partial x_j \cdots \partial x_k}$$

is called the r[th] differential of f at X*.

For example :

when $r = 1$ and $n = 3$, we have

$$df(X^*) = \sum_{i=1}^{3} h_i \frac{\partial f}{\partial x_i} = h_1 \frac{\partial f}{\partial x_1} + h_2 \frac{\partial f}{\partial x_2} + h_3 \frac{\partial f}{\partial x_3}$$

Which corresponds $df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \frac{\partial f}{\partial x_3} dx_3$

When $r = 2$ and $n = 3$, we have

$$d^2 f(\mathbf{X}^*) = d^2 f(x_1^*, x_2^*, x_3^*) = \sum_{i=1}^{3} \sum_{j=1}^{3} h_i h_j \frac{\partial^2 f(\mathbf{X}^*)}{\partial x_i \, \partial x_j}$$

$$= h_1^2 \frac{\partial^2 f}{\partial x_1^2}(\mathbf{X}^*) + h_2^2 \frac{\partial^2 f}{\partial x_2^2}(\mathbf{X}^*) + h_3^2 \frac{\partial^2 f}{\partial x_3^2}(\mathbf{X}^*)$$

$$+ 2h_1 h_2 \frac{\partial^2 f}{\partial x_1 \, \partial x_2}(\mathbf{X}^*) + 2h_2 h_3 \frac{\partial^2 f}{\partial x_2 \, \partial x_3}(\mathbf{X}^*) + 2h_1 h_3 \frac{\partial^2 f}{\partial x_1 \, \partial x_3}(\mathbf{X}^*)$$

The Taylor's series expansion of a function f (X) near a point X* is given by

$$f(\mathbf{X}) = f(\mathbf{X}^*) + df(\mathbf{X}^*) + \frac{1}{2!} d^2f(\mathbf{X}^*) + \frac{1}{3!} d^3f(\mathbf{X}^*)$$

$$+ \cdots + \frac{1}{N!} d^Nf(\mathbf{X}^*) + R_N(\mathbf{X}^*,\mathbf{h})$$

Example 3 *:* Find the second-order Taylor's series approximation of the function

$$f(x_1,x_2,x_3) = x_2^2x_3 + x_1e^{x_3}$$

near the point

$$\mathbf{X}^* = \begin{Bmatrix} 1 \\ 0 \\ -2 \end{Bmatrix}.$$

SOLUTION The second-order Taylor's series approximation of the function f about point $\mathbf{X}^*$ is given by

$$f(\mathbf{X}) = f\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} + df\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} + \frac{1}{2!} d^2f\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$$

where

$$f\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = e^{-2}$$

$$df\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = h_1 \frac{\partial f}{\partial x_1}\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} + h_2 \frac{\partial f}{\partial x_2}\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} + h_3 \frac{\partial f}{\partial x_3}\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$$

$$= [h_1e^{x_3} + h_2(2x_2x_3) + h_3x_2^2 + h_3x_1e^{x_3}]\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = h_1e^{-2} + h_3e^{-2}$$

$$d^2f\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = \sum_{i=1}^{3}\sum_{j=1}^{3} h_ih_j \frac{\partial^2 f}{\partial x_i\,\partial x_j}\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = \left(h_1^2\frac{\partial^2 f}{\partial x_1^2} + h_2^2\frac{\partial^2 f}{\partial x_2^2} + h_3^2\frac{\partial^2 f}{\partial x_3^2}\right.$$

$$\left. + 2h_1h_2\frac{\partial^2 f}{\partial x_1\,\partial x_2} + 2h_2h_3\frac{\partial^2 f}{\partial x_2\,\partial x_3} + 2h_1h_3\frac{\partial^2 f}{\partial x_1\,\partial x_3}\right)\begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$$

$$= [h_1^2(0) + h_2^2(2x_3) + h_3^2(x_1e^{x_3}) + 2h_1h_2(0) + 2h_2h_3(2x_2)$$

$$+ 2h_1h_3(e^{x_3})] \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} = -4h_2^2 + e^{-2}h_3^2 + 2h_1h_3e^{-2}$$

Thus the Taylor's series approximation is given by

$$f(\mathbf{X}) \simeq e^{-2} + e^{-2}(h_1 + h_3) + \frac{1}{2!}(-4h_2^2 + e^{-2}h_3^2 + 2h_1h_3e^{-2})$$

Theorem : Necessary Condition If $f(\mathbf{X})$ has an extreme point (maximum or minimum) at $\mathbf{X} = \mathbf{X}^*$ and if the first partial derivatives of $f(\mathbf{X})$ exist at $\mathbf{X}^*$, then

$$\frac{\partial f}{\partial x_1}(\mathbf{X}^*) = \frac{\partial f}{\partial x_2}(\mathbf{X}^*) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{X}^*) = 0$$

Proof: The proof given for Theorem can easily be extended to prove the present theorem. However, we present a different approach to prove this theorem. Suppose that one of the first partial derivatives, say the kth one, does not vanish at $\mathbf{X}^*$. Then, by Taylor's theorem,

X=X*+h

$$f(\mathbf{X}^* + \mathbf{h}) = f(\mathbf{X}^*) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(\mathbf{X}^*) + \frac{1}{2!} d^2f(\mathbf{X}^* + \theta\mathbf{h}),$$

that is,

$$f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*) = h_k \frac{\partial f}{\partial x_k}(\mathbf{X}^*) + \frac{1}{2!} d^2f(\mathbf{X}^* + \theta\mathbf{h}), \qquad 0 < \theta < 1$$

Since $d^2f(\mathbf{X}^* + \theta\mathbf{h})$ is of order h_i^2, the terms of order $\mathbf{h}$ will dominate the higher-order terms for small $\mathbf{h}$. Thus the sign of $f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*)$ is decided by the sign of $h_k \, \partial f(\mathbf{X}^*)/\partial x_k$. Suppose that $\partial f(\mathbf{X}^*)/\partial x_k > 0$. Then the sign of $f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*)$ will be positive for $h_k > 0$ and negative for $h_k < 0$. This means that $\mathbf{X}^*$ cannot be an extreme point. The same conclusion can be obtained even if we assume that $\partial f(\mathbf{X}^*)/\partial x_k < 0$. Since this conclusion is in contradiction with the original statement that $\mathbf{X}^*$ is an extreme point, we may say that $\partial f/\partial x_k = 0$ at $\mathbf{X} = \mathbf{X}^*$. Hence the theorem is proved.

Theorem : Sufficient Condition A sufficient condition for a stationary point $\mathbf{X}^*$ to be an extreme point is that the matrix of second partial derivatives (Hessian matrix) of $f(\mathbf{X})$ evaluated at $\mathbf{X}^*$ is (i) positive definite when $\mathbf{X}^*$ is a relative minimum point, and (ii) negative definite when $\mathbf{X}^*$ is a relative maximum point.

Proof: From Taylor's theorem we can write

$$f(\mathbf{X}^* + \mathbf{h}) = f(\mathbf{X}^*) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(\mathbf{X}^*) + \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left.\frac{\partial^2 f}{\partial x_i \, \partial x_j}\right|_{\mathbf{X}=\mathbf{X}^*+\theta\mathbf{h}},$$

$$0 < \theta < 1$$

Since $\mathbf{X}^*$ is a stationary point, the necessary conditions give (Theorem

$$\frac{\partial f}{\partial x_i} = 0, \qquad i = 1,2,\ldots,n$$

Thus Eq. ( ) reduces to

$$f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*) = \frac{1}{2!} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left.\frac{\partial^2 f}{\partial x_i \, \partial x_j}\right|_{\mathbf{X}=\mathbf{X}^*+\theta\mathbf{h}}, \qquad 0 < \theta < 1$$

Therefore, the sign of

$$f(\mathbf{X}^* + \mathbf{h}) - f(\mathbf{X}^*)$$

will be same as that of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left.\frac{\partial^2 f}{\partial x_i \, \partial x_j}\right|_{\mathbf{X}=\mathbf{X}^*+\theta\mathbf{h}}$$

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left.\frac{\partial^2 f}{\partial x_i \, \partial x_j}\right|_{\mathbf{X}=\mathbf{X}^*}$$

is positive. This quantity Q is a quadratic form and can be written in matrix form as

$$Q = \mathbf{h}^T \mathbf{J} \mathbf{h}|_{\mathbf{X}=\mathbf{X}^*}$$

where

$$\mathbf{J}|_{\mathbf{X}=\mathbf{X}^*} = \left[\left.\frac{\partial^2 f}{\partial x_i \, \partial x_j}\right|_{\mathbf{X}=\mathbf{X}^*} \right]$$

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3 \partial x_3} \end{bmatrix}$$

is the matrix of second partial derivatives and is called the *Hessian matrix* of $f(\mathbf{X})$.

Definition:

A matrix A will be positive definite if all its eigenvalues are positive;
that is, all the values of λ that satisfy the determinantal equation

$$|A - \lambda I| = 0$$

should be positive. Similarly, the matrix [A] will be negative definite if its
eigenvalues are negative.

Another test that can be used to find the positive definiteness of a matrix A of order n involves evaluation of the determinants

$$A = |a_{11}|,$$

$$A_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},$$

$$A_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{32} \end{vmatrix}, \ldots,$$

$$A_n = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & & & & \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}$$

The matrix **A** will be positive definite if and only if all the values A_1, A_2, A_3, . . . , A_n are positive. The matrix **A** will be negative definite if and only if the sign of A_j is $(-1)^j$ for $j = 1,2,\ldots,n$. If some of the A_j are positive and the remaining A_j are zero, the matrix **A** will be positive semidefinite.

A matrix A will be positive definite if and only if all its determinants are positive;
A matrix A will be negative definite if and only if all its determinant A_k satisfies: $(-1)^k, k = 1,2,..$
A matrix A will be semi-definite if some of its determinant are positive, and the remaining are zeros

Saddle Point

In the case of a function of two variables, $f(x,y)$, the Hessian matrix may be neither positive nor negative definite at a point (x^*,y^*) at which

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$

In such a case, the point (x*,y*) is called a *saddle point.*

The characteristic of a saddle point is that it corresponds to a relative minimum or maximum of f(x,*y)* with respect to one variable , say, *x* (the other variable being fixed at *y* = *y*)* and a relative maximum or minimum of *f(x,y)* with respect to the second variable *y* (the other variable being fixed at x*).

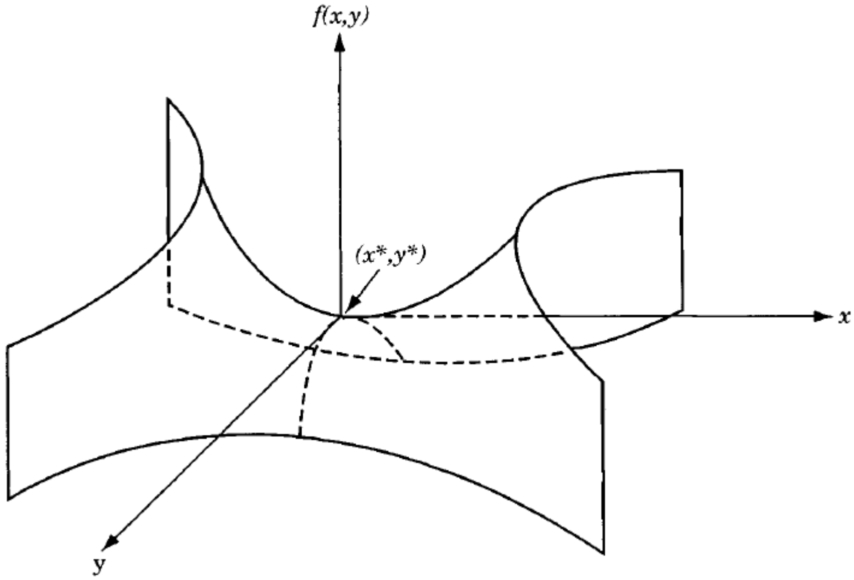


Figure 2.5 Saddle point of the function $f(x,y) = x^2 - y^2$.

1	A matrix A will be positive definite if all its eigenvalues are positive;	□	□
2	A matrix A will be positive definite if and only if all its determinants are positive;	□	□
3	A matrix A will be negative definite if and only if all its determinant A_k satisfies: $(-1)^k$	□	□
4	A matrix A will be semidefinite definite if some of its determinant are positive, and the remaining are zeros	□	□
5	A saddle point is corresponds to a relative minimum of f(x,*y)* with respect to one variable and a relative maximum with respect to the second variable	□	□

[Q3] Find the extreme points of the function

$$f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$$

The necessary condition is

$\frac{\partial f}{\partial x_1} = 3x_1^2 + 4x_1 = x_1(3x_1 + 4) = 0$ $\frac{\partial f}{\partial x_2} = 3x_2^2 + 8x_2 = x_2(3x_2 + 8) = 0$	So $x_1(3x_1 + 4) = 0$ $x_2(3x_2 + 8) = 0$

Consider finding the extreme points of the function $f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$. Answer the following

(1) The necessary condition yields

□	$x_1(3x_1 + 4) = 0,$ $x_2(3x_2 + 8) = 0$	□	$x_1(3x_1 - 4) = 0,$ $x_2(3x_2 + 8) = 0$
□	$x_1(3x_1 + 4) = 0,$ $x_2(3x_2 - 8) = 0$	□	$x_1(x_2 - 4) = 0,$ $x_2(x_1 - 8) = 0$

Consider finding the extreme points of the function $f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$. Answer the following

(2) The solutions of the necessary condition equations are

□	$x_1 = 0, \frac{4}{3}, \; x_2 = 0, \frac{-8}{3}$	□	$x_1 = 0, \frac{-4}{3}, \; x_2 = 0, \frac{-8}{3}$
□	$x_1 = 0, \frac{-4}{3}, \; x_2 = 0, \frac{8}{3}$	□	Else

Consider finding the extreme points of the function $f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$. Answer the following

(3) The necessary condition equations are satisfied at the points

□	$(0,0), \left(\frac{-8}{3}, 0\right), \left(0, \frac{-4}{3}\right), \left(\frac{-4}{3}, \frac{-8}{3}\right)$	□	$(0,0), \left(\frac{8}{3}, 0\right), \left(0, \frac{4}{3}\right), \left(\frac{4}{3}, \frac{8}{3}\right)$
□	$(0,0), \left(0, \frac{-8}{3}\right), \left(\frac{-4}{3}, 0\right), \left(\frac{-4}{3}, \frac{-8}{3}\right)$	□	$(0,1), \left(\frac{8}{3}, 1\right), \left(1, \frac{4}{3}\right), \left(\frac{4}{3}, \frac{-8}{3}\right)$

Consider finding the extreme points of the function $f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$. Answer the following

(3) The following point satisfies the necessary condition

□	$\left(0, \frac{-4}{3}\right)$	□	$\left(\frac{4}{3}, \frac{8}{3}\right)$
□	$\left(\frac{-8}{3}, 0\right)$	□	$\left(\frac{-4}{3}, \frac{-8}{3}\right)$

To find the nature of these extreme points, we have to use the sufficiency conditions. The second-order partial derivatives of f are given by

$\frac{\partial f}{\partial x_1} = 3x_1^2 + 4x_1$ $\frac{\partial f}{\partial x_2} = 3x_2^2 + 8x_2$	$\frac{\partial^2 f}{\partial {x_1}^2} = 6x_1 + 4, \frac{\partial^2 f}{\partial {x_2}^2} = 6x_2 + 8, \frac{\partial^2 f}{\partial x_1 \partial x_2} = 0$
$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} \end{bmatrix}$	The Hessian matrix of f is given by $\begin{bmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{bmatrix}$

If $J_1 = |6x_1 + 4|$ and $J_2 = \begin{vmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{vmatrix}$, the values of J_1 and J_2 and the nature of the extreme point are as given below.

$$f(x_1, x_2) = x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$$

Point X	Value of J_1	Value of J_2	Nature of J	Nature of X	$f(X)$
$(0,0)$	+4	+32	Positive definite	Relative minimum	6
$(0,-\frac{8}{3})$	+4	−32	Indefinite	Saddle point	418/27
$(-\frac{4}{3},0)$	−4	−32	Indefinite	Saddle point	194/27
$(-\frac{4}{3},-\frac{8}{3})$	−4	+32	Negative definite	Relative maximum	50/3

(4)To find the nature of these extreme points, we use the sufficiency conditions. The second-order partial derivatives of f are given by

□	$\frac{\partial^2 f}{\partial x_1{}^2}=0, \frac{\partial^2 f}{\partial x_2{}^2}=6x_2+8, \frac{\partial^2 f}{\partial x_1 \partial x_2}=6x_1+4$	□	$\frac{\partial^2 f}{\partial x_1{}^2}=6x_1+4, \frac{\partial^2 f}{\partial x_2{}^2}=0, \frac{\partial^2 f}{\partial x_1 \partial x_2}=6x_2+8$
□	$\frac{\partial^2 f}{\partial x_1{}^2}=6x_1+4, \frac{\partial^2 f}{\partial x_2{}^2}=6x_2+8, \frac{\partial^2 f}{\partial x_1 \partial x_2}=0$	□	Else

(5) The Hessian matrix of f is given by....

□	$\begin{bmatrix} 0 & 6x_1+4 \\ 0 & 6x_2+8 \end{bmatrix}$	□	$\begin{bmatrix} 6x_1+4 & 0 \\ 0 & 6x_2+8 \end{bmatrix}$
□	$\begin{bmatrix} 6x_1+4 & 0 \\ 6x_2+8 & 0 \end{bmatrix}$	□	$\begin{bmatrix} 6x_1+4 & 3x_1 \\ 2x_2 & 6x_2+8 \end{bmatrix}$

(6) The nature of the extreme point (0,0) is

□	Relative minimum	□	Saddle point	□	Relative maximum	□	indefinite points

(7) The nature of the extreme point $\left(0,\frac{-8}{3}\right)$ is

□	Relative minimum	□	Saddle point	□	Relative maximum	□	indefinite points

(8) The nature of the extreme point $\left(\frac{-4}{3},0\right)$ is

□	Relative minimum	□	Saddle point	□	Relative maximum	□	indefinite points

(9) The nature of the extreme point $\left(\frac{-4}{3},\frac{-8}{3}\right)$ is

□	Relative minimum	□	Saddle point	□	Relative maximum	□	indefinite points

(10) The Relative maximum of the function is

□	194/27	□	6	□	50/3	□	$\mathbf{\frac{418}{27}}$

[1] Answer whether each of the following quadratic forms is positive definite, negative definite, or neither.

(a) $f = x_1^2 - x_2^2$

(b) $f = 4x_1x_2$

(c) $f = x_1^2 + 2x_2^2$

(d) $f = -x_1^2 + 4x_1x_2 + 4x_2^2$

(e) $f = -x_1^2 + 4x_1x_2 - 9x_2^2 + 2x_1x_3 + 8x_2x_3 - 4x_3^2$

(2) Match the following equations and their characteristics.

(a) $f = 4x_1 - 3x_2 + 2$	Relative maximum at (1, 2)
(b) $f = (2x_1 - 2)^2 + (x_1 - 2)^2$	Saddle point at origin
(c) $f = -(x_1 - 1)^2 - (x_2 - 2)^2$	No minimum
(d) $f = x_1x_2$	Inflection point at origin
(e) $f = x^3$	Relative minimum at (1, 2)

(4) Determine whether each of the following matrices is positive definite, negative definite, or indefinite by finding its eigenvalues.

$$[A] = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

$$[B] = \begin{bmatrix} 4 & 2 & -4 \\ 2 & 4 & -2 \\ -4 & -2 & 4 \end{bmatrix}$$

$$[C] = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -2 & -2 \\ -1 & -2 & -3 \end{bmatrix}$$

$$[A] = \begin{bmatrix} -14 & 3 & 0 \\ 3 & -1 & 4 \\ 0 & 4 & 2 \end{bmatrix}$$

(5) Determine whether each of the following matrices is positive definite, negative definite, or indefinite by evaluating the signs of its submatrices.

$$[A] = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

$$[B] = \begin{bmatrix} 4 & 2 & -4 \\ 2 & 4 & -2 \\ -4 & -2 & 4 \end{bmatrix}$$

$$[C] = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -2 & -2 \\ -1 & -2 & -3 \end{bmatrix}$$

$$[A] = \begin{bmatrix} 4 & -3 & 0 \\ -3 & 0 & 4 \\ 0 & 4 & 2 \end{bmatrix}$$

(6) Express the function

$$f(x_1, x_2, x_3) = -x_1^2 - x_2^2 + 2x_1x_2 - x_3^2 + 6x_1x_3 + 4x_1 - 5x_3 + 2$$

in matrix form as

$$f(X) = \frac{1}{2} X^T [A]X + B^T X + C$$

and determine whether the matrix [A] is positive definite, negative definite, or indefinite.

(7) The profit per acre of a farm is given by

$$20x_1 + 26x_2 + 4x_1x_2 - 4x_1^2 - 3x_2^2$$

where x_1 and x_2 denote, respectively, the labor cost and the fertilizer cost. Find the values of x_1 and x_2 to maximize the profit.

where x_1 and x_2 denote, respectively, the labor cost and the fertilizer cost. Find the values of X_1 and X_2 to maximize the profit.

Multivariable Optimization With Equality Constraints

In this section we consider the optimization of continuous functions subjected to equality constraints:

$$\text{Minimize } f = f(\mathbf{X})$$

$$\text{subject to}$$

$$g_j(\mathbf{X}) = 0, \qquad j = 1,2,\ldots,m$$

Where

$$\mathbf{X} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix}$$

Here *m* is less than or equal to *n;* otherwise (if $m > n$), the problem becomes over defined and, in general, there will be no solution. There are several methods available for the solution of this problem. The methods of direct substitution, constrained variation, and Lagrange multipliers are discussed in the following sections.

(*) In the equality constraints optimization problem, the number of constraints must be ….. the number of variable.

□	<	□	≥	□	≤	□	Else

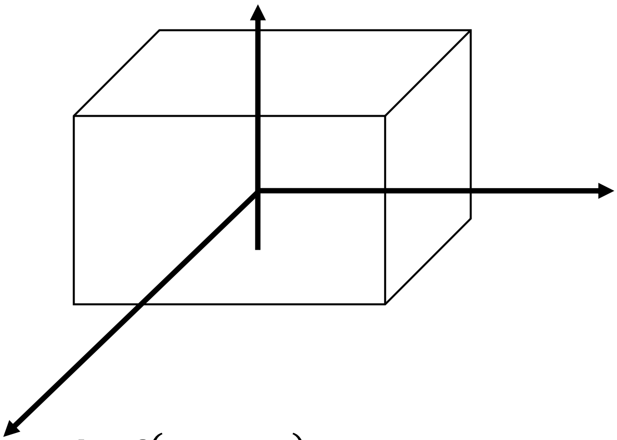
Solution by Direct Substitution

For a problem with *n* variables and *m* equality constraints, it is theoretically possible to solve simultaneously the *m* equality constraints and express any set of *m* variables in terms of the remaining *n* - *m* variables. When these expressions are substituted into the original objective function, there results a new objective function involving only *n* - *m* variables. The new objective function is not subjected to any constraint, and hence its optimum can be found by using the unconstrained optimization techniques discussed in Section 2.3.

[Q4] Find the dimensions of a box of largest volume that can be inscribed in a sphere of unit radius.

SOLUTION Let the origin of the Cartesian coordinate system x_1, x_2, x_3 be at the center of the sphere and the sides of the box be $2x_1$, $2x_2$, and $2x_3$. The

أوجد أبعاد صندوق بحيث يكون له أكبر حجم يمكن احتواؤه في كرة نصف قطرها الوحدة.



volume of the box is given by $f(x_{1,}x_2, x_3) = 8x_1x_2x_3$

Since the corners of the box lie on the surface of the sphere of unit radius, x_1, x_2, and x_3 have to satisfy the constraint

$$x_1^2 + x_2^2 + x_3^2 = 1$$

This problem has three design variables and one equality constraint. Hence the equality constraint can be used to eliminate any one of the design variables from the objective function. If we choose to eliminate x_3, Eq. (E_2) gives

$$x_3 = (1 - x_1^2 - x_2^2)^{1/2} \quad (E_3)$$

Thus the objective function becomes

$$f(x_1,x_2) = 8x_1x_2(1 - x_1^2 - x_2^2)^{1/2} \quad (E_4)$$

$$f(x_{1,}x_2) = 8x_1x_{2}\sqrt{1 - x_1^2 - x_2^2}$$

which can be maximized as an unconstrained function in two variables.

The necessary conditions for the maximum of f give

$$\frac{\partial f}{\partial x_1} = 8x_2\left[(1 - x_1^2 - x_2^2)^{1/2} - \frac{x_1^2}{(1 - x_1^2 - x_2^2)^{1/2}}\right] = 0 \qquad (E_5)$$

$$\frac{\partial f}{\partial x_2} = 8x_1\left[(1 - x_1^2 - x_2^2)^{1/2} - \frac{x_2^2}{(1 - x_1^2 - x_2^2)^{1/2}}\right] = 0 \qquad (E_6)$$

Equations (E_5) and (E_6) can be simplified to obtain

$$1 - 2x_1^2 - x_2^2 = 0$$

$$1 - x_1^2 - 2x_2^2 = 0$$

from which it follows that $x_1^* = x_2^* = 1/\sqrt{3}$ and hence $x_3^* = 1/\sqrt{3}$. This solution gives the maximum volume of the box as

$$f_{max} = \frac{8}{3\sqrt{3}}$$

For the sufficient condition, it is clear that the Hessian matrix is negative definite. Hence the point X_1 is maximum for the given function.

1	In the equality constraints optimization problem, the number of constraints must be less than or equal to the number of variable.	☐	☐
2	If the number of constraints is greater than the number of variable in the equality constraints optimization problem, the problem becomes over defined	☐	☐
3	Max. $f(x_1, x_2, x_3) = 8x_1x_2x_3$ subject to $x_1^2 + x_2^2 + x_3^2 = 1$ is equivalent to Max. $f(x_1, x_2) = 8x_1x_2\sqrt{1 - x_1^2 - x_2^2}$	☐	☐
4	The extreme points for the problem Max. $f(x_1, x_2, x_3) = 8x_1x_2x_3$ subject to $x_1^2 + x_2^2 + x_3^2 = 1$ is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.	☐	☐
5	The maximum value for the problem Max. $f(x_1, x_2, x_3) = 8x_1x_2x_3$ subject to $x_1^2 + x_2^2 + x_3^2 = 1$ is $\frac{8}{3\sqrt{3}}$.	☐	☐

[Q4]

Minimize $f = 9 - 8x_1 - 6x_2 - 4x_3 + 2x_1^2 + 2x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3$

subject to $x_1 + x_2 + 2x_3 = 3$ by direct substitution,

[Q5] Consider the problem Minimize $f(X) = \frac{x_1^2 + x_2^2 + x_3^2}{2}$

Subject to $g_1(X) = x_1 - x_2 = 0$, $g_2(X) = x_1 + x_2 + x_3 = 1$

By Direct substitution

[Q6] Find the value of x, y, and z that maximize the function $f(x, y, z) = \frac{6xyz}{x+2y+2z}$

When x, y, and z are restricted by the relation xyz = 16.

<u>Solution by the Method of Lagrange Multipliers</u>

The basic features of the Lagrange multiplier method is given initially for a simple problem of two variables with one constraint. The extension of the method to a general problem of *n* variables with *m* constraints is given later.

Problem with Two Variables and One Constraint.
Consider the problem:
Minimize $f(x_1, x_2)$ subject to $g(x_1, x_2) = 0$

1	Lagrange multiplier for the problem [Minimize $f(x_1, x_2)$ subject to $g(x_1, x_2) = 0$], evaluated at the extreme point is $\lambda = -\frac{\frac{\partial f}{\partial x_2}}{\frac{\partial g}{\partial x_2}}$	☐	☐

The necessary conditions generated by constructing a function *L*, known as the Lagrange function, as

$$L(x_1,x_2,\lambda) = f(x_1,x_2) + \lambda g(x_1,x_2)$$

By treating *L* as a function of the three variables x_1, x_2, and λ, the necessary conditions for its extremum are given by

$$\frac{\partial L}{\partial x_1}(x_1,x_2,\lambda) = \frac{\partial f}{\partial x_1}(x_1,x_2) + \lambda \frac{\partial g}{\partial x_1}(x_1,x_2) = 0$$

$$\frac{\partial L}{\partial x_2}(x_1,x_2,\lambda) = \frac{\partial f}{\partial x_2}(x_1,x_2) + \lambda \frac{\partial g}{\partial x_2}(x_1,x_2) = 0$$

$$\frac{\partial L}{\partial \lambda}(x_1,x_2,\lambda) = g(x_1,x_2) = 0$$

2	One of the necessary conditions for extremum of Lagrange multiplier solution for the problem [Minimize $f(x_1, x_2)$ subject to $g(x_1, x_2) = 0$], evaluated at the extreme point is $g(x_1, x_2) = 0$	☐	☐

[Q9] Find the solution of Minimize $f = k/xy^2$ Subject *to* $x^2 + y^2 = a^2$ using the necessary condition of Lagrange multiplier method

SOLUTION

The Lagrange function is

$$L(x,y,\lambda) = f(x,y) + \lambda g(x,y) = kx^{-1}y^{-2} + \lambda(x^2 + y^2 - a^2)$$

The necessary conditions for the minimum of *f(x, y)* give

$$\frac{\partial L}{\partial x} = -kx^{-2}y^{-2} + 2x\lambda = 0 \qquad (E_1)$$

$$\frac{\partial L}{\partial y} = -2kx^{-1}y^{-3} + 2y\lambda = 0 \qquad (E_2)$$

$$\frac{\partial L}{\partial \lambda} = x^2 + y^2 - a^2 = 0 \qquad (E_3)$$

Equations (E1) and (E2) yield

$$2\lambda = \frac{k}{x^3y^2} = \frac{2k}{xy^4}$$

$$\frac{1}{x^2} = \frac{2}{y^2}$$

from which the relation $x^* = (1/\sqrt{2})\, y^*$ can be obtained. This relation, along with Eq. (E_3), gives the optimum solution as

$$x^* = \frac{a}{\sqrt{3}} \quad \text{and} \quad y^* = \sqrt{2}\,\frac{a}{\sqrt{3}}$$

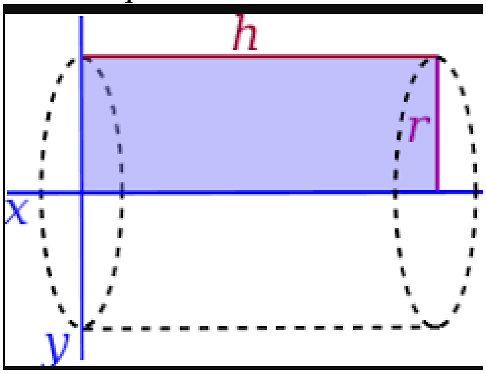
Sufficiency Conditions for a General Problem

Theorem : Sufficient Condition A sufficient condition for $f(\mathbf{X})$ to have a relative minimum at $\mathbf{X}^*$ is that the quadratic, Q, defined by

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 L}{\partial x_i\, \partial x_j}\, dx_i\, dx_j$$

evaluated at $\mathbf{X} = \mathbf{X}^*$ must be positive definite for all values of $d\mathbf{X}$ for which the constraints are satisfied.

[Q11]Find the dimensions of a cylindrical tin (with top and bottom) made up of sheet metal to maximize its volume such that the total surface area is equal to $A0 = 24\,\pi$.



Let the radius of the tin is $r = x_1$ and the length is $h = x_2$. respectively, the problem can be stated as:

$$\text{Maximize } f(x_1,x_2) = \pi x_1^2 x_2$$

subject to

$$2\pi x_1^2 + 2\pi x_1 x_2 = A_0 = 24\pi$$

The Lagrange function is

$$L(X,\lambda) = f(X) + \sum_{j=1}^{m} \lambda_j g_j(X)$$

$$L(x_1,x_2,\lambda) = \pi x_1^2 x_2 + \lambda(2\pi x_1^2 + 2\pi x_1 x_2 - A_0)$$

and the necessary conditions for the maximum of f give

$$\frac{\partial L}{\partial x_1} = 2\pi x_1 x_2 + 4\pi\lambda x_1 + 2\pi\lambda x_2 = 0 \qquad (E_1)$$

$$\frac{\partial L}{\partial x_2} = \pi x_1^2 + 2\pi\lambda x_1 = 0 \qquad (E_2)$$

$$\frac{\partial L}{\partial \lambda} = 2\pi x_1^2 + 2\pi x_1 x_2 - A_0 = 0 \qquad (E_3)$$

Equations (E_1) and (E_2) lead to

$$\lambda = -\frac{x_1 x_2}{2x_1 + x_2} = -\frac{1}{2} x_1$$

that is,

$$x_1 = \tfrac{1}{2} x_2 \qquad (E_4)$$

and Eqs. (E_3) and (E_4) give the desired solution as

$$x_1^* = \left(\frac{A_0}{6\pi}\right)^{1/2}, \quad x_2^* = \left(\frac{2A_0}{3\pi}\right)^{1/2}, \quad \text{and} \quad \lambda^* = -\left(\frac{A_0}{24\pi}\right)^{1/2}$$

This gives the maximum value of f as

$$f^* = \left(\frac{A_0^3}{54\pi}\right)^{1/2}$$

If $A_0 = 24\pi$, the optimum solution becomes

$$x_1^* = 2, \quad x_2^* = 4, \quad \lambda^* = -1, \quad \text{and} \quad f^* = 16\pi$$

To see that this solution really corresponds to the maximum of f, we apply the sufficiency condition of Eq. (2.44). In this case

$$L_{11} = \left.\frac{\partial^2 L}{\partial x_1^2}\right|_{(\mathbf{X}^*, \lambda^*)} = 2\pi x_2^* + 4\pi\lambda^* = 4\pi$$

$$L_{12} = \left.\frac{\partial^2 L}{\partial x_1\, \partial x_2}\right|_{(\mathbf{X}^*, \lambda^*)} = L_{21} = 2\pi x_1^* + 2\pi\lambda^* = 2\pi$$

$$L_{22} = \left.\frac{\partial^2 L}{\partial x_2^2}\right|_{(\mathbf{X}^*, \lambda^*)} = 0$$

Now since

$$\frac{\partial L}{\partial \lambda} = 2\pi x_1^2 + 2\pi x_1 x_2 - A_0 = 0$$

$$\frac{\partial^2 L}{\partial x_1 \partial \lambda}\bigg|_{(X^*,\lambda^*)} = 4\pi x_1^* + 2\pi x_2^* = 16\pi$$

And since

$$\frac{\partial L}{\partial \lambda} = 2\pi x_1^2 + 2\pi x_1 x_2 - A_0 = 0$$

$$\frac{\partial^2 L}{\partial x_2 \partial \lambda}\bigg|_{(X^*,\lambda^*)} = 2\pi x_1^* = 4\pi$$

And since

$$\frac{\partial L}{\partial \lambda} = 2\pi x_1^2 + 2\pi x_1 x_2 - A_0 = 0$$

$$\frac{\partial^2 L}{\partial \lambda \partial \lambda} = 0$$

$$\begin{bmatrix} \frac{\partial^2 L}{\partial x_1 \partial x_1} & \frac{\partial^2 L}{\partial x_1 \partial x_2} & \frac{\partial^2 L}{\partial x_1 \partial \lambda} \\ \frac{\partial^2 L}{\partial x_2 \partial x_1} & \frac{\partial^2 L}{\partial x_2 \partial x_2} & \frac{\partial^2 L}{\partial x_2 \partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda \partial x_1} & \frac{\partial^2 L}{\partial \lambda \partial x_2} & \frac{\partial^2 L}{\partial \lambda \partial \lambda} \end{bmatrix}$$

For test the positiveness of Hessian matrix:

$$H = \begin{bmatrix} 4\pi & 2\pi & 16\pi \\ 2\pi & 0 & 4\pi \\ 16\pi & 4\pi & 0 \end{bmatrix}$$

$$|H - \lambda I| = 0$$

$$\begin{bmatrix} 4\pi - \lambda & 2\pi & 16\pi \\ 2\pi & 0 - \lambda & 4\pi \\ 16\pi & 4\pi & 0 - \lambda \end{bmatrix} = 0$$

$$272\pi^2\lambda + 192\pi^3 = 0$$

$$\lambda = -\frac{12}{17}\pi$$

Since the value of λ is negative, the point $(x_1{}^*, x_2{}^*)$ corresponds to the maximum of f.

===========================

[Q12] Find the maximum of the function

$f(\mathbf{X}) = 2x_1 + x_2 + 10$ subject to

$g(\mathbf{X}) = x_1 + 2x_2^2 = 3$

using the Lagrange multiplier method.

SOLUTION The Lagrange function is given by

$$L(\mathbf{X},\lambda) = 2x_1 + x_2 + 10 + \lambda(3 - x_1 - 2x_2^2) \quad (E_1)$$

The necessary conditions for the solution of the problem are

$$\frac{\partial L}{\partial x_1} = 2 - \lambda = 0$$

$$\frac{\partial L}{\partial x_2} = 1 - 4\lambda x_2 = 0 \quad (E_2)$$

$$\frac{\partial L}{\partial \lambda} = 3 - x_1 - 2x_2^2 = 0$$

The solution of Eqs. (E_2) is

$$\mathbf{X}^* = \begin{Bmatrix} x_1^* \\ x_2^* \end{Bmatrix} = \begin{Bmatrix} 2.97 \\ 0.13 \end{Bmatrix} \quad (E_3)$$

$$\lambda^* = 2.0$$

Sufficient condition is Homework

[Q5]

Find the admissible and constrained variations at the point $\mathbf{X} = \begin{Bmatrix} 0 \\ 4 \end{Bmatrix}$ for the following problem:

$$\text{Minimize } f = x_1^2 + (x_2 - 1)^2$$

subject to

$$-2x_1^2 + x_2 = 4$$

$$\text{Minimize } f = 9 - 8x_1 - 6x_2 - 4x_3 + 2x_1^2 + 2x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3$$

subject to

$$x_1 + x_2 + 2x_3 = 3$$

by **(a)** direct substitution, **(b)** Lagrange multiplier method.

[Q6] Consider the problem

$$\text{Minimize } f(X) = \frac{x_1^2 + x_2^2 + x_3^2}{2}$$

Subject to

$$g_1(X) = x_1 - x_2 = 0$$
$$g_2(X) = x_1 + x_2 + x_3 = 1$$

By Lagrange multipliers method.

===

[Q7] (b) Minimize $f(X) = \frac{x_1^2 + x_2^2 + x_3^2}{2}$ (1)

Subject to $g_1(X) = x_1 - x_2 = 0,$ (2)

$g_2(X) = x_1 + x_2 + x_3 = 1$ (3)

By Lagrange multipliers method.

[Q8] find the value of x, y, and z that maximize the function

$$f(x, y, z) = \frac{6xyz}{x + 2y + 2z}$$

When x, y, and z are restricted by the relation xyz = 16.

Unconstrained Multivariable Optimization Techniques

This chapter deals with the various methods of solving the unconstrained minimization problem:

$$\text{Find } \mathbf{X} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Bmatrix} \text{ which minimizes } f(\mathbf{X})$$

As discussed in Chapter 1, a point X* will be a relative minimum of f (X) if the necessary conditions

$$\frac{\partial f}{\partial x_i}(\mathbf{X} = \mathbf{X}^*) = 0, \qquad i = 1,2,\ldots,n \tag{2}$$

are satisfied. The point $\mathbf{X}^*$ is guaranteed to be a relative minimum if the Hessian matrix is positive definite, that is,

$$\mathbf{J}_{\mathbf{X}^*} = [J]_{\mathbf{X}^*} = \left[\frac{\partial^2 f}{\partial x_i\, \partial x_j}(\mathbf{X}^*)\right] = \text{positive definite} \tag{.3}$$

Equations (2) and (3) can be used to identify the optimum point during numerical computations. However, if the function is not differentiate, Eqs. (2) and (3) cannot be applied to identify the optimum point.

Classification of Unconstrained Minimization Methods

Several methods are available for solving an unconstrained minimization problem. These methods can be classified into two broad categories as direct search methods and descent methods as indicated in Table 1.

TABLE 6.1 Unconstrained Minimization Methods

Direct Search Methods	Descent Methods
Random search method	Steepest descent (Cauchy) method
Grid search method	Fletcher–Reeves method
Univariate method	Newton's method
Pattern search methods	Marquardt method
Powell's method	Quasi-Newton methods
Hooke–Jeeves method	Davidon–Fletcher–Powell method
Rosenbrock's method	Broyden–Fletcher–Goldfarb–Shanno method
Simplex method	

Direct Search Methods: Do not require the derivatives of the function.
Descent Methods: Require the derivatives of the function.

	Direct Search Methods don't require the derivatives of the function.	□	□
	Descent Search require the derivatives of the function.	□	□

1.2 General Approach

[Q1]Draw the flowchart of general iterative scheme of unconstrained multivariable optimization

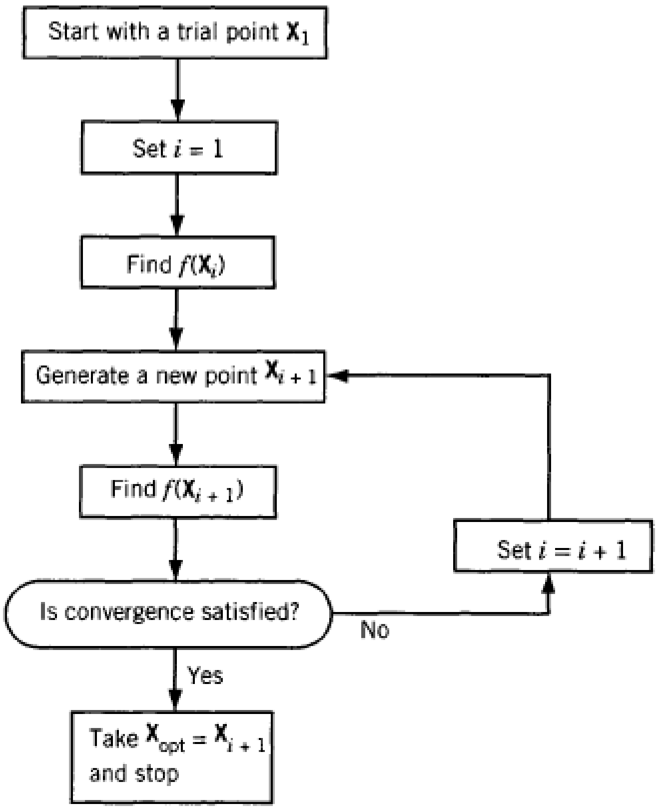


Figure 6.3 General iterative scheme of optimization.

3 Rate of Convergence

Different iterative optimization methods have different rates of convergence.
In general, an optimization method is said to have convergence of order *p* if

$$\frac{\|X_{i+1}-X^*\|}{\|X_i-X^*\|^p} \leq K, K \geq 0, p \geq 1 \qquad (4)$$

where $\mathbf{X}_i$ and $\mathbf{X}_{i+1}$ denote the points obtained at the end of iterations i and $i + 1$, respectively, $\mathbf{X}^*$ represents the optimum point, and $\|\mathbf{X}\|$ denotes the length or norm of the vector $\mathbf{X}$:

$$\|\mathbf{X}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

If p = 1 and 0 ≤ k ≤ 1, the method is said to be linearly convergent (corresponds to slow convergence), If p = 2, the method is said to be quadratically convergent (corresponds to fast convergence). An optimization method is said to have superlinear convergence (corresponds to fast convergence) if

$$\lim_{k \to \infty} \frac{\|X_{k+1} - X^*\|}{\|X_k - X^*\|} \to 0 \qquad (6)$$

The definitions of rates of convergence given in Eqs. (.4) and (.6) are applicable to single-variable as well as multivariable optimization problems. In the case of single-variable problems, the vector, $\mathbf{X}_i$, for example, degenerates to a scalar, x_i.

An iterative optimization method satisfies $\lim_{k \to \infty} \frac{\|X_{k+1} - X^*\|}{\|X_k - X^*\|} \to 0$, is *said to be ….. convergence*

□	quadratically	□	superlinear	□	linearly	□	super

	An iterative optimization method *is said to be quadratically convergent if* $\frac{\|X_{i+1} - X^*\|}{\|X_i - X^*\|^p} \leq K, K \geq 0,$ $p = 2$	□	□
	Different iterative optimization methods have the same rates of convergence.	□	□
	Some of the methods for solving constrained minimization problems require the use of unconstrained minimization techniques.	□	□
	The study of unconstrained minimization techniques provides the basic understanding necessary for the study of constrained minimization methods.	□	□
	All the unconstrained minimization methods are iterative in nature.	□	□

Indirect search (descent) methods

Gradient of a function

The gradient of a function is an n-component vector given by

$$\underset{n\times 1}{\nabla f} = \left[\frac{\partial f}{\partial x_1}\ \frac{\partial f}{\partial x_2}\ \frac{\partial f}{\partial x_3} \cdots \frac{\partial f}{\partial x_n}\right]^T$$

The gradient has a very important property. If we move along the gradient direction from any point in n-dimensional space, the function value increases at the fastest rate. Hence the gradient direction is called the *direction of steepest ascent.* Unfortunately, the direction of steepest ascent is a local property and not a global one. This is illustrated in Fig. 6.15, where the gradient

Since the gradient vector represents the direction of steepest ascent, the negative of the gradient vector denotes the direction of steepest descent. Thus any method that makes use of the gradient vector can be expected to give the minimum point faster than one that does not make use of the gradient vector. All the descent methods make use of the gradient vector, either directly or indirectly, in finding the search directions. Before considering the descent methods of minimization, we prove that the gradient vector represents the direction of steepest ascent.

[Q1] Prove that the gradient vector represents the direction of steepest ascent.

Theorem 6.3 The gradient vector represents the direction of steepest ascent.

Proof: Consider an arbitrary point **X** in the *n*-dimensional space. Let *f* denote the value of the objective function at the point **X**. Consider a neighboring point $\mathbf{X} + d\mathbf{X}$ with

$$d\mathbf{X} = \begin{Bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{Bmatrix}$$

where $dx_1, dx_2, \ldots, dx_n$ represent the components of the vector $d\mathbf{X}$. The magnitude of the vector $d\mathbf{X}$, ds, is given by

$$d\mathbf{X}^T d\mathbf{X} = (ds)^2 = \sum_{i=1}^{n} (dx_i)^2$$

If $f + df$ denotes the value of the objective function at $\mathbf{X} + d\mathbf{X}$, the change in f, df, associated with $d\mathbf{X}$ can be expressed as

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i = \nabla f^T d\mathbf{X} \qquad (1)$$

If $\mathbf{u}$ denotes the unit vector along the direction $d\mathbf{X}$ and ds the length of $d\mathbf{X}$, we can write

$$d\mathbf{X} = \mathbf{u}\, ds$$

The rate of change of the function with respect to the step length ds is given by Eq. (1) as

$$\frac{df}{ds} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{dx_i}{ds} = \nabla f^T \frac{d\mathbf{X}}{ds} = \nabla f^T \mathbf{u} \qquad (2)$$

The value of df/ds will be different for different directions and we are interested in finding the particular step $d\mathbf{X}$ along which the value of df/ds will be maximum. This will give the direction of steepest ascent.† By using the definition of the dot product, Eq. (2) can be rewritten as

$$\frac{df}{ds} = \|\nabla f\|\, \|\mathbf{u}\| \cos\theta$$

where $\|\nabla f\|$ and $\|\mathbf{u}\|$ denote the lengths of the vectors ∇f and $\mathbf{u}$, respectively, and θ indicates the angle between the vectors ∇f and $\mathbf{u}$. It can be seen that df/ds will be maximum when $\theta = 0°$ and minimum when $\theta = 180°$. This indicates that the function value increases at a maximum rate in the direction of the gradient (i.e., when $\mathbf{u}$ is along ∇f).

[Q2] Prove that the maximum rate of change of f at any point X is equal to the magnitude of the gradient vector at the same point. Then show what we can do if the Evaluation of the Gradient poses certain problem

Theorem 4 The maximum rate of change of f at any point X is equal to the magnitude of the gradient vector at the same point.

Proof: The rate of change of the function f with respect to the step length s along a direction $\mathbf{u}$ is given by Eq. (6.62). Since df/ds is maximum when $\theta = 0°$ and $\mathbf{u}$ is a unit vector, Eq. (6.62) gives

$$\left(\frac{df}{ds}\right)\bigg|_{\max} = \|\nabla f\|$$

which proves the theorem.

Evaluation of the Gradient

[Q3]"The evaluation of the gradient poses certain problems". Discuss this sentence.

The evaluation of the gradient requires the computation of the partial derivatives $\partial f/\partial x_i$, i = 1,2,. . .,ft. There are three situations where the evaluation of the gradient poses certain problems:

1. The function is differentiable at all the points, but the calculation of the components of the gradient, $\partial f/\partial x_i$, is either impractical or impossible.
2. The expressions for the partial derivatives $\partial f/\partial x_i$ can be derived, but they require large computational time for evaluation.
3. The gradient ∇f is not defined at all the points.

In the first case, we can use the forward finite-difference formula

$$\frac{\partial f}{\partial x_i}\bigg|_{\mathbf{X}_m} \simeq \frac{f(\mathbf{X}_m + \Delta x_i \mathbf{u}_i) - f(\mathbf{X}_m)}{\Delta x_i}, \qquad i = 1,2,\ldots,n \qquad (6.63)$$

Steepest descent (Cauchy) method

The use of the negative of the gradient vector as a direction for minimization was first made by Cauchy in 1847. In this method we start from an initial trial point X_1 and iteratively move along the steepest descent directions until the optimum point is found. The steepest descent method can be summarized by the following steps:

[Q4](a)Summarize the steps of steepest descent method for Multivariable Unconstrained Minimization problem.

1. Start with an arbitrary initial point $\mathbf{X}_1$. Set the iteration number as $i = 1$.
2. Find the search direction $\mathbf{S}_i$ as

$$\mathbf{S}_i = -\nabla f_i = -\nabla f(\mathbf{X}_i)$$

3. Determine the optimal step length λ_i^* in the direction $\mathbf{S}_i$ and set

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \lambda_i^* \mathbf{S}_i = \mathbf{X}_i - \lambda_i^* \nabla f_i$$

4. Test the new point, $\mathbf{X}_{i+1}$, for optimality. If $\mathbf{X}_{i+1}$ is optimum, stop the process. Otherwise, go to step 5.
5. Set the new iteration number $i = i + 1$ and go to step 2.

The method of steepest descent may appear to be the *best unconstrained minimization* technique since each one-dimensional search starts in the "best" direction. However, owing to the fact that the steepest descent direction is a local property, the method is not really effective in most problems.

[Q4](b)Use steepest descent method to Minimize the following Multivariable Unconstrained Minimization problem starting from X= {0 0)^T

$$f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1 x_2 + x_2^2$$

SOLUTION

Iteration 1

The gradient of f is given by

$$\nabla f = \begin{Bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{Bmatrix} = \begin{Bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{Bmatrix}$$

$$\nabla f_1 = \nabla f(\mathbf{X}_1) = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$$

Therefore,

$$\mathbf{S}_1 = -\nabla f_1 = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$$

To find $\mathbf{X}_2$, we need to find the optimal step length λ_1^*. For this, we minimize $f(\mathbf{X}_1 + \lambda_1\mathbf{S}_1) = f(-\lambda_1, \lambda_1) = \lambda_1^2 - 2\lambda_1$ with respect to λ_1. Since $df/d\lambda_1 = 0$ at $\lambda_1^* = 1$, we obtain

$$\mathbf{X}_2 = \mathbf{X}_1 + \lambda_1^*\mathbf{S}_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} + 1 \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$$

As $\nabla f_2 = \nabla f(\mathbf{X}_2) = \begin{Bmatrix} -1 \\ -1 \end{Bmatrix} \neq \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$, $\mathbf{X}_2$ is not optimum.

Iteration 2

$$\mathbf{S}_2 = -\nabla f_2 = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$$

To minimize

$$\begin{aligned} f(\mathbf{X}_2 + \lambda_2\mathbf{S}_2) &= f(-1 + \lambda_2, 1 + \lambda_2) \\ &= 5\lambda_2^2 - 2\lambda_2 - 1 \end{aligned}$$

we set $df/d\lambda_2 = 0$. This gives $\lambda_2^* = \frac{1}{5}$, and hence

$$\mathbf{X}_3 = \mathbf{X}_2 + \lambda_2^*\mathbf{S}_2 = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} + \frac{1}{5}\begin{Bmatrix} 1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -0.8 \\ 1.2 \end{Bmatrix}$$

Since the components of the gradient at $\mathbf{X}_3$, $\nabla f_3 = \begin{Bmatrix} 0.2 \\ -0.2 \end{Bmatrix}$, are not zero, we proceed to the next iteration.

Iteration 3

$$\mathbf{S}_3 = -\nabla f_3 = \begin{Bmatrix} -0.2 \\ 0.2 \end{Bmatrix}$$

As

$$f(\mathbf{X}_3 + \lambda_3\mathbf{S}_3) = f(-0.8 - 0.2\lambda_3, 1.2 + 0.2\lambda_3)$$

$$= 0.04\lambda_3^2 - 0.08\lambda_3 - 1.20, \qquad \frac{df}{d\lambda_3} = 0 \text{ at } \lambda_3^* = 1.0$$

Therefore,

$$\mathbf{X}_4 = \mathbf{X}_3 + \lambda_3^*\mathbf{S}_3 = \begin{Bmatrix} -0.8 \\ 1.2 \end{Bmatrix} + 1.0 \begin{Bmatrix} -0.2 \\ 0.2 \end{Bmatrix} = \begin{Bmatrix} -1.0 \\ 1.4 \end{Bmatrix}$$

The gradient at $\mathbf{X}_4$ is given by

$$\nabla f_4 = \begin{Bmatrix} -0.20 \\ -0.20 \end{Bmatrix}$$

Since $\nabla f_4 \neq \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$, $\mathbf{X}_4$ is not optimum and hence we have to proceed to the next iteration. This process has to be continued until the optimum point, $\mathbf{X}^* = \begin{Bmatrix} -1.0 \\ 1.5 \end{Bmatrix}$, is found.

Convergence Criteria.

The following criteria can be used to terminate the iterative process:

1. When the change in function value in two consecutive iterations is small:

$$\left| \frac{f(X_{i+1}) - f(X_i)}{f(X_i)} \right| \leq \varepsilon$$

2. When the partial derivatives (components of the gradient) of/are small:

$$\left| \frac{\partial f}{\partial x_i} \right| \leq \varepsilon_2, \qquad i = 1,2,\ldots,n$$

3. When the change in the design vector in two consecutive iterations is small:

$$|\mathbf{X}_{i+1} - \mathbf{X}_i| \leq \varepsilon_3$$

Conjugate Gradient (Fletcher-Reeves) Method

The convergence characteristics of the steepest descent method can be improved greatly by modifying it into a conjugate gradient method (which can be considered as a conjugate directions method involving the use of the gradient of the function).

We saw that any minimization method that makes use of the conjugate directions is quadratically convergent. T

his property of quadratic convergence is very useful because it ensures that the method will minimize a quadratic function in *n* steps or less. Since any general function can be approximated reasonably well by a quadratic near the optimum point, any quadratically convergent method is expected to find the optimum point in a finite number of iterations.

We have seen that Powell's conjugate direction method requires *n* single variable minimizations per iteration and sets up a new conjugate direction at the end of each iteration. Thus it requires, in general, n^2 single-variable minimizations to find the minimum of a quadratic function. On the other hand, if we can evaluate the gradients of the objective function, we can set up a new conjugate direction after every one-dimensional minimization, and hence we can achieve faster convergence. The construction of conjugate directions and development of the Fletcher-Reeves method are discussed in this section.

Development of the Fletcher-Reeves Method

[Q5] Develop the Fletcher-Reeves Method

Consider the development of an algorithm by modifying the steepest descent method applied to a quadratic function $f(\mathbf{X}) = \frac{1}{2}\mathbf{X}^T\mathbf{A}\mathbf{X} + \mathbf{B}^T\mathbf{X} + C$ by imposing the condition that the successive directions be mutually conjugate. Let $\mathbf{X}_1$ be the starting point for the minimization and let the first search direction be the steepest descent direction:

$$\mathbf{S}_1 = -\nabla f_1 = -\mathbf{A}\mathbf{X}_1 - \mathbf{B}$$

$$\mathbf{X}_2 = \mathbf{X}_1 + \lambda_1^*\mathbf{S}_1$$

or

$$\mathbf{S}_1 = \frac{\mathbf{X}_2 - \mathbf{X}_1}{\lambda_1^*}$$

where λ_1^* is the minimizing step length in the direction $\mathbf{S}_1$, so that

$$\mathbf{S}_1^T \nabla f|_{\mathbf{X}_2} = 0 \qquad (1)$$

Equation (1) can be expanded as

$$\mathbf{S}_1^T[\mathbf{A}(\mathbf{X}_1 + \lambda_1^*\mathbf{S}_1) + \mathbf{B}] = 0$$

from which the value of λ_1^* can be found as

$$\lambda_1^* = \frac{-\mathbf{S}_1^T(\mathbf{A}\mathbf{X}_1 + \mathbf{B})}{\mathbf{S}_1^T \mathbf{A}\mathbf{S}_1} = -\frac{\mathbf{S}_1^T \nabla f_1}{\mathbf{S}_1^T \mathbf{A}\mathbf{S}_1}$$

Now express the second search direction as a linear combination of $\mathbf{S}_1$ and $-\nabla f_2$:

$$\mathbf{S}_2 = -\nabla f_2 + \beta_2 \mathbf{S}_1$$

where β_2 is to be chosen so as to make $\mathbf{S}_1$ and $\mathbf{S}_2$ conjugate. This requires that

$$\mathbf{S}_1^T \mathbf{A}\mathbf{S}_2 = 0$$

Substituting Eq. ( ) into Eq. ( ) leads to

$$\mathbf{S}_1^T\mathbf{A}\,(-\nabla f_2 + \beta_2 \mathbf{S}_1) = 0$$

Equations ( ) and ( ) yield

$$-\frac{(\mathbf{X}_2 - \mathbf{X}_1)^T}{\lambda_1^*}\,\mathbf{A}\,(\nabla f_2 - \beta_2 \mathbf{S}_1) = 0$$

The difference of the gradients $(\nabla f_2 - \nabla f_1)$ can be expressed as

$$(\nabla f_2 - \nabla f_1) = (\mathbf{A}\mathbf{X}_2 + \mathbf{B}) - (\mathbf{A}\mathbf{X}_1 + \mathbf{B}) = \mathbf{A}(\mathbf{X}_2 - \mathbf{X}_1)$$

With the help of Eq. ( ), Eq. ( ) can be written as

$$(\nabla f_2 - \nabla f_1)^T(\nabla f_2 - \beta_2 \mathbf{S}_1) = 0$$

where the symmetricity of the matrix **A** has been used. Equation ( ) can be expanded as

$$\nabla f_2^T \nabla f_2 - \nabla f_1^T \nabla f_2 - \beta_2 \nabla f_2^T \mathbf{S}_1 + \beta_2 \nabla f_1^T \mathbf{S}_1 = 0 \qquad (\quad)$$

Since $\nabla f_1^T \nabla f_2 = -\mathbf{S}_1^T \nabla f_2 = 0$ from Eq. ( ), Eq. ( ) gives

$$\beta_2 = -\frac{\nabla f_2^T \nabla f_2}{\nabla f_1^T \mathbf{S}_1} = \frac{\nabla f_2^T \nabla f_2}{\nabla f_1^T \nabla f_1} \qquad (\quad)$$

Next we consider the third search direction as a linear combination of $\mathbf{S}_1$, $\mathbf{S}_2$, and $-\nabla f_3$ as

$$\mathbf{S}_3 = -\nabla f_3 + \beta_3 \mathbf{S}_2 + \delta_3 \mathbf{S}_1$$

where the values of β_3 and δ_3 can be found by making $\mathbf{S}_3$ conjugate to $\mathbf{S}_1$ and $\mathbf{S}_2$. By using the condition $\mathbf{S}_1^T\mathbf{A}\mathbf{S}_3 = 0$, the value of δ_3 can be found to be zero When the condition $\mathbf{S}_2^T\mathbf{A}\mathbf{S}_3 = 0$ is used, the value of β_3 can be obtained as

$$\beta_3 = \frac{\nabla f_3^T \nabla f_3}{\nabla f_2^T \nabla f_2}$$

so that Eq. ( ) becomes

$$\mathbf{S}_3 = -\nabla f_3 + \beta_3 \mathbf{S}_2$$

where β_3 is given by Eq. ( ). In fact, Eq. ( ) can be generalized as

$$\mathbf{S}_i = -\nabla f_i + \beta_i \mathbf{S}_{i-1} \qquad (\quad)$$

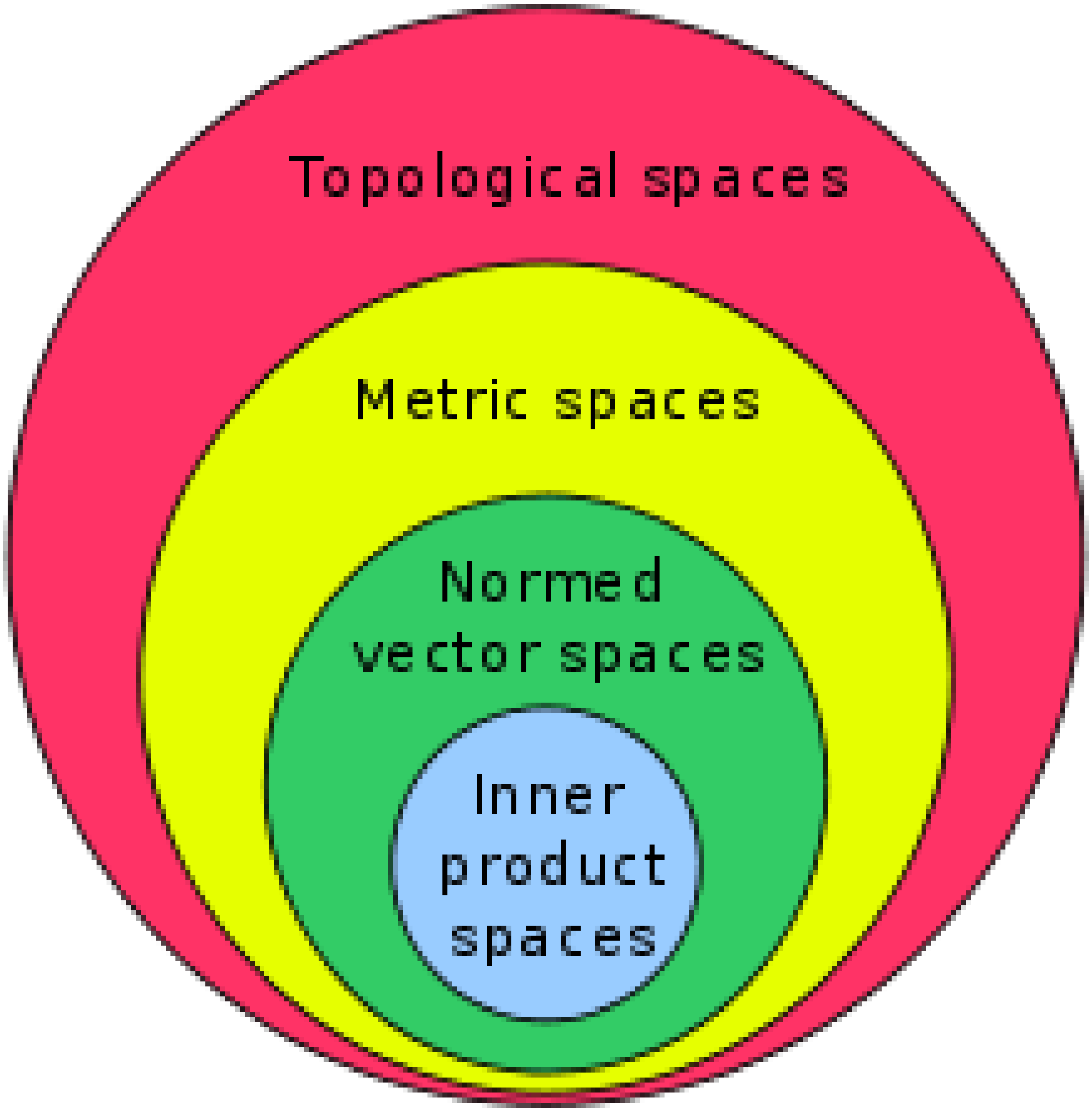
where

$$\beta_i = \frac{\nabla f_i^T \nabla f_i}{\nabla f_{i-1}^T \nabla f_{i-1}} \qquad (\quad)$$

Equations ( ) and ( ) define the search directions used in the Fletcher–Reeves method

	If we move along the gradient direction from any point in n-dimensional space, the function value increases at the fastest rate	□	□
	The gradient vector represents the direction of steepest ascent.	□	□
	The maximum rate of change of f at any point X is equal to the magnitude of the gradient vector at	□	□

	the same point		
	The convergence criteria of steepest descent method that can be used to terminate the iterative process, when the change in function value in two consecutive iterations is small is $\left\|\frac{f(X_{i+1})-f(X_i)}{f(X_i)}\right\| \leq \varepsilon..$	□	□
	All the unconstrained minimization methods are iterative in nature.	□	□



Introduction

to

Topology (1)

Chapter 1

Sets and Relations

SETS, ELEMENTS

The concept *set* appears in all branches of mathematics. Intuitively, a set is any well-defined list or collection of objects, and will be denoted by capital letters $A, B, X, Y, \ldots$. The objects comprising the set are called its *elements* or *members* and will be denoted by lower case letters $a, b, x, y, \ldots$. The statement "p is an element of A" or, equivalently, "p belongs to A" is written

$$p \in A$$

The negation of $p \in A$ is written $p \notin A$.

There are essentially two ways to specify a particular set. One way, if it is possible, is by actually listing its members. For example,

$$A = \{a, e, i, o, u\}$$

denotes the set A whose elements are the letters a, e, i, o and u. Note that the elements are separated by commas and enclosed in braces $\{\ \}$. The other way is by stating those properties which characterize the elements in the set. For example,

$$B = \{x : x \text{ is an integer, } x > 0\}$$

which reads "B is the set of x such that x is an integer and x is greater than zero," denotes the set B whose elements are the positive integers. A letter, usually x, is used to denote an arbitrary member of the set; the colon is read as 'such that' and the comma as 'and'.

Example 1.1: The set B above can also be written as $B = \{1, 2, 3, \ldots\}$. Note that $-6 \notin B$, $3 \in B$ and $\pi \notin B$.

Example 1.2: Intervals on the real line, defined below, appear very often in mathematics. Here a and b are real numbers with $a < b$.

Open interval from a to b	$= (a, b)$	$= \{x : a < x < b\}$
Closed interval from a to b	$= [a, b]$	$= \{x : a \leq x \leq b\}$
Open-closed interval from a to b	$= (a, b]$	$= \{x : a < x \leq b\}$
Closed-open interval from a to b	$= [a, b)$	$= \{x : a \leq x < b\}$

The open-closed and closed-open intervals are also called *half-open* intervals.

Two sets A and B are *equal,* written $A = B$, if they consist of the same elements, i.e. if each member of A belongs to B and each member of B belongs to A. The negation of $A = B$ is written $A \neq B$.

Example 1.3: Let $E = \{x : x^2 - 3x + 2 = 0\}$, $F = \{2, 1\}$ and $G = \{1, 2, 2, 1\}$. Then $E = F = G$. Observe that a set does not depend on the way in which its elements are displayed. A set remains the same if its elements are repeated or rearranged.

Sets can be *finite* or *infinite.* A set is finite if it consists of n different elements, where n is some positive integer; otherwise a set is infinite. In particular, a set which consists of exactly one element is called a *singleton set.*

SUBSETS, SUPERSETS

A set A is a *subset* of a set B or, equivalently, B is a *superset* of A, written

$$A \subset B \quad \text{or} \quad B \supset A$$

iff each element in A also belongs to B; that is, if $x \in A$ implies $x \in B$. We also say that A *is contained in* B or B *contains* A. The negation of $A \subset B$ is written $A \not\subset B$ or $B \not\supset A$ and states that there is an $x \in A$ such that $x \notin B$.

Example 2.1: Consider the sets

$$A = \{1, 3, 5, 7, \ldots\}, \qquad B = \{5, 10, 15, 20, \ldots\}$$

$$C = \{x : x \text{ is prime}, \; x > 2\} = \{3, 5, 7, 11, \ldots\}$$

Then $C \subset A$ since every prime number greater than 2 is odd. On the other hand, $B \not\subset A$ since $10 \in B$ but $10 \notin A$.

Example 2.2: We will let $\mathbf{N}$ denote the set of positive integers, $\mathbf{Z}$ denote the set of integers, $\mathbf{Q}$ denote the set of rational numbers and $\mathbf{R}$ denote the set of real numbers. Accordingly,

$$\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R}$$

Observe that $A \subset B$ does not exclude the possibility that $A = B$. In fact, we are able to restate the definition of equality of sets as follows:

Definition: Two sets A and B are equal if and only if $A \subset B$ and $B \subset A$.

In the case that $A \subset B$ but $A \neq B$, we say that A is a *proper subset* of B or B contains A properly. The reader should be warned that some authors use the symbol $\subseteq$ for a subset and the symbol $\subset$ only for a proper subset.

Our first theorem follows from the preceding definitions.

Theorem 1.1: Let A, B and C be any sets. Then (i) $A \subset A$; (ii) if $A \subset B$ and $B \subset A$ then $A = B$; and (iii) if $A \subset B$ and $B \subset C$ then $A \subset C$.

UNIVERSAL AND NULL SETS

In any application of the theory of sets, all sets under investigation are subsets of a fixed set. We call this set the *universal set* or *universe of discourse* and denote it in this chapter by U. It is also convenient to introduce the concept of the *empty* or *null set*, that is, a set which contains no elements. This set, denoted by $\emptyset$, is considered finite and a subset of every other set. Thus, for any set A, $\emptyset \subset A \subset U$.

Example 3.1: In plane geometry, the universal set consists of all the points in the plane.

Example 3.2: Let $A = \{x : x^2 = 4, \; x \text{ is odd}\}$. Then A is empty, i.e. $A = \emptyset$.

Example 3.3: Let $B = \{\emptyset\}$. Then $B \neq \emptyset$ for B contains one element.

CLASSES, COLLECTIONS, FAMILIES AND SPACES

Frequently, the members of a set are sets themselves. For example, each line in a set of lines is a set of points. To help clarify these situations, we use the words *"class"*, *"collection"* and *"family"* synonymously with set. Usually we use class for a set of sets, and *collection* or *family* for a set of classes. The words *subclass*, *subcollection* and *subfamily* have meanings analogous to subset.

Example 4.1: The members of the class $\{\{2, 3\}, \{2\}, \{5, 6\}\}$ are the sets $\{2, 3\}$, $\{2\}$ and $\{5, 6\}$.

Example 4.2: Consider any set A. The *power set* of A, denoted by $\mathcal{P}(A)$ or 2^A, is the class of all subsets of A. In particular, if $A = \{a, b, c\}$, then

$$\mathcal{P}(A) \quad = \quad \{A,\ \{a,b\},\ \{a,c\},\ \{b,c\},\ \{a\},\ \{b\},\ \{c\},\ \emptyset\}$$

In general, if A is finite, say A has n elements, then $\mathcal{P}(A)$ will have 2^n elements.

The word *space* shall mean a non-empty set which possesses some type of mathematical structure, e.g. vector space, metric space or topological space. In such a situation, we will call the elements in a space *points*.

SET OPERATIONS

The *union* of two sets A and B, denoted by $A \cup B$, is the set of all elements which belong to A or B, i.e.,

$$A \cup B \quad = \quad \{x : x \in A \text{ or } x \in B\}$$

Here "or" is used in the sense of "and/or".

The *intersection* of two sets A and B, denoted by $A \cap B$, is the set of elements which belong to both A and B, i.e.,

$$A \cap B \quad = \quad \{x : x \in A \text{ and } x \in B\}$$

If $A \cap B = \emptyset$, that is, if A and B do not have any elements in common, then A and B are said to be *disjoint* or *non-intersecting*. A class $\mathcal{A}$ of sets is called a *disjoint class of sets* if each pair of distinct sets in $\mathcal{A}$ is disjoint.

The *relative complement* of a set B with respect to a set A or, simply the *difference* of A and B, denoted by $A \setminus B$, is the set of elements which belong to A but which do not belong to B. In other words,

$$A \setminus B \quad = \quad \{x : x \in A,\ x \notin B\}$$

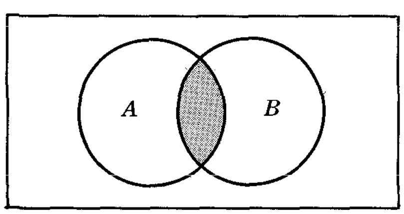
Observe that $A \setminus B$ and B are disjoint, i.e. $(A \setminus B) \cap B = \emptyset$.

The *absolute complement* or, simply, *complement* of a set A, denoted by A^c, is the set of elements which do not belong to A, i.e.,

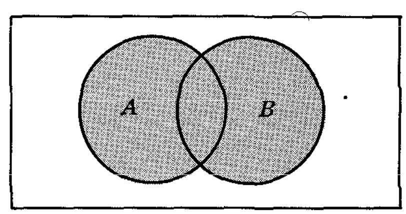
$$A^c \quad = \quad \{x : x \in U,\ x \notin A\}$$

In other words, A^c is the difference of the universal set U and A.

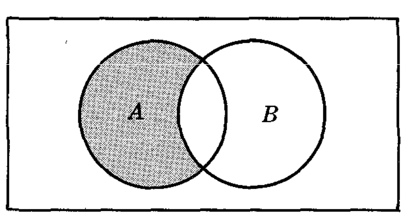
Example 5.1: The following diagrams, called Venn diagrams, illustrate the above set operations. Here sets are represented by simple plane areas and U, the universal set, by the area in the entire rectangle.



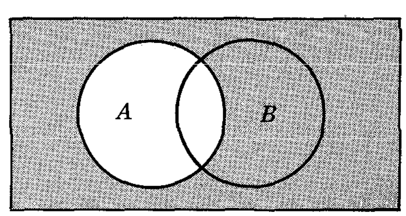
$A \cap B$ is shaded



$A \cup B$ is shaded



$A \setminus B$ is shaded



A^c is shaded

Sets under the above operations satisfy various laws or identities which are listed in the table below (Table 1). In fact, we state

Theorem 1.2: Sets satisfy the laws in Table 1.

LAWS OF THE ALGEBRA OF SETS	
Idempotent Laws	
1a. $A \cup A = A$	1b. $A \cap A = A$
Associative Laws	
2a. $(A \cup B) \cup C = A \cup (B \cup C)$	2b. $(A \cap B) \cap C = A \cap (B \cap C)$
Commutative Laws	
3a. $A \cup B = B \cup A$	3b. $A \cap B = B \cap A$
Distributive Laws	
4a. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	4b. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Identity Laws	
5a. $A \cup \emptyset = A$	5b. $A \cap U = A$
6a. $A \cup U = U$	6b. $A \cap \emptyset = \emptyset$
Complement Laws	
7a. $A \cup A^c = U$	7b. $A \cap A^c = \emptyset$
8a. $(A^c)^c = A$	8b. $U^c = \emptyset$, $\emptyset^c = U$
De Morgan's Laws	
9a. $(A \cup B)^c = A^c \cap B^c$	9b. $(A \cap B)^c = A^c \cup B^c$

Table 1

Remark: *Each of the above laws follows from an analogous logical law.* For example,

$$A \cap B = \{x : x \in A \text{ and } x \in B\} = \{x : x \in B \text{ and } x \in A\} = B \cap A$$

Here we use the fact that the composite statement "p and q", written $p \wedge q$, is logically equivalent to the composite statement "q and p", i.e. $q \wedge p$.

The relationship between set inclusion and the above set operations follows.

Theorem 1.3: Each of the following conditions is equivalent to $A \subset B$:

(i) $A \cap B = A$ (iii) $B^c \subset A^c$ (v) $B \cup A^c = U$

(ii) $A \cup B = B$ (iv) $A \cap B^c = \emptyset$

PRODUCT SETS

Let A and B be two sets. The *product set* of A and B, written $A \times B$, consists of all ordered pairs $\langle a, b \rangle$ where $a \in A$ and $b \in B$, i.e.,

$$A \times B = \{\langle a, b \rangle : a \in A,\ b \in B\}$$

The product of a set with itself, say $A \times A$, will be denoted by A^2.

Example 6.1: The reader is familiar with the Cartesian plane $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$ (Fig. 1-1 below). Here each point P represents an ordered pair $\langle a, b \rangle$ of real numbers and vice versa.

Example 6.2: Let $A = \{1, 2, 3\}$ and $B = \{a, b\}$. Then

$$A \times B = \{\langle 1, a \rangle, \langle 1, b \rangle, \langle 2, a \rangle, \langle 2, b \rangle, \langle 3, a \rangle, \langle 3, b \rangle\}$$

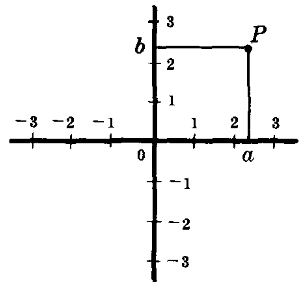


Fig. 1-1

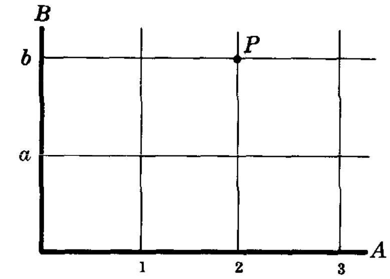


Fig. 1-2

Since A and B do not contain many elements, it is possible to represent $A \times B$ by a coordinate diagram as shown in Fig. 1-2 above. Here the vertical lines through the points of A and the horizontal lines through the points of B meet in 6 points which represent $A \times B$ in the obvious way. The point P is the ordered pair $\langle 2, b\rangle$. In general, if a set A has s elements and a set B has t elements, then $A \times B$ has s times t elements.

Remark: The notion "ordered pair" $\langle a, b\rangle$ is defined rigorously by $\langle a, b\rangle \equiv \{\{a\}, \{a, b\}\}$. From this definition, the "order" property may be proven:

$$\langle a, b\rangle = \langle c, d\rangle \quad \text{implies} \quad a = c \text{ and } b = d$$

The concept of product set can be extended to any finite number of sets in a natural way. The product set of the sets $A_1, \ldots, A_m$, denoted by

$$A_1 \times A_2 \times \cdots \times A_m \quad \text{or} \quad \prod_{i=1}^{m} A_i$$

consists of all m-tuples $\langle a_1, a_2, \ldots, a_m\rangle$ where $a_i \in A_i$ for each i.

RELATIONS

A *binary relation* (or *relation*) R from a set A to a set B assigns to each pair $\langle a, b\rangle$ in $A \times B$ exactly one of the following statements:

(i) "a is related to b", written $a\, R\, b$

(ii) "a is not related to b", written $a \not R\, b$

A relation from a set A to the same set A is called a *relation in* A.

Example 7.1: Set inclusion is a relation in any class of sets. For, given any pair of sets A and B, either $A \subset B$ or $A \not\subset B$.

Observe that any relation R from a set A to a set B uniquely defines a subset R^* of $A \times B$ as follows:

$$R^* = \{\langle a, b\rangle : a\, R\, b\}$$

On the other hand, any subset R^* of $A \times B$ defines a relation R from A to B as follows:

$$a\, R\, b \quad \text{iff} \quad \langle a, b\rangle \in R^*$$

In view of the correspondence between relations R from A to B and subsets of $A \times B$, we redefine a relation by

Definition: A relation R from A to B is a subset of $A \times B$.

The *domain* of a relation R from A to B is the set of first coordinates of the pairs in R and its *range* is the set of second coordinates, i.e.,

$$\text{domain of } R = \{a : \langle a, b\rangle \in R\}, \qquad \text{range of } R = \{b : \langle a, b\rangle \in R\}$$

The *inverse of* R, denoted by R^{-1}, is the relation from B to A defined by

$$R^{-1} = \{\langle b, a\rangle : \langle a, b\rangle \in R\}$$

Note that R^{-1} can be obtained by reversing the pairs in R.

Example 7.2: Consider the relation

$$R = \{\langle 1,2\rangle, \langle 1,3\rangle, \langle 2,3\rangle\}$$

in $A = \{1,2,3\}$. Then the domain of $R = \{1,2\}$, the range of $R = \{2,3\}$, and

$$R^{-1} = \{\langle 2,1\rangle, \langle 3,1\rangle, \langle 3,2\rangle\}$$

Observe that R and R^{-1} are identical, respectively, to the relations $<$ and $>$ in A, i.e.,

$$\langle a,b\rangle \in R \text{ iff } a < b \quad \text{and} \quad \langle a,b\rangle \in R^{-1} \text{ iff } a > b$$

The *identity relation* in any set A, denoted by Δ or Δ_A, is the set of all pairs in $A \times A$ with equal coordinates, i.e.,

$$\Delta_A = \{\langle a,a\rangle : a \in A\}$$

The identity relation is also called the *diagonal* by virtue of its position in a coordinate diagram of $A \times A$.

EQUIVALENCE RELATIONS

A relation R in a set A, i.e. a subset R of $A \times A$, is termed an *equivalence relation* iff it satisfies the following axioms:

[E₁] For every $a \in A$, $\langle a,a\rangle \in R$.

[E₂] If $\langle a,b\rangle \in R$, then $\langle b,a\rangle \in R$.

[E₃] If $\langle a,b\rangle \in R$ and $\langle b,c\rangle \in R$, then $\langle a,c\rangle \in R$.

In general, a relation is said to be *reflexive* iff it satisfies **[E₁]**, *symmetric* iff it satisfies **[E₂]** and *transitive* iff it satisfies **[E₃]**. Accordingly, a relation R is an equivalence relation iff it is reflexive, symmetric and transitive.

Example 8.1: Consider the relation $\subset$, i.e. set inclusion. Recall, by Theorem 1.1, that $A \subset A$ for every set A, and

$$\text{if } A \subset B \text{ and } B \subset C \text{ then } A \subset C$$

Hence $\subset$ is both reflexive and transitive. On the other hand,

$$A \subset B \text{ and } A \neq B \quad \text{implies} \quad B \not\subset A$$

Accordingly, $\subset$ is not symmetric and hence is not an equivalence relation.

Example 8.2: In Euclidian geometry, similarity of triangles is an equivalence relation. For if α, β and γ are any triangles then: (i) α is similar to itself; (ii) if α is similar to β, then β is similar to α; and (iii) if α is similar to β and β is similar to γ then α is similar to γ.

If R is an equivalence relation in A, then the *equivalence class* of any element $a \in A$, denoted by $[a]$, is the set of elements to which a is related:

$$[a] = \{x : \langle a,x\rangle \in R\}$$

The collection of equivalence classes of A, denoted by A/R, is called the *quotient* of A by R:

$$A/R = \{[a] : a \in A\}$$

The quotient set A/R possesses the following properties:

Theorem 1.4: Let R be an equivalence relation in A and let $[a]$ be the equivalence class of $a \in A$. Then:

(i) For every $a \in A$, $a \in [a]$.

(ii) $[a] = [b]$ if and only if $\langle a,b\rangle \in R$.

(iii) If $[a] \neq [b]$, then $[a]$ and $[b]$ are disjoint.

A class $\mathcal{A}$ of non-empty subsets of A is called a *partition* of A iff (1) each $a \in A$ belongs to some member of $\mathcal{A}$ and (2) the members of $\mathcal{A}$ are pair-wise disjoint. Accordingly, the previous theorem implies the following *fundamental theorem of equivalence relations*:

Theorem 1.5: Let R be an equivalence relation in A. Then the quotient set A/R is a partition of A.

Example 8.3: Let R_5 be the relation in $\mathbf{Z}$, the set of integers, defined by

$$x \equiv y \pmod 5$$

which reads "x is congruent to y modulo 5" and which means "$x-y$ is divisible by 5". Then R_5 is an equivalence relation in $\mathbf{Z}$. There are exactly five distinct equivalence classes in $\mathbf{Z}/R_5$:

$$\begin{aligned}
E_0 &= \{\ldots,-10,-5,0,5,10,\ldots\} = \cdots = [-10] = [-5] = [0] = [5] = \cdots \\
E_1 &= \{\ldots,-9,-4,1,6,11,\ldots\} = \cdots = [-9] = [-4] = [1] = [6] = \cdots \\
E_2 &= \{\ldots,-8,-3,2,7,12,\ldots\} = \cdots = [-8] = [-3] = [2] = [7] = \cdots \\
E_3 &= \{\ldots,-7,-2,3,8,13,\ldots\} = \cdots = [-7] = [-2] = [3] = [8] = \cdots \\
E_4 &= \{\ldots,-6,-1,4,9,14,\ldots\} = \cdots = [-6] = [-1] = [4] = [9] = \cdots
\end{aligned}$$

Observe that each integer x, which is uniquely expressible in the form $x = 5q + r$ where $0 \leq r < 5$, is a member of the equivalence class E_r where r is the remainder. Note that the equivalence classes are pairwise disjoint and that $\mathbf{Z} = E_0 \cup E_1 \cup E_2 \cup E_3 \cup E_4$.

COMPOSITION OF RELATIONS

Let U be a relation from A to B and let V be a relation from B to C, i.e. $U \subset A \times B$ and $V \subset B \times C$. Then the relation from A to C which consists of all ordered pairs $\langle a, c\rangle \in A \times C$ such that, for some $b \in B$,

$$\langle a, b\rangle \in U \quad \text{and} \quad \langle b, c\rangle \in V$$

is called the *composition* of U and V and is denoted by $V \circ U$. (The reader should be warned that some authors denote this relation by $U \circ V$.)

It is convenient to introduce some more symbols:

$\exists$, there exists s.t., such that $\forall$, for all $\Rightarrow$, implies

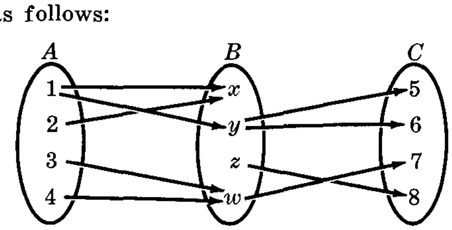
We may then write:

$$V \circ U = \{\langle x, y\rangle : x \in A,\ y \in C;\ \exists b \in B \text{ s.t. } \langle x, b\rangle \in U,\ \langle b, y\rangle \in V\}$$

Example 9.1: Let $A = \{1,2,3,4\}$, $B = \{x,y,z,w\}$ and $C = \{5,6,7,8\}$, and let

$$U = \{\langle 1,x\rangle, \langle 1,y\rangle, \langle 2,x\rangle, \langle 3,w\rangle, \langle 4,w\rangle\} \quad \text{and} \quad V = \{\langle y,5\rangle, \langle y,6\rangle, \langle z,8\rangle, \langle w,7\rangle\}$$

That is, U is a relation from A to B and V is a relation from B to C. We may illustrate U and V as follows:



Accordingly,

$$\begin{aligned}
\langle 1,5\rangle &\in V \circ U \quad \text{since} \quad y \in B \text{ and } \langle 1,y\rangle \in U,\ \langle y,5\rangle \in V \\
\langle 1,6\rangle &\in V \circ U \quad \text{since} \quad y \in B \text{ and } \langle 1,y\rangle \in U,\ \langle y,6\rangle \in V \\
\langle 3,7\rangle &\in V \circ U \quad \text{since} \quad w \in B \text{ and } \langle 3,w\rangle \in U,\ \langle w,7\rangle \in V \\
\langle 4,7\rangle &\in V \circ U \quad \text{since} \quad w \in B \text{ and } \langle 4,w\rangle \in U,\ \langle w,7\rangle \in V
\end{aligned}$$

No other ordered pairs belong to $V \circ U$, that is,

$$V \circ U = \{ \langle 1,5 \rangle, \langle 1,6 \rangle, \langle 3,7 \rangle, \langle 4,7 \rangle \}$$

Observe that $V \circ U$ consists precisely of those pairs $\langle x, y \rangle$ for which there exists, in the above diagram, a "path" from $x \in A$ to $y \in C$ composed of two arrows, one following the other.

Example 9.2: Let U and V be the relations in $\mathbf{R}$ defined by

$$U = \{\langle x, y \rangle : x^2 + y^2 = 1\} \quad \text{and} \quad V = \{\langle y, z \rangle : 2y + 3z = 4\}$$

Then the relation $V \circ U$, the composition of U and V, can be found by eliminating y from the two equations $x^2 + y^2 = 1$ and $2y + 3z = 4$. In other words,

$$V \circ U = \{\langle x, z \rangle : 4x^2 + 9z^2 - 24z + 12 = 0\}$$

Example 9.3: Let $\mathbf{N}$ denote the set of positive integers, and let R denote the relation $<$ in $\mathbf{N}$, i.e. $\langle a, b \rangle \in R$ iff $a < b$. Hence $\langle a, b \rangle \in R^{-1}$ iff $a > b$. Then

$$\begin{aligned} R \circ R^{-1} &= \{ \langle x, y \rangle : x,y \in \mathbf{N};\ \exists b \in \mathbf{N} \text{ s.t. } \langle x, b \rangle \in R^{-1}, \langle b, y \rangle \in R \} \\ &= \{ \langle x, y \rangle : x,y \in \mathbf{N};\ \exists b \in \mathbf{N} \text{ s.t. } b < x,\ b < y \} \\ &= (\mathbf{N} \setminus \{1\}) \times (\mathbf{N} \setminus \{1\}) = \{\langle x,y \rangle : x,y \in \mathbf{N};\ x,y \neq 1\} \end{aligned}$$

and

$$\begin{aligned} R^{-1} \circ R &= \{ \langle x, y \rangle : x,y \in \mathbf{N};\ \exists b \in \mathbf{N} \text{ s.t. } \langle x, b \rangle \in R, \langle b, y \rangle \in R^{-1} \} \\ &= \{ \langle x, y \rangle : x,y \in \mathbf{N};\ \exists b \in \mathbf{N} \text{ s.t. } b > x,\ b > y \} \\ &= \mathbf{N} \times \mathbf{N} \end{aligned}$$

Note that $R \circ R^{-1} \neq R^{-1} \circ R$.

Solved Problems

SETS, ELEMENTS, SUBSETS

1. Let $A = \{x : 3x = 6\}$. Does $A = 2$?

Solution:

A is the set which consists of the single element 2, i.e. $A = \{2\}$. The number 2 belongs to A; it does not equal A. There is a basic difference between an element p and the singleton set $\{p\}$.

2. Determine which of the following sets are equal: $\emptyset$, $\{0\}$, $\{\emptyset\}$.

Solution:

Each is different from the other. The set $\{0\}$ contains one element, the number zero. The set $\emptyset$ contains no elements; it is the null set. The set $\{\emptyset\}$ also contains one element, the null set.

3. Determine whether or not each of the following sets is the null set:
(i) $X = \{x : x^2 = 9, 2x = 4\}$, (ii) $Y = \{x : x \neq x\}$, (iii) $Z = \{x : x + 8 = 8\}$.

Solution:

(i) There is no number which satisfies both $x^2 = 9$ and $2x = 4$; hence $X = \emptyset$.

(ii) We assume that any object is itself, so Y is empty. In fact, some texts define the null set by $\emptyset \equiv \{x : x \neq x\}$.

(iii) The number zero satisfies $x + 8 = 8$; hence $Z = \{0\}$. Accordingly, $Z \neq \emptyset$.

4. Prove that $A = \{2, 3, 4, 5\}$ is not a subset of $B = \{x : x \text{ is even}\}$.

Solution:

It is necessary to show that at least one member of A does not belong to B. Since $3 \in A$ and $3 \notin B$, A is not a subset of B.

5. Prove Theorem 1.1 (iii): If $A \subset B$ and $B \subset C$ then $A \subset C$.

Solution:

We must show that each element in A also belongs to C. Let $x \in A$. Now $A \subset B$ implies $x \in B$. But $B \subset C$, so $x \in C$. We have therefore shown that $x \in A$ implies $x \in C$, or $A \subset C$.

6. Prove: If A is a subset of the null set $\emptyset$, then $A = \emptyset$.

Solution:

The null set $\emptyset$ is a subset of every set; in particular, $\emptyset \subset A$. But, by hypothesis, $A \subset \emptyset$; hence, by Definition 1.1, $A = \emptyset$.

7. Find the power set $\mathcal{P}(S)$ of the set $S = \{1, 2, 3\}$.

Solution:

Recall that the power set $\mathcal{P}(S)$ of S is the class of all subsets of S. The subsets of S are $\{1,2,3\}$, $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{1\}$, $\{2\}$, $\{3\}$ and the empty set $\emptyset$. Hence

$$\mathcal{P}(S) = \{S, \{1,3\}, \{2,3\}, \{1,2\}, \{1\}, \{2\}, \{3\}, \emptyset\}$$

Note that there are $2^3 = 8$ subsets of S.

8. Find the power set $\mathcal{P}(S)$ of $S = \{3, \{1,4\}\}$.

Solution:

Note first that S contains two elements, 3 and the set $\{1,4\}$. Therefore $\mathcal{P}(S)$ contains $2^2 = 4$ elements: S itself, the empty set $\emptyset$, the singleton set $\{3\}$ containing 3 and the singleton set $\{\{1,4\}\}$ containing the set $\{1,4\}$. In other words,

$$\mathcal{P}(S) = \{S, \{3\}, \{\{1,4\}\}, \emptyset\}$$

SET OPERATIONS

9. Let $U = \{1, 2, \ldots, 8, 9\}$, $A = \{1, 2, 3, 4\}$, $B = \{2, 4, 6, 8\}$ and $C = \{3, 4, 5, 6\}$. Find: (i) A^c, (ii) $(A \cap C)^c$, (iii) $B \setminus C$, (iv) $(A \cup B)^c$.

Solution:

(i) A^c consists of the elements in U that are not in A; hence $A^c = \{5, 6, 7, 8, 9\}$.

(ii) $A \cap C$ consists of the elements in both A and C; hence

$$A \cap C = \{3,4\} \quad \text{and} \quad (A \cap C)^c = \{1, 2, 5, 6, 7, 8, 9\}$$

(iii) $B \setminus C$ consists of the elements in B which are not in C; hence $B \setminus C = \{2, 8\}$.

(iv) $A \cup B$ consists of the elements in A or B (or both); hence

$$A \cup B = \{1, 2, 3, 4, 6, 8\} \quad \text{and} \quad (A \cup B)^c = \{5, 7, 9\}$$

10. Prove: $(A \setminus B) \cap B = \emptyset$.

Solution:

$$\begin{aligned} (A \setminus B) \cap B &= \{x : x \in B, \ x \in A \setminus B\} \\ &= \{x : x \in B, \ x \in A, \ x \notin B\} = \emptyset \end{aligned}$$

since there is no element x satisfying $x \in B$ and $x \notin B$.

11. Prove De Morgan's Law: $(A \cup B)^c = A^c \cap B^c$.

Solution:

$$\begin{aligned} (A \cup B)^c &= \{x : x \notin A \cup B\} \\ &= \{x : x \notin A, \ x \notin B\} \\ &= \{x : x \in A^c, \ x \in B^c\} = A^c \cap B^c \end{aligned}$$

12. Prove: $B \setminus A = B \cap A^c$.

Solution:

$$B \setminus A = \{x : x \in B, \ x \notin A\} = \{x : x \in B, \ x \in A^c\} = B \cap A^c$$

13. Prove the Distributive Law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Solution:

$$\begin{aligned} A \cap (B \cup C) &= \{x : x \in A;\ x \in B \cup C\} \\ &= \{x : x \in A;\ x \in B \text{ or } x \in C\} \\ &= \{x : x \in A,\ x \in B;\ \text{or } x \in A,\ x \in C\} \\ &= \{x : x \in A \cap B \text{ or } x \in A \cap C\} \\ &= (A \cap B) \cup (A \cap C) \end{aligned}$$

Observe that in the third step above we used the analogous logical law

$$p \wedge (q \vee r) = (p \wedge q) \vee (p \wedge r)$$

where $\wedge$ reads "and" and $\vee$ reads "or".

14. Prove: For any sets A and B, $A \cap B \subset A \subset A \cup B$.

Solution:

Let $x \in A \cap B$; then $x \in A$ and $x \in B$. In particular, $x \in A$. Accordingly, $A \cap B \subset A$. If $x \in A$, then $x \in A$ or $x \in B$, i.e. $x \in A \cup B$. Hence $A \subset A \cup B$. In other words, $A \cap B \subset A \subset A \cup B$.

15. Prove Theorem 1.3 (i): $A \subset B$ if and only if $A \cap B = A$.

Solution:

Suppose $A \subset B$. Let $x \in A$; then by hypothesis, $x \in B$. Hence $x \in A$ and $x \in B$, i.e. $x \in A \cap B$. Accordingly, $A \subset A \cap B$. But by the previous problem, $A \cap B \subset A$. Hence $A \cap B = A$.

On the other hand, suppose $A \cap B = A$. Then in particular, $A \subset A \cap B$. But, by the previous problem, $A \cap B \subset B$. Hence, by Theorem 1.1, $A \subset B$.

PRODUCT SETS, RELATIONS, COMPOSITION OF RELATIONS

16. Let $A = \{a, b\}$, $B = \{2, 3\}$ and $C = \{3, 4\}$. Find: (i) $A \times (B \cup C)$, (ii) $(A \times B) \cup (A \times C)$.

Solution:

(i) First compute $B \cup C = \{2, 3, 4\}$. Then

$$A \times (B \cup C) = \{ \langle a,2\rangle, \langle a,3\rangle, \langle a,4\rangle, \langle b,2\rangle, \langle b,3\rangle, \langle b,4\rangle \}$$

(ii) First find $A \times B$ and $A \times C$:

$$A \times B = \{ \langle a,2\rangle, \langle a,3\rangle, \langle b,2\rangle, \langle b,3\rangle \}, \quad A \times C = \{ \langle a,3\rangle, \langle a,4\rangle, \langle b,3\rangle, \langle b,4\rangle \}$$

Then compute the union of the two sets:

$$(A \times B) \cup (A \times C) = \{ \langle a,2\rangle, \langle a,3\rangle, \langle b,2\rangle, \langle b,3\rangle, \langle a,4\rangle, \langle b,4\rangle \}$$

Observe, from (i) and (ii), that $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

17. Prove: $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

Solution:

$$\begin{aligned} A \times (B \cap C) &= \{ \langle x,y\rangle : x \in A,\ y \in B \cap C \} \\ &= \{ \langle x,y\rangle : x \in A,\ y \in B,\ y \in C \} \\ &= \{ \langle x,y\rangle : \langle x,y\rangle \in A \times B,\ \langle x,y\rangle \in A \times C \} \\ &= (A \times B) \cap (A \times C) \end{aligned}$$

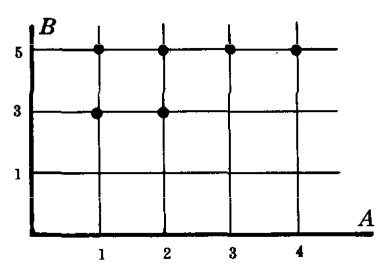
18. Let R be the relation $<$ from $A = \{1, 2, 3, 4\}$ to $B = \{1, 3, 5\}$, i.e., $\langle a, b\rangle \in R$ iff $a < b$.

(i) Write R as a set of ordered pairs.

(ii) Plot R on a coordinate diagram of $A \times B$.

(iii) Find domain of R, range of R and R^{-1}.

(iv) Find $R \circ R^{-1}$.



Solution:

(i) R consists of those ordered pairs $\langle a, b\rangle \in A \times B$ such that $a < b$; hence

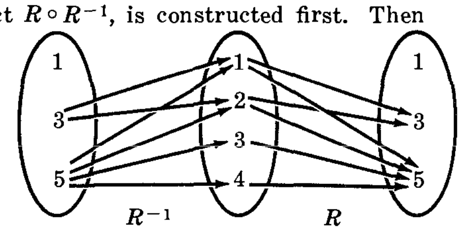
$$R = \{\langle 1,3\rangle, \langle 1,5\rangle, \langle 2,3\rangle, \langle 2,5\rangle, \langle 3,5\rangle, \langle 4,5\rangle\}$$

(ii) R is displayed on the coordinate diagram of $A \times B$ as shown above.

(iii) The domain of R is the set of first coordinates of the pairs in R; hence domain of $R = \{1,2,3,4\}$. The range of R is the set of second coordinates of the pairs in R; hence range of $R = \{3,5\}$. R^{-1} can be obtained from R by reversing the pairs in R; hence

$$R^{-1} = \{\langle 3,1\rangle, \langle 5,1\rangle, \langle 3,2\rangle, \langle 5,2\rangle, \langle 5,3\rangle, \langle 5,4\rangle\}$$

(iv) To find $R \circ R^{-1}$, construct diagrams of R^{-1} and R as shown below. Observe that R^{-1}, the second factor in the product $R \circ R^{-1}$, is constructed first. Then



$$R \circ R^{-1} = \{\langle 3,3\rangle, \langle 3,5\rangle, \langle 5,3\rangle, \langle 5,5\rangle\}$$

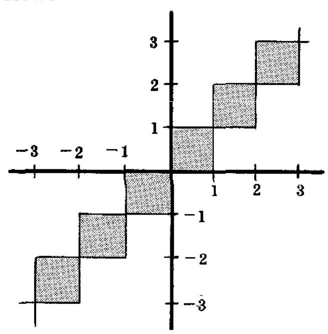
19. Let T be the relation in the set of real numbers $\mathbf{R}$ defined by

$$x\,T\,y \text{ if both } x \in [n, n+1] \text{ and } y \in [n, n+1] \text{ for some integer } n$$

Graph the relation T.

Solution:

T consists of the shaded squares below.



20. Let T be the relation in the set of real numbers $\mathbf{R}$ defined by $x\,T\,y$ iff $0 \leq x - y \leq 1$.

(i) Express T and T^{-1} as subsets of $\mathbf{R} \times \mathbf{R}$ and graph.

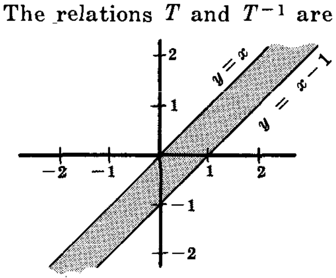
(ii) Show that $T \circ T^{-1} = \{\langle x, z\rangle : |x - z| \leq 1\}$.

Solution:

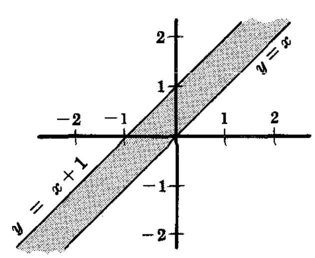
(i) $T = \{\langle x,y\rangle : x,y \in \mathbf{R},\ 0 \leq x - y \leq 1\}$

$$T^{-1} = \{\langle x,y\rangle : \langle y,x\rangle \in T\} = \{\langle x,y\rangle : x,y \in \mathbf{R},\ 0 \leq y - x \leq 1\}$$

The relations T and T^{-1} are graphed below.



Graph of T



Graph of T^{-1}

(ii) By definition of composition of relations,

$$\begin{aligned} T \circ T^{-1} &= \{\langle x,z\rangle : \exists y \in \mathbf{R} \text{ s.t. } \langle x,y\rangle \in T^{-1},\ \langle y,z\rangle \in T\} \\ &= \{\langle x,z\rangle : \exists y \in \mathbf{R} \text{ s.t. } \langle y,x\rangle,\ \langle y,z\rangle \in T\} \\ &= \{\langle x,z\rangle : \exists y \in \mathbf{R} \text{ s.t. } 0 \leq y-x \leq 1,\ 0 \leq y-z \leq 1\} \end{aligned}$$

Let $S = \{\langle x,z\rangle : |x-z| \leq 1\}$. We want to show that $T \circ T^{-1} = S$.

Let $\langle x,z\rangle$ belong to $T \circ T^{-1}$. Then $\exists y$ s.t. $0 \leq y-x,\ y-z \leq 1$. But

$$\begin{aligned} 0 \leq y-x,\ y-z \leq 1 &\Rightarrow y-z \leq 1 \\ &\Rightarrow y-z \leq 1+y-x \\ &\Rightarrow x-z \leq 1 \end{aligned}$$

Also,

$$\begin{aligned} 0 \leq y-x,\ y-z \leq 1 &\Rightarrow y-x \leq 1 \\ &\Rightarrow y-x \leq 1+y-z \\ &\Rightarrow -1 \leq x-z \end{aligned}$$

In other words, $0 \leq y-x,\ y-z \leq 1 \ \Rightarrow\ -1 \leq x-z \leq 1$ iff $|x-z| \leq 1$

Accordingly, $\langle x,z\rangle \in S$, i.e. $T \circ T^{-1} \subset S$.

Now let $\langle x,z\rangle$ belong to S; then $|x-z| \leq 1$.

Let $y = \max(x,z)$; then $0 \leq y-x \leq 1$ and $0 \leq y-z \leq 1$.

Thus $\langle x,z\rangle$ also belongs to $T \circ T^{-1}$, i.e. $S \subset T \circ T^{-1}$. Hence $T \circ T^{-1} = S$.

21. Prove: For any two relations $R \subset X \times Y$ and $S \subset Y \times Z$, $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$.

Solution:

$$\begin{aligned} (S \circ R)^{-1} &= \{\langle z,x\rangle : \langle x,z\rangle \in S \circ R\} \\ &= \{\langle z,x\rangle : \exists y \in Y \text{ s.t. } \langle x,y\rangle \in R,\ \langle y,z\rangle \in S\} \\ &= \{\langle z,x\rangle : \exists y \in Y \text{ s.t. } \langle z,y\rangle \in S^{-1},\ \langle y,x\rangle \in R^{-1}\} \\ &= R^{-1} \circ S^{-1} \end{aligned}$$

22. Prove: For any three relations $R \subset W \times X$, $S \subset X \times Y$ and $T \subset Y \times Z$, $(T \circ S) \circ R = T \circ (S \circ R)$.

Solution:

$$\begin{aligned} (T \circ S) \circ R &= \{\langle w,z\rangle : \exists x \in X \text{ s.t. } \langle w,x\rangle \in R,\ \langle x,z\rangle \in T \circ S\} \\ &= \{\langle w,z\rangle : \exists x \in X,\ \exists y \in Y \text{ s.t. } \langle w,x\rangle \in R,\ \langle x,y\rangle \in S,\ \langle y,z\rangle \in T\} \\ &= \{\langle w,z\rangle : \exists y \in Y \text{ s.t. } \langle w,y\rangle \in S \circ R,\ \langle y,z\rangle \in T\} \\ &= T \circ (S \circ R) \end{aligned}$$

REFLEXIVE, SYMMETRIC, TRANSITIVE AND EQUIVALENCE RELATIONS

23. Prove: Let R be a relation in A, i.e. $R \subset A \times A$. Then:

(i) R is reflexive iff $\Delta_A \subset R$;
(ii) R is symmetric iff $R = R^{-1}$;
(iii) R is transitive iff $R \circ R \subset R$;
(iv) R reflexive implies $R \circ R \supset R$ and $R \circ R$ is reflexive;
(v) R symmetric implies $R \circ R^{-1} = R^{-1} \circ R$;
(vi) R transitive implies $R \circ R$ is transitive.

Solution:

(i) Recall that the diagonal $\Delta_A = \{\langle a,a\rangle : a \in A\}$. Now R is reflexive iff, for every $a \in A$, $\langle a,a\rangle \in R$ iff $\Delta_A \subset R$.

(ii) Follows directly from the definition of R^{-1} and symmetric.

(iii) Let $\langle a,c\rangle \in R \circ R$; then $\exists b \in A$ such that $\langle a,b\rangle \in R$ and $\langle b,c\rangle \in R$. But, by transitivity, $\langle a,b\rangle, \langle b,c\rangle \in R$ implies $\langle a,c\rangle \in R$. Consequently, $R \circ R \subset R$.

On the other hand, suppose $R \circ R \subset R$. If $\langle a,b\rangle, \langle b,c\rangle \in R$, then $\langle a,c\rangle \in R \circ R \subset R$. In other words, R is transitive.

(iv) Let $\langle a,b\rangle \in R$. Now, $R \circ R = \{\langle a,c\rangle : \exists b \in A \text{ s.t. } \langle a,b\rangle \in R,\ \langle b,c\rangle \in R\}$.

But $\langle a,b\rangle \in R$ and, since R is reflexive, $\langle b,b\rangle \in R$. Thus $\langle a,b\rangle \in R \circ R$, i.e. $R \subset R \circ R$. Furthermore, $\Delta_A \subset R \subset R \circ R$ implies $R \circ R$ is also reflexive.

(v)
$$\begin{aligned} R \circ R^{-1} &= \{\langle a,c\rangle : \exists b \in A \text{ s.t. } \langle a,b\rangle \in R^{-1},\ \langle b,c\rangle \in R\} \\ &= \{\langle a,c\rangle : \exists b \in A \text{ s.t. } \langle a,b\rangle \in R,\ \langle b,c\rangle \in R^{-1}\} \\ &= R^{-1} \circ R \end{aligned}$$

(vi) Let $\langle a,b\rangle, \langle b,c\rangle \in R \circ R$. By (iii), $R \circ R \subset R$; hence $\langle a,b\rangle, \langle b,c\rangle \in R$. So $\langle a,c\rangle \in R \circ R$, i.e. $R \circ R$ is transitive.

24. Consider the relation $R = \{\langle 1,1\rangle, \langle 2,3\rangle, \langle 3,2\rangle\}$ in $X = \{1,2,3\}$. Determine whether or not R is *(i) reflexive, (ii) symmetric, (iii) transitive.*

Solution:

(i) R is not reflexive since $2 \in X$ but $\langle 2,2\rangle \notin R$.

(ii) R is symmetric since $R^{-1} = R$.

(iii) R is not transitive since $\langle 3,2\rangle \in R$ and $\langle 2,3\rangle \in R$ but $\langle 3,3\rangle \notin R$.

25. Consider the set $\mathbf{N} \times \mathbf{N}$, i.e. the set of ordered pairs of positive integers. Let R be the relation $\simeq$ in $\mathbf{N} \times \mathbf{N}$ which is defined by

$$\langle a,b\rangle \simeq \langle c,d\rangle \quad \text{iff} \quad ad = bc$$

Prove that R is an equivalence relation.

Solution:

Note that, for every $\langle a,b\rangle \in \mathbf{N} \times \mathbf{N}$, $\langle a,b\rangle \simeq \langle a,b\rangle$ since $ab = ba$; hence R is reflexive.

Suppose $\langle a,b\rangle \simeq \langle c,d\rangle$. Then $ad = bc$, which implies $cb = da$. Hence $\langle c,d\rangle \simeq \langle a,b\rangle$ and, therefore R is symmetric.

Now suppose $\langle a,b\rangle \simeq \langle c,d\rangle$ and $\langle c,d\rangle \simeq \langle e,f\rangle$. Then $ad = bc$ and $cf = de$. Thus

$$(ad)(cf) = (bc)(de)$$

and, by cancelling from both sides, $af = be$. Accordingly, $\langle a,b\rangle \simeq \langle e,f\rangle$ and R is transitive.

Since R is reflexive, symmetric and transitive, R is an equivalence relation.

Observe that if the ordered pair $\langle a,b\rangle$ is written as a fraction $\frac{a}{b}$, then the above relation R is, in fact, the usual definition of the equality of two fractions, i.e. $\frac{a}{b} = \frac{c}{d}$, iff $ad = bc$.

26. Prove Theorem 1.4: Let R be an equivalence relation in A and let $[a]$ be the equivalence class of $a \in A$. Then:

(i) For every $a \in A$, $a \in [a]$.

(ii) $[a] = [b]$ if and only if $\langle a,b\rangle \in R$.

(iii) If $[a] \neq [b]$, then $[a]$ and $[b]$ are disjoint.

Solution:

Proof of (i). Since R is reflexive, $\langle a,a\rangle \in R$ for every $a \in A$ and therefore $a \in [a]$.

Proof of (ii). Suppose $\langle a,b\rangle \in R$. We want to show that $[a] = [b]$. Let $x \in [b]$; then $\langle b,x\rangle \in R$. But by hypothesis, $\langle a,b\rangle \in R$; hence by transitivity, $\langle a,x\rangle \in R$. Accordingly, $x \in [a]$, i.e. $[b] \subset [a]$. To prove that $[a] \subset [b]$, we observe that $\langle a,b\rangle \in R$ implies, by symmetry, that $\langle b,a\rangle \in R$. Then by a similar argument, we get $[a] \subset [b]$. So $[a] = [b]$.

On the other hand, if $[a] = [b]$, then by reflexivity, $b \in [b] = [a]$, i.e. $\langle a,b\rangle \in R$.

Proof of (iii). We prove the equivalent contrapositive statement, i.e. if $[a] \cap [b] \neq \emptyset$, then $[a] = [b]$. If $[a] \cap [b] \neq \emptyset$, there exists an element $x \in A$ with $x \in [a] \cap [b]$. Hence $\langle a,x\rangle \in R$ and $\langle b,x\rangle \in R$. By symmetry, $\langle x,b\rangle \in R$ and, by transitivity, $\langle a,b\rangle \in R$. Consequently by (ii), $[a] = [b]$.

Supplementary Problems

SETS, ELEMENTS, SUBSETS

27. Determine which of the following sets is the empty set:

(i) $\{x : 1 < x < 2,\ x \in \mathbf{R}\}$ (iii) $\{x : x \in \emptyset\}$

(ii) $\{x : 1 < x < 2,\ x \in \mathbf{N}\}$ (iv) $\{x : x^2 < x,\ x \in \mathbf{R}\}$

28. Let $A = \{1, 2, \ldots, 8, 9\}$, $B = \{2, 4, 6, 8\}$, $C = \{1, 3, 5, 7, 9\}$, $D = \{3, 4, 5\}$ and $E = \{3, 5\}$. Which of these sets can equal X if we are given the following information?

(i) X and B are disjoint, (ii) $X \subset D$ and $X \not\subset B$, (iii) $X \subset A$ and $X \not\subset C$, (iv) $X \subset C$ and $X \not\subset A$.

29. State whether each of the following statements is true or false.

(i) Every subset of a finite set is finite. (ii) Every subset of an infinite set is infinite.

30. Discuss *all* inclusions and membership relations among the following three sets: $\emptyset$, $\{\emptyset\}$, $\{\emptyset, \{\emptyset\}\}$.

31. Prove that the closed interval $[a, b]$ is not a subset of the open interval (a, b).

32. Find the power set $\mathcal{P}(U)$ of $U = \{0, 1, 2\}$ and the power set $\mathcal{P}(V)$ of $V = \{0, \{1, 2\}\}$.

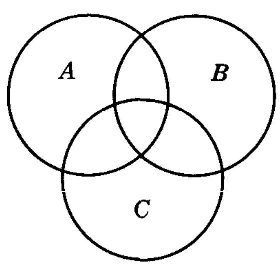
33. State whether each of the following is true or false. Here S is any non-empty set and 2^S is the power set of S.

(i) $S \in 2^S$ (ii) $S \subset 2^S$ (iii) $\{S\} \in 2^S$ (iv) $\{S\} \subset 2^S$

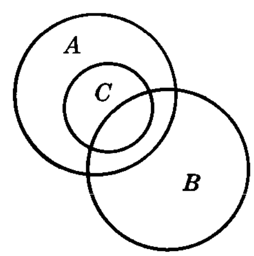
SET OPERATIONS

34. Let $A = \{1, 2, 3, \{1, 2, 3\}\}$, $B = \{1, 2, \{1, 2\}\}$. Find: $A \cup B$, $A \cap B$, $A \setminus B$, $B \setminus A$.

35. In each of the Venn diagrams below shade: (i) $A \cap (B \cup C)$, (ii) $C \setminus (A \cap B)$.



(*a*)



(*b*)

36. Prove and show by Venn diagrams: $A^c \setminus B^c = B \setminus A$.

37. (i) Prove $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C)$.

(ii) Give an example to show that $A \cup (B \setminus C) \neq (A \cup B) \setminus (A \cup C)$.

38. Prove: $2^A \cap 2^B = 2^{A \cap B}$; $2^A \cup 2^B \subset 2^{A \cup B}$. Give an example to show that $2^A \cup 2^B \neq 2^{A \cup B}$.

39. Prove Theorem 1.3: Each of the following conditions is equivalent to $A \subset B$:

(i) $A \cap B = A$, (ii) $A \cup B = B$, (iii) $B^c \subset A^c$, (iv) $A \cap B^c = \emptyset$, (v) $B \cup A^c = U$

(*Note.* $A \cap B = A$ was already proven equivalent to $A \subset B$ in Problem 15.)

40. Prove that $A \subset B$ iff $(B \cap C) \cup A = B \cap (C \cup A)$ for any C.

PRODUCT SETS, RELATIONS, COMPOSITION OF RELATIONS

41. Prove: $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

42. Using the definition of ordered pair, i.e. $\langle a, b\rangle = \{\{a\}, \{a, b\}\}$, prove that $\langle a, b\rangle = \langle c, d\rangle$ iff $a = c$ and $b = d$.

43. Determine the number of distinct relations from a set with m elements to a set with n elements, where m and n are positive integers.

44. Let R be the relation in the positive integers $\mathbf{N}$ defined by

$$R = \{ \langle x, y \rangle : x, y \in \mathbf{N},\ x + 2y = 12 \}$$

(i) Write R as a set of ordered pairs. (ii) Find domain of R, range of R and R^{-1}. (iii) Find $R \circ R$. (iv) Find $R^{-1} \circ R$.

45. Consider the relation $R = \{ \langle 4,5 \rangle, \langle 1,4 \rangle, \langle 4,6 \rangle, \langle 7,6 \rangle, \langle 3,7 \rangle \}$ in $\mathbf{N}$.
(i) Find domain of R, range of R and R^{-1}. (ii) Find $R \circ R$. (iii) Find $R^{-1} \circ R$.

46. Let U and V be the relations in $\mathbf{R}$ defined by $U = \{\langle x, y \rangle : x^2 + 2y = 5\}$ and $V = \{\langle x, y \rangle : 2x - y = 3\}$.
(i) Find $V \circ U$. (ii) Find $U \circ V$.

47. Consider the relations $<$ and $\leq$ in $\mathbf{R}$. Show that $< \cup \Delta = \leq$ where Δ is the diagonal.

EQUIVALENCE RELATIONS

48. State whether each of the following statements is true or false. Assume R and S are (non-empty) relations in a set A.

(1) If R is symmetric, then R^{-1} is symmetric.
(2) If R is reflexive, then $R \cap R^{-1} \neq \emptyset$.
(3) If R is symmetric, then $R \cap R^{-1} \neq \emptyset$.
(4) If R and S are transitive, then $R \cup S$ is transitive.
(5) If R and S are transitive, then $R \cap S$ is transitive.
(6) If R and S are symmetric, then $R \cup S$ is symmetric.
(7) If R and S are symmetric, then $R \cap S$ is symmetric.
(8) If R and S are reflexive, then $R \cap S$ is reflexive.

49. Consider $\mathbf{N} \times \mathbf{N}$, the set of ordered pairs of positive integers. Let $\simeq$ be the relation in $\mathbf{N} \times \mathbf{N}$ defined by

$$\langle a, b \rangle \simeq \langle c, d \rangle \quad \text{iff} \quad a + d = b + c$$

(i) Prove $\simeq$ is an equivalence relation. (ii) Find the equivalence class of $\langle 2, 5 \rangle$, i.e. $[\langle 2, 5 \rangle]$.

50. Let $\sim$ be the relation in $\mathbf{R}$ defined by $x \sim y$ iff $x - y$ is an integer. Prove that $\sim$ is an equivalence relation.

51. Let $\sim$ be the relation in the Cartesian plane $\mathbf{R}^2$ defined by $\langle x, y \rangle \sim \langle w, z \rangle$ iff $x = w$.
Prove that $\sim$ is an equivalence relation and graph several equivalence classes.

52. Let a and b be arbitrary real numbers. Furthermore, let $\sim$ be the relation in $\mathbf{R}^2$ defined by

$$\langle x, y \rangle \sim \langle w, z \rangle \quad \text{iff} \quad \exists k \in \mathbf{Z} \text{ s.t. } x - w = ka,\ y - z = kb$$

Prove that $\sim$ is an equivalence relation and graph several equivalence classes.

Answers to Supplementary Problems

27. The sets in (ii) and (iii) are empty.

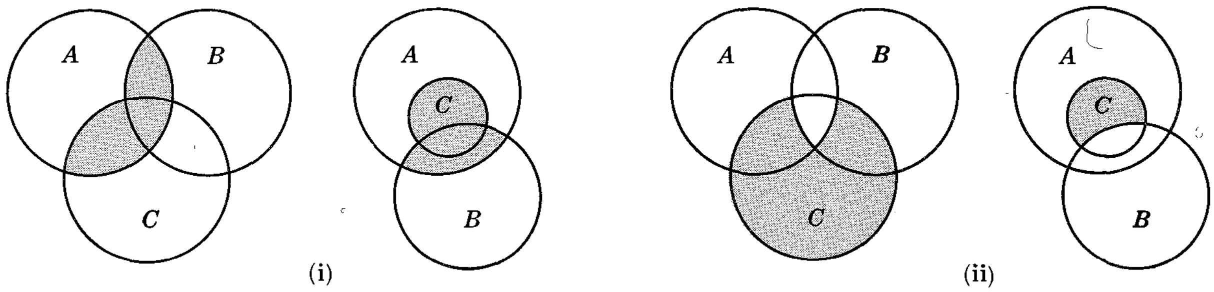
31. $a \in [a, b]$ but $a \notin (a, b)$.

32. $\mathcal{P}(V) = \{V, \{0\}, \{\{1, 2\}\}, \emptyset\}$

33. (i) T, (ii) F, (iii) F, (iv) T

34. $A \cup B = \{1, 2, 3, \{1, 2\}, \{1, 2, 3\}\}$, $A \cap B = \{1, 2\}$, $A \setminus B = \{3, \{1, 2, 3\}\}$, $B \setminus A = \{\{1, 2\}\}$.

35.



37. (ii) $C = \emptyset$, $A = B \neq \emptyset$

38. Example: $A = \{1\}$, $B = \{2\}$

43. 2^{mn}

44. (i) $R = \{\langle 10,1\rangle, \langle 8,2\rangle, \langle 6,3\rangle, \langle 4,4\rangle, \langle 2,5\rangle\}$

(ii) domain of $R = \{10,8,6,4,2\}$, range of $R = \{1,2,3,4,5\}$,
$R^{-1} = \{\langle 1,10\rangle, \langle 2,8\rangle, \langle 3,6\rangle, \langle 4,4\rangle, \langle 5,2\rangle\}$

(iii) $R \circ R = \{\langle 8,5\rangle, \langle 4,4\rangle\}$

(iv) $R^{-1} \circ R = \{\langle 10,10\rangle, \langle 8,8\rangle, \langle 6,6\rangle, \langle 4,4\rangle, \langle 2,2\rangle\}$

45. (i) domain of $R = \{4,1,7,3\}$, range of $R = \{5,4,6,7\}$, $R^{-1} = \{\langle 5,4\rangle, \langle 4,1\rangle, \langle 6,4\rangle, \langle 6,7\rangle, \langle 7,3\rangle\}$

(ii) $R \circ R = \{\langle 1,5\rangle, \langle 1,6\rangle, \langle 3,6\rangle\}$

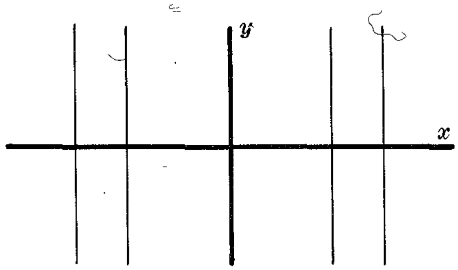
(iii) $R^{-1} \circ R = \{\langle 4,4\rangle, \langle 1,1\rangle, \langle 4,7\rangle, \langle 7,4\rangle, \langle 7,7\rangle, \langle 3,3\rangle\}$

46. $V \circ U = \{\langle x,y\rangle : x^2 + y = 2\}$, $U \circ V = \{\langle x,y\rangle : 4x^2 - 12x + 2y + 4 = 0\}$

48. (1) T, (2) T, (3) T, (4) F, (5) T, (6) T, (7) T, (8) T

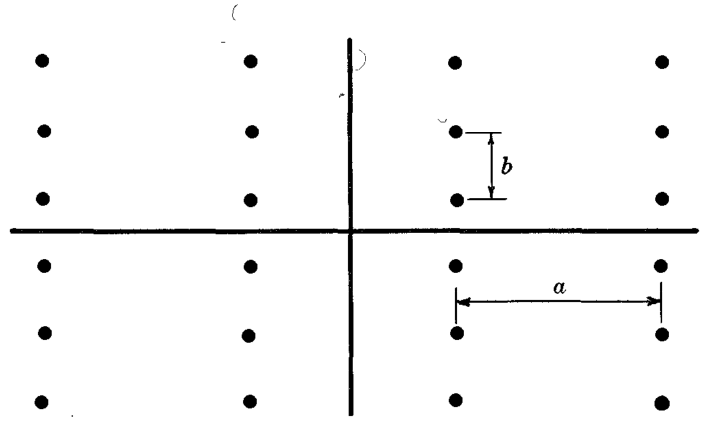
49. (ii) $[\langle 2,5\rangle] = \{\langle 1,4\rangle, \langle 2,5\rangle, \langle 3,6\rangle, \langle 4,7\rangle, \ldots, \langle n, n+3\rangle, \ldots\}$

51.



The equivalence classes are the vertical lines.

52.



The above gives a typical equivalence class. The distance between adjacent horizontal points is a and the distance between adjacent vertical points is b.

Chapter 2

Functions

FUNCTIONS

Suppose that to each element of a set A there is assigned a unique element of a set B; the collection, f, of such assignments is called a *function* (or *mapping*) *from* (or *on*) A *into* B and is written

$$f: A \longrightarrow B \qquad \text{or} \qquad A \xrightarrow{f} B$$

The unique element in B assigned to $a \in A$ by f is denoted by $f(a)$, and called the *value* of f at a or the *image* of a under f. The *domain* of f is A, the *co-domain* is B. To each function $f: A \to B$ there corresponds the relation in $A \times B$ given by

$$\{\, \langle a, f(a) \rangle : a \in A \,\}$$

We call this set the *graph* of f. The *range* of f, denoted by $f[A]$, is the set of images, i.e. $f[A] = \{f(a) : a \in A\}$.

Two functions $f: A \to B$ and $g: A \to B$ are defined to be equal, written $f = g$, iff $f(a) = g(a)$ for every $a \in A$, i.e. iff they have the same graph. Accordingly, we do not distinguish between a function and its graph. A subset f of $A \times B$, i.e. a relation from A to B, is a function iff it possesses the following property:

[F] Each $a \in A$ appears as the first coordinate in exactly one ordered pair $\langle a, b \rangle$ in f.

The negation of $f = g$ is written $f \neq g$ and is the statement: $\exists a \in A$ for which $f(a) \neq g(a)$.

Example 1.1: Let $f: \mathbf{R} \to \mathbf{R}$ be the function which assigns to each real number its square, i.e. for each $x \in \mathbf{R}$, $f(x) = x^2$. Here f is a *real-valued function.* Its graph, $\{\, \langle x, x^2 \rangle : x \in \mathbf{R} \,\}$, is displayed in Fig. 2-1 below. The range of f is the set of non-negative real numbers, i.e. $f[\mathbf{R}] = \{x : x \in \mathbf{R},\ x \geqq 0\}$.

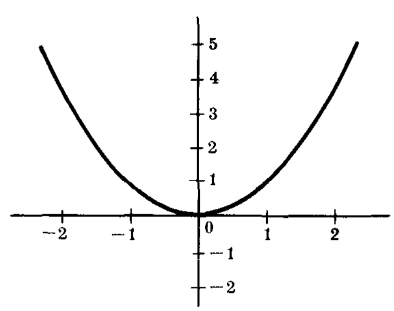


Fig. 2-1

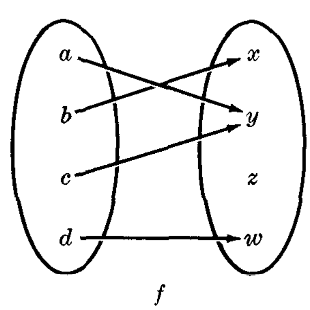


Fig. 2-2

Example 1.2: Let $A = \{a, b, c, d\}$ and $B = \{x, y, z, w\}$. Then the diagram in Fig. 2-2 above defines a function f from A into B. Here $f[A] = \{x, y, w\}$. The graph of f is the relation

$$\{\, \langle a, y \rangle,\ \langle b, x \rangle,\ \langle c, y \rangle,\ \langle d, w \rangle \,\}$$

Example 1.3: A function $f: A \to B$ is called a *constant function* if, for some $b_0 \in B$, $f(a) = b_0$ for all $a \in A$. Hence the range $f[A]$ of any constant function f is a singleton set, i.e. $f[A] = \{b_0\}$.

Consider now functions $f: A \to B$ and $g: B \to C$, illustrated below:

$$(A) \xrightarrow{f} (B) \xrightarrow{g} (C)$$

The function from A into C which maps the element $a \in A$ into the element $g(f(a))$ of C is called the *composition* or *product* of f and g and is denoted by $g \circ f$. Hence, by definition,

$$(g \circ f)(a) = g(f(a))$$

We remark that, if we view $f \subset A \times B$ and $g \subset B \times C$ as relations, we have already defined a product $g \cdot f$ (Chapter 1). However, these two products are the same in that if f and g are functions then $g \cdot f$ is a function and $g \cdot f = g \circ f$.

If $f: X \to Y$ and $A \subset X$, then the *restriction* of f to A, denoted by $f|A$, is the function from A into Y defined by

$$f|A(a) \equiv f(a) \quad \text{for all } a \in A$$

Equivalently, $f|A = f \cap (A \times Y)$. On the other hand, if $f: X \to Y$ is the restriction of some function $g: X^* \to Y$ where $X \subset X^*$, then g is called an *extension* of f.

ONE-ONE, ONTO, INVERSE AND IDENTITY FUNCTIONS

A function $f: A \to B$ is said to be *one-to-one* (or *one-one*, or 1-1) if distinct elements in A have distinct images, i.e. if

$$f(a) = f(a') \Rightarrow a = a'$$

A function $f: A \to B$ is said to be *onto* (or f is a function from A *onto* B, or f maps A *onto* B) if every $b \in B$ is the image of some $a \in A$, i.e. if

$$b \in B \Rightarrow \exists a \in A \text{ for which } f(a) = b$$

Hence if f is onto, $f[A] = B$.

In general, the inverse relation f^{-1} of a function $f \subset A \times B$ need not be a function. However, if f is both one-one and onto, then f^{-1} is a function from B onto A and is called the *inverse function.*

The diagonal $\Delta_A \subset A \times A$ is a function and called the *identity function* on A. It is also denoted by 1_A or 1. Here, $1_A(a) = a$ for every $a \in A$. Clearly, if $f: A \to B$, then

$$1_B \circ f = f = f \circ 1_A$$

Furthermore, if f is one-one and onto, and so has an inverse function f^{-1}, then

$$f^{-1} \circ f = 1_A \quad \text{and} \quad f \circ f^{-1} = 1_B$$

The converse is also true:

Proposition 2.1: Let $f: A \to B$ and $g: B \to A$ satisfy

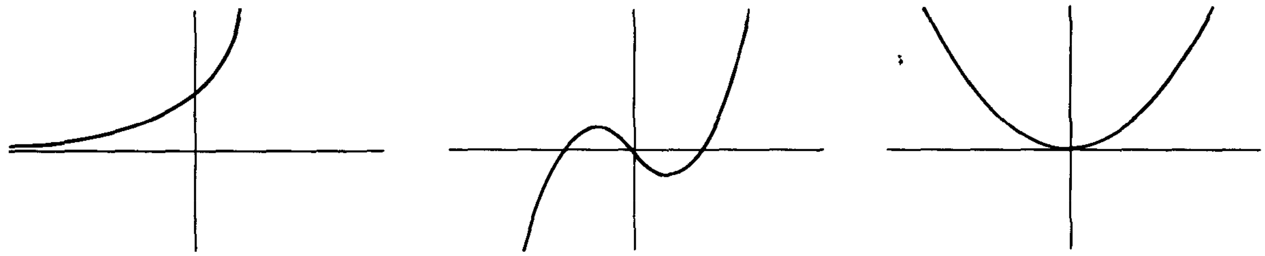
$$g \circ f = 1_A \quad \text{and} \quad f \circ g = 1_B$$

Then $f^{-1}: B \to A$ exists and $g = f^{-1}$.

Example 2.1: Let $f: \mathbf{R} \to \mathbf{R}$, $g: \mathbf{R} \to \mathbf{R}$ and $h: \mathbf{R} \to \mathbf{R}$ be defined by

$$f(x) = e^x, \quad g(x) = x^3 - x \quad \text{and} \quad h(x) = x^2$$

The function f shown in Fig. 2-3(a) below is one-one; geometrically, this means that each horizontal line does not contain more than one point of f. The function g shown in Fig. 2-3(b) below is onto; geometrically this means that each horizontal line contains at least one point of g. The function h shown in Fig. 2-3(c) below is neither one-one nor onto, for $h(2) = h(-2) = 4$ and $h[\mathbf{R}]$ is a proper subset of $\mathbf{R}$, e.g. $-16 \notin h[\mathbf{R}]$.



(a) $f(x) = e^x$ (b) $g(x) = x^3 - x$ (c) $h(x) = x^2$

Fig. 2-3

INDEXED SETS, CARTESIAN PRODUCTS

An *indexed class of sets*, denoted by

$$\{A_i : i \in I\}, \quad \{A_i\}_{i \in I} \quad \text{or simply} \quad \{A_i\}$$

assigns a set A_i to each $i \in I$, i.e. is a function from I into a class of sets. The set I is called the *index set*, the sets A_i are called *indexed sets*, and each $i \in I$ is called an *index*. When the index set I is the set of positive integers, the indexed class $\{A_1, A_2, \ldots\}$ is called a *sequence* (of sets).

Example 3.1: For each $n \in \mathbf{N}$, the positive integers, let

$$D_n = \{x : x \in \mathbf{N},\ x \text{ is a multiple of } n\}$$

Then $D_1 = \{1, 2, 3, \ldots\}$, $D_2 = \{2, 4, 6, \ldots\}$, $D_3 = \{3, 6, 9, \ldots\}$, ...

The *Cartesian product* of an indexed class of sets, $\mathcal{A} = \{A_i : i \in I\}$, denoted by

$$\prod\{A_i : i \in I\} \quad \text{or} \quad \prod_{i \in I} A_i \quad \text{or simply} \quad \prod_i A_i$$

is the set of all functions $p : I \to \cup_i A_i$ such that $p(i) = a_i \in A_i$. We denote such an element of the Cartesian product by $p = \langle a_i : i \in I \rangle$. For each $i_0 \in I$ there exists a function π_{i_0}, called the i_0th *projection function*, from the product set $\prod_i A_i$ into the i_0th *coordinate set* A_{i_0} defined by

$$\pi_{i_0}(\langle a_i : i \in I \rangle) = a_{i_0}$$

Example 3.2: Recall that $\mathbf{R}^3 = \mathbf{R} \times \mathbf{R} \times \mathbf{R}$ consists of all 3-tuples $p = \langle a_1, a_2, a_3 \rangle$ of real numbers. Now let R_1, R_2 and R_3 denote copies of $\mathbf{R}$. Then p can be viewed as a function on $I = \{1, 2, 3\}$ where $p(1) = a_1 \in R_1$, $p(2) = a_2 \in R_2$ and $p(3) = a_3 \in R_3$. In other words,

$$\mathbf{R}^3 = \prod\{R_i : i \in I,\ R_i = \mathbf{R}\}$$

GENERALIZED OPERATIONS

The notion of union and intersection, originally defined for two sets, may be generalized to any class $\mathcal{A}$ of subsets of a universal set U. The union of the sets in $\mathcal{A}$, denoted by $\bigcup\{A : A \in \mathcal{A}\}$, is the set of elements which belong to at least one set in $\mathcal{A}$:

$$\bigcup\{A : A \in \mathcal{A}\} = \{x : x \in U,\ \exists A \in \mathcal{A} \text{ s.t. } x \in A\}$$

The intersection of the sets in $\mathcal{A}$, denoted by $\bigcap\{A : A \in \mathcal{A}\}$, is the set of elements which belong to every set in $\mathcal{A}$:

$$\bigcap\{A : A \in \mathcal{A}\} = \{x : x \in U,\ x \in A \text{ for every } A \in \mathcal{A}\}$$

For an indexed class of subsets of U, say $\mathcal{A} = \{A_i : i \in I\}$, we write

$$\bigcup\{A_i : i \in I\}, \quad \bigcup_{i \in I} A_i \quad \text{or} \quad \cup_i A_i$$

for the union of the sets in $\mathcal{A}$, and

$$\bigcap\{A_i : i \in I\}, \quad \bigcap_{i \in I} A_i \quad \text{or} \quad \cap_i A_i$$

for the intersection of the sets in $\mathcal{A}$. We will also write

$$\cup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots \quad \text{and} \quad \cap_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots$$

for the union and intersection, respectively, of a sequence $\{A_1, A_2, \ldots\}$ of subsets of U.

Example 4.1: For each $n \in \mathbf{N}$, the positive integers, let $D_n = \{x : x \in \mathbf{N},\ x \text{ is a multiple of } n\}$ (see Example 3.1). Then

$$\bigcup\{D_i : i \geqq 10\} = \{10, 11, 12, \ldots\} \quad \text{and} \quad \cap_{i=1}^{\infty} D_i = \emptyset$$

Example 4.2: Let $I = [0,1]$ and, for each $i \in I$, let $A_i = [0, i]$. Then

$$\cup_i A_i = [0,1] \quad \text{and} \quad \cap_i A_i = \{0\}$$

The distributive laws and De Morgan's laws also hold for these generalized operations:

Theorem 2.2: For any class of sets $\mathcal{A} = \{A_i\}$ and any set B,

$$\text{(i)}\ B \cup (\cap_i A_i) = \cap_i (B \cup A_i) \qquad \text{(ii)}\ B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$$

Theorem 2.3: Let $\mathcal{A} = \{A_i\}$ be any class of subsets of U. Then:

$$\text{(i)}\ (\cup_i A_i)^c = \cap_i A_i^c \qquad \text{(ii)}\ (\cap_i A_i)^c = \cup_i A_i^c$$

The following theorem will be used frequently.

Theorem 2.4: Let A be any set and, for each $p \in A$, let G_p be a subset of A such that $p \in G_p \subset A$. Then $A = \bigcup\{G_p : p \in A\}$.

Remark: In the case of an empty class $\emptyset$ of subsets of a universal set U, it is convenient to define

$$\bigcup\{A : A \in \emptyset\} = \emptyset \quad \text{and} \quad \bigcap\{A : A \in \emptyset\} = U$$

Hence

$$\bigcup\{A_i : i \in \emptyset\} = \emptyset \quad \text{and} \quad \bigcap\{A_i : i \in \emptyset\} = U$$

ASSOCIATED SET FUNCTIONS

Let $f : X \to Y$. Then the *image* $f[A]$ of any subset A of X is the set of images of points in A, and the *inverse image* $f^{-1}[B]$ of any subset B of Y is the set of points in X whose images lie in B. That is,

$$f[A] = \{f(x) : x \in A\} \quad \text{and} \quad f^{-1}[B] = \{x : x \in X,\ f(x) \in B\}$$

Example 5.1: Let $f : \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x^2$. Then

$$f[\{1,3,4,7\}] = \{1, 9, 16, 49\}, \quad f[(1,2)] = (1,4)$$

Also, $$f^{-1}[\{4,9\}] = \{-3, -2, 2, 3\}, \quad f^{-1}[(1,4)] = (1,2) \cup (-2,-1)$$

Thus a function $f : X \to Y$ induces a function, also denoted by f, from the power set $\mathcal{P}(X)$ of X into the power set $\mathcal{P}(Y)$ of Y, and a function f^{-1} from $\mathcal{P}(Y)$ into $\mathcal{P}(X)$. The induced functions f and f^{-1} are called *set functions* since they are maps of classes (of sets) into classes.

We remark that the associated set function f^{-1} is not in general the inverse of the associated set function f. For example, if f is the function in Example 5.1, then

$$f^{-1} \circ f[(1,2)] = f^{-1}[(1,4)] = (1,2) \cup (-2,-1)$$

Observe that different brackets are used to distinguish between a function and its associated set functions, i.e. $f(a)$ denotes a value of the original function, and $f[A]$ and $f^{-1}[B]$ denote values of the associated set functions.

The associated set functions possess various properties. In particular we state:

Theorem 2.5: Let $f: X \to Y$. Then, for any subsets A and B of X,

(i) $f[A \cup B] = f[A] \cup f[B]$ (iii) $f[A \setminus B] \supset f[A] \setminus f[B]$

(ii) $f[A \cap B] \subset f[A] \cap f[B]$ (iv) $A \subset B$ implies $f[A] \subset f[B]$

and, more generally, for any indexed class $\{A_i\}$ of subsets of X,

(i′) $f[\cup_i A_i] = \cup_i f[A_i]$ (ii′) $f[\cap_i A_i] \subset \cap_i f[A_i]$

The following example shows that the inclusions of (ii) and (iii) cannot in general be replaced by equality.

Example 5.2: Consider the subsets

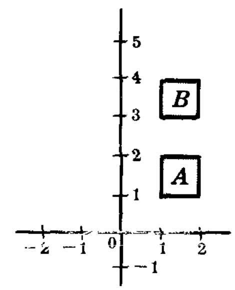
$$A = [1,2] \times [1,2] \quad \text{and} \quad B = [1,2] \times [3,4]$$

of the plane $\mathbf{R}^2$ and the projection $\pi: \mathbf{R}^2 \to \mathbf{R}$, into the first coordinate set, i.e. the x-axis. Observe that $\pi[A] = [1,2]$ and $\pi[B] = [1,2]$, and that $A \cap B = \emptyset$ implies $\pi[A \cap B] = \emptyset$. Hence

$$\pi[A] \cap \pi[B] = [1,2] \neq \pi[A \cap B] = \emptyset$$

Furthermore, $A \setminus B = A$, so

$$\pi[A \setminus B] = [1,2] \neq \emptyset = \pi[A] \setminus \pi[B]$$



On the other hand, the inverse set function is much more "well-behaved" in the sense that equality holds in both cases. Namely,

Theorem 2.6: Let $f: X \to Y$. Then for any subsets A and B of Y,

(i) $f^{-1}[A \cup B] = f^{-1}[A] \cup f^{-1}[B]$

(ii) $f^{-1}[A \cap B] = f^{-1}[A] \cap f^{-1}[B]$

(iii) $f^{-1}[A \setminus B] = f^{-1}[A] \setminus f^{-1}[B]$

(iv) $A \subset B$ implies $f^{-1}[A] \subset f^{-1}[B]$

and, more generally, for any indexed class $\{A_i\}$ of subsets of Y,

(i′) $f^{-1}[\cup_i A_i] = \cup_i f^{-1}[A_i]$

(ii′) $f^{-1}[\cap_i A_i] = \cap_i f^{-1}[A_i]$

Since $f^{-1}[Y] = X$, we have, as a special case of (iii),

Corollary 2.7: Let $f: X \to Y$ and let $A \subset Y$. Then $f^{-1}[A^c] = (f^{-1}[A])^c$.

Next follows an important relationship between the two set functions.

Theorem 2.8: Let $f: X \to Y$ and let $A \subset X$ and $B \subset Y$. Then:

(i) $A \subset f^{-1} \circ f[A]$ (ii) $B \supset f \circ f^{-1}[B]$

As shown previously, the inclusion in (i) cannot in general be replaced by equality.

ALGEBRA OF REAL-VALUED FUNCTIONS

Let $\mathcal{F}(X, \mathbf{R})$ denote the collection of all real-valued functions defined on some set X. Many operations are inherited by $\mathcal{F}(X, \mathbf{R})$ from corresponding operations in $\mathbf{R}$. Specifically, let $f: X \to \mathbf{R}$ and $g: X \to \mathbf{R}$ and let $k \in \mathbf{R}$: then we define

$$(f+g): X \to \mathbf{R} \quad \text{by} \quad (f+g)(x) \equiv f(x) + g(x)$$

$$(k \cdot f): X \to \mathbf{R} \quad \text{by} \quad (k \cdot f)(x) \equiv k(f(x))$$

$$(|f|): X \to \mathbf{R} \quad \text{by} \quad (|f|)(x) \equiv |f(x)|$$

$$(fg): X \to \mathbf{R} \quad \text{by} \quad (fg)(x) \equiv f(x)\,g(x)$$

It is also convenient to identify the real number $k \in \mathbf{R}$ with the constant function $f(x) = k$ for every $x \in \mathbf{R}$. Then $(f+k): X \to \mathbf{R}$ is the function

$$(f+k)(x) \equiv f(x) + k$$

Observe that $(fg): X \to \mathbf{R}$ is not the composition of f and g discussed previously.

Example 6.1: Consider the functions

$$f = \{\langle a,1\rangle, \langle b,3\rangle\} \quad \text{and} \quad g = \{\langle a,2\rangle, \langle b,-1\rangle\}$$

with domain $X = \{a, b\}$. Then

$$(3f-2g)(a) \equiv 3f(a) - 2g(a) = 3(1) - 2(2) = -1$$
$$(3f-2g)(b) \equiv 3f(b) - 2g(b) = 3(3) - 2(-1) = 11$$

that is, $$3f - 2g = \{\langle a,-1\rangle, \langle b,11\rangle\}$$

Also, since $|g|(x) \equiv |g(x)|$ and $(g+3)(x) \equiv g(x) + 3$,

$$|g| = \{\langle a,2\rangle, \langle b,1\rangle\} \quad \text{and} \quad g+3 = \{\langle a,5\rangle, \langle b,2\rangle\}$$

The collection $\mathcal{F}(X, \mathbf{R})$ with the above operations possesses various properties of which some are included in the next theorem.

Theorem 2.9: The collection $\mathcal{F}(X, \mathbf{R})$ of all real-valued functions defined on a non-empty set X together with the above operations satisfies the following axioms of a *real linear vector space*:

[V_1] The operation of addition of functions f and g satisfies:

(1) $(f+g)+h = f+(g+h)$

(2) $f+g = g+f$

(3) $\exists 0 \in \mathcal{F}(X, \mathbf{R})$, i.e. $0: X \to \mathbf{R}$, such that $f+0 = f$.

(4) For each $f \in \mathcal{F}(X, \mathbf{R})$, $\exists -f \in \mathcal{F}(X, \mathbf{R})$, i.e. $-f: X \to \mathbf{R}$, such that $f+(-f) = 0$.

[V_2] The operation of scalar multiplication $k \cdot f$ of a function f by a real number k satisfies:

(1) $k \cdot (k' \cdot f) = (kk') \cdot f$

(2) $1 \cdot f = f$

[V_3] The operations of addition and scalar multiplication satisfy:

(1) $k \cdot (f+g) = k \cdot f + k \cdot g$

(2) $(k+k') \cdot f = k \cdot f + k' \cdot f$

Example 6.2: Let $X = \{1, 2, \ldots, m\}$. Then each function $f \in \mathcal{F}(X, \mathbf{R})$ may be written as an ordered m-tuple $\langle f(1), \ldots, f(m)\rangle$. Furthermore, if

$$f = \langle a_1, \ldots, a_m\rangle \quad \text{and} \quad g = \langle b_1, \ldots, b_m\rangle$$

then $$f + g = \langle a_1 + b_1,\ a_2 + b_2,\ \ldots,\ a_m + b_m\rangle$$

and, for any $k \in \mathbf{R}$, $$k \cdot f = \langle ka_1, \ldots, ka_m\rangle$$

In this case, the real linear (vector) space $\mathcal{F}(X, \mathbf{R})$ is called *m-dimensional Euclidean space.*

Example 6.3: A function $f \in \mathcal{F}(X, \mathbf{R})$ is said to be *bounded* iff

$$\exists M \in \mathbf{R} \text{ such that } |f(x)| \leq M \text{ for every } x \in X$$

Let $\beta(X, \mathbf{R})$ denote the collection of all bounded functions in $\mathcal{F}(X, \mathbf{R})$. Then $\beta(X, \mathbf{R})$ possesses the following properties:

(i) If $f, g \in \beta(X, \mathbf{R})$, then $f+g \in \beta(X, \mathbf{R})$.

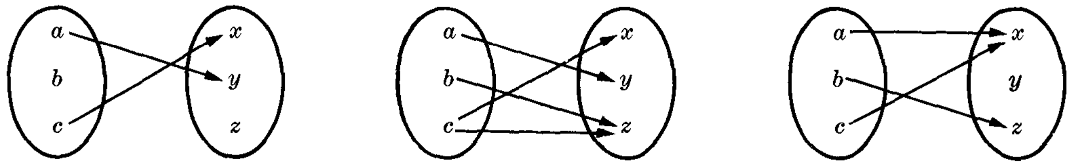
(ii) If $f \in \beta(X, \mathbf{R})$ and $k \in \mathbf{R}$, then $k \cdot f \in \beta(X, \mathbf{R})$.

Any subset of $\mathcal{F}(X, \mathbf{R})$ satisfying (i) and (ii) is called a (linear) *subspace* of $\mathcal{F}(X, \mathbf{R})$.

Solved Problems

FUNCTIONS

1. State whether or not each of the diagrams defines a function from $A = \{a, b, c\}$ into $B = \{x, y, z\}$.



(i) (ii) (iii)

Solution:

(i) No. There is nothing assigned to the element $b \in A$.

(ii) No. Two elements, x and z, are assigned to $c \in A$.

(iii) Yes.

2. Let $X = \{1, 2, 3, 4\}$. State whether or not each of the following relations is a function from X into X.

(i) $f = \{ \langle 2,3\rangle, \langle 1,4\rangle, \langle 2,1\rangle, \langle 3,2\rangle, \langle 4,4\rangle \}$

(ii) $g = \{ \langle 3,1\rangle, \langle 4,2\rangle, \langle 1,1\rangle \}$

(iii) $h = \{ \langle 2,1\rangle, \langle 3,4\rangle, \langle 1,4\rangle, \langle 2,1\rangle, \langle 4,4\rangle \}$

Solution:

Recall that a subset f of $X \times X$ is a function $f: X \to X$ iff each $x \in X$ appears as the first coordinate in exactly one ordered pair in f.

(i) No. Two different ordered pairs $\langle 2,3\rangle$ and $\langle 2,1\rangle$ in f have the same first coordinate.

(ii) No. The element $2 \in X$ does not appear as the first coordinate in any ordered pair in g.

(iii) Yes. Although $2 \in X$ appears as the first coordinate in two ordered pairs in h, these two ordered pairs are equal.

3. Consider the functions

$$f = \{ \langle 1,3\rangle, \langle 2,5\rangle, \langle 3,3\rangle, \langle 4,1\rangle, \langle 5,2\rangle \}$$
$$g = \{ \langle 1,4\rangle, \langle 2,1\rangle, \langle 3,1\rangle, \langle 4,2\rangle, \langle 5,3\rangle \}$$

from $X = \{1, 2, 3, 4, 5\}$ into X.

(i) Determine the range of f and of g.

(ii) Find the composition functions $g \circ f$ and $f \circ g$.

Solution:

(i) Recall that the range of a function is the set of image values, i.e. the set of second coordinates. Hence

$$\text{range of } f = \{3, 5, 1, 2\} \quad \text{and} \quad \text{range of } g = \{4, 1, 2, 3\}$$

(ii) Use the definition of the composition function and compute:

$$\begin{array}{ll}
(g \circ f)(1) \equiv g(f(1)) = g(3) = 1 & (f \circ g)(1) \equiv f(g(1)) = f(4) = 1 \\
(g \circ f)(2) \equiv g(f(2)) = g(5) = 3 & (f \circ g)(2) \equiv f(g(2)) = f(1) = 3 \\
(g \circ f)(3) \equiv g(f(3)) = g(3) = 1 & (f \circ g)(3) \equiv f(g(3)) = f(1) = 3 \\
(g \circ f)(4) \equiv g(f(4)) = g(1) = 4 & (f \circ g)(4) \equiv f(g(4)) = f(2) = 5 \\
(g \circ f)(5) \equiv g(f(5)) = g(2) = 1 & (f \circ g)(5) \equiv f(g(5)) = f(3) = 3
\end{array}$$

In other words,

$$g \circ f = \{ \langle 1,1\rangle, \langle 2,3\rangle, \langle 3,1\rangle, \langle 4,4\rangle, \langle 5,1\rangle \}$$
$$f \circ g = \{ \langle 1,1\rangle, \langle 2,3\rangle, \langle 3,3\rangle, \langle 4,5\rangle, \langle 5,3\rangle \}$$

Observe that $f \circ g \neq g \circ f$.

4. Let the functions $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ be defined by

$$f(x) = 2x + 1, \quad g(x) = x^2 - 2$$

Find formulas defining the product functions $g \circ f$ and $f \circ g$.

Solution:

Compute $g \circ f: \mathbf{R} \to \mathbf{R}$ as follows:

$$(g \circ f)(x) \equiv g(f(x)) = g(2x+1) = (2x+1)^2 - 2 = 4x^2 + 4x - 1$$

Observe that the same answer can be found by writing

$$y = f(x) = 2x + 1, \qquad z = g(y) = y^2 - 2$$

and then eliminating y from the two equations:

$$z = y^2 - 2 = (2x+1)^2 - 2 = 4x^2 + 4x - 1$$

Now compute $f \circ g: \mathbf{R} \to \mathbf{R}$:

$$(f \circ g)(x) \equiv f(g(x)) = f(x^2 - 2) = 2(x^2 - 2) + 1 = 2x^2 - 3$$

5. Prove the associative law for composition of functions, i.e. if $f: A \to B$, $g: B \to C$ and $h: C \to D$, then $(h \circ g) \circ f = h \circ (g \circ f)$.

Solution:

Since the associative law was proven for composition of relations in general, this result follows. We also give a direct proof:

$$((h \circ g) \circ f)(a) = (h \circ g)(f(a)) = h(g(f(a))), \quad \forall a \in A$$
$$(h \circ (g \circ f))(a) = h((g \circ f)(a)) = h(g(f(a))), \quad \forall a \in A$$

Hence $(h \circ g) \circ f = h \circ (g \circ f)$.

ONE-ONE AND ONTO FUNCTIONS

6. Let $f: A \to B$, $g: B \to C$. Prove:

(i) If f and g are onto, then $g \circ f: A \to C$ is onto.

(ii) If f and g are one-one, then $g \circ f: A \to C$ is one-one.

Solution:

(i) Let $c \in C$. Since g is onto, $\exists b \in B$ s.t. $g(b) = c$. Since f is onto, $\exists a \in A$ s.t. $f(a) = b$. But then $(g \circ f)(a) = g(f(a)) = c$, i.e. $g \circ f$ is also onto.

(ii) Suppose $(g \circ f)(a) = (g \circ f)(a')$; i.e. $g(f(a)) = g(f(a'))$. So $f(a) = f(a')$ since g is one-one; hence $a = a'$ since f is one-one. Accordingly, $g \circ f$ is also one-one.

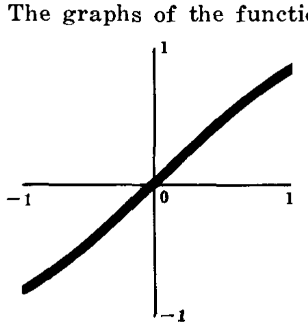
7. Let $A = [-1, 1]$ and let $f: A \to A$, $g: A \to A$ and $h: A \to A$ be defined by

$$f(x) = \sin x, \quad g(x) = \sin \pi x, \quad h(x) = \sin \frac{\pi}{2} x$$

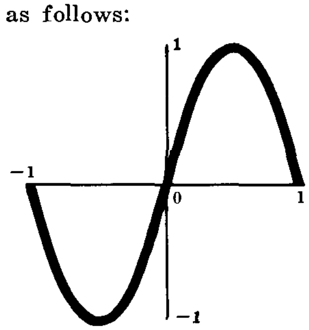
State whether or not each of the functions is (i) one-one, (ii) onto, (iii) bijective (i.e. one-one and onto).

Solution:

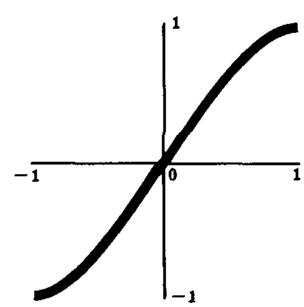
The graphs of the functions are as follows:



$f(x)$



$g(x)$



$h(x)$

The function f is one-one; each horizontal line does not contain more than one point of f. It is not onto since, for example, $\sin x \neq 1$ for any $x \in A$. On the other hand, g is onto; each horizontal line contains at least one point of f. But g is not one-one since, for example, $g(-1) = g(0) = 0$. The function h is both one-one and onto; each horizontal line contains exactly one point of h.

8. Prove: Let $f: A \to B$ and $g: B \to C$ be one-one and onto; then $(g \circ f)^{-1}: C \to A$ exists and equals $f^{-1} \circ g^{-1}: C \to A$.

Solution:

Utilizing Proposition 2.1, we show that:

$$(f^{-1} \circ g^{-1}) \circ (g \circ f) = 1_A \quad \text{and} \quad (g \circ f) \circ (f^{-1} \circ g^{-1}) = 1_B$$

Using the associative law for composition of functions,

$$\begin{aligned}(f^{-1} \circ g^{-1}) \circ (g \circ f) &= f^{-1} \circ (g^{-1} \circ (g \circ f)) \\ &= f^{-1} \circ ((g^{-1} \circ g) \circ f) \\ &= f^{-1} \circ (1 \circ f) \\ &= f^{-1} \circ f \\ &= 1_A\end{aligned}$$

since $g^{-1} \circ g = 1$ and $1 \circ f = f = f \circ 1$. Similarly,

$$\begin{aligned}(g \circ f) \circ (f^{-1} \circ g^{-1}) &= g \circ (f \circ (f^{-1} \circ g^{-1})) \\ &= g \circ ((f \circ f^{-1}) \circ g^{-1}) \\ &= g \circ (1 \circ g^{-1}) \\ &= g \circ g^{-1} \\ &= 1_B\end{aligned}$$

9. When will a projection function $\pi_{i_0}: \prod \{A_i : i \in I\} \to A_{i_0}$, $A_{i_0} \neq \emptyset$, be an onto function?

Solution:

A projection function is always onto, providing the Cartesian product $\prod \{A_i : i \in I\}$ is non-empty, i.e. provided no A_i is the empty set.

INDEXED SETS, GENERALIZED OPERATIONS

10. Let $A_n = \{x : x \text{ is a multiple of } n\}$, where $n \in \mathbf{N}$, the positive integers, and let $B_i = [i, i+1]$, where $i \in \mathbf{Z}$, the integers. Find: (i) $A_3 \cap A_5$; (ii) $\bigcup\{A_i : i \in P\}$, where P is the set of prime numbers; (iii) $B_3 \cap B_4$; (iv) $\bigcup\{B_i : i \in \mathbf{Z}\}$; (v) $(\bigcup\{B_i : i \geqq 7\}) \cap A_5$.

Solution:

(i) Those numbers which are multiples of both 3 and 5 are the multiples of 15; hence $A_3 \cap A_5 = A_{15}$.

(ii) Every positive integer except 1 is a multiple of at least one prime number; hence $\bigcup\{A_i : i \in P\} = \{2, 3, 4, \ldots\} = \mathbf{N} \setminus \{1\}$.

(iii) $B_3 \cap B_4 = \{x : 3 \leqq x \leqq 4,\ 4 \leqq x \leqq 5\} = \{4\}$

(iv) Since every real number belongs to at least one interval $[i, i+1]$, $\bigcup\{B_i : i \in \mathbf{Z}\} = \mathbf{R}$, the set of real numbers.

(v) $(\bigcup\{B_i : i \geqq 7\}) \cap A_5 = \{x : x \text{ is a multiple of } 5,\ x \geqq 7\} = A_5 \setminus \{5\} = \{10, 15, 20, \ldots\}$.

11. Let $D_n = (0, 1/n)$, where $n \in \mathbf{N}$, the positive integers. Find:

(i) $D_3 \cup D_7$ (iii) $D_s \cup D_t$ (v) $\bigcup\{D_i : i \in A \subset \mathbf{N}\}$

(ii) $D_3 \cap D_{20}$ (iv) $D_s \cap D_t$ (vi) $\bigcap\{D_i : i \in \mathbf{N}\}$

Solution:

(i) Since $(0, 1/7) \subset (0, 1/3)$, $D_3 \cup D_7 = D_3$.

(ii) Since $(0, 1/20) \subset (0, 1/3)$, $D_3 \cap D_{20} = D_{20}$.

(iii) Let $m = \min\{s, t\}$, i.e. the smaller of the two numbers s and t; then D_m equals D_s or D_t and contains the other. So $D_s \cup D_t = D_m$.

(iv) Let $M = \max\{s, t\}$, i.e. the larger of the two numbers. Then $D_s \cap D_t = D_M$.

(v) Let $a \in A$ be the smallest number in A. Then $\bigcup\{D_i : i \in A \subset \mathbf{N}\} = D_a$.

(vi) If $x \in \mathbf{R}$, then $\exists i \in \mathbf{N}$ s.t. $x \notin (0, 1/i)$. Hence $\bigcap\{D_i : i \in \mathbf{N}\} = \emptyset$.

12. Prove (Distributive Law) Theorem 2.2 (ii): $B \cap (\cup_{i \in I} A_i) = \cup_{i \in I}(B \cap A_i)$.

Solution:

$$\begin{aligned} B \cap (\cup_{i \in I} A_i) &= \{x : x \in B,\ x \in \cup_{i \in I} A_i\} \\ &= \{x : x \in B,\ \exists i_0 \in I \text{ s.t. } x \in A_{i_0}\} \\ &= \{x : \exists i_0 \in I \text{ s.t. } x \in B \cap A_{i_0}\} \\ &= \cup_{i \in I}(B \cap A_i) \end{aligned}$$

13. Prove: Let $\{A_i : i \in I\}$ be an indexed class of sets and let $i_0 \in I$. Then

$$\cap_{i \in I} A_i \subset A_{i_0} \subset \cup_{i \in I} A_i$$

Solution:

Let $x \in \cap_{i \in I} A_i$; then $x \in A_i$ for every $i \in I$. In particular, $x \in A_{i_0}$. Hence $\cap_{i \in I} A_i \subset A_{i_0}$. Now let $y \in A_{i_0}$. Since $i_0 \in I$, $y \in \cup_{i \in I} A_i$. Hence $A_{i_0} \subset \cup_{i \in I} A_i$.

14. Prove Theorem 2.4: Let A be any set and, for each $p \in A$, let G_p be a subset of A such that $p \in G_p \subset A$. Then $A = \bigcup\{G_p : p \in A\}$.

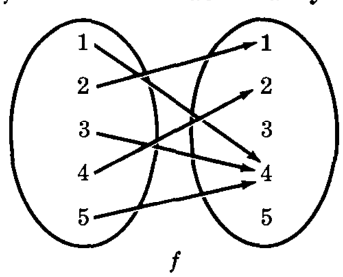
Solution:

Let $x \in \bigcup\{G_p : p \in A\}$. Then $\exists p_0 \in A$ s.t. $x \in G_{p_0} \subset A$; hence $x \in A$, so $\bigcup\{G_p : p \in A\} \subset A$. (In other words, if each G_p is a subset of A, then the union of the G_p is also a subset of A.)

Now let $y \in A$. Then $y \in G_y$, so $y \in \bigcup\{G_p : p \in A\}$. Thus $A \subset \bigcup\{G_p : p \in A\}$ and the two sets are equal.

ASSOCIATED SET FUNCTIONS

15. Let $A = \{1, 2, 3, 4, 5\}$ and let $f : A \to A$ be defined by the diagram:



Find (i) $f[\{1, 3, 5\}]$, (ii) $f^{-1}[\{2, 3, 4\}]$, (iii) $f^{-1}[\{3, 5\}]$.

Solution:

(i) $f[\{1, 3, 5\}] = \{f(1), f(3), f(5)\} = \{4\}$

(ii) $f^{-1}[\{2, 3, 4\}] = \{4, 1, 3, 5\}$

(iii) $f^{-1}[\{3, 5\}] = \emptyset$ since no element of A has 3 or 5 as an image.

16. Consider the function $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = x^2$. Find:
(i) $f^{-1}[\{25\}]$, (ii) $f^{-1}[\{-9\}]$, (iii) $f^{-1}[\{x : x \leq 0\}]$, (iv) $f^{-1}[\{x : 4 \leq x \leq 25\}]$.

Solution:

(i) $f^{-1}[\{25\}] = \{5, -5\}$ since $f(5) = 25$, $f(-5) = 25$ and since the square of no other number is 25.

(ii) $f^{-1}[\{-9\}] = \emptyset$ since the square of no real number is -9.

(iii) $f^{-1}[\{x : x \leq 0\}] = \{0\}$ since $f(0) = 0 \leq 0$ and since the square of every other real number is greater than 0.

(iv) $f^{-1}[\{x : 4 \leq x \leq 25\}]$ consists of those numbers x such that $4 \leq x^2 \leq 25$. Accordingly,

$$f^{-1}[\{x : 4 \leq x \leq 25\}] = [2,5] \cup [-5,-2]$$

17. Prove: Let $f : X \to Y$ be one-one. Then the associated set function $f : \mathcal{P}(X) \to \mathcal{P}(Y)$ is also one-one.

Solution:

If $X = \emptyset$, then $\mathcal{P}(X) = \{\emptyset\}$; hence $f : \mathcal{P}(X) \to \mathcal{P}(Y)$ is one-one, for no two different members of $\mathcal{P}(X)$ can have the same image, as there are no two different members in $\mathcal{P}(X)$.

If $X \neq \emptyset$, $\mathcal{P}(X)$ has at least two members. Let $A, B \in \mathcal{P}(X)$, but $A \neq B$. Then $\exists p \in X$ s.t. $p \in A$, $p \notin B$ (or $p \in B$, $p \notin A$). Thus $f(p) \in f[A]$ and, since f is one-one, $f(p) \notin f[B]$ (or $f(p) \in f[B]$ and $f(p) \notin f[A]$). Hence $f[A] \neq f[B]$, and so the induced set function is also one-one.

18. Prove (Theorem 2.5, (i) and (iii)):
(*a*) $f[A \cup B] = f[A] \cup f[B]$, (*b*) $f[A] \setminus f[B] \subset f[A \setminus B]$.

Solution:

(*a*) We first show $f[A \cup B] \subset f[A] \cup f[B]$. Let $y \in f[A \cup B]$, i.e. $\exists x \in A \cup B$ s.t. $f(x) = y$. Then either $x \in A$ or $x \in B$, but

$$x \in A \quad \text{implies} \quad f(x) = y \in f[A]$$

or

$$x \in B \quad \text{implies} \quad f(x) = y \in f[B]$$

In either case, $y \in f[A] \cup f[B]$.

We now prove the reverse inclusion, i.e. $f[A] \cup f[B] \subset f[A \cup B]$. Let $y \in f[A] \cup f[B]$. Then $y \in f[A]$ or $y \in f[B]$, but

$$y \in f[A] \quad \text{implies} \quad \exists x \in A \text{ s.t. } f(x) = y$$

$$y \in f[B] \quad \text{implies} \quad \exists x \in B \text{ s.t. } f(x) = y$$

In either case, $y = f(x)$ with $x \in A \cup B$, i.e. $y \in f[A \cup B]$.

(*b*) Let $y \in f[A] \setminus f[B]$. Then $\exists x \in A$ s.t. $f(x) = y$, but $y \notin \{f(x) : x \in B\}$. Hence $x \notin B$, or $x \in B \setminus A$. Accordingly, $y \in f[A \setminus B]$.

19. Prove (Theorem 2.6, (ii) and (iii)):
(*a*) $f^{-1}[A \cap B] = f^{-1}[A] \cap f^{-1}[B]$, (*b*) $f^{-1}[A \setminus B] = f^{-1}[A] \setminus f^{-1}[B]$.

Solution:

(*a*) We first show $f^{-1}[A \cap B] \subset f^{-1}[A] \cap f^{-1}[B]$. Let $x \in f^{-1}[A \cap B]$. Then $f(x) \in A \cap B$ so $f(x) \in A$ and $f(x) \in B$, or $x \in f^{-1}[A]$ and $x \in f^{-1}[B]$. Hence $x \in f^{-1}[A] \cap f^{-1}[B]$.

For the reverse inclusion, let $x \in f^{-1}[A] \cap f^{-1}[B]$. Then $f(x) \in A$ and $f(x) \in B$, i.e. $f(x) \in A \cap B$. Hence $x \in f^{-1}[A \cap B]$.

(*b*) To show $f^{-1}[A \setminus B] \subset f^{-1}[A] \setminus f^{-1}[B]$, assume $x \in f^{-1}[A \setminus B]$. Then $f(x) \in A \setminus B$, i.e. $f(x) \in A$ and $f(x) \notin B$. Thus $x \in f^{-1}[A]$ but $x \notin f^{-1}[B]$, i.e. $x \in f^{-1}[A] \setminus f^{-1}[B]$.

For the reverse inclusion, let $x \in f^{-1}[A] \setminus f^{-1}[B]$. Then $f(x) \in A$ but $f(x) \notin B$, i.e. $f(x) \in A \setminus B$. Hence $x \in f^{-1}[A \setminus B]$.

ALGEBRA OF REAL-VALUED FUNCTIONS

20. Let $X = \{a, b, c\}$ and let $f, g \in \mathcal{F}(X, \mathbf{R})$ be as follows:

$$f = \{\langle a, 1\rangle, \langle b, -2\rangle, \langle c, 3\rangle\}, \qquad g = \{\langle a, -2\rangle, \langle b, 0\rangle, \langle c, 1\rangle\}$$

Find: (i) $f + 2g$, (ii) $fg - 2f$, (iii) $f + 4$, (iv) $|f|$, (v) f^2.

Solution:

(i) Compute as follows:

$$(f + 2g)(a) \equiv f(a) + 2g(a) = 1 - 4 = -3$$

$$(f + 2g)(b) \equiv f(b) + 2g(b) = -2 + 0 = -2$$

$$(f + 2g)(c) \equiv f(c) + 2g(c) = 3 + 2 = 5$$

In other words, $f + 2g = \{\langle a, -3\rangle, \langle b, -2\rangle, \langle c, 5\rangle\}$.

(ii) Similarly,

$$(fg - 2f)(a) \equiv f(a)\,g(a) - 2f(a) = (1)(-2) - 2(1) = -4$$
$$(fg - 2f)(b) \equiv f(b)\,g(b) - 2f(b) = (-2)(0) - 2(-2) = 4$$
$$(fg - 2f)(c) \equiv f(c)\,g(c) - 2f(c) = (3)(1) - 2(3) = -3$$

That is,
$$fg - 2f = \{\langle a, -4\rangle, \langle b, 4\rangle, \langle c, -3\rangle\}$$

(iii) Since, by definition, $(f+4)(x) \equiv f(x) + 4$, add 4 to each image value, i.e. to the second coordinate in each pair in f. Thus
$$f + 4 = \{\langle a, 5\rangle, \langle b, 2\rangle, \langle c, 7\rangle\}$$

(iv) Since $|f|(x) \equiv |f(x)|$, replace the second coordinate of each pair in f by its absolute value. Thus
$$|f| = \{\langle a, 1\rangle, \langle b, 2\rangle, \langle c, 3\rangle\}$$

(v) Since $f^2(x) = (ff)(x) \equiv f(x)\,f(x) = (f(x))^2$, replace the second coordinate of each pair in f by its square. Thus
$$f^2 = \{\langle a, 1\rangle, \langle b, 4\rangle, \langle c, 9\rangle\}$$

21. Let $\hat{0} \in \mathcal{F}(X, \mathbf{R})$ be defined by $\hat{0}(x) = 0$ for all $x \in X$. Prove: For any $f \in \mathcal{F}(X, \mathbf{R})$, (i) $f + \hat{0} = f$ and (ii) $f\hat{0} = \hat{0}$.

Solution:

(i) $(f + \hat{0})x \equiv f(x) + \hat{0}(x) = f(x) + 0 = f(x)$ for every $x \in X$; hence $f + \hat{0} = f$. Observe that $\hat{0}$ satisfies the conditions of the 0 in the axiom $[\mathbf{V}_1]$ of Theorem 2.9.

(ii) $(f\hat{0})(x) \equiv f(x)\,\hat{0}(x) = f(x)\cdot(0) = 0 = \hat{0}(x)$ for all $x \in X$; hence $f\hat{0} = \hat{0}$.

22. Prove: $\mathcal{F}(X, \mathbf{R})$ satisfies the axiom $[\mathbf{V}_3]$ of Theorem 2.9, i.e. if $f, g \in \mathcal{F}(X, \mathbf{R})$ and $k, k' \in \mathbf{R}$, then:

(i) $k\cdot(f+g) = k\cdot f + k\cdot g$, (ii) $(k+k')\cdot f = k\cdot f + k'\cdot f$.

Solution:

(i)
$$[k\cdot(f+g)](x) = k[(f+g)(x)] = k[f(x) + g(x)] = k(f(x)) + k(g(x))$$
$$(k\cdot f + k\cdot g)(x) = (k\cdot f)(x) + (k\cdot g)(x) = k(f(x)) + k(g(x))$$

for all $x \in X$; hence $k\cdot(f+g) = k\cdot f + k\cdot g$. Observe that we use the fact that k, $f(x)$ and $g(x)$ are real numbers and satisfy the distributive law.

(ii)
$$((k+k')\cdot f)(x) = (k+k')\,f(x) = k(f(x)) + k'(f(x))$$
$$(k\cdot f + k'\cdot f)(x) = (k\cdot f)(x) + (k'\cdot f)(x) = k(f(x)) + k'(f(x))$$

for all $x \in X$; so $(k+k')\cdot f = k\cdot f + k'\cdot f$.

Supplementary Problems

FUNCTIONS

23. Let $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = \begin{cases} 2x - 5 & \text{if } x > 2 \\ x^2 - 2|x| & \text{if } x \leq 2 \end{cases}$, $g(x) = 3x + 1$.

Find (i) $f(-2)$, (ii) $g(-3)$, (iii) $f(4)$, (iv) $(g \circ f)(1)$, (v) $(f \circ g)(2)$, (vi) $(f \circ f)(3)$.

24. Let $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x^2 + 3x + 1$, $g(x) = 2x - 3$. Find formulas which define the composition functions (i) $f \circ g$, (ii) $g \circ f$, (iii) $f \circ f$.

25. Let $k: X \to X$ be a constant function. Prove that for any function $f: X \to X$, $k \circ f = k$. What can be said about $f \circ k$?

26. Consider the function $f(x) = x$ where $x \in \mathbf{R}$, $x \geqq 0$. State whether or not each of the following functions is an extension of f.

(i) $g_1(x) = |x|$ for all $x \in \mathbf{R}$

(ii) $g_2(x) = x$ where $x \in [-1, 1]$

(iii) $g_3(x) = (x + |x|)/2$ for all $x \in \mathbf{R}$

(iv) $1_{\mathbf{R}}: \mathbf{R} \to \mathbf{R}$

27. Let $A \subset X$ and let $f: X \to Y$. The *inclusion function* j from A into X, denoted by $j: A \subset X$, is defined by $j(a) = a$ for all $a \in A$. Show that $f|A$, the restriction of f to A, equals the composition $f \circ j$, i.e. $f|A = f \circ j$.

ONE-ONE, ONTO, INVERSE AND IDENTITY FUNCTIONS

28. Prove: For any function $f: A \to B$, $f \circ 1_A = f = 1_B \circ f$.

29. Prove: If $f: A \to B$ is both one-one and onto, then $f^{-1} \circ f = 1_A$ and $f \circ f^{-1} = 1_B$.

30. Prove: If $f: A \to B$ and $g: B \to A$ satisfy $g \circ f = 1_A$, then f is one-one and g is onto.

31. Prove Proposition 2.1: Let $f: A \to B$ and $g: B \to A$ satisfy $g \circ f = 1_A$ and $f \circ g = 1_B$. Then $f^{-1}: B \to A$ exists and $g = f^{-1}$.

32. Under what conditions will the projection $\pi_{i_0}: \prod\{A_i : i \in I\} \to A_{i_0}$ be one-to-one?

33. Let $f: (-1, 1) \to \mathbf{R}$ be defined by $f(x) = x/(1 - |x|)$. Prove that f is both one-one and onto.

34. Let R be an equivalence relation in a non-empty set A. The natural function η from A into the quotient set A/R is defined by $\eta(a) = [a]$, the equivalence class of a. Prove that η is an onto function.

35. Let $f: A \to B$. The relation R in A defined by $a\,R\,a'$ iff $f(a) = f(a')$ is an equivalence relation. Let $\hat{f}$ denote the correspondence from the quotient set A/R into the range $f[A]$ of f by $\hat{f}: [a] \to f(a)$.

(i) Prove that $\hat{f}: A/R \to f[A]$ is a function which is both one-one and onto.

(ii) Prove that $f = j \circ \hat{f} \circ \eta$, where $\eta: A \to A/R$ is the natural function and $j: f[A] \subset B$ is the inclusion function.

$$A \xrightarrow{\eta} A/R \xrightarrow{\hat{f}} f[A] \xrightarrow{j} B$$

INDEXED SETS AND GENERALIZED OPERATIONS

36. Let $A_n = \{x : x \text{ is a multiple of } n\} = \{n, 2n, 3n, \ldots\}$, where $n \in \mathbf{N}$, the positive integers. Find: (i) $A_2 \cap A_7$; (ii) $A_6 \cap A_8$; (iii) $A_3 \cup A_{12}$; (iv) $A_3 \cap A_{12}$; (v) $A_s \cup A_{st}$, where $s, t \in \mathbf{N}$; (vi) $A_s \cap A_{st}$, where $s, t \in \mathbf{N}$. (vii) Prove: If $J \subset \mathbf{N}$ is infinite, then $\cap\{A_i : i \in J\} = \emptyset$.

37. Let $B_i = (i, i+1]$, an open-closed interval, where $i \in \mathbf{Z}$, the integers. Find:

(i) $B_4 \cup B_5$

(ii) $B_6 \cap B_7$

(iii) $\cup_{i=4}^{20} B_i$

(iv) $B_s \cup B_{s+1} \cup B_{s+2}$, $s \in \mathbf{Z}$

(v) $\cup_{i=0}^{15} B_{s+i}$

(vi) $\cup_{i \in \mathbf{Z}} B_{s+i}$

38. Let $D_n = [0, 1/n]$, $S_n = (0, 1/n]$ and $T_n = [0, 1/n)$ where $n \in \mathbf{N}$, the positive integers. Find: (i) $\bigcap\{D_n : n \in \mathbf{N}\}$, (ii) $\bigcap\{S_n : n \in \mathbf{N}\}$, (iii) $\bigcap\{T_n : n \in \mathbf{N}\}$.

39. Prove DeMorgan's Laws: (i) $(\cup_i A_i)^c = \cap_i A_i^c$, (ii) $(\cap_i A_i)^c = \cup_i A_i^c$.

40. Let $\mathcal{A} = \{A_i : i \in I\}$ be an indexed class of sets and let $J \subset K \subset I$. Prove:

(i) $\bigcup\{A_i : i \in J\} \subset \bigcup\{A_i : i \in K\}$, (ii) $\bigcap\{A_i : i \in J\} \supset \bigcap\{A_i : i \in K\}$

ASSOCIATED SET FUNCTIONS

41. Let $f : \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x^2 + 1$. Find: (i) $f[\{-1, 0, 1\}]$, (ii) $f^{-1}[\{10, 17\}]$, (iii) $f[(-2, 2)]$, (iv) $f^{-1}[(5, 10)]$, (v) $f[\mathbf{R}]$, (vi) $f^{-1}[\mathbf{R}]$.

42. Prove: A function $f : X \to Y$ is one-one if and only if $f[A \cap B] = f[A] \cap f[B]$, for all subsets A and B of X.

43. Prove: Let $f : X \to Y$. Then, for any subsets A and B of X,

$$(a)\ f[A \cap B] \subset f[A] \cap f[B], \qquad (b)\ A \subset B \text{ implies } f[A] \subset f[B]$$

44. Prove: Let $f : X \to Y$. Then, for any subsets A and B of Y,

$$(a)\ f^{-1}[A \cup B] = f^{-1}[A] \cup f^{-1}[B], \qquad (b)\ A \subset B \text{ implies } f^{-1}[A] \subset f^{-1}[B]$$

45. Prove Theorem 2.8: Let $f : X \to Y$ and let $A \subset X$ and $B \subset Y$. Then

$$\text{(i)}\ A \subset f^{-1} \circ f[A], \qquad \text{(ii)}\ B \supset f \circ f^{-1}[B]$$

46. Prove: Let $f : X \to Y$ be onto. Then the associated set function $f : \mathcal{P}(X) \to \mathcal{P}(Y)$ is also onto.

47. Prove: A function $f : X \to Y$ is both one-one and onto if and only if $f[A^c] = (f[A])^c$ for every subset A of X.

48. Prove: A function $f : X \to Y$ is one-one if and only if $A = f^{-1} \circ f[A]$ for every subset A of X.

ALGEBRA OF REAL-VALUED FUNCTIONS

49. Let $X = \{a, b, c\}$ and let f and g be the following real valued functions on X:

$$f = \{\langle a, 2\rangle, \langle b, -3\rangle, \langle c, -1\rangle\}, \qquad g = \{\langle a, -2\rangle, \langle b, 0\rangle, \langle c, 1\rangle\}$$

Find (i) $3f$, (ii) $2f - 5g$, (iii) fg, (iv) $|f|$, (v) f^3, (vi) $|3f - fg|$.

50. Let A be any subset of a universal set U. Then the real-valued function $\chi_A : U \to \mathbf{R}$ defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

is called the *characteristic function* of A. Prove:

(i) $\chi_{A \cap B} = \chi_A \chi_B$, (ii) $\chi_{A \cup B} = \chi_A + \chi_B - \chi_{A \cap B}$, (iii) $\chi_{A \setminus B} = \chi_A - \chi_{A \cap B}$.

51. Prove: $\mathcal{F}(X, \mathbf{R})$ satisfies the axiom $[\mathbf{V}_2]$ of Theorem 2.9; i.e. if $f \in \mathcal{F}(X, \mathbf{R})$ and $k, k' \in \mathbf{R}$, then (i) $k \cdot (k' \cdot f) = (kk') \cdot f$, (ii) $1 \cdot f = f$

52. For each $k \in \mathbf{R}$, let $\hat{k} \in \mathcal{F}(X, \mathbf{R})$ denote the constant function $\hat{k}(x) = k$ for all $x \in X$.

(i) Show that the collection $\mathcal{C}$ of constant functions, i.e. $\mathcal{C} = \{\hat{k} : k \in \mathbf{R}\}$, is a linear subspace of $\mathcal{F}(X, \mathbf{R})$.

(ii) Let $\alpha : \mathcal{C} \to \mathbf{R}$ be defined by $\alpha(\hat{k}) = k$. Show that α is both one-one and onto and that, for any k, $k' \in \mathbf{R}$,

$$\alpha(\hat{k} + \hat{k}') = \alpha(\hat{k}) + \alpha(\hat{k}')$$

Answers to Supplementary Problems

23. (i) 0, (ii) −8, (iii) 3, (iv) −2, (v) 9, (vi) −1

24. (i) $(f \circ g)(x) = 4x^2 - 6x + 1$, (ii) $(g \circ f)(x) = 2x^2 + 6x - 1$, (iii) $(f \circ f)(x) = x^4 + 6x^3 + 14x^2 + 15x + 5$

25. The function $f \circ k$ is a constant function.

26. (i) yes, (ii) no, (iii) yes, (iv) yes

32. A_i is a singleton set, say $A_i = \{a_i\}$, for $i \neq i_0$.

36. (i) A_{14}, (ii) A_{24}, (iii) A_3, (iv) A_{12}, (v) A_s, (vi) A_{st}

37. (i) $(4, 6]$, (ii) $\emptyset$, (iii) $(4, 21]$, (iv) $(s, s+3]$, (v) $(s, s+16]$, (vi) R

38. (i) $\{0\}$, (ii) $\emptyset$, (iii) $\{0\}$

41. (i) $\{1, 2\}$, (ii) $\{3, -3, 4, -4\}$, (iii) $(1, 5)$, (iv) $(-3, -2)$, $(2, 3)$, (v) $\{x : x \geqq 1\}$, (vi) $\mathbf{R}$

49. (i) $3f = \{ \langle a, 6\rangle, \langle b, -9\rangle, \langle c, -3\rangle \}$

(ii) $2f - 5g = \{ \langle a, 14\rangle, \langle b, -6\rangle, \langle c, -7\rangle \}$

(iii) $fg = \{ \langle a, -4\rangle, \langle b, 0\rangle, \langle c, -1\rangle \}$

(iv) $|f| = \{ \langle a, 2\rangle, \langle b, 3\rangle, \langle c, 1\rangle \}$

(v) $f^3 = \{ \langle a, 8\rangle, \langle b, -27\rangle, \langle c, -1\rangle \}$

(vi) $|3f - fg| = \{ \langle a, 10\rangle, \langle b, 9\rangle, \langle c, 2\rangle \}$

Chapter 3

Cardinality, Order

EQUIVALENT SETS

A set A is called *equivalent* to a set B, written $A \sim B$, if there exists a function $f: A \to B$ which is one-one and onto. The function f is then said to define a *one-to-one correspondence* between the sets A and B.

A set is *finite* iff it is empty or equivalent to $\{1, 2, \ldots, n\}$ for some $n \in \mathbf{N}$; otherwise it is said to be *infinite*. Clearly two finite sets are equivalent iff they contain the same number of elements. Hence, for finite sets, equivalence corresponds to the usual meaning of two sets containing the same number of elements.

Example 1.1: Let $\mathbf{N} = \{1, 2, 3, \ldots\}$ and $E = \{2, 4, 6, \ldots\}$. The function $f: \mathbf{N} \to E$ defined by $f(x) = 2x$ is both one-one and onto; hence $\mathbf{N}$ is equivalent to E.

Example 1.2: The function $f: (-1, 1) \to \mathbf{R}$ defined by $f(x) = x/(1 - |x|)$ is both one-one and onto. Hence the open interval $(-1, 1)$ is equivalent to $\mathbf{R}$, the set of real numbers.

Observe that an infinite set can be equivalent to a proper subset of itself. This property is true of infinite sets generally.

Proposition 3.1: The relation in any collection of sets defined by $A \sim B$ is an equivalence relation.

DENUMERABLE AND COUNTABLE SETS

Let $\mathbf{N}$ be the set of positive integers $\{1, 2, 3, \ldots\}$. A set X is called *denumerable* and is said to have cardinality $\aleph_0$ (read: *aleph-null*) iff it is equivalent to $\mathbf{N}$. A set is called *countable* iff it is finite or denumerable.

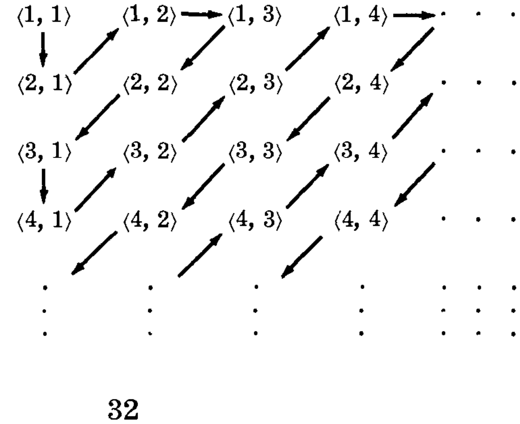
Example 2.1: The set of terms in any infinite sequence

$$a_1, a_2, a_3, \ldots$$

of distinct terms is denumerable, for a sequence is essentially a function $f(n) = a_n$ whose domain is $\mathbf{N}$. So if the a_n are distinct, the function is one-one and onto. Accordingly, each of the following sets is denumerable:

$$\{1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots\}, \quad \{1, -2, 3, -4, \ldots\}, \quad \{\langle 1,1\rangle, \langle 4,8\rangle, \langle 9,27\rangle, \ldots, \langle n^2, n^3\rangle, \ldots\}$$

Example 2.2: Consider the product set $\mathbf{N} \times \mathbf{N}$ as exhibited below.



The set $\mathbf{N} \times \mathbf{N}$ can be written in an infinite sequence of distinct elements as follows:

$$\langle 1,1\rangle, \langle 2,1\rangle, \langle 1,2\rangle, \langle 1,3\rangle, \langle 2,2\rangle, \ldots$$

(Note that the sequence is determined by "following the arrows" in the above diagram.) Thus we see that $\mathbf{N} \times \mathbf{N}$ is denumerable.

Example 2.3: Let $M = \{0,1,2,3,\ldots\} = \mathbf{N} \cup \{0\}$. Now each positive integer $a \in \mathbf{N}$ can be written uniquely in the form $a = 2^r(2s+1)$ where $r,s \in M$. The function $f: \mathbf{N} \to M \times M$ defined by

$$f(a) = \langle r, s\rangle$$

where r and s are as above, is one-one and onto. Hence $M \times M$ is denumerable. Note that $\mathbf{N} \times \mathbf{N}$ is a subset of $M \times M$.

The following theorems concern denumerable and countable sets.

Theorem 3.2: Every infinite set contains a denumerable subset.

Theorem 3.3: Every subset of a countable set is countable.

Lemma 3.4: Let $\{A_1 A_2, \ldots\}$ be a denumerable disjoint class of denumerable sets. Then $\cup_{i=1}^{\infty} A_i$ is also denumerable.

Theorem 3.5: Let $\{A_i : i \in I\}$ be a countable class of countable sets, i.e. I is countable and A_i is countable for each $i \in I$. Then $\bigcup\{A_i : i \in I\}$ is countable.

A set which is neither finite nor denumerable is said to be *non-denumerable* or *non-countable*.

THE CONTINUUM

Not every infinite set is denumerable; in fact, the next theorem gives a specific and extremely important example.

Theorem 3.6: The unit interval $[0,1]$ is non-denumerable.

A set X is said to have the *power of the continuum* or is said to have *cardinality* $\mathbf{c}$ iff it is equivalent to the unit interval $[0,1]$.

We show, in a solved problem, that every interval, open or closed, has cardinality $\mathbf{c}$. By Example 1.2, the open interval $(-1,1)$ is equivalent to $\mathbf{R}$. Hence,

Proposition 3.7: $\mathbf{R}$, the set of real numbers, has cardinality $\mathbf{c}$.

SCHROEDER-BERNSTEIN THEOREM

We write $A \precsim B$ if A is equivalent to a subset of B, i.e.,

$$A \precsim B \quad \text{iff} \quad \exists\, B^* \subset B \quad \text{such that} \quad A \sim B^*$$

We also write $A \prec B$ if $A \precsim B$ but $A \not\sim B$, i.e. A is not equivalent to B.

Example 3.1: Since $\mathbf{N}$ is a subset of $\mathbf{R}$, we may write $\mathbf{N} \precsim \mathbf{R}$. On the other hand, by Proposition 3.7, $\mathbf{R}$ is not denumerable, i.e. $\mathbf{R} \not\sim \mathbf{N}$. Accordingly, $\mathbf{N} \prec \mathbf{R}$.

Given any pair of sets A and B, then at least one of the following must be true:
(i) $A \sim B$, (ii) $A \prec B$ or $B \prec A$, (iii) $A \precsim B$ and $B \precsim A$, (iv) $A \not\prec B$, $A \not\sim B$ and $B \not\prec A$

The celebrated Schroeder-Bernstein Theorem states that, in Case (iii) above, A is equivalent to B. Namely,

Theorem (Schroeder-Bernstein) 3.8: If $A \precsim B$ and $B \precsim A$, then $A \sim B$.

The Schroeder-Bernstein Theorem can be restated as follows:

Theorem 3.8: Let $X \supset Y \supset X_1$ and let $X \sim X_1$. Then $X \sim Y$.

We remark that Case (iv) above is impossible. That is,

Theorem (Law of Trichotomy) 3.9: Given any pair of sets A and B, either $A \prec B$, $A \sim B$ or $B \prec A$.

CONCEPT OF CARDINALITY

If A is equivalent to B, i.e. $A \sim B$, then we say that A and B have the same *cardinal number* or *cardinality*. We write $\#(A)$ for "the cardinal number (or cardinality) of A". So

$$\#(A) = \#(B) \quad \text{iff} \quad A \sim B$$

On the other hand, if $A \prec B$ then we say that A has *cardinality less than* B or B has *cardinality greater than* A. That is,

$$\#(A) < \#(B) \quad \text{iff} \quad A \prec B$$

So $\#(A) \leqq \#(B)$ iff $A \precsim B$. Accordingly, the Schroeder-Bernstein Theorem can be re-stated as follows:

Theorem 3.8: If $\#(A) \leqq \#(B)$ and $\#(B) \leqq \#(A)$, then $\#(A) = \#(B)$.

The cardinal number of each of the sets

$$\emptyset, \quad \{\emptyset\}, \quad \{\emptyset, \{\emptyset\}\}, \quad \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \quad \ldots$$

is denoted by $0, 1, 2, 3, \ldots$, respectively, and is called a *finite* cardinal. The cardinal numbers of $\mathbf{N}$ and $[0,1]$ are denoted by

$$\aleph_0 = \#(N), \qquad \mathbf{c} = \#([0,1])$$

Accordingly, we may write $\quad 0 < 1 < 2 < 3 < \cdots < \aleph_0 < \mathbf{c}$

CANTOR'S THEOREM AND THE CONTINUUM HYPOTHESIS

It is natural to ask if there are infinite cardinal numbers other than $\aleph_0$ and $\mathbf{c}$. The answer is yes. In fact, Cantor's Theorem determines a set with cardinality greater than any given set. Namely,

Theorem (Cantor) 3.10: The power set $\mathcal{P}(A)$ of any set A has cardinality greater than A.

It is also natural to ask if there exists a set whose cardinality lies between $\aleph_0$ and $\mathbf{c}$. The conjecture that the answer to this question is negative is known as the Continuum Hypothesis. That is,

Continuum Hypothesis: There does not exist a set A with the property that $\aleph_0 < \#(A) < \mathbf{c}$.

In 1963 it was shown that the Continuum Hypothesis is independent of our axioms of set theory in somewhat the same sense that Euclid's Fifth Postulate on parallel lines is independent of the other axioms of geometry.

PARTIALLY ORDERED SETS

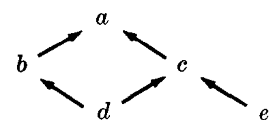
A relation $\precsim$ in a set A is called a *partial order* (or *order*) on A iff, for every $a, b, c \in A$: (i) $a \precsim a$; (ii) $a \precsim b$ and $b \precsim a$ implies $a = b$; and (iii) $a \precsim b$ and $b \precsim c$ implies $a \precsim c$. The set A together with the partial order, i.e. the pair $(A, \precsim)$, is called a *partially ordered set*.

Recall that a relation is reflexive iff it satisfies (i), and transitive iff it satisfies (iii). A relation is said to be *anti-symmetric* iff it satisfies (ii). In other words, a partial order is a reflexive, anti-symmetric, transitive relation.

Example 4.1: Set inclusion is a partial order in any class of sets since: (i) $A \subset A$ for any set A; (ii) $A \subset B$ and $B \subset A$ implies $A = B$; and (iii) $A \subset B$ and $B \subset C$ implies $A \subset C$.

Example 4.2: Let A be any set of real numbers. Then the relation in A defined by $x \leq y$ is a partial order and is called the *natural order* in A.

Example 4.3: Let $X = \{a, b, c, d, e\}$. Then the diagram



defines a partial order in X as follows: $x \lesssim y$ iff $x = y$ or if one can go from x to y in the diagram, always moving in the indicated direction, i.e. upward.

If $a \lesssim b$ in an ordered set, then we say that a *precedes* or is *smaller* than b and that b *follows* or *dominates* or is *larger* than a. Furthermore, we write $a \prec b$ if $a \lesssim b$ but $a \neq b$.

A partially ordered set A is said to be *totally* (or *linearly*) *ordered* if, for every $a, b \in A$, either $a \lesssim b$ or $b \lesssim a$. **R**, the set of real numbers, with the natural order defined by $x \leq y$ is an example of a totally ordered set.

Example 4.4: Let A and B be totally ordered. Then the product set $A \times B$ can be totally ordered as follows:

$$\langle a, b \rangle \prec \langle a', b' \rangle \quad \text{if} \quad a \prec a' \quad \text{or if} \quad a = a' \text{ and } b \prec b'$$

This order is called the *lexicographical order* of $A \times B$ since it is similar to the way words are arranged in a dictionary.

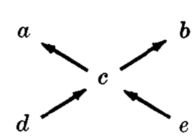
Remark: If a relation R in a set A defines a partial order, i.e. is reflexive, anti-symmetric and transitive, then the inverse relation R^{-1} is also a partial order; it is called the *inverse order*.

SUBSETS OF ORDERED SETS

Let A be a subset of a partially ordered set X. Then the order in X induces an order in A in the following natural way: If $a, b \in A$, then $a \lesssim b$ as elements in A iff $a \lesssim b$ as elements in X. More precisely, if R is a partial order in X, then the relation $R_A = R \cap (A \times A)$, called the *restriction* of R to A, is a partial order in A. The ordered set (A, R_A) is called a *(partially ordered) subset* of the ordered set (X, R).

Some subsets of a partially ordered set X may, in fact, be totally ordered. Clearly if X itself is totally ordered, every subset of X will also be totally ordered.

Example 5.1: Consider the partial order in $W = \{a, b, c, d, e\}$ defined by the diagram

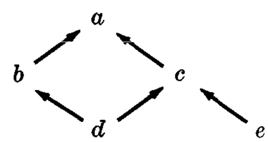


The sets $\{a, c, d\}$ and $\{b, e\}$ are totally ordered subsets; the sets $\{a, b, c\}$ and $\{d, e\}$ are not totally ordered subsets.

FIRST AND LAST ELEMENTS

Let X be an ordered set. An element $a_0 \in X$ is a *first* or *smallest* element of X iff $a_0 \lesssim x$ for all $x \in X$. Analogously, an element $b_0 \in X$ is a *last* or *largest* element of X iff $x \lesssim b_0$ for all $x \in X$.

Example 6.1: Let $X = \{a, b, c, d, e\}$ be ordered by the diagram



Then a is a last element since a follows every element. Note that X has no first element. The element d is not a first element since d does not precede e.

Example 6.2: The positive integers **N** with the natural order has 1 as a first element. The integers **Z** with the natural order has no first element and no last element.

MAXIMAL AND MINIMAL ELEMENTS

Let X be an ordered set. An element $a_0 \in X$ is *maximal* iff $a_0 \precsim x$ implies $x = a$, i.e. if no element follows a_0 except itself. Similarly, an element $b_0 \in X$ is *minimal* iff $x \precsim b_0$ implies $x = b_0$, i.e. if no element precedes b_0 except itself.

Example 7.1: Let $X = \{a, b, c, d, e\}$ be ordered by the diagram in Example 6.1. Then both d and e are minimal elements. The element a is a maximal element.

Example 7.2: Although **R** with the natural order is totally ordered it has no minimal and no maximal elements.

Example 7.3: Let $A = \{a_1, a_2, \ldots, a_m\}$ be a finite totally ordered set. Then A contains precisely one minimal element and precisely one maximal element, denoted respectively by

$$\min\{a_1, \ldots, a_m\} \quad \text{and} \quad \max\{a_1, \ldots, a_m\}$$

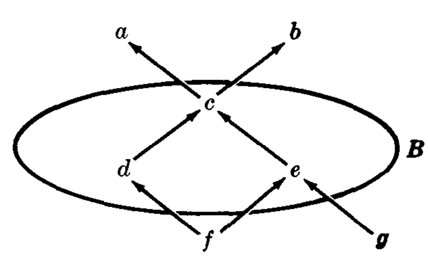
UPPER AND LOWER BOUNDS

Let A be a subset of a partially ordered set X. An element $m \in X$ is a *lower bound* of A iff $m \precsim x$ for all $x \in A$, i.e. if m precedes every element in A. If some lower bound of A follows every other lower bound of A, then it is called the *greatest lower bound* (g.l.b.) or *infimum* of A and is denoted by $\inf(A)$.

Similarly, an element $M \in X$ is an *upper bound* of A iff $x \precsim M$ for all $x \in A$, i.e. if M follows every element in A. If some upper bound of A precedes every other upper bound of A, then it is called the *least upper bound* (l.u.b.) or *supremum* of A and is denoted by $\sup(A)$.

A is said to be *bounded above* if it has an upper bound, and *bounded below* if it has a lower bound. If A has both an upper and lower bound, then it is said to be *bounded.*

Example 8.1: Let $X = \{a, b, c, d, e, f, g\}$ be ordered by the following diagram:



Let $B = \{c, d, e\}$. Then a, b and c are upper bounds of B, and f is the only lower bound of B. Note that g is not a lower bound of B since g does not precede d. Furthermore, $c = \sup(B)$ belongs to B, while $f = \inf(B)$ does not belong to B.

Example 8.2: Let A be a bounded set of real numbers. Then a fundamental theorem about real numbers states that, under the natural order, $\inf(A)$ and $\sup(A)$ exist.

Example 8.3: Let **Q** be the set of rational numbers. Let

$$B = \{x : x \in \mathbf{Q},\ x > 0,\ 2 < x^2 < 3\}$$

that is, B consists of those rational points which lie between $\sqrt{2}$ and $\sqrt{3}$ on the real line. Then B has an infinite number of upper and lower bounds, but $\inf(B)$ and $\sup(B)$ do not exist. Note that the real numbers $\sqrt{2}$ and $\sqrt{3}$ do not belong to **Q** and cannot be considered as upper or lower bounds of B.

ZORN'S LEMMA

Zorn's Lemma is one of the most important tools in mathematics; it asserts the existence of certain types of elements although no constructive process is given to find these elements.

Zorn's Lemma 3.11: Let X be a non-empty partially ordered set in which every totally ordered subset has an upper bound. Then X contains at least one maximal element.

Remark. Zorn's Lemma is equivalent to the classical Axiom of Choice and the Well-ordering Principle. The proof of this fact, which uses the concept of ordinal numbers, is beyond the scope of this text.

Solved Problems

EQUIVALENT SETS, DENUMERABLE SETS

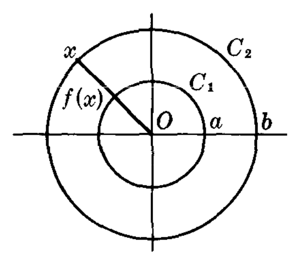
1. Consider the concentric circles

$$C_1 = \{\langle x,y\rangle : x^2+y^2 = a^2\}, \quad C_2 = \{\langle x,y\rangle : x^2+y^2 = b^2\}$$

where, say $0 < a < b$. Establish, geometrically, a one-to-one correspondence between C_1 and C_2.

Solution:

Let $x \in C_2$. Consider the function $f: C_2 \to C_1$, where $f(x)$ is the point of intersection of the radius from the center of C_2 (and C_1) to x, and C_1, as shown in the adjacent diagram.



Note that f is both one-one and onto. Thus f defines a one-to-one correspondence between C_1 and C_2.

2. Prove: The set of rational numbers is denumerable.

Let $\mathbf{Q}^+$ be the set of positive rational numbers and let $\mathbf{Q}^-$ be the set of negative rational numbers. Then $\mathbf{Q} = \mathbf{Q}^- \cup \{0\} \cup \mathbf{Q}^+$ is the set of rational numbers.

Let the function $f: \mathbf{Q}^+ \to \mathbf{N} \times \mathbf{N}$ be defined by

$$f(p/q) = \langle p, q\rangle$$

where p/q is any positive rational number expressed as the ratio of two positive integers. Note f is one-one; hence $\mathbf{Q}^+$ is equivalent to a subset of $\mathbf{N} \times \mathbf{N}$. But $\mathbf{N} \times \mathbf{N}$ is denumerable (see Example 2.2); hence $\mathbf{Q}^+$ is also denumerable. Similarly $\mathbf{Q}^-$ is denumerable. Accordingly, by Theorem 3.5, the union of $\mathbf{Q}^-$, $\{0\}$ and $\mathbf{Q}^+$, i.e. the set of rational numbers, is also denumerable.

3. Prove Proposition 3.1: The relation in any collection of sets defined by $A \sim B$ is an equivalence relation. That is, (i) $A \sim A$ for any set A; (ii) if $A \sim B$ then $B \sim A$; and (iii) if $A \sim B$ and $B \sim C$ then $A \sim C$.

Solution:

(i) The identity function $1_A: A \to A$ is one-one and onto; hence $A \sim A$.

(ii) If $A \sim B$, then there exists $f: A \to B$ which is one-one and onto. But then f has an inverse $f^{-1}: B \to A$ which is also one-one and onto. Hence

$$A \sim B \quad \text{implies} \quad B \sim A$$

(iii) If $A \sim B$ and $B \sim C$, then there exist functions $f: A \to B$ and $g: B \to C$ which are one-one and onto. Thus the composition function $g \circ f: A \to C$ is also one-one and onto. Hence

$$A \sim B \text{ and } B \sim C \quad \text{implies} \quad A \sim C$$

4. Prove: The collection P of all polynomials

$$p(x) = a_0 + a_1x + \cdots + a_mx^m$$

with integral coefficients, i.e. where $a_0, a_1, \ldots, a_m$ are integers, is denumerable.

Solution:

For each pair of positive integers $\langle n, m\rangle \in \mathbf{N} \times \mathbf{N}$, let P_{nm} denote the set of polynomials $p(x)$ of degree m in which

$$|a_0| + |a_1| + \cdots + |a_m| = n$$

Observe that P_{nm} is finite. Accordingly,

$$P = \bigcup\{P_{nm} : \langle n, m\rangle \in \mathbf{N} \times \mathbf{N}\}$$

is countable since it is a countable union of countable sets. In particular, since P is not finite, P is denumerable.

5. A real number r is called an *algebraic number* if r is a solution to a polynomial equation

$$p(x) = a_0 + a_1x + \cdots + a_mx^m$$

with integral coefficients. Prove that the set A of algebraic numbers is denumerable.

Solution:

Note, by the preceding problem, that the set E of polynomial equations is denumerable:

$$E = \{p_1(x) = 0,\ p_2(x) = 0,\ p_3(x) = 0,\ \ldots\}$$

Let

$$A_i = \{x : x \text{ is a solution of } p_i(x) = 0\}$$

Since a polynomial of degree n can have at most n roots, each A_i is finite. Hence $A = \bigcup\{A_i : i \in \mathbf{N}\}$ is denumerable.

6. Prove Theorem 3.2: Every infinite set X contains a subset D which is denumerable.

Solution:

Let $f: \mathcal{P}(X) \to X$ be a choice function, i.e. for each non-empty subset A of X, $f(A) \in A$. (Such a function exists by virtue of the Axiom of Choice.) Consider the following sequence:

$$\begin{aligned} a_1 &= f(X) \\ a_2 &= f(X \setminus \{a_1\}) \\ a_3 &= f(X \setminus \{a_1, a_2\}) \\ &\cdots\cdots\cdots\cdots \\ a_n &= f(X \setminus \{a_1, \ldots, a_{n-1}\}) \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

Since X is infinite, $X \setminus \{a_1, \ldots, a_{n-1}\}$ is not empty for every $n \in N$. Furthermore, since f is a choice function,

$$a_n \neq a_i \quad \text{for} \quad i < n$$

Thus the a_n are distinct and $D = \{a_1, a_2, \ldots\}$ is a denumerable subset of X.

Essentially, the choice function f "chooses" an element $a_1 \in X$, then chooses an element a_2 from those elements which "remain" in X, etc. Since X is infinite, the set of elements which "remain" in X is non-empty.

7. Prove: Let X be any set and let $C(X)$ be the collection of characteristic functions on X, i.e. the collection of functions $f: X \to \{1, 0\}$. Then the power set of X is equivalent to $C(X)$, i.e. $\mathcal{P}(X) \sim C(X)$.

Solution:

Let A be any subset of X, i.e. $A \in \mathcal{P}(X)$. Let $f: \mathcal{P}(X) \to C(X)$ be defined by

$$f(A) \;=\; \chi_A \;=\; \begin{cases} 0 & \text{if } x \notin A \\ 1 & \text{if } x \in A \end{cases}$$

Then f is one-one and onto. Hence $\mathcal{P}(X) \sim C(X)$.

8. Prove: A subset of a denumerable set is either finite or denumerable, i.e. is countable.

Solution:

Let $X = \{a_1, a_2, \ldots\}$ be any denumerable set and let A be a subset of X. If $A = \emptyset$, then A is finite. If $A \neq \emptyset$, then let n_1 be the least positive integer such that $a_{n_1} \in A$; let n_2 be the least positive integer such that $n_2 > n_1$ and $a_{n_2} \in A$; etc. Then $A = \{a_{n_1}, a_{n_2}, \ldots\}$. If the set of integers $\{n_1, n_2, \ldots\}$ is bounded, then A is finite. Otherwise A is denumerable.

9. Prove Theorem 3.3: Every subset of a countable set is countable.

Solution:

If X is countable, then X is either finite or denumerable. In either case, its subsets are countable.

10. Prove Lemma 3.4: Let $\{A_1, A_2, \ldots\}$ be a denumerable disjoint class of denumerable sets. Then $\bigcup_{i=1}^{\infty} A_i$ is denumerable.

Solution:

Since the sets A_i are denumerable, we can write

$$\begin{aligned} A_1 &= \{a_{11}, a_{12}, a_{13}, \ldots\} \\ A_2 &= \{a_{21}, a_{22}, a_{23}, \ldots\} \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ A_n &= \{a_{n1}, a_{n2}, a_{n3}, \ldots\} \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

Then $\bigcup_{i=1}^{\infty} A_i = \{a_{ij} : \langle i, j\rangle \in \mathbf{N} \times \mathbf{N}\}$. The function $f: \bigcup_{i=1}^{\infty} A_i \to \mathbf{N} \times \mathbf{N}$ defined by $f(a_{ij}) = \langle i, j\rangle$ is clearly one-one and onto. Hence $\bigcup_{i=1}^{\infty} A_i$ is denumerable since $\mathbf{N} \times \mathbf{N}$ is denumerable.

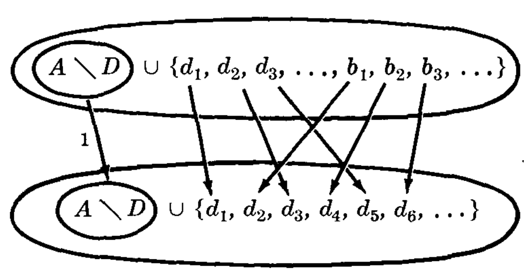
11. Prove: Let A be an infinite set, let $B = \{b_1, b_2, \ldots\}$ be denumerable, and let A and B be disjoint. Then $A \cup B \sim A$.

Solution:

Since A is infinite, A contains a denumerable subset $D = \{d_1, d_2, \ldots\}$. Let $f: A \cup B \to A$ be defined by the following diagram:

$A \cup B \;=\; (A \setminus D) \cup (D \cup B) \;=$

$A \;=\; (A \setminus D) \cup D \;=$



In other words,

$$f(x) \;=\; \begin{cases} x & \text{if } x \in A \setminus D \\ d_{2n-1} & \text{if } x = d_n \\ d_{2n} & \text{if } x = b_n \end{cases}$$

Observe that f is one-one and onto; hence $A \cup B \sim A$.

CONTINUUM, CARDINALITY

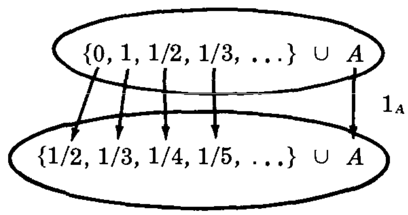
12. Prove that the intervals $(0,1)$, $[0,1)$ and $(0,1]$ have cardinality **c**, i.e. is equivalent to $[0,1]$.

Solution:

(i) Note that $[0,1] = \{0, 1, 1/2, 1/3, \ldots\} \cup A$, $(0,1) = \{1/2, 1/3, 1/4, \ldots\} \cup A$

where $A = [0,1] \setminus \{0, 1, 1/2, 1/3, \ldots\} = (0,1) \setminus \{1/2, 1/3, 1/4, \ldots\}$

Consider the function $f : [0,1] \to (0,1)$ defined by the following diagram



In other words,

$$f(x) = \begin{cases} 1/2 & \text{if } x = 0 \\ 1/(n+2) & \text{if } x = 1/n,\ n \in \mathbf{N} \\ x & \text{if } x \neq 0,\ 1/n,\ n \in \mathbf{N}, \text{ i.e. if } x \in A \end{cases}$$

The function f is one-one and onto. Accordingly, $[0,1] \sim (0,1)$.

(ii) The function $f : [0,1] \to [0,1)$ defined by

$$f(x) = \begin{cases} 1/(n+1) & \text{if } x = 1/n,\ n \in \mathbf{N} \\ x & \text{if } x \neq 1/n,\ n \in \mathbf{N} \end{cases}$$

is one-one and onto. (It is similar to the function in Part (i)). Hence $[0,1] \sim [0,1)$.

(iii) Let $f : [0,1) \to (0,1]$ be defined by $f(x) = 1 - x$. Then f is one-one and onto. Hence $[0,1) \sim (0,1]$ and, by transitivity, $[0,1] \sim (0,1]$.

In other words, $(0,1)$, $[0,1)$ and $(0,1]$ have cardinality **c**.

13. Prove: Each of the following intervals has the power of the continuum, i.e. has cardinality **c**: $[a,b]$, (a,b), $[a,b)$ and $(a,b]$. Here $a < b$.

Solution:

Let each of the following functions be defined by $f(x) = a + (b-a)x$:

$$[0,1] \xrightarrow{f} [a,b] \qquad [0,1) \xrightarrow{f} [a,b) \qquad (0,1) \xrightarrow{f} (a,b) \qquad (0,1] \xrightarrow{f} (a,b]$$

Each function is one-one and onto. Hence by the preceding problem and Proposition 3.1, each interval is equivalent to $[0,1]$, i.e. has cardinality **c**.

14. Prove Theorem 3.6: The unit interval $A = [0,1]$ is non-denumerable.

Solution:

Method 1. Assume the contrary; then

$$A = \{x_1, x_2, x_3, \ldots\}$$

i.e. the elements of A can be written in a sequence.

Each element in A can be written in the form of an infinite decimal as follows:

$$\begin{aligned} x_1 &= 0.\ a_{11}\, a_{12}\, a_{13} \ldots a_{1n} \ldots \\ x_2 &= 0.\ a_{21}\, a_{22}\, a_{23} \ldots a_{2n} \ldots \\ x_3 &= 0.\ a_{31}\, a_{32}\, a_{33} \ldots a_{3n} \ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n &= 0.\ a_{n1}\, a_{n2}\, a_{n3} \ldots a_{nn} \ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

where $a_{ij} \in \{0, 1, \ldots, 9\}$ and where each decimal contains an infinite number of non-zero elements, i.e. for those numbers which can be written in the form of a decimal in two ways, e.g.,

$$1/2 = .5000\ldots = .4999\ldots$$

we write the infinite decimal in which all except a finite set of digits are nines.

Now construct the real number

$$y = 0.\,b_1\,b_2\,b_3\,\ldots\,b_n\,\ldots$$

which will belong to A, in the following way: choose b_1 so $b_1 \neq a_{11}$ and $b_1 \neq 0$, choose b_2 so $b_2 \neq a_{22}$ and $b_2 \neq 0$, etc.

Observe that $y \neq x_1$ since $b_1 \neq a_{11}$, $y_2 \neq x_2$ since $b_2 \neq a_{22}$, etc., that is, $y \neq x_n$, for $n \in \mathbf{N}$. Hence $y \notin A$, which is impossible. Thus the assumption that A is denumerable has led to a contradiction. Consequently, A is non-denumerable.

Method 2. Assume the contrary. Then, as above,

$$A = \{x_1, x_2, x_3, \ldots\}$$

Now construct a sequence of closed intervals as follows: Consider the following three closed sub-intervals of $A = [0, 1]$,

$$[0, \tfrac{1}{3}], \quad [\tfrac{1}{3}, \tfrac{2}{3}], \quad [\tfrac{2}{3}, 1] \tag{1}$$

each having length $\frac{1}{3}$. Now x_1 cannot belong to all three intervals. Let $I_1 = [a_1, b_1]$ be one of the intervals in (*1*) such that $x_1 \notin I_1$.

Now consider the following three closed sub-intervals of $I_1 = [a_1, b_1]$,

$$[a_1, a_1 + \tfrac{1}{9}], \quad [a_1 + \tfrac{1}{9}, a_1 + \tfrac{2}{9}], \quad [a_1 + \tfrac{2}{9}, b_1] \tag{2}$$

each having length $\frac{1}{9}$. Similarly, let I_2 be one of the intervals in (*2*) such that $x_2 \notin I_2$.

By continuing in this manner, we obtain a sequence of closed intervals

$$I_1 \supset I_2 \supset I_3 \supset \cdots \tag{3}$$

such that $x_n \notin I_n$ for all $n \in N$. By the *Nested Interval Property* (see Appendix A) of the real numbers, there exists a real number $y \in A = [0, 1]$ such that y belongs to every interval in (*3*). But

$$y \in A = \{x_1, x_2, \ldots\} \quad \text{implies} \quad y = x_{m_0} \text{ for some } m_0 \in \mathbf{N}$$

Then by our construction $y = x_{m_0} \notin I_{m_0}$, which contradicts the fact that y belongs to every interval in (*3*). Thus our assumption that A is denumerable has led to a contradiction. In other words, A is non-denumerable.

15. Prove Theorem (Schroeder-Bernstein) 3.8: Let $X \supset Y \supset X_1$ and let $X \sim X_1$; then $X \sim Y$.

Solution:

Since $X \sim X_1$, there exists a function $f: X \to X_1$ which is one-one and onto. But $X \supset Y$; hence the restriction of f to Y, which we shall also denote by f, is also one-one. So Y is equivalent to a subset of X_1, i.e. $Y \sim Y_1$ where

$$X \supset Y \supset X_1 \supset Y_1$$

and $f: Y \to Y_1$ is one-one and onto. But now $Y \supset X_1$; hence, for similar reasons, $X \sim X_2$ where

$$X \supset Y \supset X_1 \supset Y_1 \supset X_2$$

and $f: X_1 \to X_2$ is one-one and onto. Consequently, there exist equivalent sets $X_1, X_2, X_3, \ldots$ and equivalent sets $Y_1, Y_2, Y_3, \ldots$ such that

$$X \supset Y \supset X_1 \supset Y_1 \supset X_2 \supset Y_2 \supset \cdots$$

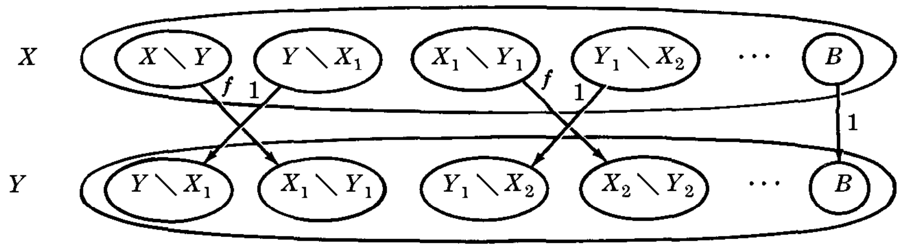
Let
$$B = X \cap Y \cap X_1 \cap Y_1 \cap X_2 \cap Y_2 \cap \cdots$$

Then
$$X = (X \setminus Y) \cup (Y \setminus X_1) \cup (X_1 \setminus Y_1) \cup \cdots \cup B$$
$$Y = (Y \setminus X_1) \cup (X_1 \setminus Y_1) \cup (Y_1 \setminus X_2) \cup \cdots \cup B$$

Note further that
$$(X \setminus Y) \sim (X_1 \setminus Y_1) \sim (X_2 \setminus Y_2) \sim \cdots$$

Specifically, the function $f: (X_n \setminus Y_n) \to (X_{n+1} \setminus Y_{n+1})$ is one-one and onto.

Consider the function $g : X \to Y$ defined by the following diagram:



In other words,

$$g(x) = \begin{cases} f(x) & \text{if } x \in X_i \setminus Y_i \text{ or } x \in X \setminus Y \\ x & \text{if } x \in Y_i \setminus X_i \text{ or } x \in B \end{cases}$$

Then g is one-one and onto. Therefore $X \sim Y$.

16. Prove Theorem (Cantor) 3.10: The power set $\mathcal{P}(A)$ of any arbitrary set A has cardinality greater than A, i.e. $A \prec \mathcal{P}(A)$ and hence $\#(A) < \#(\mathcal{P}(A))$.

Solution:

The function $g : A \to \mathcal{P}(A)$ which sends each element $a \in A$ into the singleton set $\{a\}$, i.e. $g(a) = \{a\}$, is one-one; hence $A \precsim \mathcal{P}(A)$.

If we show that A is not equivalent to $\mathcal{P}(A)$, then the theorem will follow. Suppose the contrary, i.e. let there exist a function $f : A \to \mathcal{P}(A)$ which is one-one and onto. Call $a \in A$ a "bad" element if a is not a member of the set which is its image, i.e. if $a \notin f(a)$. Let B be the set of "bad" elements, i.e.,

$$B = \{x : x \in A,\ x \notin f(x)\}$$

Observe that B is a subset of A, that is, $B \in \mathcal{P}(A)$. Since $f : A \to \mathcal{P}(A)$ is onto, there exists an element $b \in A$ with the property that $f(b) = B$. Question: Is b "bad" or "good"? If $b \in B$ then, by definition of B, $b \notin f(b) = B$ which is a contradiction. Likewise, if $b \notin B$, then $b \in f(b) = B$ which is also a contradiction. Thus the original assumption, that $A \sim \mathcal{P}(A)$, has led to a contradiction. Accordingly $A \sim \mathcal{P}(A)$ is false, and so the theorem is true.

ORDERED SETS AND SUBSETS

17. Let **N**, the positive integers, be ordered as follows: each pair of elements $a, a' \in \mathbf{N}$ can be written uniquely in the form

$$a = 2^r(2s+1), \quad a' = a^{r'}(as'+1)$$

where $r, r', s, s' \in \{0, 1, 2, 3, \ldots\}$. Let

$$a \prec a' \text{ if } r < r' \text{ or if } r = r' \text{ but } s < s'$$

Insert the correct symbol, $\prec$ or $\succ$, between each of the following pairs of numbers. (Here $x \succ y$ iff $y \prec x$.)

(i) 5__14, (ii) 6__9, (iii) 3__20, (iv) 14__21

Solution:

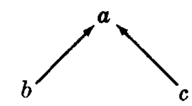
The elements in **N** can be written as follows:

r \ s	0	1	2	3	4	5	6	7	
0	1	3	5	7	9	11	13	15	...
1	2	6	10	14	18	22	26	30	...
2	4	12	20	28	36	44	52	60	...
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Then a number in a higher row precedes a number in a lower row and, if two numbers are in the same row, the number to the left precedes the number to the right. Accordingly,

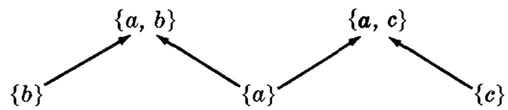
(i) $5 \prec 14$, (ii) $6 \succ 9$, (iii) $3 \prec 20$, (iv) $14 \succ 21$

18. Let $A = \{a, b, c\}$ be ordered as in the diagram on the right. Let $\mathcal{A}$ be the collection of all non-empty totally ordered subsets of A, and let $\mathcal{A}$ be partially ordered by set inclusion. Construct a diagram of the order of $\mathcal{A}$.



Solution:

The totally ordered subsets of A are: $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{a, c\}$. Since $\mathcal{A}$ is ordered by set inclusion, the order of $\mathcal{A}$ is the following:



19. Let $A = \{2, 3, 4, \ldots\} = \mathbf{N} \setminus \{1\}$, and let A be ordered by "x divides y". (i) Determine the minimal elements of A. (ii) Determine the maximal elements of A.

Solution:

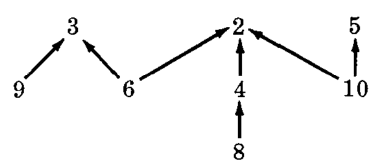
(i) If $p \in A$ is a prime number, then only p divides p (since $1 \notin A$); hence all prime numbers are minimal elements. Furthermore, if $a \in A$ is not prime, then there is a number $b \in A$ such that b divides a, i.e. $b \prec a$; hence a is not minimal. In other words, the minimal elements are precisely the prime numbers.

(ii) There are no maximal elements since, for every $a \in A$, a divides $2a$, for example.

20. Let $B = \{2, 3, 4, 5, 6, 8, 9, 10\}$ be ordered by "x is a multiple of y". (i) Find all maximal elements of B. (ii) Find all minimal elements of B.

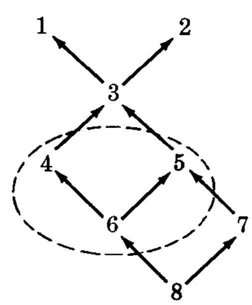
Solution:

Construct a diagram of the order of B as follows:



(i) The maximal elements are 2, 3 and 5. (ii) The minimal elements are 6, 8, 9 and 10.

21. Let $W = \{1, 2, \ldots, 7, 8\}$ be ordered as follows:



Consider the subset $V = \{4, 5, 6\}$ of W. (i) Find the set of upper bounds of V. (ii) Find the set of lower bounds of V. (iii) Does $\sup(V)$ exist? (iv) Does $\inf(V)$ exist?

Solution:

(i) Each of the elements in $\{1, 2, 3\}$, and only these elements, follows every element in V and hence is an upper bound.

(ii) Only 6 and 8 precede every element in V; hence $\{6, 8\}$ is the set of lower bounds.

(iii) Since 3 is a first element in the set of upper bounds of V, $\sup(V) = 3$. Note that $3 \notin V$.

(iv) Since 6 is a last element in the set of lower bounds of V, $\inf(V) = 6$. Note that $6 \in V$.

22. Let $\mathcal{A}$ be a collection of sets partially ordered by set inclusion, and let $\mathcal{B}$ be a subcollection of $\mathcal{A}$. (i) Prove that if $A \in \mathcal{A}$ is an upper bound of $\mathcal{B}$, then $\bigcup\{B : B \in \mathcal{B}\} \subset A$. (ii) Is $\bigcup\{B : B \in \mathcal{B}\}$ an upper bound of $\mathcal{B}$?

Solution:

(i) Let $x \in \bigcup\{B : B \in \mathcal{B}\}$; then $\exists B_0 \in \mathcal{B}$ s.t. $x \in B_0$. But A is an upper bound of $\mathcal{B}$; so $B_0 \subset A$ and hence $x \in A$. Accordingly, $\bigcup\{B : B \in \mathcal{B}\} \subset A$.

(ii) Even though $\mathcal{B}$ is a subcollection of $\mathcal{A}$, it need not be true that the union of members of $\mathcal{B}$, i.e. $\bigcup\{B : B \in \mathcal{B}\}$, is a member of $\mathcal{A}$. In other words, $\bigcup\{B : B \in \mathcal{B}\}$ is an upper bound of $\mathcal{B}$ if and only if it belongs to $\mathcal{A}$.

APPLICATIONS OF ZORN'S LEMMA

23. Prove: Let X be a partially ordered set. Then there exists a totally ordered subset of X which is not a proper subset of any other totally ordered subset of X.

Solution:

Let $\mathcal{A}$ be the class of all totally ordered subsets of X. Let $\mathcal{A}$ be partially ordered by set inclusion. We want to show, by Zorn's Lemma, that $\mathcal{A}$ possesses a maximal element. So suppose $\mathcal{B} = \{B_i : i \in I\}$ is a totally ordered subclass of $\mathcal{A}$. Let $A = \bigcup\{B_i : i \in I\}$.

Observe that $\quad B_i \subset X$ for all $B_i \in \mathcal{B}$ implies $A \subset X$

We next show that A is totally ordered. Let $a, b \in A$; then

$$\exists B_j,\ B_k \in \mathcal{B} \quad \text{such that} \quad a \in B_j,\ b \in B_k$$

But $\mathcal{B}$ is totally ordered by set inclusion; hence one of them, say B_j, is a subset of the other. Consequently, $a, b \in B_k$. Recall that $B_k \in \mathcal{B}$ is a totally ordered subset of X; so either $a \precsim b$ or $b \precsim a$. Then A is a totally ordered subset of X, and so $A \in \mathcal{A}$.

But $B_i \subset A$ for all $B_i \in \mathcal{B}$; hence A is an upper bound of $\mathcal{B}$. Since every totally ordered subset of $\mathcal{A}$ has an upper bound in $\mathcal{A}$, by Zorn's Lemma, $\mathcal{A}$ has a maximal element, i.e. a totally ordered subset of X which is not a proper subset of any other totally ordered subset of X.

24. Prove: Let R be a relation from A to B, i.e. $R \subset A \times B$, and suppose the domain of R is A. Then there exists a subset f^* of R such that f^* is a function from A into B.

Solution:

Let $\mathcal{A}$ be the class of subsets of R such that each $f \in \mathcal{A}$ is a function from a subset of A into B. Partially order $\mathcal{A}$ by set inclusion. Observe that if $f : A_1 \to B$ is a subset of $g : A_2 \to B$ then $A_1 \subset A_2$.

Now suppose $\mathcal{B} = \{f_i : A_i \to B\}_{i \in I}$ is a totally ordered subset of $\mathcal{A}$. Then (see Problem 44) $f = \cup_i f_i$ is a function from $\cup_i A_i$ into B. Furthermore, $f \subset R$. Hence f is an upper bound of $\mathcal{B}$. By Zorn's Lemma, $\mathcal{A}$ possesses a maximal element $f^* : A^* \to B$. If we show that $A^* = A$, then the theorem is proven.

Suppose $A^* \neq A$. Then $\exists a \in A$ s.t. $a \notin A^*$. By hypothesis, the domain of R is A; hence there exists an ordered pair $\langle a, b\rangle \in R$. Then $f^* \cup \{\langle a, b\rangle\}$ is a function from $A^* \cup \{a\}$ into B. But this contradicts the fact that f^* is a maximal element in $\mathcal{A}$. So $A^* = A$, and the theorem is proven.

Supplementary Problems

EQUIVALENT SETS, CARDINALITY

25. Prove: Every infinite set is equivalent to a proper subset of itself.

26. Prove: If A and B are denumerable, then $A \times B$ is denumerable.

27. Prove: The set of points in the plane R^2 with rational coordinates is denumerable.

28. A real number x is called *transcendental* if x is not algebraic, i.e. if x is not a solution to a polynomial equation

$$p(x) = a_0 + a_1x + \cdots + a_mx^m = 0$$

with integral coefficients (see Problem 5). For example, π and e are transcendental numbers.

(i) Prove that the set T of transcendental numbers is non-denumerable.

(ii) Prove that T has the power of the continuum, i.e. has cardinality $\mathbf{c}$.

29. An operation of multiplication is defined for cardinal numbers as follows:

$$\#(A)\,\#(B) = \#(A \times B)$$

(i) Show that the operation is well-defined, i.e.,

$$\#(A) = \#(A') \text{ and } \#(B) = \#(B') \quad \text{implies} \quad \#(A)\,\#(B) = \#(A')\,\#(B')$$

or, equivalently, $A \sim A'$ and $B \sim B'$ implies $(A \times B) \sim (A' \times B')$

(ii) Prove: (*a*) $\aleph_0\aleph_0 = \aleph_0$, (*b*) $\aleph_0\mathbf{c} = \mathbf{c}$, (*c*) $\mathbf{c}\,\mathbf{c} = \mathbf{c}$.

30. An operation of addition is defined for cardinal numbers as follows:

$$\#(A) + \#(B) = \#(A \times \{1\} \cup B \times \{2\})$$

(i) Show that if $A \cap B = \emptyset$, then $\#(A) + \#(B) = \#(A \cup B)$.

(ii) Show that the operation is well-defined, i.e.,

$$\#(A) = \#(A') \text{ and } \#(B) = \#(B') \quad \text{implies} \quad \#(A) + \#(B) = \#(A') + \#(B')$$

31. An operation of powers is defined for cardinal numbers as follows:

$$\#(A)^{\#(B)} = \#(\{f : f : B \to A\})$$

(i) Show that if $\#(A) = m$ and $\#(B) = n$ are finite cardinals, then

$$\#(A)^{\#(B)} = m^n$$

i.e. the operation of powers for cardinals corresponds, in the case of finite cardinals, to the usual operation of powers of positive integers.

(ii) Show that the operation is well-defined, i.e.,

$$\#(A) = \#(A') \text{ and } \#(B) = \#(B') \quad \text{implies} \quad \#(A)^{\#(B)} = \#(A')^{\#(B')}$$

(iii) Prove: For any set A, $\#(\mathcal{P}(A)) = 2^{\#(A)}$.

32. Let $\sim$ be the equivalence relation in $\mathbf{R}$ defined by $x \sim y$ iff $x - y$ is rational. Determine the cardinality of the quotient set $\mathbf{R}/\sim$.

33. Prove: The cardinal number of the class of all functions from $[0,1]$ into $\mathbf{R}$ is $2^{\mathbf{c}}$.

34. Prove that the following two statements of the Schroeder-Bernstein Theorem 3.8 are equivalent:

(i) If $A \precsim B$ and $B \precsim A$, then $A \sim B$.

(ii) If $X \supset Y \supset X_1$ and $X \sim X_1$, then $X \sim Y$.

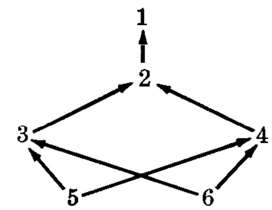
35. Prove Theorem 3.9: Given any pair of sets A and B, either $A \prec B$, $A \sim B$ or $B \prec A$. (*Hint.* Use Zorn's Lemma.)

ORDERED SETS AND SUBSETS

36. Let $A = (\mathbf{N}, \leqq)$, the positive integers with the natural order; and let $B = (\mathbf{N}, \geqq)$, the positive integers with the inverse order. Furthermore, let $A \times B$ denote the lexicographical ordering of $\mathbf{N} \times \mathbf{N}$ according to the order of A and then B. Insert the correct symbol, $\prec$ or $\succ$, between each pair of elements of $\mathbf{N} \times \mathbf{N}$.

(i) $\langle 3,8\rangle$____$\langle 1,1\rangle$, (ii) $\langle 2,1\rangle$____$\langle 2,8\rangle$, (iii) $\langle 3,3\rangle$____$\langle 3,1\rangle$, (iv) $\langle 4,9\rangle$____$\langle 7,15\rangle$.

37. Let $X = \{1, 2, 3, 4, 5, 6\}$ be ordered as in the adjacent diagram. Consider the subset $A = \{2, 3, 4\}$ of X. (i) Find the maximal elements of X. (ii) Find the minimal elements of X. (iii) Does X have a first element? (iv) Does X have a last element? (v) Find the set of upper bounds of A. (vi) Find the set of lower bounds of A. (vii) Does $\sup(A)$ exist? (viii) Does $\inf(A)$ exist?



38. Consider Q, the set of rational numbers, with the natural order, and its subset $A = \{x : x \in Q, x^3 < 3\}$. (i) Is A bounded above? (ii) Is A bounded below? (iii) Does $\sup(A)$ exist? (iv) Does $\inf(A)$ exist?

39. Let $\mathbf{N}$, the positive integers, be ordered by "x divides y", and let $A \subset \mathbf{N}$. (i) Does $\inf(A)$ exist? (ii) Does $\sup(A)$ exist?

40. Prove: Every finite partially ordered set has a maximal element.

41. Give an example of an ordered set which has exactly one maximal element but does not have a last element.

42. Prove: If R is a partial order on A, then R^{-1} is also a partial order on A.

ZORN'S LEMMA

43. Consider the proof of the following statement: There exists a finite set of positive integers which is not a proper subset of any other finite set of positive integers.

Proof. Let $\mathcal{A}$ be the class of all finite sets of positive integers. Partially order $\mathcal{A}$ by set inclusion. Now let $\mathcal{B} = \{B_i : i \in I\}$ be a totally ordered subclass of $\mathcal{A}$. Consider the set $A = \cup_i B_i$. Observe that $B_i \subset A$ for every $B_i \in \mathcal{B}$; hence A is an upper bound of $\mathcal{B}$.

Since every totally ordered subset of $\mathcal{A}$ has an upper bound, by Zorn's Lemma, $\mathcal{A}$ has a maximal element, a finite set which is not a proper subset of another finite set.

Question: Since the statement is clearly false, which step in the proof is incorrect?

44. Prove the following fact which was assumed in the proof in Problem 24: Let $\{f_i : A_i \to B\}$ be a class of functions which is totally ordered by set inclusion. Then $\cup_i f_i$ is a function from $\cup_i A_i$ into B.

45. Prove that the following two statements are equivalent:

(i) (Axiom of Choice.) The product $\prod\{A_i : i \in I\}$ of a non-empty class of non-empty sets is non-empty.

(ii) If $\mathcal{A}$ is a non-empty class of non-empty disjoint sets, then there exists a subset $B \subset \bigcup\{A : A \in \mathcal{A}\}$ such that the intersection of B and each set $A \in \mathcal{A}$ consists of exactly one element.

46. Prove: If every totally ordered subset of an ordered set X has a lower bound in X, then X has a minimal element.

Answers to Supplementary Problems

32. *c*

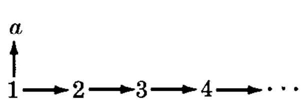
36. (i) $\succ$, (ii) $\succ$, (iii) $\prec$, (iv) $\prec$

37. (i) $\{1\}$; (ii) $\{5, 6\}$; (iii) No; (iv) Yes, 1; (v) $\{1, 2\}$; (vi) $\{5, 6\}$; (vii) Yes, 2; (viii) No

38. (i) Yes, (ii) No, (iii) No, (iv) No

39. (i) $\inf(A)$ exists iff $A \neq \emptyset$. (ii) $\sup(A)$ exists iff A is finite.

41.



Here a is maximal but a is not a last element.

Chapter 4

Topology of the Line and Plane

REAL LINE

The set of *real numbers*, denoted by **R**, plays a dominant role in mathematics and, in particular, in analysis. In fact, many concepts in topology are abstractions of properties of sets of real numbers. The set **R** can be characterized by the statement that **R** is a *complete, Archimedean ordered field.* These notions are explained in the Appendix. Here we use the order relation in **R** to define the "usual topology" for **R**.

We assume the reader is familiar with the geometric representation of **R** by means of the points on a straight line. As in Fig. 4-1, a point, called the origin, is chosen to represent 0 and another point, usually to the right of 0, to represent 1. Then there is a natural way to pair off the points on the line and the real numbers, i.e. each point will represent a unique real number and each real number will be represented by a unique point. For this reason we refer to the line as the *real line* or *real axis.* Furthermore, we will use the words point and number interchangeably.

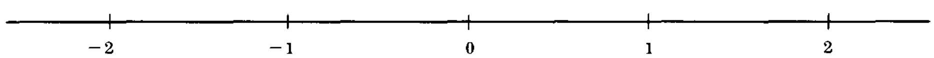


Fig. 4-1

OPEN SETS IN R

Let A be a set of real numbers. A point $p \in A$ is an *interior point* of A iff p belongs to some open interval S_p which is contained in A:

$$p \in S_p \subset A$$

The set A is *open* (or $\mathcal{U}$-open) iff each of its points is an interior point. (The significance of $\mathcal{U}$ in $\mathcal{U}$-open will appear in the next chapter.)

Example 1.1: An open interval $A = (a, b)$ is an open set, for we may choose $S_p = A$ for each $p \in A$.

Example 1.2: The real line **R**, itself, is open since any open interval S_p must be a subset of **R**, i.e. $p \in S_p \subset \mathbf{R}$.

Observe that a set is not open iff there exists a point in the set that is not an interior point.

Example 1.3: The closed interval $B = [a, b]$ is not an open set, for any open interval containing a or b must contain points outside of B. Hence the end points a and b are not interior points of B.

Example 1.4: The empty set $\emptyset$ is open since there is no point in $\emptyset$ which is not an interior point.

Example 1.5: The infinite open intervals, i.e. the subsets of **R** defined and denoted by

$$\{x : x \in \mathbf{R},\ x > a\} = (a, \infty), \quad \{x : x \in \mathbf{R},\ x < a\} = (-\infty, a),$$
$$\{x : x \in \mathbf{R}\} = \mathbf{R} = (-\infty, \infty)$$

are open sets. On the other hand, the infinite closed intervals, i.e. the subsets of **R** defined and denoted by

$$\{x : x \in \mathbf{R},\ x \geqq a\} = [a, \infty), \quad \{x : x \in \mathbf{R},\ x \leqq a\} = (-\infty, a]$$

are not open sets, since $a \in \mathbf{R}$ is not an interior point of either $[a, \infty)$ or $(-\infty, a]$.

We state two fundamental theorems about open sets.

Theorem 4.1: The union of any number of open sets in **R** is open.

Theorem 4.2: The intersection of any finite number of open sets in **R** is open.

The next example shows that the finiteness condition in the preceding theorem cannot be removed.

Example 1.6: Consider the class of open intervals and, hence, open sets

$$\{A_n = (-1/n, 1/n) : n \in \mathbf{N}\}, \quad \text{i.e.} \quad \{(-1,1),\ (-\tfrac{1}{2},\tfrac{1}{2}),\ (-\tfrac{1}{3},\tfrac{1}{3}),\ \ldots\}$$

Observe that the intersection

$$\cap_{n=1}^{\infty} A_n \;=\; \{0\}$$

of the open intervals consists of the single point 0 which is not an open set. In other words, an arbitrary intersection of open sets need not be open.

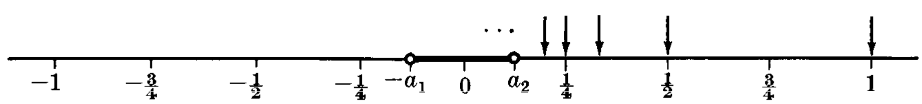
ACCUMULATION POINTS

Let A be a subset of **R**, i.e. a set of real numbers. A point $p \in \mathbf{R}$ is an *accumulation point* or *limit point* of A iff every open set G containing p contains a point of A different from p; i.e.,

$$G \text{ open},\ p \in G \quad \text{implies} \quad A \cap (G \setminus \{p\}) \neq \emptyset$$

The set of accumulation points of A, denoted by A', is called the *derived set* of A.

Example 2.1: Let $A = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$. The point 0 is an accumulation point of A since any open set G with $0 \in G$ contains an open interval $(-a_1, a_2) \subset G$ with $-a_1 < 0 < a_2$ which contains points of A.



Observe that the limit point 0 of A does not belong to A. Observe also that A does not contain any other limit points; hence the derived set of A is the singleton set $\{0\}$, i.e. $A' = \{0\}$.

Example 2.2: Consider the set **Q** of rational numbers. Every real number $p \in \mathbf{R}$ is a limit point of **Q** since every open set contains rational numbers, i.e. points of **Q**.

Example 2.3: The set of integers $\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ does not have any points of accumulation. In other words, the derived set of **Z** is the empty set $\emptyset$.

Remark: The reader should not confuse the concept "limit point of a set" with the different, though related, concept "limit of a sequence". Some of the solved and supplementary problems will show the relationship between these two concepts.

BOLZANO-WEIERSTRASS THEOREM

The existence or non-existence of accumulation points for various sets is an important question in topology. Not every set, even if it is infinite as in Example 2.3, has a limit point. There does exist, however, an important general case which gives a positive answer.

Theorem (Bolzano-Weierstrass) 4.3: Let A be a bounded, infinite set of real numbers. Then A has at least one accumulation point.

CLOSED SETS

A subset A of **R**, i.e. a set of real numbers, is a *closed set* iff its complement A^c is an open set. A closed set can also be described in terms of its accumulation points.

Theorem 4.4: A subset A of **R** is closed if and only if A contains each of its points of accumulation.

Example 3.1: The closed interval $[a, b]$ is a closed set since its complement $(-\infty, a) \cup (b, \infty)$, the union of two open infinite intervals, is open.

Example 3.2: The set $A = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$ is not closed since, as seen in Example 2.1, 0 is a limit point of A but does not belong to A.

Example 3.3: The empty set $\emptyset$ and the entire line $\mathbf{R}$ are closed sets since their complements $\mathbf{R}$ and $\emptyset$, respectively, are open sets.

Sets may be neither open nor closed as seen in the next example.

Example 3.4: Consider the open-closed interval $A = (a, b]$. Note that A is not open since $b \in A$ is not an interior point of A, and is not closed since $a \notin A$ but is a limit point of A.

HEINE-BOREL THEOREM

One of the most important properties of a closed and bounded interval is given in the next theorem. Here a class of sets, $\mathcal{A} = \{A_i\}$, is said to *cover* a set A if A is contained in the union of the members of $\mathcal{A}$, i.e. $A \subset \cup_i A_i$.

Theorem (Heine-Borel) 4.5: Let $A = [c, d]$ be a closed and bounded interval, and let $\mathcal{G} = \{G_i : i \in I\}$ be a class of open intervals which covers A, i.e. $A \subset \cup_i G_i$. Then $\mathcal{G}$ contains a finite subclass, say $\{G_{i_1}, \cdots, G_{i_m}\}$, which also covers A, i.e.,

$$A \subset G_{i_1} \cup G_{i_2} \cup \ldots \cup G_{i_m}$$

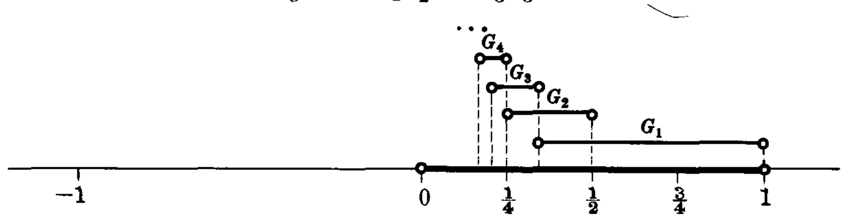
Both conditions, closed and bounded, must be satisfied by A or else the theorem is not true. We show this by the next two examples.

Example 4.1: Consider the open, bounded unit interval $A = (0, 1)$. Observe that the class

$$\mathcal{G} = \left\{G_n = \left(\frac{1}{n+2}, \frac{1}{n}\right) : n \in \mathbf{N}\right\}$$

of open intervals covers A, i.e.,

$$A \subset (\tfrac{1}{3}, 1) \cup (\tfrac{1}{4}, \tfrac{1}{2}) \cup (\tfrac{1}{5}, \tfrac{1}{3}) \cup \cdots$$

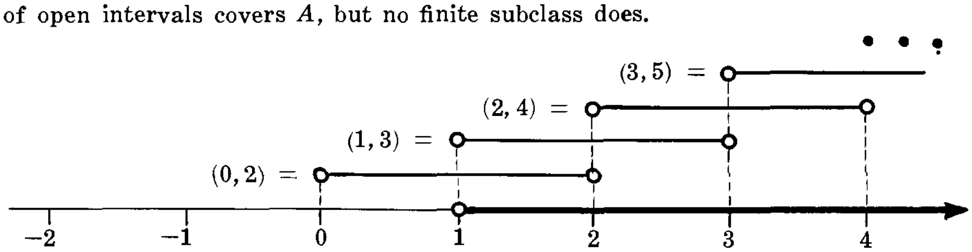


But the union of no finite subclass of $\mathcal{G}$ contains A.

Example 4.2: Consider the closed infinite interval $A = [1, \infty)$. The class

$$\mathcal{G} = \{(0, 2), (1, 3), (2, 4), \ldots\}$$

of open intervals covers A, but no finite subclass does.



SEQUENCES

A *sequence*, denoted by

$$\langle s_1, s_2, \ldots \rangle, \quad \langle s_n : n \in \mathbf{N} \rangle \quad \text{or} \quad \langle s_n \rangle$$

is a function whose domain is $\mathbf{N} = \{1, 2, 3, \ldots\}$, i.e. a sequence assigns a point s_n to each positive integer $n \in \mathbf{N}$. The image s_n or $s(n)$ of $n \in \mathbf{N}$ is called the nth *term* of the sequence.

Example 5.1: The sequences

$$\langle s_n\rangle = \langle 1, 3, 5, \ldots\rangle, \quad \langle t_n\rangle = \langle -\tfrac{1}{2}, \tfrac{1}{4}, -\tfrac{1}{8}, \tfrac{1}{16}, \ldots\rangle, \quad \langle u_n\rangle = \langle 1, 0, 1, 0, \ldots\rangle$$

can be defined, respectively, by the formulas

$$s(n) = 2n-1, \quad t(n) = (-1)^n/2^n, \quad u(n) = \tfrac{1}{2}(1+(-1)^{n+1}) = \begin{cases} 1 & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}$$

A sequence $\langle s_n : n \in \mathbf{N}\rangle$ is said to be bounded if its range $\{s_n : n \in \mathbf{N}\}$ is a bounded set.

Example 5.2: Consider the three sequences in Example 5.1. The range of $\langle s_n\rangle$ is $\{1, 3, 5, \ldots\}$; so $\langle s_n\rangle$ is not a bounded sequence. The range of $\langle t_n\rangle$ is $\{-\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}, \ldots\}$ which is bounded; hence $\langle t_n\rangle$ is a bounded sequence. The range of $\langle u_n\rangle$ is the finite set $\{0, 1\}$; so $\langle u_n\rangle$ is also a bounded sequence.

Observe that $\langle s_n : n \in \mathbf{N}\rangle$ denotes a sequence and is a function. On the other hand, $\{s_n : n \in \mathbf{N}\}$ denotes the range of the sequence and is a set.

CONVERGENT SEQUENCES

The usual definition of a convergent sequence is stated as follows:

Definition: The sequence $\langle a_1, a_2, \ldots\rangle$ of real numbers converges to $b \in R$ or, equivalently, the real number b is the limit of the sequence $\langle a_n : n \in \mathbf{N}\rangle$, denoted by

$$\lim_{n\to\infty} a_n = b, \quad \lim a_n = b \quad \text{or} \quad a_n \to b$$

if for every $\epsilon > 0$ there exists a positive integer n_0 such that

$$n > n_0 \quad \text{implies} \quad |a_n - b| < \epsilon$$

Observe that $|a_n - b| < \epsilon$ means that $b - \epsilon < a_n < b + \epsilon$, or, equivalently, that a_n belongs to the open interval $(b-\epsilon, b+\epsilon)$ containing b. Furthermore, since each term after the n_0th lies inside the interval $(b-\epsilon, b+\epsilon)$, only the terms before a_{n_0}, and there are only a finite number of them, can lie outside the interval $(b-\epsilon, b+\epsilon)$. Hence we can restate the preceding definition as follows.

Definition: The sequence $\langle a_n : n \in \mathbf{N}\rangle$ converges to b if every open set containing b contains *almost all*, i.e. all but a finite number, of the terms of the sequence.

Example 6.1: A constant sequence $\langle a_0, a_0, a_0, \ldots\rangle$, such as $\langle 1, 1, 1, \ldots\rangle$ or $\langle -3, -3, -3, \ldots\rangle$, converges to a_0 since each open set containing a_0 contains every term of the sequence.

Example 6.2: Each of the sequences

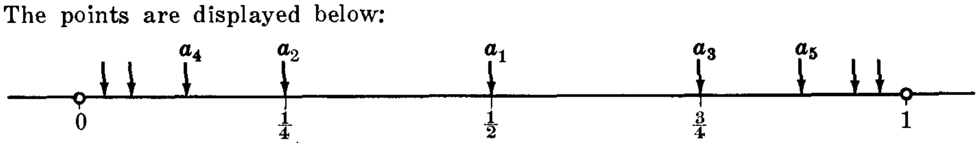
$$\langle 1, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{4}, \ldots\rangle, \quad \langle 1, 0, \tfrac{1}{2}, 0, \tfrac{1}{3}, 0, \tfrac{1}{4}, 0, \ldots\rangle, \quad \langle 1, -\tfrac{1}{2}, \tfrac{1}{3}, -\tfrac{1}{4}, \ldots\rangle$$

converges to 0 since any open interval containing 0 contains almost all of the terms of each of the sequences.

Example 6.3: Consider the sequence $\langle \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{7}{8}, \frac{1}{16}, \frac{15}{16}, \ldots\rangle$, i.e. the sequence

$$a_n = \begin{cases} \dfrac{1}{2^{(n+2)/2}} & \text{if } n \text{ is even} \\[2ex] 1 - \dfrac{1}{2^{(n+1)/2}} & \text{if } n \text{ is odd} \end{cases}$$

The points are displayed below:



Observe that any open interval containing either 0 or 1 contains an infinite number of the terms of the sequence. Neither 0 nor 1, however, is a limit of the sequence. Observe, though, that 0 and 1 are accumulation points of the *range* of the sequence, that is, of the *set* $\{\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{7}{8}, \ldots\}$.

SUBSEQUENCES

Consider a sequence $\langle a_1, a_2, a_3, \ldots\rangle$. If $\langle i_n\rangle$ is a sequence of positive integers such that $i_1 < i_2 < \cdots$, then

$$\langle a_{i_1}, a_{i_2}, a_{i_3}, \ldots\rangle$$

is called a *subsequence* of $\langle a_n : n \in \mathbf{N}\rangle$.

Example 7.1: Consider the sequence $\langle a_n\rangle = \langle 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\rangle$. Observe that $\langle 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\rangle$ is a subsequence of $\langle a_n\rangle$, but that $\langle \frac{1}{2}, 1, \frac{1}{4}, \frac{1}{3}, \frac{1}{6}, \frac{1}{5}, \ldots\rangle$ is not a subsequence of $\langle a_n\rangle$ since 1 appears before $\frac{1}{2}$ in the original sequence.

Example 7.2: Although the sequence $\langle \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{7}{8}, \ldots\rangle$ of Example 6.3 does not converge, it does have convergent subsequences such as $\langle \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\rangle$ and $\langle \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots\rangle$. On the other hand, the sequence $\langle 1, 3, 5, \ldots\rangle$ does not have any convergent subsequences.

As seen in the preceding example, sequences may or may not have convergent subsequences. There does exist a very important general case which gives a positive answer.

Theorem 4.6: Every bounded sequence of real numbers contains a convergent subsequence.

CAUCHY SEQUENCES

A sequence $\langle a_n : n \in \mathbf{N}\rangle$ of real numbers is a *Cauchy sequence* iff for every $\epsilon > 0$ there exists a positive integer n_0 such that

$$n, m > n_0 \quad \text{implies} \quad |a_n - a_m| < \epsilon$$

In other words, a sequence is a Cauchy sequence iff the terms of the sequence become arbitrarily close to each other as n gets large.

Example 8.1: Let $\langle a_n : n \in \mathbf{N}\rangle$ be a Cauchy sequence of integers, i.e. each term of the sequence belongs to $\mathbf{Z} = \{\ldots, -1, 0, 1, \ldots\}$. Then the sequence must be of the form

$$\langle a_1, a_2, \ldots, a_{n_0}, b, b, b, \ldots\rangle$$

i.e. the sequence is constant after some n_0th term. For if we choose $\epsilon = \frac{1}{2}$, then

$$a_n, a_m \in \mathbf{Z} \text{ and } |a_n - a_m| < \tfrac{1}{2} \quad \text{implies} \quad a_n = a_m$$

Example 8.2: We show that every convergent sequence is a Cauchy sequence. Let $a_n \to b$ and let $\epsilon > 0$. Then there exists $n_0 \in \mathbf{N}$ sufficiently large such that

$$n > n_0 \text{ implies } |a_n - b| < \tfrac{1}{2}\epsilon \quad \text{and} \quad m > n_0 \text{ implies } |a_m - b| < \tfrac{1}{2}\epsilon$$

Consequently, $n, m > n_0$ implies

$$|a_n - a_m| = |a_n - b + b - a_m| \leq |a_n - b| + |b - a_m| < \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon = \epsilon$$

Hence $\langle a_n\rangle$ is a Cauchy sequence.

COMPLETENESS

A set A of real numbers is said to be *complete* if every Cauchy sequence $\langle a_n \in A : n \in \mathbf{N}\rangle$ of points in A converges to a point in A.

Example 9.1: The set $\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ of integers is complete. For, as seen in Example 8.1, a Cauchy sequence $\langle a_n : n \in \mathbf{N}\rangle$ of points in $\mathbf{Z}$ is of the form

$$\langle a_1, a_2, \ldots, a_{n_0}, b, b, b, \ldots\rangle$$

which converges to the point $b \in \mathbf{Z}$.

Example 9.2: The set $\mathbf{Q}$ of rational numbers is not complete. For we can choose a sequence of rational numbers, such as $\langle 1, 1.4, 1.41, 1.412, \ldots\rangle$ which converges to the real number $\sqrt{2}$, which is not rational, i.e. which does not belong to $\mathbf{Q}$.

A fundamental property of the entire set **R** of real numbers is that **R** is complete. Namely,

Theorem (Cauchy) 4.7: Every Cauchy sequence of real numbers converges to a real number.

CONTINUOUS FUNCTIONS

The usual $\epsilon - \delta$ definition of a continuous function is stated as follows:

Definition: A function $f: \mathbf{R} \to \mathbf{R}$ is *continuous* at a point x_0 if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|x - x_0| < \delta \quad \text{implies} \quad |f(x) - f(x_0)| < \epsilon$$

The function f is a *continuous function* if it is continuous at every point.

Observe that $|x - x_0| < \delta$ means that $x_0 - \delta < x < x_0 + \delta$, or equivalently that x belongs to the open interval $(x_0 - \delta,\ x_0 + \delta)$. Similarly, $|f(x) - f(x_0)| < \epsilon$ means that $f(x)$ belongs to the open interval $(f(x_0) - \epsilon,\ f(x_0) + \epsilon)$. Accordingly, the statement

$$|x - x_0| < \delta \quad \text{implies} \quad |f(x) - f(x_0)| < \epsilon$$

is equivalent to the statement

$$x \in (x_0 - \delta,\ x_0 + \delta) \quad \text{implies} \quad f(x) \in (f(x_0) - \epsilon,\ f(x_0) + \epsilon)$$

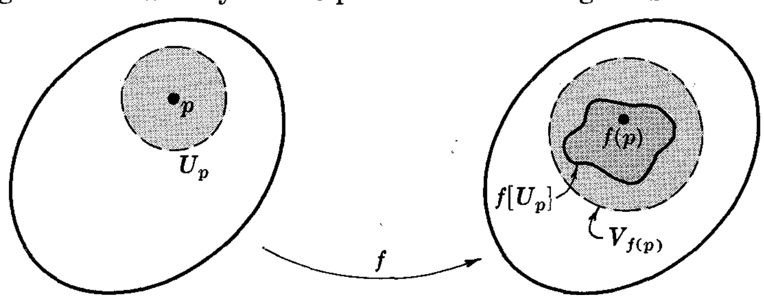
which is equivalent to the statement

$$f[(x_0 - \delta,\ x_0 + \delta)] \quad \text{is contained in} \quad (f(x_0) - \epsilon,\ f(x_0) + \epsilon)$$

Hence we can restate the previous definition as follows.

Definition: A function $f: \mathbf{R} \to \mathbf{R}$ is continuous at a point $p \in \mathbf{R}$ if for any open set $V_{f(p)}$ containing $f(p)$ there exists an open set U_p containing p such that $f[U_p] \subset V_{f(p)}$. The function f is a continuous function if it is continuous at every point.

The Venn diagram below may be helpful in visualizing this definition.



A continuous function can be completely characterized in terms of open sets as follows:

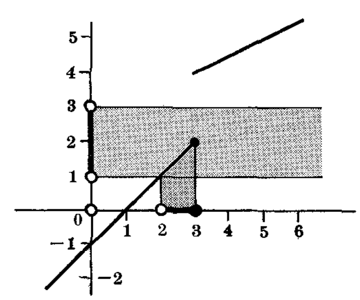
Theorem 4.8: A function is continuous if and only if the inverse image of every open set is open.

Observe that Theorem 4.8 also states that a function is not continuous iff there exists an open set whose inverse image is not open.

Example 10.1: Consider the function $f: \mathbf{R} \to \mathbf{R}$ defined by

$$f(x) = \begin{cases} x - 1 & \text{if } x \leqq 3 \\ \frac{1}{2}(x + 5) & \text{if } x > 3 \end{cases}$$

and illustrated in the adjacent diagram. Note that the inverse of the open interval $(1, 3)$ is the open-closed interval $(2, 3]$ which is not an open set. Hence the function f is not continuous.



We now state one important property of continuous functions which we will refer to later in the text.

Theorem 4.9: Let $f: \mathbf{R} \to \mathbf{R}$ be continuous on a closed interval $[a,b]$. Then the function assumes every value between $f(a)$ and $f(b)$.

In other words, if y_0 is a real number for which $f(a) \leq y_0 \leq f(b)$ or $f(b) \leq y_0 \leq f(a)$, then

$$\exists x_0 \in \mathbf{R} \quad \text{such that} \quad a \leq x_0 \leq b \text{ and } f(x_0) = y_0$$

This theorem is known as the *Weierstrass Intermediate Value Theorem.*

Remark: A function $f: \mathbf{R} \to \mathbf{R}$ is said to be continuous on a subset D of $\mathbf{R}$ if it is continuous at each point in D.

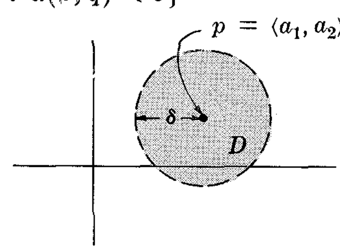
TOPOLOGY OF THE PLANE

An *open disc* D in the plane $\mathbf{R}^2$ is the set of points inside a circle, say, with center $p = \langle a_1, a_2 \rangle$ and radius $\delta > 0$, i.e.,

$$D = \{\langle x,y \rangle : (x-a_1)^2 + (y-a_2)^2 < \delta^2\} = \{q \in \mathbf{R}^2 : d(p,q) < \delta\}$$

Here $d(p,q)$ denotes the usual distance between two points $p = \langle a_1, a_2 \rangle$ and $q = \langle b_1, b_2 \rangle$ in $\mathbf{R}^2$:

$$d(p,q) = \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2}$$



The open disc plays a role in the topology of the plane $\mathbf{R}^2$ that is analogous to the role of the open interval in the topology of the line $\mathbf{R}$.

Let A be a subset of $\mathbf{R}^2$. A point $p \in A$ is an *interior point* of A iff p belongs to some open disc D_p which is contained in A:

$$p \in D_p \subset A$$

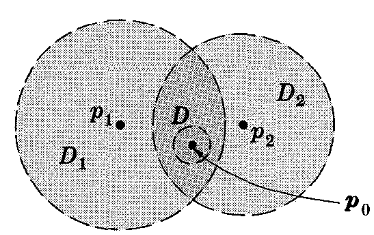
The set A is *open* (or *U-open*) iff each of its points is an interior point.

Example 11.1: Clearly an open disc, the entire plane $\mathbf{R}^2$ and the empty set $\emptyset$ are open subsets of $\mathbf{R}^2$. We now show that the intersection of any two open discs, say

$$D_1 = \{q \in \mathbf{R}^2 : d(p_1,q) < \delta_1\} \quad \text{and} \quad D_2 = \{q \in \mathbf{R}^2 : d(p_2,q) < \delta_2\}$$

is also an open set. For let $p_0 \in D_1 \cap D_2$ so

$$d(p_1,p_0) < \delta_1 \quad \text{and} \quad d(p_2,p_0) < \delta_2$$



Set $$r = \min\{\delta_1 - d(p_1,p_0),\ \delta_2 - d(p_2,p_0)\} > 0$$

and let $$D = \{q \in \mathbf{R}^2 : d(p_0,q) < \tfrac{1}{2}r\}$$

Then $p_0 \in D \subset D_1 \cap D_2$ or, p_0 is an interior point of $D_1 \cap D_2$.

A point $p \in \mathbf{R}^2$ is an *accumulation point* or *limit point* of a subset A of $\mathbf{R}^2$ iff every open set G containing p contains a point of A different from p, i.e.,

$$G \subset \mathbf{R}^2 \text{ open}, \quad p \in G \quad \text{implies} \quad A \cap (G \setminus \{p\}) \neq \emptyset$$

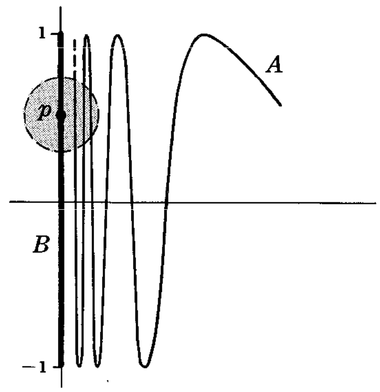
Example 11.2: Consider the following subset of $\mathbf{R}^2$:

$$A = \left\{\langle x, y\rangle : y = \sin\frac{1}{x},\ x > 0\right\}$$

The set A is illustrated in the adjacent diagram. Observe that the curve, going from right to left, fluctuates faster and faster, i.e. that the points where the curve crosses the x-axis become closer and closer. The point $p = \langle 0, \frac{1}{2}\rangle$ is a limit point of A since A will eventually pass through any open disc containing p. In fact, each point on the y-axis between -1 and 1, i.e. each point in the set

$$B = \{\langle x, y\rangle : x = 0,\ -1 \leqq y \leqq 1\}$$

is a limit point of A.



A subset A of $\mathbf{R}^2$ is *closed* iff its complement A^c is an open subset of $\mathbf{R}^2$.

A sequence $\langle p_1, p_2, \ldots\rangle$ of points in $\mathbf{R}^2$ *converges* to the point $q \in \mathbf{R}^2$ iff every open set containing q contains almost all of the terms of the sequence. Convergence in the plane $\mathbf{R}^2$ can be characterized in terms of convergence in $\mathbf{R}$ as follows.

Proposition 4.10: Consider the sequence $\langle p_1 = \langle a_1, b_1\rangle,\ p_2 = \langle a_2, b_2\rangle,\ \ldots\rangle$ of point in $\mathbf{R}^2$ and the point $q = \langle a, b\rangle \in \mathbf{R}^2$. Then

$$p_n \to q \quad \text{if and only if} \quad a_n \to a \text{ and } b_n \to b$$

A function $f: \mathbf{R}^2 \to \mathbf{R}^2$ is *continuous* at a point $p \in \mathbf{R}^2$ iff for any open set $V_{f(p)}$ containing $f(p)$ there exists an open set U_p containing p such that $f[U_p] \subset V_{f(p)}$.

We list theorems for the plane $\mathbf{R}^2$ which are analagous to theorems for the line $\mathbf{R}$ stated earlier in this chapter.

Theorem 4.1*: The union of any number of open subsets of $\mathbf{R}^2$ is open.

Theorem 4.2*: The intersection of any finite number of open subsets of $\mathbf{R}^2$ is open.

Theorem 4.4*: A subset A of $\mathbf{R}^2$ is closed if and only if A contains each of its accumulation points.

Theorem 4.8*: A function $f: \mathbf{R}^2 \to \mathbf{R}^2$ is continuous if and only if the inverse image of every open set is open.

Solved Problems

OPEN SETS, ACCUMULATION POINTS

1. Determine the accumulation points of each set of real numbers:

   (i) $\mathbf{N}$; (ii) $(a, b]$; (iii) $\mathbf{Q}^c$, the set of irrational points.

   Solution:

   (i) $\mathbf{N}$, the set of positive integers, does not have any limit points. For if a is any real number, we can find a $\delta > 0$ so small that the open set $(a - \delta,\ a + \delta)$ contains no point of $\mathbf{N}$ other than a.

   (ii) Every point p in the closed interval $[a, b]$ is a limit point of the open-closed interval $(a, b]$, since every open interval containing $p \in [a, b]$ will contain points of $(a, b]$ other than p.

   (iii) Every real number $p \in \mathbf{R}$ is a limit point of $\mathbf{Q}^c$ since every open interval containing $p \in \mathbf{R}$ will contain points of $\mathbf{Q}^c$, i.e. irrational numbers, other than p.

2. Recall that A' denotes the derived set, i.e. set of limit points, of a set A. Find sets A such that (i) A and A' are disjoint, (ii) A is a proper subset of A', (iii) A' is a proper subset of A, (iv) $A = A'$.

Solution:

(i) The set $A = \{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$ has 0 as its only point of accumulation. Hence $A' = \{0\}$ and A and A' are disjoint.

(ii) Let $A = (a, b]$, an open-closed interval. As seen in the preceding problem $A' = [a, b]$, the closed interval, and so $A \subset A'$.

(iii) Let $A = \{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\}$. Then 0, which belongs to A, is the only limit point of A. Hence $A' = \{0\}$ and $A' \subset A$.

(iv) Let $A = [a, b]$, a closed interval. Then each point in A is a limit point of A and they are the only limit points. So $A = A' = [a, b]$.

3. Prove Theorem 4.1*: The union of any number of open subsets of $\mathbf{R}^2$ is open.

Solution:

Let $\mathcal{A}$ be a class of open subsets of $\mathbf{R}^2$, let $H = \bigcup\{G : G \in \mathcal{A}\}$, and let $p \in H$. The theorem is proved if we show that p is an interior point of H, i.e. there exists an open disc D_p containing p such that D_p is contained in H.

Since $p \in H = \bigcup\{G : G \in \mathcal{A}\}$,

$$\exists G_0 \in \mathcal{A} \quad \text{such that} \quad p \in G_0$$

But G_0 is an open set; hence there exists an open disc D_p containing p such that

$$p \in D_p \subset G_0$$

Since G_0 is a subset of $H = \bigcup\{G : G \in \mathcal{A}\}$, D_p is also a subset of H. Thus H is open.

4. Prove: Every open subset G of the plane $\mathbf{R}^2$ is the union of open discs.

Solution:

Since G is open, for each point $p \in G$ there is an open disc D_p such that $p \in D_p \subset G$. Then $G = \bigcup\{D_p : p \in G\}$.

5. Prove Theorem 4.2*: The intersection of any finite number of open subsets of $\mathbf{R}^2$ is open.

Solution:

We prove the theorem in the case of two open subsets of $\mathbf{R}^2$. The theorem will then follow by induction.

Let G and H be open subsets of $\mathbf{R}^2$ and let $p \in G \cap H$; so $p \in G$ and $p \in H$. Hence there exist open discs D_1 and D_2 such that

$$p \in D_1 \subset G \quad \text{and} \quad p \in D_2 \subset H$$

Then $p \in D_1 \cap D_2 \subset G \cap H$. By Example 11.1, the intersection of any two open discs is open; so there exists an open disc D such that

$$p \in D \subset D_1 \cap D_2 \subset G \cap H$$

Hence p is an interior point of $G \cap H$ and, so, $G \cap H$ is open.

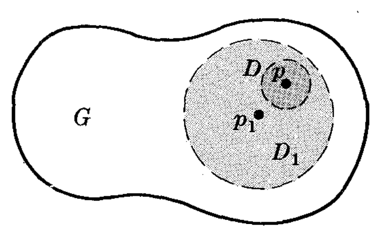
6. Prove: Let $p \in G$, an open subset of $\mathbf{R}^2$. Then there exists an open disc D with center p such that $p \in D \subset G$.

Solution:

By definition of an interior point, there exists an open disc $D_1 = \{q \in \mathbf{R}^2 : d(p_1, q) < \delta\}$, with center p_1 and radius δ, such that $p \in D_1 \subset G$. So $d(p_1, p) < \delta$. Set

$$r = \delta - d(p_1, p) > 0$$

and let $$D = \{q \in \mathbf{R}^2 : d(p, q) < \tfrac{1}{2}r\}$$



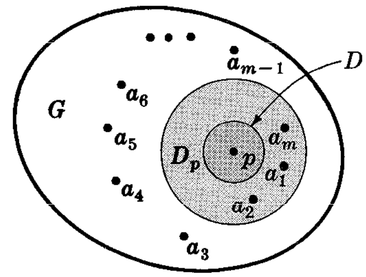
Then, as indicated in the diagram, $p \in D \subset D_1 \subset G$ where D is an open disc with center p.

7. Prove: Let p be an accumulation point of a subset A of the plane $\mathbf{R}^2$. Then every open set containing p contains an infinite number of points of A.

Solution:

Suppose G is an open set containing p and containing only a finite number of points, say $a_1, \ldots, a_m$, of A different from p. By the preceding problem, there exists an open disc D_p with center p and, say, radius δ such that $p \in D_p \subset G$. Choose $r > 0$ to be less than δ and less than the distance from p to any of the points $a_1, \ldots, a_m$; and let

$$D = \{q \in \mathbf{R}^2 : d(p,q) < \tfrac{1}{2}r\}$$



Then the open disc D containing p does not contain $a_1, \ldots, a_m$ and, since $D \subset D_p \subset G$, does not contain any other points of A different from p.

The last statement contradicts the fact that p is a limit point of A. Hence every open set containing p contains an infinite number of points of A.

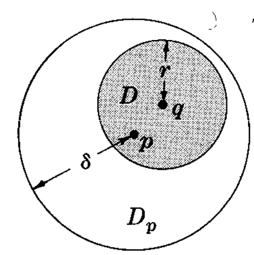
Remark: A similar statement is true for the real line $\mathbf{R}$, i.e. if $a \in \mathbf{R}$ is a limit point of $A \subset \mathbf{R}$, then every open subset of $\mathbf{R}$ containing a contains an infinite number of points of A.

8. Prove: Consider any open disc D_p with center $p \in \mathbf{R}^2$ and radius δ. Then there exists an open disc D such that (i) the center of D has rational coordinates, (ii) the radius of D is rational, and (iii) $p \in D \subset D_p$.

Solution:

Suppose $p = \langle a, b\rangle$. Then there exist rational numbers c and d such that

$$a < c < a + \tfrac{1}{6}\delta \quad \text{and} \quad b < d < b + \tfrac{1}{6}\delta$$



Let $q = \langle c, d\rangle$. Note that $d(p,q) < \frac{1}{3}\delta$. Now choose a rational number r such that $\frac{1}{3}\delta < r < \frac{2}{3}\delta$; and let D be the open disc with center q, which has rational coordinates, and radius r which is rational. Then, as indicated in the diagram, $p \in D \subset D_p$.

9. Prove: Every open subset G of the plane $\mathbf{R}^2$ is the union of a countable number of open discs.

Solution:

Since G is open, for each point $p \in G$ there exists an open disc D_p with center p such that $p \in D_p \subset G$. But, by the preceding problem, for each disc D_p there exists an open disc E_p such that (i) the center of E_p has rational coordinates, (ii) the radius of E_p is rational, and (iii) $p \in E_p \subset D_p$. So

$$p \in E_p \subset D_p \subset G$$

Accordingly,
$$G = \bigcup\{E_p : p \in G\}$$

The theorem now follows from the fact that there are only a countable number of open discs whose center has rational coordinates and whose radius is rational.

10. Prove Theorem (Bolzano-Weierstrass) 4.3: Let A be a bounded infinite set of real numbers. Then A contains at least one accumulation point.

Solution:

Since A is bounded, A is a subset of a closed interval $I_1 = [a_1, b_1]$. Bisect I_1 at $\frac{1}{2}(a_1 + b_1)$. Note that both of the closed subintervals of I_1,

$$[a_1, \tfrac{1}{2}(a_1 + b_1)] \quad \text{and} \quad [\tfrac{1}{2}(a_1 + b_1), b_1] \tag{1}$$

cannot contain a finite number of points of A since A is infinite. Let $I_2 = [a_2, b_2]$ be one of the intervals in (*1*) which contains an infinite number of points of A.

Now bisect I_2. As before, one of the two closed intervals

$$[a_2, \tfrac{1}{2}(a_2+b_2)] \quad \text{and} \quad [\tfrac{1}{2}(a_2+b_2), b_2]$$

must contain an infinite number of points of A. Call that interval I_3.

Continuing this procedure we obtain a sequence of nested closed intervals

$$I_1 \supset I_2 \supset I_3 \supset \cdots$$

such that each interval I_n contains an infinite number of points of A and

$$\lim |I_n| = 0$$

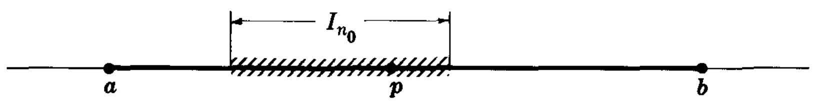
where $|I_n|$ denotes the length of the interval I_n.

By the Nested Interval Property of the real numbers (see Appendix A), there exists a point p in each interval I_n. We show that p is a limit point of A and then the theorem will follow.

Let $S_p = (a, b)$ be an open interval containing p. Since $\lim |I_n| = 0$,

$$\exists n_0 \in \mathbf{N} \quad \text{such that} \quad |I_{n_0}| < \min(p-a, b-p)$$

Then the interval I_{n_0} is a subset of the open interval $S_p = (a, b)$ as indicated in the diagram below.



Since I_{n_0} contains an infinite number of points of A, so does the open interval S_p. Thus each open interval containing p contains points of A other than p, i.e. p is a limit point of A.

CLOSED SETS

11. Prove: A set F is closed if and only if its complement F^c is open.

Solution:

Note that $(F^c)^c = F$; so F is the complement of F^c. Thus, by definition, F is closed iff F^c is open.

12. Prove: The union of a finite number of closed sets is closed.

Solution:

Let $F_1, \ldots, F_m$ be closed sets and let $F = F_1 \cup \cdots \cup F_m$. By DeMorgan's Law,

$$F^c = (F_1 \cup \cdots \cup F_m)^c = F_1^c \cap F_2^c \cap \cdots \cap F_m^c$$

So F^c is the intersection of a finite number of open sets F_i^c, and thus F^c is also open. Hence its complement $F^{cc} = F$ is closed.

13. Prove: The intersection of any number of closed sets is closed.

Solution:

Let $\{F_i\}$ be a class of closed sets and let $F = \cap_i F_i$. By DeMorgan's Law,

$$F^c = (\cap_i F_i)^c = \cup_i F_i^c$$

So F^c is the union of open sets and, hence, is open itself. Consequently, $F^{cc} = F$ is closed.

14. Prove Theorem 4.4*: A subset of $\mathbf{R}^2$ is closed if and only if it contains each of its accumulation points.

Solution:

Suppose p is a limit point of a closed set F. Then every open disc containing p contains points of F other than p. Hence there cannot be an open disc D_p containing p which is completely contained in the complement of F. In other words, p is not an interior point of F^c. But F^c is open since F is closed; so p does not belong to F^c, i.e. $p \in F$.

On the other hand, suppose a set A contains each of its limit points. We claim that A is closed or, equivalently, that its complement A^c is open. Let $p \in A^c$. Since A contains each of its limit points, p is not a limit point of A. Hence there exists at least one open disc D_p containing p such that D_p does not contain any points of A. So $D_p \subset A^c$, and hence p is an interior point of A^c. Since each point $p \in A^c$ is an interior point, A^c is open and so A is closed.

15. Prove: The derived set A', i.e. set of accumulation points, of an arbitrary subset A of $\mathbf{R}^2$ is closed.

Solution:

Let p be a limit point of A'. By Theorem 4.4*, the theorem is proved if we show that $p \in A'$, that is, that p is also a limit point of A.

Let G_p be an open set containing p. Since p is a limit point of A', G_p contains at least one point $q \in A'$ different from p. But G_p is an open set containing $q \in A'$; hence G_p contains (infinitely many) points of A. So,

$$\exists a \in A \quad \text{such that} \quad a \neq p, \;\; a \neq q, \;\; \text{and} \;\; a \in G_p$$

That is, each open set containing p contains points of A other than p; so $p \in A'$.

16. Prove: Let A be a closed and bounded set of real numbers and let $\sup(A) = p$. Then $p \in A$.

Solution:

Suppose $p \notin A$. Let G be an open set containing p. Then G contains an open interval (b, c) containing p, i.e. such that $b < p < c$. Since $\sup(A) = p$ and $p \notin A$,

$$\exists a \in A \quad \text{such that} \quad b < a < p < c$$

for otherwise b would be an upper bound for A. So $a \in (b, c) \subset G$. Thus each open set containing p contains a point of A different from p; hence p is a limit point of A. But A is closed; hence, by Theorem 4.4*, $p \in A$.

17. Prove Theorem (Heine-Borel) 4.5:

Let $I_1 = [c_1, d_1]$ be covered by a class $\mathcal{G} = \{(a_i, b_i) : i \in I\}$ of open intervals. Then $\mathcal{G}$ contains a finite subclass which also covers I_1.

Solution:

Assume that no finite subclass of $\mathcal{G}$ covers I_1. We bisect $I_1 = [c_1, d_1]$ at $\frac{1}{2}(c_1 + d_1)$ and consider the two closed intervals

$$[c_1, \tfrac{1}{2}(c_1 + d_1)] \quad \text{and} \quad [\tfrac{1}{2}(c_1 + d_1), d_1] \tag{1}$$

At least one of these two intervals cannot be covered by a finite subclass of $\mathcal{G}$ or else the whole interval I_1 will be covered by a finite subclass of $\mathcal{G}$. Let $I_2 = [c_2, d_2]$ be one of the two intervals in (*1*) which cannot be covered by a finite subclass of $\mathcal{G}$. We now bisect I_2. As before, one of the two closed intervals

$$[c_2, \tfrac{1}{2}(c_2 + d_2)] \quad \text{and} \quad [\tfrac{1}{2}(c_2 + d_2), d_2]$$

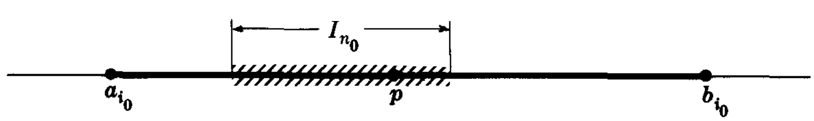
cannot be covered by a finite subclass of $\mathcal{G}$. Call that interval I_3.

We continue this procedure and obtain a sequence of nested closed intervals $I_1 \supset I_2 \supset I_3 \supset \cdots$ such that each interval I_n cannot be covered by a finite subclass of $\mathcal{G}$ and $\lim |I_n| = 0$ where $|I_n|$ denotes the length of the interval I_n.

By the Nested Interval Property of the real numbers (see Appendix), there exists a point p in each interval I_n. In particular, $p \in I_1$. Since $\mathcal{G}$ is a cover of I_1, there exists an open interval (a_{i_0}, b_{i_0}) in $\mathcal{G}$ which contains p. Hence $a_{i_0} < p < b_{i_0}$. Since $\lim |I_n| = 0$,

$$\exists n_0 \in N \quad \text{such that} \quad |I_{n_0}| < \min(p - a_{i_0}, b_{i_0} - p)$$

Then, as indicated in the diagram below, the interval I_{n_0} is a subset of the one interval (a_{i_0}, b_{i_0}) in $\mathcal{G}$.



But this contradicts our choice of I_{n_0}. Thus the original assumption that no finite subclass of $\mathcal{G}$ covers I_1 is false and the theorem is true.

SEQUENCES

18. Write the first six terms of each of the following sequences:

(i) $s(n) = \begin{cases} n-1 & \text{if } n \text{ is odd} \\ n^2 & \text{if } n \text{ is even} \end{cases}$ (ii) $t(n) = \begin{cases} 1 & \text{if } n = 1 \\ 2 & \text{if } n = 2 \\ t(n-1) + t(n-2) & \text{if } n > 2 \end{cases}$

Solution:

(i) Two formulas are used to define this function. Substitute 1, 3 and 5 into $s(n) = n-1$ to get $s_1 = 0$, $s_3 = 2$ and $s_5 = 4$. Then substitute 2, 4 and 6 into $s(n) = n^2$ to get $s_2 = 4$, $s_4 = 16$ and $s_6 = 36$. Thus we have $\langle 0, 4, 2, 16, 4, 36, \ldots \rangle$.

(ii) Here the function is defined recursively. Each term after the second is found by adding the two previous terms. Thus:

$$t_1 = 1 \qquad t_4 = t_3 + t_2 = 3 + 2 = 5$$
$$t_2 = 2 \qquad t_5 = t_4 + t_3 = 5 + 3 = 8$$
$$t_3 = t_2 + t_1 = 2 + 1 = 3 \qquad t_6 = t_5 + t_4 = 8 + 5 = 13$$

Hence we have $\langle 1, 2, 3, 5, 8, 13, \ldots \rangle$.

19. Consider the sequence $\langle a_n = (-1)^{n-1}(2n-1) \rangle$:

$$\langle 1, -3, 5, -7, 9, -11, 13, -15, \ldots \rangle$$

Determine whether or not each of the following sequences is a subsequence of $\langle a_n \rangle$.

(i) $\langle b_n \rangle = \langle 1, 5, -3, -7, 9, 13, -11, -15, \ldots \rangle$

(ii) $\langle c_n \rangle = \langle 1, 3, 5, 7, 9, 11, 13, \ldots \rangle$

(iii) $\langle d_n \rangle = \langle -3, -7, -11, -15, -19, -23, \ldots \rangle$

Solution:

(i) Note that 5 appears before -3 in $\langle b_n \rangle$, but -3 appears before 5 in $\langle a_n \rangle$. Hence $\langle b_n \rangle$ is not a subsequence of $\langle a_n \rangle$.

(ii) The terms 3, 7 and 11 do not even appear in $\langle a_n \rangle$; hence $\langle c_n \rangle$ is not a subsequence of $\langle a_n \rangle$.

(iii) The sequence $\langle d_n \rangle$ is a subsequence of $\langle a_n \rangle$, for $\langle i_n = 2n \rangle = \langle 2, 4, 6, \ldots \rangle$ is a sequence of positive integers such that $i_1 < i_2 < i_3 < \cdots$; so

$$\langle a_{i_1}, a_{i_2}, \ldots \rangle = \langle a_2, a_4, a_6, \ldots \rangle = \langle -3, -7, -11, \ldots \rangle$$

is a subsequence of $\langle a_n \rangle$.

20. Determine the range of each sequence:

(i) $\langle 1, \frac{1}{2}, 1, \frac{1}{3}, 1, \frac{1}{4}, 1, \frac{1}{5}, \ldots \rangle$ (iii) $\langle 2, 4, 6, 8, 10, \ldots \rangle$

(ii) $\langle 1, 0, -1, 0, 1, 0, -1, 0, 1, 0, -1, 0, \ldots \rangle$

Solution:

The range of a sequence is the set of image points. Hence the ranges of the sequences are

(i) $\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$, (ii) $\{1, 0, -1\}$, (iii) $\{2, 4, 6, 8, \ldots\}$

21. Prove: If the range of a sequence $\langle a_n \rangle$ is finite, then the sequence has a convergent subsequence.

Solution:

If the range $\{a_n\}$ of $\langle a_n \rangle$ is finite, then one of the image points, say b, appears an infinite number of times in the sequence. Hence $\langle b, b, b, b, \ldots \rangle$ is a subsequence of $\langle a_n \rangle$ and it converges.

22. Prove: If $\lim a_n = b$ and $\lim a_n = c$, then $b = c$.

Solution:

Suppose that b and c are distinct. Let $\delta = |b-c| > 0$. Then the open intervals $B = (b - \frac{1}{2}\delta, b + \frac{1}{2}\delta)$ and $C = (c - \frac{1}{2}\delta, c + \frac{1}{2}\delta)$, containing b and c respectively, are disjoint. Since $\langle a_n \rangle$ converges to b, B must contain all except a finite number of the terms of the sequence. Hence C can only contain a finite number of the terms of the sequence. But this contradicts the fact that $\langle a_n \rangle$ converges to c. Accordingly, b and c are not distinct.

23. Prove: If the range $\{a_n\}$ of a sequence $\langle a_n \rangle$ contains an accumulation point b, then the sequence $\langle a_n \rangle$ contains a subsequence $\langle a_{i_n} \rangle$ which converges to b.

Solution:

Since b is a limit point of $\{a_n\}$, each of the open intervals

$$S_1 = (b-1, b+1), \quad S_2 = (b - \tfrac{1}{2}, b + \tfrac{1}{2}), \quad S_3 = (b - \tfrac{1}{3}, b + \tfrac{1}{3}), \quad \ldots$$

contains an infinite number of elements of the set $\{a_n\}$ and, hence, an infinite number of the terms of the sequence $\langle a_n \rangle$. We choose a sequence $\langle a_{i_n} \rangle$ as follows:

Choose a_{i_1} to be a point in S_1.

Choose a_{i_2} to be a point in S_2 such that $i_2 > i_1$, i.e. such that a_{i_2} appears after a_{i_1} in the sequence $\langle a_n \rangle$.

Choose a_{i_3} to be a point in S_3 such that $i_3 > i_2$.

We continue in the same manner.

Observe that we are always able to choose the next term in the sequence $\langle a_{i_n} \rangle$ since there are an infinite number of the terms of the original sequence $\langle a_n \rangle$ in each interval S_n.

We claim that $\langle a_{i_n} \rangle$ satisfies the conditions of the theorem. Recall that we choose the terms of the sequence $\langle a_{i_n} \rangle$ so that $i_1 < i_2 < i_3 < \cdots$; hence $\langle a_{i_n} \rangle$ is a subsequence of $\langle a_n \rangle$. We need to show that $\lim a_{i_n} = b$. Let G be an open set containing b. Then G contains an open interval (d_1, d_2) containing b; so $d_1 < b < d_2$. Let $\delta = \min(b - d_1, d_2 - b) > 0$; then

$$\exists n_0 \in N \quad \text{such that} \quad 1/n_0 < \delta$$

Hence $S_{n_0} \subset (d_1, d_2) \subset G$, and so

$$n > n_0 \quad \text{implies} \quad a_{i_n} \in S_n \subset S_{n_0} \subset (d_1, d_2) \subset G$$

Thus G contains almost all the terms of the sequence $\langle a_{i_n} \rangle$; that is, $\lim a_{i_n} = b$.

24. Prove Theorem 4.6: Every bounded sequence $\langle a_n \rangle$ of real numbers contains a convergent subsequence.

Solution:

Consider the range $\{a_n\}$ of the sequence $\langle a_n \rangle$. If the range is finite, then by Problem 21 the sequence contains a convergent subsequence. On the other hand, if the range is infinite, then, by the Bolzano-Weierstrass Theorem, the bounded infinite set $\{a_n\}$ contains a limit point. But then, by the previous problem, the sequence in this case also contains a convergent subsequence.

25. Prove: Every Cauchy sequence $\langle a_n \rangle$ of real numbers is bounded.

Solution:

Let $\epsilon = 1$. Then, by definition of a Cauchy sequence,

$$\exists n_0 \in \mathbf{N} \quad \text{such that} \quad n, m \geqq n_0 \quad \text{implies} \quad |a_n - a_m| < 1$$

In particular, $\quad m \geqq n_0 \quad$ implies $\quad |a_{n_0} - a_m| < 1$, or, $a_{n_0} - 1 < a_m < a_{n_0} + 1$

Let

$$\alpha = \max(a_1, a_2, \ldots, a_{n_0}, a_{n_0} + 1)$$

$$\beta = \min(a_1, a_2, \ldots, a_{n_0}, a_{n_0} - 1)$$

Then α is an upper bound for the range $\{a_n\}$ of the sequence $\langle a_n \rangle$ and β is a lower bound. Accordingly, $\langle a_n \rangle$ is a bounded sequence.

26. Prove: Let $\langle a_n \rangle$ be a Cauchy sequence. If a subsequence $\langle a_{i_n} \rangle$ of $\langle a_n \rangle$ converges to a point b, then the Cauchy sequence itself converges to b.

Solution:

Let $\epsilon > 0$. We need to find a positive integer n_0 such that

$$n > n_0 \quad \text{implies} \quad |a_n - b| < \epsilon$$

Since $\langle a_n \rangle$ is a Cauchy sequence,

$$\exists n_0 \in \mathbf{N} \quad \text{such that} \quad n, m > n_0 \quad \text{implies} \quad |a_n - a_m| < \tfrac{1}{2}\epsilon$$

Also, since the subsequence $\langle a_{i_n} \rangle$ converges to b,

$$\exists i_m \in \mathbf{N} \quad \text{such that} \quad |a_{i_m} - b| < \tfrac{1}{2}\epsilon$$

Observe that we can choose i_m so that $i_m > n_0$. Accordingly,

$$\begin{aligned} n > n_0 \quad \text{implies} \quad |a_n - b| &= |a_n - a_{i_m} + a_{i_m} - b| \\ &\le |a_n - a_{i_m}| + |a_{i_m} - b| \\ &< \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon = \epsilon \end{aligned}$$

Hence $\langle a_n \rangle$ converges to b.

Observe that we need $i_m > n_0$ in order to state that: $n > n_0$ implies $|a_n - a_{i_m}| < \tfrac{1}{2}\epsilon$.

27. Prove Theorem (Cauchy) 4.7: Every Cauchy sequence $\langle a_n \rangle$ of real numbers converges to a real number.

Solution:

By Problem 25, the Cauchy sequence $\langle a_n \rangle$ is bounded. Hence, by Theorem 4.6, the bounded sequence $\langle a_n \rangle$ contains a convergent subsequence $\langle a_{i_n} \rangle$. But, by the preceding problem, the Cauchy sequence $\langle a_n \rangle$ converges to the same limit as its subsequence $\langle a_{i_n} \rangle$. In other words, the Cauchy sequence $\langle a_n \rangle$ converges to a real number.

28. Determine whether or not each of the following subsets of $\mathbf{R}$ is complete: (i) $\mathbf{N}$, the set of positive integers; (ii) $\mathbf{Q}^c$, the set of irrational numbers.

Solution:

(i) Let $\langle a_n \rangle$ be a Cauchy sequence of positive integers. If $\epsilon = \frac{1}{2}$, then

$$|a_n - a_m| < \epsilon = \tfrac{1}{2} \quad \text{implies} \quad a_n = a_m$$

Therefore, the Cauchy sequence $\langle a_n \rangle$ is of the form $\langle a_1, a_2, \ldots, a_{n_0}, b, b, b, \ldots \rangle$ which converges to the positive integer b. Hence $\mathbf{N}$ is complete.

(ii) Observe that each of the open intervals

$$(-1, 1), \ (-\tfrac{1}{2}, \tfrac{1}{2}), \ (-\tfrac{1}{3}, \tfrac{1}{3}), \ \ldots$$

contains irrational points. Hence there exists a sequence $\langle a_n \rangle$ of irrational numbers such that a_n belongs to the open interval $(-1/n, 1/n)$. The sequence $\langle a_n \rangle$ will be a Cauchy sequence of points in $\mathbf{Q}^c$ and it will converge to the rational number 0. Hence $\mathbf{Q}^c$ is not complete.

CONTINUITY

29. Prove: If the function $f: \mathbf{R} \to \mathbf{R}$ is constant, say $f(x) = a$ for every $x \in \mathbf{R}$, then f is continuous.

Solution:

Method 1. The function f is continuous iff the inverse $f^{-1}[G]$ of any open set G is also open. Since $f(x) = a$ for every $x \in \mathbf{R}$,

$$f^{-1}[G] = \begin{cases} \emptyset & \text{if } a \notin G \\ \mathbf{R} & \text{if } a \in G \end{cases}$$

for any open set G. In either case, $f^{-1}[G]$ is open since both $\mathbf{R}$ and $\emptyset$ are open sets.

Method 2. We show that f is continuous at any point x_0 using the $\epsilon - \delta$ definition of continuity. Let $\epsilon > 0$. Then for any $\delta > 0$, say $\delta = 1$,

$$|x - x_0| < 1 \quad \text{implies} \quad |f(x) - f(x_0)| = |a - a| = 0 < \epsilon$$

Hence f is continuous.

30. Prove: The identity function $f: \mathbf{R} \to \mathbf{R}$, that is, the function defined by $f(x) = x$, is continuous.

Solution:

Method 1. Let G be any open set. Then $f^{-1}[G] = G$ is also an open set. Accordingly, f is continuous.

Method 2. We show that f is continuous at any point x_0 using the $\epsilon - \delta$ definition of continuity. Let $\epsilon > 0$. Then choosing $\epsilon = \delta$,

$$|x - x_0| < \delta \quad \text{implies} \quad |f(x) - f(x_0)| = |x - x_0| < \delta = \epsilon$$

Accordingly, f is continuous.

31. Prove: Let the functions $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ be continuous. Then the composition function $g \circ f: \mathbf{R} \to \mathbf{R}$ is also continuous.

Solution:

We show that the inverse $(g \circ f)^{-1}[G]$ of any open set G is also open. Since g is continuous, the inverse $g^{-1}[G]$ is an open set. But since f is continuous, the inverse $f^{-1}[g^{-1}[G]]$ of $g^{-1}[G]$ is also open. Recall that

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}$$

Hence

$$(g \circ f)^{-1}[G] = (f^{-1} \circ g^{-1})[G] = f^{-1}[g^{-1}[G]]$$

is an open set. Thus the composition function $g \circ f: \mathbf{R} \to \mathbf{R}$ is continuous.

32. Prove: Let $f: \mathbf{R} \to \mathbf{R}$ be continuous and let $f(q) = 0$ for every rational number $q \in Q$. Then $f(x) = 0$ for every real number $x \in \mathbf{R}$.

Solution:

Suppose $f(p)$ is not zero for some real number $p \in \mathbf{R}$, i.e. suppose

$$\exists p \in \mathbf{R} \quad \text{such that} \quad f(p) = \gamma, \ |\gamma| > 0$$

Choose $\epsilon = \frac{1}{2}|\gamma|$. Since f is continuous,

$$\exists \delta > 0 \quad \text{such that} \quad |x - p| < \delta \quad \text{implies} \quad |f(x) - f(p)| < \epsilon = \tfrac{1}{2}|\gamma|$$

Now there are rational points in every open interval. In particular,

$$\exists q \in \mathbf{Q} \quad \text{such that} \quad q \in \{x : |x - p| < \delta\}$$

which implies

$$|f(q) - f(p)| = |f(p)| = |\gamma| < \epsilon = \tfrac{1}{2}|\gamma|$$

an impossibility. Hence $f(x) = 0$ for every $x \in \mathbf{R}$.

33. Prove Theorem 4.8: A function $f: \mathbf{R}^2 \to \mathbf{R}^2$ is continuous if and only if the inverse *image of every open set is open.*

Solution:

Let $f: \mathbf{R}^2 \to \mathbf{R}^2$ be continuous and let V be an open subset of $\mathbf{R}^2$. We want to show that $f^{-1}[V]$ is also an open set. Let $p \in f^{-1}[V]$. Then $f(p) \in V$. By definition of continuity, there exists an open set U_p containing p such that $f[U_p] \subset V$. Hence (as indicated in the diagram below)

$$U_p \subset f^{-1}[f[U_p]] \subset f^{-1}[V]$$

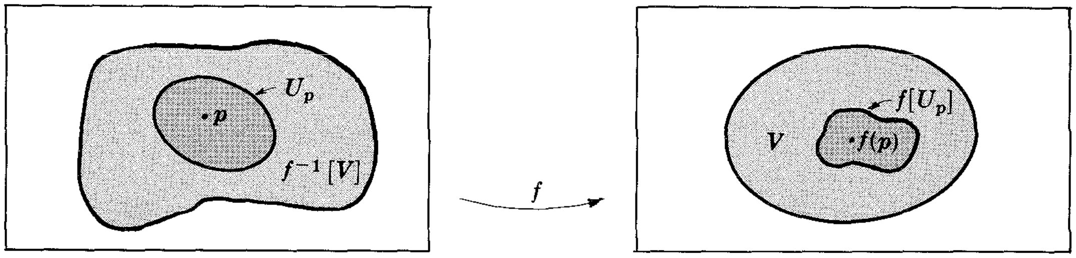
We have shown that, for every point $p \in f^{-1}[V]$, there exists an open set U_p such that

$$p \in U_p \subset f^{-1}[V]$$

Accordingly,

$$f^{-1}[V] = \bigcup\{U_p : p \in f^{-1}[V]\}$$

So $f^{-1}[V]$ is the union of open sets and is, therefore, open itself.



On the other hand, suppose the inverse of every open set is open. We want to show that f is continuous at any point $p \in \mathbf{R}$. Let V be an open set containing $f(p)$, i.e. $f(p) \in V$. Then $f^{-1}[V]$ is an open set containing p with the property that $f[f^{-1}[V]] \subset V$. Hence f is continuous at p.

34. Give an example of two functions $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ such that f and g are each discontinuous (not continuous) at every point and such that the sum $f+g$ is continuous at every point in R.

Solution:

Consider the functions f and g defined by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational} \\ 1 & \text{if } x \text{ is irrational} \end{cases}, \qquad g(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

The functions f and g are discontinuous at every point in $\mathbf{R}$, but the sum $f+g$ is the constant function $(f+g)(x) = 1$ which is continuous.

35. Prove: Let the function $f: \mathbf{R} \to \mathbf{R}$ be continuous at a point $p \in \mathbf{R}$.

(i) If $f(p)$ is positive, i.e. $f(p) > 0$, then there exists an open interval S containing p such that f is positive at every point in S.

(ii) If $f(p)$ is negative, i.e. $f(p) < 0$, then there exists an open interval S containing p such that f is negative at every point in S.

Solution:

We prove (i). The proof of (ii) is similar and will be omitted. Suppose $f(p) = \epsilon > 0$. Since f is continuous at p,

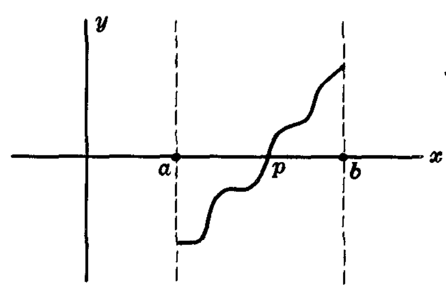
$$\exists \delta > 0 \quad \text{such that} \quad |x - p| < \delta \quad \text{implies} \quad |f(x) - f(p)| < \epsilon$$

or, equivalently,

$$x \in (p - \delta, p + \delta) \quad \text{implies} \quad f(x) \in (f(p) - \epsilon, f(p) + \epsilon) = (0, 2\epsilon)$$

Thus for every point x in the open interval $(p - \delta, p + \delta)$, $f(x)$ is positive.

36. Prove: Let $f: \mathbf{R} \to \mathbf{R}$ be continuous at every point in a closed interval $[a, b]$, and let $f(a) < 0 < f(b)$. Then there exists a point $p \in [a, b]$ such that $f(p) = 0$. (In other words, the graph of a continuous function defined on a closed interval which lies both below and above the x-axis must cross the x-axis at at least one point, as indicated in the diagram.)



Solution:

Let A be the set of points in $[a, b]$ at which f is negative, i.e.,

$$A = \{x : x \in [a, b],\ f(x) < 0\}$$

Observe that A is not empty since, for example, $a \in A$. Let $p = \sup(A)$ be the least upper bound for A. Since $a \in A$, $a \leq p$; and since b is an upper bound for A, $p \leq b$. So p belongs to the interval $[a, b]$.

We claim that $f(p) = 0$. If $f(p) < 0$, then, by the preceding problem, there is an open interval $(p-\delta,\ p+\delta)$ in which f is negative, i.e.,

$$(p-\delta, p+\delta) \subset A$$

So p cannot be an upper bound for A. On the other hand, if $f(p) > 0$, then there exists an interval $(p-\delta,\ p+\delta)$ in which f is positive; so

$$(p-\delta, p+\delta) \cap A = \emptyset$$

which implies that p cannot be a least upper bound for A. Thus $f(p)$ can only be zero, i.e. $f(p) = 0$.

Remark. The theorem is also true and proved similarly in the case $f(b) < 0 < f(a)$.

37. Prove Theorem (Weierstrass) 4.9: Let $f: \mathbf{R} \to \mathbf{R}$ be continuous on a closed interval $[a, b]$. Then the function assumes every value between $f(a)$ and $f(b)$.

Solution:

Suppose $f(a) < f(b)$ and let y_0 be a real number such that $f(a) < y_0 < f(b)$. We want to prove that there is a point p such that $f(p) = y_0$. Consider the function $g(x) = f(x) - y_0$ which is also continuous. Observe that $g(a) < 0 < g(b)$.

By the preceding problem, there exists a point p such that $g(p) = f(p) - y_0 = 0$. Hence $f(p) = y_0$.

The case when $f(b) < f(a)$ is proved similarly.

Supplementary Problems

OPEN SETS, CLOSED SETS, ACCUMULATION POINTS

38. Prove: If A is a finite subset of $\mathbf{R}$, then the derived set A' of A is empty, i.e. $A' = \emptyset$.

39. Prove: Every finite subset of $\mathbf{R}$ is closed.

40. Prove: If $A \subset B$, then $A' \subset B'$.

41. Prove: A subset B of $\mathbf{R}^2$ is closed if and only if $d(p, B) = 0$ implies $p \in B$, where $d(p, B) = \inf\{d(p, q) : q \in B\}$.

42. Prove: $A \cup A'$ is closed for any set A.

43. Prove: $A \cup A'$ is the smallest closed set containing A, i.e. if F is closed and $A \subset F \subset A \cup A'$ then $F = A \cup A'$.

44. Prove: The set of interior points of any set A, written $\text{int}(A)$, is an open set.

45. Prove: The set of interior points of A is the largest open set contained in A, i.e. if G is open and $\text{int}(A) \subset G \subset A$, then $\text{int}(A) = G$.

46. Prove: The only subsets of $\mathbf{R}$ which are both open and closed are $\emptyset$ and $\mathbf{R}$.

SEQUENCES

47. Prove: If the sequence $\langle a_n \rangle$ converges to $b \in \mathbf{R}$, then the sequence $\langle |a_n - b| \rangle$ converges to 0.

48. Prove: If the sequence $\langle a_n \rangle$ converges to 0, and the sequence $\langle b_n \rangle$ is bounded, then the sequence $\langle a_n b_n \rangle$ also converges to 0.

49. Prove: If $a_n \to a$ and $b_n \to b$, then the sequence $\langle a_n + b_n \rangle$ converges to $a + b$.

50. Prove: If $a_n \to a$ and $b_n \to b$, then the sequence $\langle a_n b_n \rangle$ converges to ab.

51. Prove: If $a_n \to a$ and $b_n \to b$ where $b_n \neq 0$ and $b \neq 0$, then the sequence $\langle a_n/b_n \rangle$ converges to a/b.

52. Prove: If the sequence $\langle a_n \rangle$ converges to b, then every subsequence $\langle a_{i_n} \rangle$ of $\langle a_n \rangle$ also converges to b.

53. Prove: If the sequence $\langle a_n \rangle$ converges to b, then either the range $\{a_n\}$ of the sequence $\langle a_n \rangle$ is finite, or b is an accumulation point of the range $\{a_n\}$.

54. Prove: If the sequence $\langle a_n \rangle$ of distinct elements is bounded and the range $\{a_n\}$ of $\langle a_n \rangle$ has exactly one limit point b, then the sequence $\langle a_n \rangle$ converges to b.

(*Remark*: The sequence $\langle 1, \frac{1}{2}, 2, \frac{1}{3}, 3, \frac{1}{4}, 4, \ldots \rangle$ shows that the condition of boundedness cannot be removed from this theorem.)

CONTINUITY

55. Prove: A function $f: \mathbf{R} \to \mathbf{R}$ is continuous at $a \in \mathbf{R}$ if and only if for every sequence $\langle a_n \rangle$ converging to a, the sequence $\langle f(a_n) \rangle$ converges to $f(a)$.

56. Prove: Let the function $f: \mathbf{R} \to \mathbf{R}$ be continuous at $p \in \mathbf{R}$. Then there exists an open interval S containing p such that f is bounded on the open interval S.

57. Give an example of a function $f: \mathbf{R} \to \mathbf{R}$ which is continuous at every point in the open interval $S = (0, 1)$ but which is not bounded on the open interval S.

58. Prove: Let $f: \mathbf{R} \to \mathbf{R}$ be continuous at every point in a closed interval $A = [a, b]$. Then f is bounded on A. (*Remark*: By the preceding problem, this theorem is not true if A is not closed.)

59. Prove: Let $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ be continuous. Then the sum $(f+g): \mathbf{R} \to \mathbf{R}$ is continuous, where $f+g$ is defined by $(f+g)(x) \equiv f(x) + g(x)$.

60. Prove: Let $f: \mathbf{R} \to \mathbf{R}$ be continuous, and let k be any real number. Then the function $(kf): \mathbf{R} \to \mathbf{R}$ is continuous, where kf is defined by $(kf)(x) \equiv k(f(x))$.

61. Prove: Let $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ be continuous. Then $\{x \in \mathbf{R} : f(x) = g(x)\}$ is a closed set.

62. Prove: The projection $\pi_x: \mathbf{R}^2 \to \mathbf{R}$ is continuous where π_x is defined by $\pi_x(\langle a, b \rangle) = a$.

63. Consider the functions $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ defined by

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}, \qquad g(x) = \begin{cases} x \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Prove g is continuous at 0 but f is not continuous at 0.

64. Recall that every rational number $q \in \mathbf{Q}$ can be written uniquely in the form $q = a/b$ where $a \in \mathbf{Z}$, $b \in \mathbf{N}$, and a and b are relatively prime. Consider the function $f: \mathbf{R} \to \mathbf{R}$ defined by

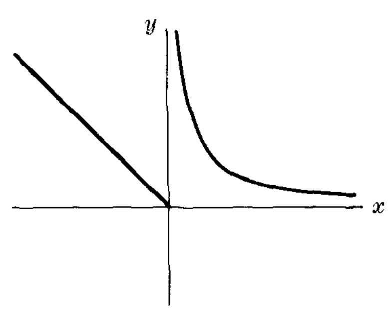
$$f(x) = \begin{cases} 0 & \text{if } x \text{ is irrational} \\ 1/b & \text{if } x \text{ is rational and } x = a/b \text{ as above} \end{cases}$$

Prove that f is continuous at every irrational point, but f is discontinuous at every rational point.

Answers to Supplementary Problems

57. Consider the function

$$f(x) = \begin{cases} -x & \text{if } x \leq 0 \\ 1/x & \text{if } x > 0 \end{cases}$$



The function f is continuous at every point in $\mathbf{R}$ except at 0 as indicated in the adjacent graph of f. Hence f is continuous at every point in the open interval $(0, 1)$. But f is not bounded on $(0, 1)$.

58. *Hint*. Use the result stated in Problem 56 and the Heine-Borel Theorem.

Chapter 5

Topological Spaces: Definitions

TOPOLOGICAL SPACES

Let X be a non-empty set. A class $\mathcal{T}$ of subsets of X is a *topology* on X iff $\mathcal{T}$ satisfies the following axioms.

[O_1] X and $\emptyset$ belong to $\mathcal{T}$.

[O_2] The union of any number of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

[O_3] The intersection of any two sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

The members of $\mathcal{T}$ are then called *$\mathcal{T}$-open sets,* or simply *open sets,* and X together with $\mathcal{T}$, i.e. the pair $(X, \mathcal{T})$ is called a *topological space.*

Example 1.1: Let $\mathcal{U}$ denote the class of all open sets of real numbers discussed in Chapter 4. Then $\mathcal{U}$ is a topology on $\mathbf{R}$; it is called the *usual topology* on $\mathbf{R}$. Similarly, the class $\mathcal{U}$ of all open sets in the plane $\mathbf{R}^2$ is a topology, and also called the *usual topology,* on $\mathbf{R}^2$. We shall always assume the usual topology on $\mathbf{R}$ and $\mathbf{R}^2$ unless otherwise specified.

Example 1.2: Consider the following classes of subsets of $X = \{a, b, c, d, e\}$.

$$\mathcal{T}_1 = \{X,\ \emptyset,\ \{a\},\ \{c,d\},\ \{a,c,d\},\ \{b,c,d,e\}\}$$
$$\mathcal{T}_2 = \{X,\ \emptyset,\ \{a\},\ \{c,d\},\ \{a,c,d\},\ \{b,c,d\}\}$$
$$\mathcal{T}_3 = \{X,\ \emptyset,\ \{a\},\ \{c,d\},\ \{a,c,d\},\ \{a,b,d,e\}\}$$

Observe that $\mathcal{T}_1$ is a topology on X since it satisfies the necessary three axioms **[O_1]**, **[O_2]** and **[O_3]**. But $\mathcal{T}_2$ is not a topology on X since the union

$$\{a,c,d\} \cup \{b,c,d\} = \{a,b,c,d\}$$

of two members of $\mathcal{T}_2$ does not belong to $\mathcal{T}_2$, i.e. $\mathcal{T}_2$ does not satisfy the axiom **[O_2]**.

Also, $\mathcal{T}_3$ is not a topology on X since the intersection

$$\{a,c,d\} \cap \{a,b,d,e\} = \{a,d\}$$

of two sets in $\mathcal{T}_3$ does not belong to $\mathcal{T}_3$, i.e. $\mathcal{T}_3$ does not satisfy the axiom **[O_3]**.

Example 1.3: Let $\mathcal{D}$ denote the class of all subsets of X. Observe that $\mathcal{D}$ satisfies the axioms for a topology on X. This topology is called the *discrete topology*; and X together with its discrete topology, i.e. the pair $(X, \mathcal{D})$, is called a *discrete topological space* or simply a *discrete space.*

Example 1.4: As seen by axiom **[O_1]**, a topology on X must contain the sets X and $\emptyset$. The class $\mathcal{J} = \{X, \emptyset\}$, consisting of X and $\emptyset$ alone, is itself a topology on X. It is called the *indiscrete topology*; and X together with its indiscrete topology, i.e. $(X, \mathcal{J})$, is called an *indiscrete topological space* or simply an *indiscrete space.*

Example 1.5: Let $\mathcal{T}$ denote the class of all subsets of X whose complements are finite together with the empty set $\emptyset$. This class $\mathcal{T}$ is also a topology on X. It is called the *cofinite topology* or the *T_1-topology* on X. (The significance of the T_1 will appear in a later chapter.)

Example 1.6: The intersection $\mathcal{T}_1 \cap \mathcal{T}_2$ of any two topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ on X is also a topology on X. For, by $[\mathbf{O}_1]$, X and $\emptyset$ each belongs to both $\mathcal{T}_1$ and $\mathcal{T}_2$; hence X and $\emptyset$ each belongs to the intersection $\mathcal{T}_1 \cap \mathcal{T}_2$, i.e. $\mathcal{T}_1 \cap \mathcal{T}_2$ satisfies $[\mathbf{O}_1]$. Furthermore, if $G,H \in \mathcal{T}_1 \cap \mathcal{T}_2$ then, in particular, $G,H \in \mathcal{T}_1$ and $G,H \in \mathcal{T}_2$. But since $\mathcal{T}_1$ and $\mathcal{T}_2$ are topologies, $G \cap H \in \mathcal{T}_1$ and $G \cap H \in \mathcal{T}_2$. Accordingly,

$$G \cap H \in \mathcal{T}_1 \cap \mathcal{T}_2$$

In other words $\mathcal{T}_1 \cap \mathcal{T}_2$ satisfies $[\mathbf{O}_3]$. Similarly, $\mathcal{T}_1 \cap \mathcal{T}_2$ satisfies $[\mathbf{O}_2]$.

The statement in the preceding example can, in fact, be generalized to any collection of topologies. Namely,

Theorem 5.1: Let $\{\mathcal{T}_i : i \in I\}$ be any collection of topologies on a set X. Then the intersection $\cap_i \mathcal{T}_i$ is also a topology on X.

In our last example, we show that the union of topologies need not be a topology.

Example 1.7: Each of the classes

$$\mathcal{T}_1 = \{X, \emptyset, \{a\}\} \quad \text{and} \quad \mathcal{T}_2 = \{X, \emptyset, \{b\}\}$$

is a topology on $X = \{a, b, c\}$. But the union

$$\mathcal{T}_1 \cup \mathcal{T}_2 = \{X, \emptyset, \{a\}, \{b\}\}$$

is not a topology on X since it violates $[\mathbf{O}_2]$. That is, $\{a\} \in \mathcal{T}_1 \cup \mathcal{T}_2$, $\{b\} \in \mathcal{T}_1 \cup \mathcal{T}_2$ but $\{a\} \cup \{b\} = \{a, b\}$ does not belong to $\mathcal{T}_1 \cup \mathcal{T}_2$.

If G is an open set containing a point $p \in X$, then G is called an *open neighborhood* of p. Also, G without p, i.e. $G \setminus \{p\}$, is called a *deleted open neighborhood* of p.

Remark: The axioms $[\mathbf{O}_1]$, $[\mathbf{O}_2]$ and $[\mathbf{O}_3]$ are equivalent to the following two axioms:

$[\mathbf{O}_1^*]$ The union of any number of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

$[\mathbf{O}_2^*]$ The intersection of any finite number of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

For $[\mathbf{O}_1^*]$ implies that $\emptyset$ belongs to $\mathcal{T}$ since

$$\cup\{G \in \mathcal{T} : G \in \emptyset\} = \emptyset$$

i.e. the empty union of sets is the empty set. Furthermore, $[\mathbf{O}_2^*]$ implies that X belongs to $\mathcal{T}$ since

$$\cap\{G \in \mathcal{T} : G \in \emptyset\} = X$$

i.e. the empty intersection of subsets of X is X itself.

ACCUMULATION POINTS

Let X be a topological space. A point $p \in X$ is an *accumulation point* or *limit point* (also called *cluster point* or *derived point*) of a subset A of X iff every open set G containing p contains a point of A different from p, i.e.,

$$G \text{ open}, \ p \in G \quad \text{implies} \quad (G \setminus \{p\}) \cap A \neq \emptyset$$

The set of accumulation points of A, denoted by A', is called the *derived set* of A.

Example 2.1: The class $\mathcal{T} = \{X, \emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d\, e\}\}$

defines a topology on $X = \{a, b, c, d, e\}$. Consider the subset $A = \{a, b, c\}$ of X. Observe that $b \in X$ is a limit point of A since the open sets containing b are $\{b, c, d, e\}$ and X, and each contains a point of A different from b, i.e. c. On the other hand, the point $a \in X$ is not a limit point of A since the open set $\{a\}$, which contains a, does not contain a point of A different from a. Similarly, the points d and e are limit points of A and the point c is not a limit point of A. So $A' = \{b, d, e\}$ is the derived set of A.

Example 2.2: Let X be an indiscrete topological space, i.e. X and $\emptyset$ are the only open subsets of X. Then X is the only open set containing any point $p \in X$. Hence p is an accumulation point of every subset of X except the empty set $\emptyset$ and the set consisting of p alone, i.e. the singleton set $\{p\}$. Accordingly, the derived set A' of any subset A of X is as follows:

$$A' = \begin{cases} \emptyset & \text{if } A = \emptyset \\ \{p\}^c = X \setminus \{p\} & \text{if } A = \{p\} \\ X & \text{if } A \text{ contains two or more points} \end{cases}$$

Observe that, for the usual topology on the line $\mathbf{R}$ and the plane $\mathbf{R}^2$, the above definition of an accumulation point is the same as that given in Chapter 4.

CLOSED SETS

Let X be a topological space. A subset A of X is a *closed set* iff its complement A^c is an open set.

Example 3.1: The class $\mathcal{T} = \{X, \emptyset, \{a\}, \{c,d\}, \{a,c,d\}, \{b,c,d,e\}\}$

defines a topology on $X = \{a,b,c,d,e\}$. The closed subsets of X are

$$\emptyset,\ X,\ \{b,c,d,e\},\ \{a,b,e\},\ \{b,e\},\ \{a\}$$

that is, the complements of the open subsets of X. Note that there are subsets of X, such as $\{b,c,d,e\}$, which are both open and closed, and there are subsets of X, such as $\{a,b\}$, which are neither open nor closed.

Example 3.2: Let X be a discrete topological space, i.e. every subset of X is open. Then every subset of X is also closed since its complement is always open. In other words, all subsets of X are both open and closed.

Recall that $A^{cc} = A$, for any subset A of a space X. Hence

Proposition 5.2: In a topological space X, a subset A of X is open if and only if its complement is closed.

The axioms $[\mathbf{O}_1]$, $[\mathbf{O}_2]$ and $[\mathbf{O}_3]$ of a topological space and DeMorgan's Laws give

Theorem 5.3: Let X be a topological space. Then the class of closed subsets of X possesses the following properties:

(i) X and $\emptyset$ are closed sets.

(ii) The intersection of any number of closed sets is closed.

(iii) The union of any two closed sets is closed.

Closed sets can also be characterized in terms of their limit points as follows:

Theorem 5.4: A subset A of a topological space X is closed if and only if A contains each of its accumulation points.

In other words, a set A is closed if and only if the derived set A' of A is a subset of A, i.e. $A' \subset A$.

CLOSURE OF A SET

Let A be a subset of a topological space X. The *closure of* A, denoted by

$$\bar{A} \quad \text{or} \quad A^-$$

is the intersection of all closed supersets of A. In other words, if $\{F_i : i \in I\}$ is the class of all closed subsets of X containing A, then

$$\bar{A} = \cap_i F_i$$

Observe first that $\bar{A}$ is a closed set since it is the intersection of closed sets. Furthermore, $\bar{A}$ is the smallest closed superset of A, that is, if F is a closed set containing A, then

$$A \subset \bar{A} \subset F$$

Accordingly, a set A is closed if and only if $A = \bar{A}$. We state these results formally:

Proposition 5.5: Let $\bar{A}$ be the closure of a set A. Then: (i) $\bar{A}$ is closed; (ii) if F is a closed superset of A, then $A \subset \bar{A} \subset F$; and (iii) A is closed iff $A = \bar{A}$.

Example 4.1: Consider the topology $\mathcal{T}$ on $X = \{a, b, c, d, e\}$ of **Example 3.1** where the closed subsets of X are

$$\emptyset,\ X,\ \{b, c, d, e\},\ \{a, b, e\},\ \{b, e\},\ \{a\}$$

Accordingly, $\overline{\{b\}} = \{b, e\}$, $\overline{\{a, c\}} = X$, $\overline{\{b, d\}} = \{b, c, d, e\}$

Example 4.2: Let X be a cofinite topological space, i.e. the complements of finite sets and $\emptyset$ are the open sets. Then the closed sets are precisely the finite subsets of X together with X. Hence if $A \subset X$ is finite, its closure $\bar{A}$ is A itself since A is closed. On the other hand, if $A \subset X$ is infinite then X is the only closed superset of A; so $\bar{A}$ is X. More concisely, for any subset A of a cofinite space X,

$$\bar{A} = \begin{cases} A & \text{if } A \text{ is finite} \\ X & \text{if } A \text{ is infinite} \end{cases}$$

The closure of a set can be completely described in terms of its limit points as follows:

Theorem 5.6: Let A be a subset of a topological space X. Then the closure of A is the union of A and its set of accumulation points, i.e.,

$$\bar{A} = A \cup A'$$

A point $p \in X$ is called a *closure point* or *adherent point* of $A \subset X$ iff p belongs to the closure of A, i.e. $p \in \bar{A}$. In view of the preceding theorem, $p \in X$ is a closure point of $A \subset X$ iff $p \in A$ or p is a limit point of A.

Example 4.3: Consider the set $\mathbf{Q}$ of rational numbers. As seen previously, in the usual topology for $\mathbf{R}$, every real number $a \in \mathbf{R}$ is a limit point of $\mathbf{Q}$. Hence the closure of $\mathbf{Q}$ is the entire set $\mathbf{R}$ of real numbers, i.e. $\bar{\mathbf{Q}} = \mathbf{R}$.

A subset A of a topological space X is said to be *dense* in $B \subset X$ if B is contained in the closure of A, i.e. $B \subset \bar{A}$. In particular, A is dense in X or is a dense subset of X iff $\bar{A} = X$.

Example 4.4: Observe in **Example 4.1** that

$$\overline{\{a, c\}} = X \quad \text{and} \quad \overline{\{b, d\}} = \{b, c, d, e\}$$

where $X = \{a, b, c, d, e\}$. Hence the set $\{a, c\}$ is a dense subset of X but the set $\{b, d\}$ is not.

Example 4.5: As noted in **Example 4.3**, $\bar{\mathbf{Q}} = \mathbf{R}$. In other words, in the usual topology, the set $\mathbf{Q}$ of rational numbers is dense in $\mathbf{R}$.

The operator "closure", assigning to each subset A of X its closure $\bar{A} \subset X$ satisfies the four properties appearing in the proposition below, called the Kuratowski Closure Axioms. In fact, these axioms may be used to define a topology on X, as we shall prove subsequently.

Proposition 5.7: (i) $\bar{\emptyset} = \emptyset$; (ii) $A \subset \bar{A}$; (iii) $\overline{A \cup B} = \bar{A} \cup \bar{B}$; and (iv) $(A^-)^- = \bar{A}$.

INTERIOR, EXTERIOR, BOUNDARY

Let A be a subset of a topological space X. A point $p \in A$ is called an *interior point* of A if p belongs to an open set G contained in A:

$$p \in G \subset A \quad \text{where } G \text{ is open}$$

The set of interior points of A, denoted by

$$\text{int}(A), \quad \overset{\circ}{A} \quad \text{or} \quad A^\circ$$

is called the *interior* of A. The interior of A can also be characterized as follows:

Proposition 5.8: The interior of a set A is the union of all open subsets of A. Furthermore: (i) A° is open; (ii) A° is the largest open subset of A, i.e. if G is an open subset of A then $G \subset A^\circ \subset A$; and (iii) A is open iff $A = A^\circ$.

The *exterior* of A, written $\text{ext}(A)$, is the interior of the complement of A, i.e. $\text{int}(A^c)$. The *boundary* of A, written $\text{b}(A)$, is the set of points which do not belong to the interior or the exterior of A. Next follows an important relationship between interior, exterior and closure.

Theorem 5.9: Let A be any subset of a topological space X. Then the closure of A is the union of the interior and boundary of A, i.e. $\bar{A} = A^\circ \cup \text{b}(A)$.

Example 5.1: Consider the four intervals $[a,b]$, (a,b), $(a,b]$ and $[a,b)$ whose endpoints are a and b. The interior of each is the open interval (a,b) and the boundary of each is the set of endpoints, i.e. $\{a,b\}$.

Example 5.2: Consider the topology

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{c,d\}, \{a,c,d\}, \{b,c,d,e\}\}$$

on $X = \{a,b,c,d,e\}$, and the subset $A = \{b,c,d\}$ of X. The points c and d are each interior points of A since

$$c,d \in \{c,d\} \subset A$$

where $\{c,d\}$ is an open set. The point $b \in A$ is not an interior point of A; so $\text{int}(A) = \{c,d\}$. Only the point $a \in X$ is exterior to A, i.e. interior to the complement $A^c = \{a,e\}$ of A; hence $\text{int}(A^c) = \{a\}$. Accordingly the boundary of A consists of the points b and e, i.e. $\text{b}(A) = \{b,e\}$.

Example 5.3: Consider the set $\mathbf{Q}$ of rational numbers. Since every open subset of $\mathbf{R}$ contains both rational and irrational points, there are no interior or exterior points of $\mathbf{Q}$; so $\text{int}(\mathbf{Q}) = \emptyset$ and $\text{int}(\mathbf{Q}^c) = \emptyset$. Hence the boundary of $\mathbf{Q}$ is the entire set of real numbers, i.e. $\text{b}(\mathbf{Q}) = \mathbf{R}$.

A subset A of a topological space X is said to be *nowhere dense* in X if the interior of the closure of A is empty, i.e. $\text{int}(\bar{A}) = \emptyset$.

Example 5.4: Consider the subset $A = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$ of $\mathbf{R}$. As noted previously, A has exactly one limit point, 0. Hence $\bar{A} = \{0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$. Observe that $\bar{A}$ has no interior points; so A is nowhere dense in $\mathbf{R}$.

Example 5.5: Let A consist of the rational points between 0 and 1, i.e. $A = \{x : x \in Q, 0 < x < 1\}$. Observe that the interior of A is empty, i.e. $\text{int}(A) = \emptyset$. But A is not nowhere dense in $\mathbf{R}$; for the closure of A is $[0,1]$, and so

$$\text{int}(\bar{A}) = \text{int}([0,1]) = (0,1)$$

is not empty.

NEIGHBORHOODS AND NEIGHBORHOOD SYSTEMS

Let p be a point in a topological space X. A subset N of X is a *neighborhood* of p iff N is a superset of an open set G containing p:

$$p \in G \subset N \quad \text{where } G \text{ is an open set}$$

In other words, the relation "N is a neighborhood of a point p" is the inverse of the relation "p is an interior point of N". The class of all neighborhoods of $p \in X$, denoted by $\mathcal{N}_p$, is called the *neighborhood system* of p.

Example 6.1: Let a be any real number, i.e. $a \in \mathbf{R}$. Then each closed interval $[a-\delta, a+\delta]$, with center a, is a neighborhood of a since it contains the open interval $(a-\delta, a+\delta)$ containing a. Similarly, if p is a point in the plane $\mathbf{R}^2$, then every closed disc $\{q \in \mathbf{R}^2 : d(p,q) < \delta \neq 0\}$, with center p, is a neighborhood of p since it contains the open disc with center p.

The central facts about the neighborhood system $\mathcal{N}_p$ of any point $p \in X$ are the four properties appearing in the proposition below, called the Neighborhood Axioms. In fact, these axioms may be used to define a topology on X, as we shall note subsequently.

Proposition 5.10: (i) $\mathcal{N}_p$ is not empty and p belongs to each member of $\mathcal{N}_p$.

(ii) The intersection of any two members of $\mathcal{N}_p$ belongs to $\mathcal{N}_p$.

(iii) Every superset of a member of $\mathcal{N}_p$ belongs to $\mathcal{N}_p$.

(iv) Each member $N \in \mathcal{N}_p$ is a superset of a member $G \in \mathcal{N}_p$ where G is a neighborhood of each of its points, i.e. $G \in \mathcal{N}_g$ for every $g \in G$.

CONVERGENT SEQUENCES

A sequence $\langle a_1, a_2, \ldots \rangle$ of points in a topological space X *converges* to a point $b \in X$, or b is the *limit* of the sequence $\langle a_n \rangle$, denoted by

$$\lim_{n \to \infty} a_n = b, \quad \lim a_n = b \quad \text{or} \quad a_n \to b$$

iff for each open set G containing b there exists a positive integer $n_0 \in N$ such that

$$n > n_0 \quad \text{implies} \quad a_n \in G$$

that is, if G contains almost all, i.e. all except a finite number, of the terms of the sequence.

Example 7.1: Let $\langle a_1, a_2, \ldots \rangle$ be a sequence of points in an indiscrete topological space $(X, \mathcal{J})$. Note that: (i) X is the only open set containing any point $b \in X$; and (ii) X contains every term of the sequence $\langle a_n \rangle$. Accordingly, the sequence $\langle a_1, a_2, \ldots \rangle$ converges to every point $b \in X$.

Example 7.2: Let $\langle a_1, a_2, \ldots \rangle$ be a sequence of points in a discrete topological space $(X, \mathcal{D})$. Now for every point $b \in X$, the singleton set $\{b\}$ is an open set containing b. So, if $a_n \to b$, then the set $\{b\}$ must contain almost all of the terms of the sequence. In other words, the sequence $\langle a_n \rangle$ converges to a point $b \in X$ iff the sequence is of the form $\langle a_1, a_2, \ldots, a_{n_0}, b, b, b, \ldots \rangle$.

Example 7.3: Let $\mathcal{T}$ be the topology on an infinite set X which consists of $\emptyset$ and the complements of countable sets (see Problem 56). We claim that a sequence $\langle a_1, a_2, \ldots \rangle$ in X converges to $b \in X$ iff the sequence is also of the form $\langle a_1, a_2, \ldots, a_{n_0}, b, b, b, \ldots \rangle$, i.e. the set A consisting of the terms of $\langle a_n \rangle$ different from b is finite. Now A is countable and so A^c is an open set containing b. Hence if $a_n \to b$ then A^c contains all except a finite number of the terms of the sequence, and so A is finite.

COARSER AND FINER TOPOLOGIES

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be topologies on a non-empty set X. Suppose that each $\mathcal{T}_1$-open subset of X is also a $\mathcal{T}_2$-open subset of X. That is, suppose that $\mathcal{T}_1$ is a subclass of $\mathcal{T}_2$, i.e. $\mathcal{T}_1 \subset \mathcal{T}_2$. Then we say that $\mathcal{T}_1$ is *coarser* or *smaller* (sometimes called *weaker*) than $\mathcal{T}_2$ or that $\mathcal{T}_2$ is *finer* or *larger* than $\mathcal{T}_1$. Observe that the collection $\mathbf{T} = \{\mathcal{T}_i\}$ of all topologies on X is partially ordered by class inclusion; so we shall also write

$$\mathcal{T}_1 \precsim \mathcal{T}_2 \quad \text{for} \quad \mathcal{T}_1 \subset \mathcal{T}_2$$

and we shall say that two topologies on X are *not comparable* if neither is coarser than the other.

Example 8.1: Consider the discrete topology $\mathcal{D}$, the indiscrete topology $\mathcal{J}$ and any other topology $\mathcal{T}$ on any set X. Then $\mathcal{T}$ is coarser than $\mathcal{D}$ and $\mathcal{T}$ is finer than $\mathcal{J}$. That is, $\mathcal{J} \precsim \mathcal{T} \precsim \mathcal{D}$.

Example 8.2: Consider the cofinite topology $\mathcal{T}$ and the usual topology $\mathcal{U}$ on the plane $\mathbf{R}^2$. Recall that every finite subset of $\mathbf{R}^2$ is a $\mathcal{U}$-closed set; hence the complement of any finite subset of $\mathbf{R}^2$, i.e. any member of $\mathcal{T}$, is also a $\mathcal{U}$-open set. In other words, $\mathcal{T}$ is coarser than $\mathcal{U}$, i.e. $\mathcal{T} \precsim \mathcal{U}$.

SUBSPACES, RELATIVE TOPOLOGIES

Let A be a non-empty subset of a topological space $(X, \mathcal{T})$. The class $\mathcal{T}_A$ of all intersections of A with $\mathcal{T}$-open subsets of X is a topology on A; it is called the *relative topology* on A or the *relativization* of $\mathcal{T}$ to A, and the topological space $(A, \mathcal{T}_A)$ is called a *subspace* of $(X, \mathcal{T})$. In other words, a subset H of A is a $\mathcal{T}_A$-open set, i.e. open relative to A, if and only if there exists a $\mathcal{T}$-open subset G of X such that

$$H = G \cap A$$

Example 9.1: Consider the topology

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{c,d\}, \{a,c,d\}, \{b,c,d,e\}\}$$

on $X = \{a,b,c,d,e\}$, and the subset $A = \{a,d,e\}$ of X. Observe that

$$X \cap A = A, \quad \{a\} \cap A = \{a\}, \quad \{a,c,d\} \cap A = \{a,d\}$$
$$\emptyset \cap A = \emptyset, \quad \{c,d\} \cap A = \{d\}, \quad \{b,c,d,e\} \cap A = \{d,e\}$$

Hence the relativization of $\mathcal{T}$ to A is

$$\mathcal{T}_A = \{A, \emptyset, \{a\}, \{d\}, \{a,d\}, \{d,e\}\}$$

Example 9.2: Consider the usual topology $\mathcal{U}$ on $\mathbf{R}$ and the relative topology $\mathcal{T}_A$ on the closed interval $A = [3,8]$. Note that the closed-open interval $[3,5)$ is open in the relative topology on A, i.e. is a $\mathcal{T}_A$-open set, since

$$[3,5) = (2,5) \cap A$$

where $(2,5)$ is a $\mathcal{T}$-open subset of $\mathbf{R}$. Thus we see that a set may be open relative to a subspace but be neither open nor closed in the entire space.

EQUIVALENT DEFINITIONS OF TOPOLOGIES

Our definition of a topological space gave axioms for the open sets in the topological space, that is, we used the open set as the primitive notion for the topology. We now state two theorems which exhibit alternate methods of defining a topology on a set, using as primitives the notions of "neighborhood of a point" and "closure of a set".

Theorem 5.11: Let X be a non-empty set and let there be assigned to each point $p \in X$ a class $\mathcal{A}_p$ of subsets of X satisfying the following axioms:

[A₁] $\mathcal{A}_p$ is not empty and p belongs to each member of $\mathcal{A}_p$.

[A₂] The intersection of any two members of $\mathcal{A}_p$ belongs to $\mathcal{A}_p$.

[A₃] Every superset of a member of $\mathcal{A}_p$ belongs to $\mathcal{A}_p$.

[A₄] Each member $N \in \mathcal{A}_p$ is a superset of a member $G \in \mathcal{A}_p$ such that $G \in \mathcal{A}_g$ for every $g \in G$.

Then there exists one and only one topology $\mathcal{T}$ on X such that $\mathcal{A}_p$ is the $\mathcal{T}$-neighborhood system of the point $p \in X$.

Theorem 5.12: Let X be a non-empty set and let k be an operation which assigns to each subset A of X the subset A^k of X, satisfying the following axioms, called the Kuratowski Closure Axioms:

[K_1] $\emptyset^k = \emptyset$

[K_2] $A \subset A^k$

[K_3] $(A \cup B)^k = A^k \cup B^k$

[K_4] $(A^k)^k = A^k$

Then there exists one and only one topology $\mathcal{T}$ on X such that A^k will be the $\mathcal{T}$-closure of the subset A of X.

Solved Problems

TOPOLOGIES, OPEN SETS

1. Let $X = \{a, b, c, d, e\}$. Determine whether or not each of the following classes of subsets of X is a topology on X.

(i) $\mathcal{T}_1 = \{X, \emptyset, \{a\}, \{a, b\}, \{a, c\}\}$

(ii) $\mathcal{T}_2 = \{X, \emptyset, \{a, b, c\}, \{a, b, d\}, \{a, b, c, d\}\}$

(iii) $\mathcal{T}_3 = \{X, \emptyset, \{a\}, \{a, b\}, \{a, c, d\}, \{a, b, c, d\}\}$

Solution:

(i) $\mathcal{T}_1$ is not a topology on X since

$$\{a, b\}, \{a, c\} \in \mathcal{T}_1 \quad \text{but} \quad \{a, b\} \cup \{a, c\} = \{a, b, c\} \notin \mathcal{T}_1$$

(ii) $\mathcal{T}_2$ is not a topology on X since

$$\{a, b, c\}, \{a, b, d\} \in \mathcal{T}_2 \quad \text{but} \quad \{a, b, c\} \cap \{a, b, d\} = \{a, b\} \notin \mathcal{T}_2$$

(iii) $\mathcal{T}_3$ is a topology on X since it satisfies the necessary axioms.

2. Let $\mathcal{T}$ be the class consisting of $\mathbf{R}$, $\emptyset$ and all infinite open intervals $A_q = (q, \infty)$ with $q \in \mathbf{Q}$, the rationals. Show that $\mathcal{T}$ is not a topology on $\mathbf{R}$.

Solution:

Observe that $$A = \cup\{A_q : q \in \mathbf{Q},\ q > \sqrt{2}\} = (\sqrt{2}, \infty)$$

is the union of members of $\mathcal{T}$, but $A \notin \mathcal{T}$ since $\sqrt{2}$ is irrational. Hence $\mathcal{T}$ violates **[O$_2$]** and is therefore not a topology on $\mathbf{R}$.

3. Let $\mathcal{T}$ be a topology on a set X consisting of four sets, i.e.

$$\mathcal{T} = \{X, \emptyset, A, B\}$$

where A and B are non-empty distinct proper subsets of X. What conditions must A and B satisfy?

Solution:

Since $A \cap B$ must also belong to $\mathcal{T}$, there are two possibilities:

Case I. $A \cap B = \emptyset$

Then $A \cup B$ cannot be A or B; hence $A \cup B = X$. Thus the class $\{A, B\}$ is a partition of X.

Case II. $A \cap B = A$ or $A \cap B = B$

In either case, one of the sets is a subset of the other, and the members of $\mathcal{T}$ are totally ordered by inclusion: $\emptyset \subset A \subset B \subset X$ or $\emptyset \subset B \subset A \subset X$.

4. List all topologies on $X = \{a, b, c\}$ which consist of exactly four members.

Solution:

Each topology $\mathcal{T}$ on X with four members is of the form $\mathcal{T} = \{X, \emptyset, A, B\}$ where A and B correspond to Case I or Case II of the preceding problem.

Case I. $\{A, B\}$ is a partition of X.

The topologies in this case are the following:

$$\mathcal{T}_1 = \{X, \emptyset, \{a\}, \{b, c\}\}, \quad \mathcal{T}_2 = \{X, \emptyset, \{b\}, \{a, c\}\}, \quad \mathcal{T}_3 = \{X, \emptyset, \{c\}, \{a, b\}\}$$

Case II. The members of $\mathcal{T}$ are totally ordered by inclusion.

The topologies in this case are the following:

$$\mathcal{T}_4 = \{X, \emptyset, \{a\}, \{a, b\}\} \qquad \mathcal{T}_7 = \{X, \emptyset, \{b\}, \{a, b\}\}$$
$$\mathcal{T}_5 = \{X, \emptyset, \{a\}, \{a, c\}\} \qquad \mathcal{T}_8 = \{X, \emptyset, \{c\}, \{a, c\}\}$$
$$\mathcal{T}_6 = \{X, \emptyset, \{b\}, \{b, c\}\} \qquad \mathcal{T}_9 = \{X, \emptyset, \{c\}, \{b, c\}\}$$

5. Let $f: X \to Y$ be a function from a non-empty set X into a topological space $(Y, \mathcal{U})$. Furthermore, let $\mathcal{T}$ be the class of inverses of open subsets of Y:

$$\mathcal{T} = \{f^{-1}[G] : G \in \mathcal{U}\}$$

Show that $\mathcal{T}$ is a topology on X.

Solution:

Since $\mathcal{U}$ is a topology, $Y, \emptyset \in \mathcal{U}$. But $X = f^{-1}[Y]$ and $\emptyset = f^{-1}[\emptyset]$, so $X, \emptyset \in \mathcal{T}$ and $\mathcal{T}$ satisfies **[O₁]**.

Let $\{A_i\}$ be a class of sets in $\mathcal{T}$. By definition, there exist $G_i \in \mathcal{U}$ for which $A_i = f^{-1}[G_i]$. But

$$\cup_i A_i = \cup_i f^{-1}[G_i] = f^{-1}[\cup_i G_i]$$

Since $\mathcal{U}$ is a topology, $\cup_i G_i \in \mathcal{U}$, so $\cup_i A_i \in \mathcal{T}$, and $\mathcal{T}$ satisfies **[O_2]**.

Lastly, let $A_1, A_2 \in \mathcal{T}$. Then

$$\exists\, G_1, G_2 \in \mathcal{U} \quad \text{such that} \quad A_1 = f^{-1}[G_1],\ A_2 = f^{-1}[G_2]$$

But

$$A_1 \cap A_2 = f^{-1}[G_1] \cap f^{-1}[G_2] = f^{-1}[G_1 \cap G_2]$$

and $G_1 \cap G_2 \in \mathcal{U}$. Thus $A_1 \cap A_2 \in \mathcal{T}$ and **[O_3]** is also satisfied.

6. Consider the second axiom for a topology $\mathcal{T}$ on a set X:

[O_2] The union of any number of sets in $\mathcal{T}$ belongs to $\mathcal{T}$.

Show that **[O_2]** can be replaced by the following weaker axiom:

[O_2'] The union of any number of sets in $\mathcal{T} \setminus \{X, \emptyset\}$ belongs to $\mathcal{T}$.

In other words, show that the axioms **[O_1]**, **[O_2']** and **[O_3]** are equivalent to the axioms **[O_1]**, **[O_2]** and **[O_3]**.

Solution:

Let $\mathcal{T}$ be a class of subsets of X satisfying **[O_1]**, **[O_2']** and **[O_3]**, and let $\mathcal{A}$ be a subclass of $\mathcal{T}$. We want to show that $\mathcal{T}$ also satisfies **[O_2]**, i.e. that $\bigcup\{E : E \in \mathcal{A}\} \in \mathcal{T}$.

Case I. $X \in \mathcal{A}$.

Then $\bigcup\{E : E \in \mathcal{A}\} = X$ and therefore belongs to $\mathcal{T}$ by **[O_1]**.

Case II. $X \notin \mathcal{A}$.

Then

$$\bigcup\{E : E \in \mathcal{A}\} = \bigcup\{E : E \in \mathcal{A} \setminus \{X\}\}$$

But the empty set $\emptyset$ does not contribute any elements to a union of sets; hence

$$\bigcup\{E : E \in \mathcal{A}\} = \bigcup\{E : E \in \mathcal{A} \setminus \{X\}\} = \bigcup\{E : E \in \mathcal{A} \setminus \{X, \emptyset\}\} \qquad (1)$$

Since $\mathcal{A}$ is a subclass of $\mathcal{T}$, $\mathcal{A} \setminus \{X, \emptyset\}$ is a subclass of $\mathcal{T} \setminus \{X, \emptyset\}$, so by **[$O_2'$]** the union in (1) belongs to $\mathcal{T}$.

7. Prove: Let A be a subset of a topological space X with the property that each point $p \in A$ belongs to an open set G_p contained in A. Then A is open.

Solution:

For each point $p \in A$, $p \in G_p \subset A$. Hence $\bigcup\{G_p : p \in A\} = A$ and so A is a union of open sets and, by **[O_2]**, is open.

8. Let $\mathcal{T}$ be a class of subsets of X totally ordered by set inclusion. Show that $\mathcal{T}$ satisfies **[O_3]**, i.e. the intersection of any two members of $\mathcal{T}$ belongs to $\mathcal{T}$.

Solution:

Let $A, B \in \mathcal{T}$. Since $\mathcal{T}$ is totally ordered by set inclusion,

$$\text{either} \quad A \cap B = A \quad \text{or} \quad A \cap B = B$$

In either case $A \cap B \in \mathcal{T}$, and so $\mathcal{T}$ satisfies **[O_3]**.

9. Let $\mathcal{T}$ be the class of subsets of $\mathbf{R}$ consisting of $\mathbf{R}$, $\emptyset$ and all open infinite intervals $E_a = (a, \infty)$ with $a \in \mathbf{R}$. Show that $\mathcal{T}$ is a topology on $\mathbf{R}$.

Solution:

Since $\mathbf{R}$ and $\emptyset$ belong to $\mathcal{T}$, $\mathcal{T}$ satisfies $[\mathbf{O}_1]$. Observe that $\mathcal{T}$ is totally ordered by set inclusion; hence $\mathcal{T}$ satisfies $[\mathbf{O}_3]$.

Now let $\mathcal{A}$ be a subclass of $\mathcal{T} \setminus \{X, \emptyset\}$, that is $\mathcal{A} = \{E_i : i \in I\}$ where I is some set of real numbers. We want to show that $\cup_i E_i$ belongs to $\mathcal{T}$. If I is not bounded from below, i.e. if $\inf(I) = -\infty$, then $\cup_i E_i = \mathbf{R}$. If I is bounded from below, say $\inf(I) = i_0$, then $\cup_i E_i = (i_0, \infty) = E_{i_0}$. In either case, $\cup_i E_i \in \mathcal{T}$, and $\mathcal{T}$ satisfies $[\mathbf{O}_2']$.

10. Let $\mathcal{T}$ be the class of subsets of $\mathbf{N}$ consisting of $\emptyset$ and all subsets of $\mathbf{N}$ of the form $E_n = \{n, n+1, n+2, \ldots\}$ with $n \in \mathbf{N}$.

(i) Show that $\mathcal{T}$ is a topology on $\mathbf{N}$.

(ii) List the open sets containing the positive integer 6.

Solution:

(i) Since $\emptyset$ and $E_1 = \{1, 2, 3, \ldots\} = \mathbf{N}$ belong to $\mathcal{T}$, $\mathcal{T}$ satisfies $[\mathbf{O}_1]$. Furthermore, since $\mathcal{T}$ is totally ordered by set inclusion, $\mathcal{T}$ also satisfies $[\mathbf{O}_3]$.

Now let $\mathcal{A}$ be a subclass of $\mathcal{T} \setminus \{\mathbf{N}, \emptyset\}$, that is, $\mathcal{A} = \{E_n : n \in I\}$ where I is some set of positive integers. Note that I contains a smallest positive integer n_0 and

$$\cup \{E_n : n \in I\} = \{n_0, n_0 + 1, n_0 + 2, \ldots\} = E_{n_0}$$

which belongs to $\mathcal{T}$. Hence $\mathcal{T}$ satisfies $[\mathbf{O}_2']$, and so $\mathcal{T}$ is a topology on $\mathbf{N}$.

(ii) Since the non-empty open sets are of the form

$$E_n = \{n, n+1, n+2, \ldots\}$$

with $n \in \mathbf{N}$, the open sets containing the positive integer 6 are the following:

$$E_1 = \mathbf{N} = \{1, 2, 3, \ldots\} \qquad E_4 = \{4, 5, 6, \ldots\}$$
$$E_2 = \{2, 3, 4, \ldots\} \qquad E_5 = \{5, 6, 7, \ldots\}$$
$$E_3 = \{3, 4, 5, \ldots\} \qquad E_6 = \{6, 7, 8, \ldots\}$$

ACCUMULATION POINTS, DERIVED SETS

11. Let $\mathcal{T}$ be the topology on $\mathbf{N}$ which consists of $\emptyset$ and all subsets of $\mathbf{N}$ of the form $E_n = \{n, n+1, n+2, \ldots\}$ where $n \in \mathbf{N}$ as in Problem 10.

(i) Find the accumulation points of the set $A = \{4, 13, 28, 37\}$.

(ii) Determine those subsets E of $\mathbf{N}$ for which $E' = \mathbf{N}$.

Solution:

(i) Observe that the open sets containing any point $p \in \mathbf{N}$ are the sets E_i where $i \leq p$. If $n_0 \leq 36$, then every open set containing n_0 also contains $37 \in A$ which is different from n_0; hence $n_0 \leq 36$ is a limit point of A. On the other hand, if $n_0 > 36$ then the open set $E_{n_0} = \{n_0, n_0 + 1, n_0 + 2, \ldots\}$ contains no point of A different from n_0. So $n_0 > 36$ is not a limit point of A. Accordingly, the derived set of A is $A' = \{1, 2, 3, \ldots, 34, 35, 36\}$.

(ii) If E is an infinite subset of $\mathbf{N}$ then E is not bounded from above. So every open set containing any point $p \in \mathbf{N}$ will contain points of E other than p. Hence $E' = \mathbf{N}$.

On the other hand, if E is finite then E is bounded from above, say, by $n_0 \in \mathbf{N}$. Then the open set E_{n_0+1} contains no point of E. Hence $n_0 + 1 \in \mathbf{N}$ is not a limit point of E, and so $E' \neq \mathbf{N}$.

12. Let A be a subset of a topological space $(X, \mathcal{T})$. When will a point $p \in X$ not be a limit point of A?

Solution:

The point $p \in X$ is a limit point of A iff every open neighborhood of p contains a point of A other than p, i.e.,

$$p \in G \text{ and } G \in \mathcal{T} \quad \text{implies} \quad (G \setminus \{p\}) \cap A \neq \emptyset$$

So p is not a limit point of A if there exists an open set G such that

$$p \in G \quad \text{and} \quad (G \setminus \{p\}) \cap A = \emptyset$$

or, equivalently, $$p \in G \quad \text{and} \quad G \cap A = \emptyset \text{ or } G \cap A = \{p\}$$

or, equivalently, $$p \in G \quad \text{and} \quad G \cap A \subset \{p\}$$

13. Let A be any subset of a discrete topological space X. Show that the derived set A' of A is empty.

Solution:

Let p be any point in X. Recall that every subset of a discrete space is open. Hence, in particular, the singleton set $G = \{p\}$ is an open subset of X. But

$$p \in G \quad \text{and} \quad G \cap A = (\{p\} \cap A) \subset \{p\}$$

Hence, by the above problem, $p \notin A'$ for every $p \in X$, i.e. $A' = \emptyset$.

14. Consider the topology

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{a,b\}, \{a,c,d\}, \{a,b,c,d\}, \{a,b,e\}\}$$

on $X = \{a,b,c,d,e\}$. Determine the derived sets of (i) $A = \{c,d,e\}$ and (ii) $B = \{b\}$.

Solution:

(i) Note that $\{a,b\}$ and $\{a,b,e\}$ are open subsets of X and that

$$a, b \in \{a,b\} \quad \text{and} \quad \{a,b\} \cap A = \emptyset$$

$$e \in \{a,b,e\} \quad \text{and} \quad \{a,b,e\} \cap A = \{e\}$$

Hence a, b and e are not limit points of A. On the other hand, every other point in X is a limit point of A since every open set containing it also contains a point of A different from it. Accordingly, $A' = \{c,d\}$.

(ii) Note that $\{a\}$, $\{a,b\}$ and $\{a,c,d\}$ are open subsets of X and that

$$a \in \{a\} \quad \text{and} \quad \{a\} \cap B = \emptyset$$

$$b \in \{a,b\} \quad \text{and} \quad \{a,b\} \cap B = \{b\}$$

$$c, d \in \{a,c,d\} \quad \text{and} \quad \{a,c,d\} \cap B = \emptyset$$

Hence a, b, c and d are not limit points of $B = \{b\}$. But e is a limit point of B since the open sets containing e are $\{a,b,e\}$ and X and each contains the point $b \in B$ different from e. Thus $B' = \{e\}$.

15. Prove: If A is a subset of B, then every limit point of A is also a limit point of B, i.e., $A \subset B$ implies $A' \subset B'$.

Solution:

Recall that $p \in A'$ iff $(G \setminus \{p\}) \cap A \neq \emptyset$ for every open set G containing p. But $B \supset A$; hence

$$(G \setminus \{p\}) \cap B \supset (G \setminus \{p\}) \cap A \neq \emptyset$$

So $p \in A'$ implies $p \in B'$, i.e. $A' \subset B'$.

16. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be topologies on X such that $\mathcal{T}_1 \subset \mathcal{T}_2$, i.e. every $\mathcal{T}_1$-open subset of X is also a $\mathcal{T}_2$-open subset of X. Furthermore, let A be any subset of X.

(i) Show that every $\mathcal{T}_2$-limit point of A is also a $\mathcal{T}_1$-limit point of A.

(ii) Construct a space in which a $\mathcal{T}_1$-limit point is not a $\mathcal{T}_2$-limit point.

Solution:

(i) Let p be a $\mathcal{T}_2$-limit point of A; i.e. $(G \setminus \{p\}) \cap A \neq \emptyset$ for every $G \in \mathcal{T}_2$ such that $p \in G$. But $\mathcal{T}_1 \subset \mathcal{T}_2$; so, in particular, $(G \setminus \{p\}) \cap A \neq \emptyset$ for every $G \in \mathcal{T}_1$ such that $p \in G$, i.e. p is a $\mathcal{T}_1$-limit point of A.

(ii) Consider the usual topology $\mathcal{U}$ and the discrete topology $\mathcal{D}$ on $\mathbf{R}$. Note that $\mathcal{U} \subset \mathcal{D}$ since $\mathcal{D}$ contains every subset of $\mathbf{R}$. By Problem 13, 0 is not a $\mathcal{D}$-limit point of the set $A = \{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$ since A' is empty. But 0 is a limit point of A with respect to the usual topology on $\mathbf{R}$.

17. Prove: Let A and B be subsets of a topological space $(X, \mathcal{T})$. Then $(A \cup B)' = A' \cup B'$.

Solution:

Utilizing Problem 15, $\quad A \subset A \cup B$ implies $A' \subset (A \cup B)'$

$B \subset A \cup B$ implies $B' \subset (A \cup B)'$

So $A' \cup B' \subset (A \cup B)'$, and we need only show that

$$(A \cup B)' \subset A' \cup B'$$

Assume $p \notin A' \cup B'$; thus $\exists G, H \in \mathcal{T}$ such that

$$p \in G \text{ and } G \cap A \subset \{p\} \quad \text{and} \quad p \in H \text{ and } H \cap B \subset \{p\}$$

But $G \cap H \in \mathcal{T}$, $p \in G \cap H$ and

$$(G \cap H) \cap (A \cup B) = (G \cap H \cap A) \cup (G \cap H \cap B) \subset (G \cap A) \cup (H \cap B) \subset \{p\} \cup \{p\} = \{p\}$$

Thus $p \notin (A \cup B)'$, and so $(A \cup B)' \subset (A' \cup B')$.

CLOSED SETS, CLOSURE OPERATION, DENSE SETS

18. Consider the following topology on $X = \{a, b, c, d, e\}$:

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{a, b\}, \{a, c, d\}, \{a, b, c, d\}, \{a, b, e\}\}$$

(i) List the closed subsets of X.

(ii) Determine the closure of the sets $\{a\}$, $\{b\}$ and $\{c, e\}$.

(iii) Which sets in (ii) are dense in X?

Solution:

(i) A set is closed iff its complement is open. Hence write the complement of each set in $\mathcal{T}$:

$$\emptyset,\ X,\ \{b, c, d, e\},\ \{c, d, e\},\ \{b, e\},\ \{e\},\ \{c, d\}$$

(ii) The closure $\bar{A}$ of any set A is the intersection of all closed supersets of A. The only closed superset of $\{a\}$ is X; the closed supersets of $\{b\}$ are $\{b, e\}$, $\{b, c, d, e\}$ and X; and the closed supersets of $\{c, e\}$ are $\{c, d, e\}$, $\{b, c, d, e\}$ and X. Thus,

$$\overline{\{a\}} = X, \quad \overline{\{b\}} = \{b, e\}, \quad \overline{\{c, e\}} = \{c, d, e\}$$

(iii) A set A is dense in X iff $\bar{A} = X$; so $\{a\}$ is the only dense set.

19. Let $\mathcal{T}$ be the topology on $\mathbf{N}$ which consists of $\emptyset$ and all subsets of $\mathbf{N}$ of the form $E_n = \{n, n+1, n+2, \ldots\}$ where $n \in \mathbf{N}$ as in Problem 10.

(i) Determine the closed subsets of $(\mathbf{N}, \mathcal{T})$.

(ii) Determine the closure of the sets $\{7, 24, 47, 85\}$ and $\{3, 6, 9, 12, \ldots\}$.

(iii) Determine those subsets of $\mathbf{N}$ which are dense in $\mathbf{N}$.

Solution:

(i) A set is closed iff its complement is open. Hence the closed subsets of $\mathbf{N}$ are as follows:

$$\mathbf{N},\ \emptyset,\ \{1\},\ \{1, 2\},\ \{1, 2, 3\},\ \ldots,\ \{1, 2, \ldots, m\},\ \ldots$$

(ii) The closure of a set is the smallest closed superset. So

$$\overline{\{7,24,47,85\}} = \{1,2,\ldots,84,85\}, \qquad \overline{\{3,6,9,12,\ldots\}} = \{1,2,3,\ldots\} = \mathbf{N}$$

(iii) If a subset A of $\mathbf{N}$ is infinite, or equivalently unbounded, then $\bar{A} = \mathbf{N}$, i.e. A is dense in $\mathbf{N}$. If A is finite then its closure is not $\mathbf{N}$, i.e. A is not dense in $\mathbf{N}$.

20. Let $\mathcal{T}$ be the topology on $\mathbf{R}$ consisting of $\mathbf{R}$, $\emptyset$ and all open infinite intervals $E_a = (a, \infty)$ where $a \in \mathbf{R}$.

(i) Determine the closed subsets of $(\mathbf{R}, \mathcal{T})$.

(ii) Determine the closure of the sets $[3,7)$, $\{7,24,47,85\}$ and $\{3,6,9,12,\ldots\}$.

Solution:

(i) A set is closed iff its complement is open. Hence the closed subsets of $(\mathbf{R}, \mathcal{T})$ are $\emptyset$, $\mathbf{R}$ and all closed infinite intervals $E_a^c = (-\infty, a]$.

(ii) The closure of a set is the smallest closed superset. Hence

$$\overline{[3,7)} = (-\infty, 7], \quad \overline{\{7,24,47,85\}} = (-\infty, 85], \quad \overline{\{3,6,9,12,\ldots\}} = (-\infty,\infty) = \mathbf{R}$$

21. Let X be a discrete topological space. (i) Determine the closure of any subset A of X. (ii) Determine the dense subsets of X.

Solution:

(i) Recall that in a discrete space X any $A \subset X$ is closed; hence $\bar{A} = A$.

(ii) A is dense in X iff $\bar{A} = X$. But $\bar{A} = A$, so X is the only dense subset of X.

22. Let X be an indiscrete space. (i) Determine the closed subsets of X. (ii) Determine the closure of any subset A of X. (iii) Determine the dense subsets of X.

Solution:

(i) Recall that the only open subsets of an indiscrete space X are X and $\emptyset$; hence the closed subsets of X are also X and $\emptyset$.

(ii) If $A = \emptyset$, then $\bar{A} = \emptyset$. If $A \neq \emptyset$, then X is the only closed superset of A; so $\bar{A} = X$. That is, for any $A \subset X$,

$$\bar{A} = \begin{cases} \emptyset & \text{if } A = \emptyset \\ X & \text{if } A \neq \emptyset \end{cases}$$

(iii) $A \subset X$ is dense in X iff $\bar{A} = X$; hence every non-empty subset of X is dense in X.

23. Prove Theorem 5.4: A subset A of a topological space X is closed if and only if A contains each of its accumulation points, i.e. $A' \subset A$.

Solution:

Suppose A is closed, and let $p \notin A$, i.e. $p \in A^c$. But A^c, the complement of a closed set, is open; hence $p \notin A'$ for A^c is an open set such that

$$p \in A^c \quad \text{and} \quad A^c \cap A = \emptyset$$

Thus $A' \subset A$ if A is closed.

Now assume $A' \subset A$; we show that A^c is open. Let $p \in A^c$; then $p \notin A'$, so $\exists$ an open set G such that

$$p \in G \quad \text{and} \quad (G \setminus \{p\}) \cap A = \emptyset$$

But $p \notin A$; hence

$$G \cap A = (G \setminus \{p\}) \cap A = \emptyset$$

So $G \subset A^c$. Thus p is an interior point of A^c, and so A^c is open.

24. Prove: If F is a closed superset of any set A, then $A' \subset F$.

Solution:

By Problem 15, $A \subset F$ implies $A' \subset F'$. But $F' \subset F$, by Theorem 5.4, since F is closed. Thus $A' \subset F' \subset F$, which implies $A' \subset F$.

25. Prove: $A \cup A'$ is a closed set.

Solution:

Let $p \in (A \cup A')^c$. Since $p \notin A'$, $\exists$ an open set G such that

$$p \in G \quad \text{and} \quad G \cap A = \emptyset \text{ or } \{p\}$$

However, $p \notin A$; hence, in particular, $G \cap A = \emptyset$.

We also claim that $G \cap A' = \emptyset$. For if $g \in G$, then

$$g \in G \quad \text{and} \quad G \cap A = \emptyset$$

where G is an open set. So $g \notin A'$ and thus $G \cap A' = \emptyset$. Accordingly,

$$G \cap (A \cup A') = (G \cap A) \cup (G \cap A') = \emptyset \cup \emptyset = \emptyset$$

and so $G \subset (A \cup A')^c$. Thus p is an interior point of $(A \cup A')^c$ which is therefore an open set. Hence $A \cup A'$ is closed.

26. Prove Theorem 5.6: $\bar{A} = A \cup A'$.

Solution:

Since $A \subset \bar{A}$ and $\bar{A}$ is closed, $A' \subset (\bar{A})' \subset \bar{A}$ and hence $A \cup A' \subset \bar{A}$. But $A \cup A'$ is a closed set containing A, so $A \subset \bar{A} \subset A \cup A'$. Thus $\bar{A} = A \cup A'$.

27. Prove: If $A \subset B$ then $\bar{A} \subset \bar{B}$.

Solution:

If $A \subset B$, then by Problem 15, $A' \subset B'$. So $A \cup A' \subset B \cup B'$ or, by the preceding problem, $\bar{A} \subset \bar{B}$.

28. Prove: $\overline{A \cup B} = \bar{A} \cup \bar{B}$.

Solution:

Utilizing the preceding problem, $\bar{A} \subset \overline{A \cup B}$ and $\bar{B} \subset \overline{A \cup B}$; hence $(\bar{A} \cup \bar{B}) \subset \overline{A \cup B}$. But $(A \cup B) \subset (\bar{A} \cup \bar{B})$, a closed set since it is the union of two closed sets. Then (Proposition 5.5) $(A \cup B) \subset \overline{A \cup B} \subset (\bar{A} \cup \bar{B})$ and therefore $\overline{A \cup B} = \bar{A} \cup \bar{B}$.

29. Prove Proposition 5.7: (i) $\bar{\emptyset} = \emptyset$; (ii) $A \subset \bar{A}$; (iii) $\overline{A \cup B} = \bar{A} \cup \bar{B}$; and (iv) $(A^-)^- = A^-$.

Solution:

(i) and (iv): $\emptyset$ and $\bar{A}$ are closed; hence they are equal to their closures. (ii) $A \subset A \cup A' = \bar{A}$ (Problem 26). (iii) Preceding problem.

INTERIOR, EXTERIOR, BOUNDARY

30. Consider the following topology on $X = \{a, b, c, d, e\}$:

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{a, b\}, \{a, c, d\}, \{a, b, c, d\}, \{a, b, e\}\}$$

(i) Find the interior points of the subset $A = \{a, b, c\}$ of X. (ii) Find the exterior points of A. (iii) Find the boundary points of A.

Solution:

(i) The points a and b are interior points of A since

$$a, b \in \{a, b\} \subset A = \{a, b, c\}$$

where $\{a, b\}$ is an open set, i.e. since each belongs to an open set contained in A. Note that c is not an interior point of A since c does not belong to any open set contained in A. Hence $\text{int}(A) = \{a, b\}$ is the interior of A.

(ii) The complement of A is $A^c = \{d, e\}$. Neither d nor e are interior points of A^c since neither belongs to any open subset of $A^c = \{d, e\}$. Hence $\text{int}(A^c) = \emptyset$, i.e. there are no exterior points of A.

(iii) The boundary $\text{b}(A)$ of A consists of those points which are neither interior nor exterior to A. So $\text{b}(A) = \{c, d, e\}$.

31. Prove Proposition 5.8: The interior of a set A is the union of all open subsets of A. Furthermore: (i) A° is open; (ii) A° is the largest open subset of A, i.e. if G is an open subset of A then $G \subset A^\circ \subset A$; and (iii) A is open iff $A = A^\circ$.

Solution:

Let $\{G_i\}$ be the class of all open subsets of A. If $x \in A^\circ$, then x belongs to an open subset of A, i.e.,

$$\exists i_0 \quad \text{such that} \quad x \in G_{i_0}$$

Hence $x \in \cup_i G_i$ and so $A^\circ \subset \cup_i G_i$. On the other hand, if $y \in \cup_i G_i$, then $y \in G_{i_0}$ for some i_0. Thus $y \in A^\circ$, and $\cup_i G_i \subset A^\circ$. Accordingly, $A^\circ = \cup_i G_i$.

(i) $A^\circ = \cup_i G_i$ is open since it is the union of open sets.

(ii) If G is an open subset of A then $G \in \{G_i\}$; so $G \subset \cup_i G_i = A^\circ \subset A$.

(iii) If A is open then $A \subset A^\circ \subset A$ or $A = A^\circ$. If $A = A^\circ$ then A is open since A° is open.

32. Let A be a non-empty proper subset of an indiscrete space X. Find the interior, exterior and boundary of A.

Solution:

X and $\emptyset$ are the only open subsets of X. Since $X \neq A$, $\emptyset$ is the only open subset of A; hence $\text{int}(A) = \emptyset$. Similarly, $\text{int}(A^c) = \emptyset$, i.e. the exterior of A is empty. Thus $\text{b}(A) = X$.

33. Let $\mathcal{T}$ be the topology on $\mathbf{R}$ consisting of $\mathbf{R}$, $\emptyset$ and all open infinite intervals $E_a = (a, \infty)$ where $a \in \mathbf{R}$. Find the interior, exterior and boundary of the closed infinite interval $A = [7, \infty)$.

Solution:

Since the interior of A is the largest open subset of A, $\text{int}(A) = (7, \infty)$. Note that $A^c = (-\infty, 7)$ contains no open set except $\emptyset$; so $\text{int}(A^c) = \text{ext}(A) = \emptyset$. The boundary consists of those points which do not belong to $\text{int}(A)$ or $\text{ext}(A)$; hence $\text{b}(A) = (-\infty, 7]$.

34. Prove Theorem 5.9: $\bar{A} = \text{int}(A) \cup \text{b}(A)$

Solution:

Since $X = \text{int}(A) \cup \text{b}(A) \cup \text{ext}(A)$, $(\text{int}(A) \cup \text{b}(A))^c = \text{ext}(A)$ and it suffices to show $(\bar{A})^c = \text{ext}(A)$.

Let $p \in \text{ext}(A)$; then $\exists$ an open G such that

$$p \in G \subset A^c \quad \text{which implies} \quad G \cap A = \emptyset$$

So p is not a limit point of A, i.e. $p \notin A'$, and $p \notin A$. Hence $p \notin A' \cup A = \bar{A}$. In other words, $\text{ext}(A) \subset (\bar{A})^c$.

Now assume $p \in (\bar{A})^c = (A \cup A')^c$. Thus $p \notin A'$, so $\exists$ an open set G such that

$$p \in G \quad \text{and} \quad (G \setminus \{p\}) \cap A = \emptyset$$

But also $p \notin A$, so $G \cap A = \emptyset$ and $p \in G \subset A^c$. Thus $p \in \text{ext}(A)$, and $(\bar{A})^c = \text{ext}(A)$.

35. Show by a counterexample that the function f which assigns to each set its interior, i.e. $f(A) = \text{int}(A)$, does not commute with the function g which assigns to each set its closure, i.e. $g(A) = \bar{A}$.

Solution:

Consider $\mathbf{Q}$, the set of rational numbers, as a subset of $\mathbf{R}$ with the usual topology. Recall (Example 5.3) that the interior of $\mathbf{Q}$ is empty; hence

$$(g \circ f)(\mathbf{Q}) = g(f(\mathbf{Q})) = g(\text{int}(\mathbf{Q})) = g(\emptyset) = \bar{\emptyset} = \emptyset$$

On the other hand, $\bar{\mathbf{Q}} = \mathbf{R}$ and the interior of $\mathbf{R}$ is $\mathbf{R}$ itself. So

$$(f \circ g)(\mathbf{Q}) = f(g(\mathbf{Q})) = f(\bar{\mathbf{Q}}) = f(\mathbf{R}) = \mathbf{R}$$

Thus $g \circ f \neq f \circ g$, or f and g do not commute.

NEIGHBORHOODS, NEIGHBORHOOD SYSTEMS

36. Consider the following topology on $X = \{a, b, c, d, e\}$:

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{a, b\}, \{a, c, d\}, \{a, b, c, d\}, \{a, b, e\}\}$$

List the neighborhoods (i) of the point e, (ii) of the point c.

Solution:

(i) A neighborhood of e is any superset of an open set containing e. The open sets containing e are $\{a, b, e\}$ and X. The supersets of $\{a, b, e\}$ are $\{a, b, e\}$, $\{a, b, c, e\}$, $\{a, b, d, e\}$ and X; the only superset of X is X. Accordingly, the class of neighborhoods of e, i.e. the neighborhood system of e, is

$$\mathcal{N}_e = \{\{a, b, e\}, \{a, b, c, e\}, \{a, b, d, e\}, X\}$$

(ii) The open sets containing c are $\{a, c, d\}$, $\{a, b, c, d\}$ and X. Hence the neighborhood system of c is

$$\mathcal{N}_c = \{\{a, c, d\}, \{a, b, c, d\}, \{a, c, d, e\}, X\}$$

37. Determine the neighborhood system of a point p in an indiscrete space X.

Solution:

X and $\emptyset$ are the only open subsets of X; hence X is the only open set containing p. In addition, X is the only superset of X. Hence $\mathcal{N}_p = \{X\}$.

38. Prove: The intersection $N \cap M$ of any two neighborhoods N and M of a point p is also a neighborhood of p.

Solution:

N and M are neighborhoods of p, so $\exists$ open sets G, H such that

$$p \in G \subset N \quad \text{and} \quad p \in H \subset M$$

Hence $p \in G \cap H \subset N \cap M$, and $G \cap H$ is open, or $N \cap M$ is a neighborhood of p.

39. Prove: Any superset M of a neighborhood N of a point p is also a neighborhood of p.

Solution:

N is a neighborhood of p, so $\exists$ an open set G such that $p \in G \subset N$. By hypothesis, $N \subset M$, so

$$p \in G \subset N \subset M \quad \text{which implies} \quad p \in G \subset M$$

and hence M is a neighborhood of p.

40. Determine whether or not each of the following intervals is a neighborhood of 0 under the usual topology for the real line **R**. (i) $(-\frac{1}{2}, \frac{1}{2}]$, (ii) $(-1, 0]$, (iii) $[0, \frac{1}{2})$, (iv) $(0, 1]$.

Solution:

(i) Note that $0 \in (-\frac{1}{2}, \frac{1}{2}) \subset (-\frac{1}{2}, \frac{1}{2}]$ and $(-\frac{1}{2}, \frac{1}{2})$ is open; so $(-\frac{1}{2}, \frac{1}{2}]$ is a neighborhood of 0.

(ii) and (iii) Any $\mathcal{U}$-open set G containing 0 contains an open interval (a, b) containing 0, i.e. $a < 0 < b$; hence G contains points both greater and less than 0. So neither $(-1, 0]$ nor $[0, \frac{1}{2})$ is a neighborhood of 0.

(iv) The interval $(0, 1]$ does not even contain 0 and hence is not a neighborhood of 0.

41. Prove: A set G is open if and only if it is a neighborhood of each of its points.

Solution:

Suppose G is open; then each point $p \in G$ belongs to the open set G contained in G. Hence G is a neighborhood of each of its points.

Conversely, suppose G is a neighborhood of each of its points. So, for each point $p \in G$, $\exists$ an open set G_p such that $p \in G_p \subset G$. Hence $G = \cup\{G_p : p \in G\}$ and is open since it is a union of open sets.

42. Prove Proposition 5.10: Let $\mathcal{N}_p$ be the neighborhood system of a point p in a topological space X. Then:

(i) $\mathcal{N}_p$ is not empty and p belongs to each member of $\mathcal{N}_p$.

(ii) The intersection of any two members of $\mathcal{N}_p$ belongs to $\mathcal{N}_p$.

(iii) Every superset of a member of $\mathcal{N}_p$ belongs to $\mathcal{N}_p$.

(iv) Each member $N \in \mathcal{N}_p$ is a superset of a member $G \in \mathcal{N}_p$ where G is a neighborhood of each of its points.

Solution:

(i) If $N \in \mathcal{N}_p$, then $\exists$ an open set G such that $p \in G \subset N$; hence $p \in N$. Note $X \in \mathcal{N}_p$ since X is an open set containing p; so $\mathcal{N}_p \neq \emptyset$.

(ii) Proven in Problem 38. (iii) Proven in Problem 39.

(iv) If $N \in \mathcal{N}_p$, then N is a neighborhood of p, so $\exists$ an open set G such that $p \in G \subset N$. But by the preceding problem $G \in \mathcal{N}_p$ and G is a neighborhood of each of its points.

SUBSPACES, RELATIVE TOPOLOGIES

43. Consider the following topology on $X = \{a, b, c, d, e\}$:

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{a,b\}, \{a,c,d\}, \{a,b,c,d\}, \{a,b,e\}\}$$

List the members of the relative topology $\mathcal{T}_A$ on $A = \{a, c, e\}$.

Solution:

$\mathcal{T}_A = \{A \cap G : G \in \mathcal{T}\}$, so the members of $\mathcal{T}_A$ are:

$A \cap X = A$ $\quad$ $A \cap \{a\} = \{a\}$ $\quad$ $A \cap \{a,c,d\} = \{a,c\}$ $\quad$ $A \cap \{a,b,e\} = \{a,e\}$

$A \cap \emptyset = \emptyset$ $\quad$ $A \cap \{a,b\} = \{a\}$ $\quad$ $A \cap \{a,b,c,d\} = \{a,c\}$

In other words, $\mathcal{T}_A = \{A, \emptyset, \{a\}, \{a,c\}, \{a,e\}\}$. Observe that $\{a,c\}$ is not open in X, but is relatively open in A, i.e. is $\mathcal{T}_A$-open.

44. Consider the usual topology $\mathcal{U}$ on the real line $\mathbf{R}$. Describe the relative topology $\mathcal{U}_N$ on the set $\mathbf{N}$ of positive integers.

Solution:

Observe that, for each positive integer $n_0 \in \mathbf{N}$,

$$\{n_0\} = \mathbf{N} \cap (n_0 - \tfrac{1}{2}, n_0 + \tfrac{1}{2})$$

and $(n_0 - \frac{1}{2}, n_0 + \frac{1}{2})$ is a $\mathcal{U}$-open set; so every singleton subset $\{n_0\}$ of $\mathbf{N}$ is open relative to $\mathbf{N}$. Hence every subset of $\mathbf{N}$ is open relative to $\mathbf{N}$ since it is a union of singleton sets. In other words, $\mathcal{U}_N$ is the discrete topology on $\mathbf{N}$.

45. Let A be a $\mathcal{T}$-open subset of $(X, \mathcal{T})$ and let $A \subset Y \subset X$. Show that A is also open relative to the relative topology on Y, i.e. A is a $\mathcal{T}_Y$-open subset of Y.

Solution:

$\mathcal{T}_Y = \{Y \cap G : G \in \mathcal{T}\}$. But $A \subset Y$ and $A \in \mathcal{T}$; so $A = Y \cap A \in \mathcal{T}_Y$.

46. Consider the usual topology $\mathcal{U}$ on the real line $\mathbf{R}$. Determine whether or not each of the following subsets of $I = [0,1]$ are open relative to I, i.e. $\mathcal{T}_I$-open: (i) $(\frac{1}{2}, 1]$, (ii) $(\frac{1}{2}, \frac{2}{3})$, (iii) $(0, \frac{1}{2}]$.

Solution:

(i) Note that $(\frac{1}{2}, 1] = I \cap (\frac{1}{2}, 3)$ and $(\frac{1}{2}, 3)$ is open in $\mathbf{R}$; hence $(\frac{1}{2}, 1]$ is open relative to I.

(ii) Since $(\frac{1}{2}, \frac{2}{3})$ is open in $\mathbf{R}$, i.e. $(\frac{1}{2}, \frac{2}{3}) \in \mathcal{U}$, it is open relative I by the preceding problem. In fact, $(\frac{1}{2}, \frac{2}{3}) = I \cap (\frac{1}{2}, \frac{2}{3})$.

(iii) Since $(0, \frac{1}{2}]$ is not the intersection of I with any $\mathcal{U}$-open subset of $\mathbf{R}$, it is not $\mathcal{U}_I$-open.

47. Let A be a subset of a topological space $(X, \mathcal{T})$. Show that the relative topology $\mathcal{T}_A$ is well-defined. In other words, show that $\mathcal{T}_A = \{A \cap G : G \in \mathcal{T}\}$ is a topology on A.

Solution:

Since $\mathcal{T}$ is a topology, X and $\emptyset$ belong to $\mathcal{T}$. Hence $A \cap X = A$ and $A \cap \emptyset = \emptyset$ both belong to $\mathcal{T}_A$, which then satisfies $[\mathbf{O}_1]$.

Now let $\{H_i : i \in I\}$ be a subclass of $\mathcal{T}_A$. By definition of $\mathcal{T}_A$, for each $i \in I$ $\exists$ a $\mathcal{T}$-open set G_i such that $H_i = A \cap G_i$. By the distributive law of intersection over union,

$$\cup_i H_i = \cup_i (A \cap G_i) = A \cap (\cup_i G_i)$$

But $\cup_i G_i \in \mathcal{T}$ as it is the union of $\mathcal{T}$-open sets; hence $\cup_i H_i \in \mathcal{T}_A$. Thus $\mathcal{T}_A$ satisfies $[\mathbf{O}_2]$.

Now suppose $H_1, H_2 \in \mathcal{T}_A$. Then $\exists$ $G_1, G_2 \in \mathcal{T}$ such that $H_1 = A \cap G_1$ and $H_2 = A \cap G_2$. But $G_1 \cap G_2 \in \mathcal{T}$ since $\mathcal{T}$ is a topology. Hence

$$H_1 \cap H_2 = (A \cap G_1) \cap (A \cap G_2) = A \cap (G_1 \cap G_2)$$

belongs to $\mathcal{T}_A$. Accordingly, $\mathcal{T}_A$ satisfies $[\mathbf{O}_3]$ and is a topology on A.

48. Let $(X, \mathcal{T})$ be a subspace of $(Y, \mathcal{T}^*)$ and let $(Y, \mathcal{T}^*)$ be a subspace of $(Z, \mathcal{T}^{**})$. Show that $(X, \mathcal{T})$ is also a subspace of $(Z, \mathcal{T}^{**})$.

Solution:

Since $X \subset Y \subset Z$, $(X, \mathcal{T})$ is a subspace of $(Z, \mathcal{T}^{**})$ if and only if $\mathcal{T}^{**}_X = \mathcal{T}$. Let $G \in \mathcal{T}$; now $\mathcal{T}^*_X = \mathcal{T}$, so $\exists G^* \in \mathcal{T}^*_X$ for which $G = X \cap G^*$. But $\mathcal{T}^* = \mathcal{T}^{**}_Y$, so $\exists G^{**} \in \mathcal{T}^{**}$ such that $G^* = Y \cap G^{**}$. Thus

$$G = X \cap G^* = X \cap Y \cap G^{**} = X \cap G^{**}$$

since $X \subset Y$; so $G \in \mathcal{T}^{**}_X$. Accordingly, $\mathcal{T} \subset \mathcal{T}^{**}_X$.

Now assume $G \in \mathcal{T}^{**}_X$, i.e. $\exists H \in \mathcal{T}^{**}$ such that $G = X \cap H$. But $Y \cap H \in \mathcal{T}^{**}_Y = \mathcal{T}^*$ so $X \cap (Y \cap H) \in \mathcal{T}^*_X = \mathcal{T}$. Since

$$X \cap (Y \cap H) = X \cap H = G$$

we have $G \in \mathcal{T}$. Accordingly, $\mathcal{T}^{**}_X \subset \mathcal{T}$ and the theorem is proved.

MISCELLANEOUS PROBLEMS

49. Let $\mathcal{P}(X)$ be the power set, i.e. class of subsets, of a non-empty set X. Furthermore, let $k : \mathcal{P}(X) \to \mathcal{P}(X)$ be the identity mapping, i.e. for each $A \subset X$, $k(A) = A$.

(i) Verify that k satisfies the Kuratowski Closure Axioms of Theorem 5.12.

(ii) Determine the topology on X induced by k.

Solution:

(i) $k(\emptyset) = \emptyset$, so $[\mathbf{K}_1]$ is satisfied. $k(A \cup B) = A \cup B = k(A) \cup k(B)$, so $[\mathbf{K}_3]$ is satisfied.
$k(A) = A \supset A$, so $[\mathbf{K}_2]$ is satisfied. $k(k(A)) = k(A)$, so $[\mathbf{K}_4]$ is satisfied.

(ii) A subset $F \subset X$ is closed in the topology induced by k if and only if $k(F) = F$. But $k(A) = A$ for every $A \subset X$, so every set is closed and k induces the discrete topology.

50. Let $\mathcal{T}$ be the cofinite topology on the real line $\mathbf{R}$, and let $\langle a_1, a_2, \ldots \rangle$ be a sequence in $\mathbf{R}$ with distinct terms. Show that $\langle a_n \rangle$ converges to every real number $p \in \mathbf{R}$.

Solution:

Let G be any open set containing $p \in \mathbf{R}$. By definition of the cofinite topology, G^c is a finite set and hence can contain only a finite number of the terms of the sequence $\langle a_n \rangle$ since the terms are distinct. Thus G contains almost all of the terms of $\langle a_n \rangle$, and so $\langle a_n \rangle$ converges to p.

51. Let $\mathbf{T}$ be the collection of all topologies on a non-empty set X, partially ordered by class inclusion. Show that $\mathbf{T}$ is a complete lattice, i.e. if $\mathbf{S}$ is a non-empty subcollection of $\mathbf{T}$ then $\sup(\mathbf{S})$ and $\inf(\mathbf{S})$ exist.

Solution:

Let $\mathcal{T}_1 = \cap\{\mathcal{T} : \mathcal{T} \in \mathbf{S}\}$. By Theorem 5.1, $\mathcal{T}_1$ is a topology; so $\mathcal{T}_1 \in \mathbf{T}$ and $\mathcal{T}_1 = \inf(\mathbf{S})$.

Now let $\mathbf{B}$ be the collection of all upper bounds of $\mathbf{S}$. Observe that $\mathbf{B}$ is non-empty since, for example, the discrete topology $\mathcal{D}$ on X belongs to $\mathbf{B}$. Let $\mathcal{T}_2 = \cap\{\mathcal{T} : \mathcal{T} \in \mathbf{B}\}$. Again by Theorem 5.1, $\mathcal{T}_2$ is a topology on X and, furthermore, $\mathcal{T}_2 = \sup(\mathbf{S})$.

52. Let X be a non-empty set and, for each point $p \in X$, let $\mathcal{A}_p$ denote the class of subsets of X containing p.

(i) Verify that $\mathcal{A}_p$ satisfies the Neighborhood Axioms of Theorem 5.11.

(ii) Determine the induced topology on X.

Solution:

(i) Since $p \in X$, $X \in \mathcal{A}_p$ and, so, $\mathcal{A}_p \neq \emptyset$. By hypothesis, p belongs to each member of $\mathcal{A}_p$. Hence $[\mathbf{A}_1]$ is satisfied.

If $M, N \in \mathcal{A}_p$, then $p \in M$ and $p \in N$, and so $p \in M \cap N$. Hence $M \cap N \in \mathcal{A}_p$ and so $[\mathbf{A}_2]$ is satisfied.

If $N \in \mathcal{A}_p$ and $N \subset M$, i.e. if $p \in N \subset M$, then $p \in M$. Hence $M \in \mathcal{A}_p$ and so $[\mathbf{A}_3]$ is satisfied.

By definition of $\mathcal{A}_p$, every $A \subset X$ has the property that $A \in \mathcal{A}_p$ for every $p \in A$. Hence $[\mathbf{A}_4]$ is satisfied.

(ii) A subset $A \subset X$ is open in the induced topology if and only if $A \in \mathcal{A}_p$ for every $p \in A$. Since every subset of X has this property, the induced topology on X is the discrete topology.

Supplementary Problems

TOPOLOGICAL SPACES

53. List all possible topologies on the set $X = \{a, b\}$.

54. Prove Theorem 5.1: Let $\{\mathcal{T}_i : i \in I\}$ be any collection of topologies on a set X. Then the intersection $\cap_i \mathcal{T}_i$ is also a topology on X.

55. Let X be an infinite set and let $\mathcal{T}$ be a topology on X in which all infinite subsets of X are open. Show that $\mathcal{T}$ is the discrete topology on X.

56. Let X be an infinite set and let $\mathcal{T}$ consist of $\emptyset$ and all subsets of X whose complements are countable.

(i) Prove that $(X, \mathcal{T})$ is a topological space.

(ii) If X is countable, describe the topology determined by $\mathcal{T}$.

57. Let $\mathcal{T} = \{\mathbf{R}^2, \emptyset\} \cup \{G_k : k \in \mathbf{R}\}$ be the class of subsets of the plane $\mathbf{R}^2$ where

$$G_k = \{\langle x, y\rangle : x, y \in \mathbf{R},\ x > y + k\}$$

(i) Prove that $\mathcal{T}$ is a topology on $\mathbf{R}^2$.

(ii) Is $\mathcal{T}$ a topology on $\mathbf{R}^2$ if "$k \in \mathbf{R}$" is replaced by "$k \in \mathbf{N}$"? by "$k \in \mathbf{Q}$"?

58. Prove that $(\mathbf{R}^2, \mathcal{T})$ is a topological space where the elements of $\mathcal{T}$ are $\emptyset$ and the complements of finite sets of lines and points.

59. Let $\{p\}$ be an arbitrary singleton set such that $p \notin \mathbf{R}$; e.g. $\{\mathbf{R}\}$. Furthermore, let $\mathbf{R}^* = \mathbf{R} \cup \{p\}$ and let $\mathcal{T}$ be the class of subsets of $\mathbf{R}^*$ consisting of all $\mathcal{U}$-open subsets of $\mathbf{R}$ and the complements (relative to $\mathbf{R}^*$) of all bounded $\mathcal{U}$-closed subsets of $\mathbf{R}$. Prove that $\mathcal{T}$ is a topology on $\mathbf{R}^*$.

60. Let $\{p\}$ be an arbitrary singleton set such that $p \notin \mathbf{R}$; and let $\mathbf{R}^* = \mathbf{R} \cup \{p\}$. Furthermore, let $\mathcal{T}$ be the class of subsets of $\mathbf{R}^*$ consisting of all subsets of $\mathbf{R}$ and the complements (relative to $\mathbf{R}^*$) of all finite subsets of $\mathbf{R}$. Prove that $\mathcal{T}$ is a topology on $\mathbf{R}^*$.

ACCUMULATION POINTS, DERIVED SETS

61. Prove: $A' \cup B' = (A \cup B)'$.

62. Prove: If p is a limit point of the set A, then p is also a limit point of $A \setminus \{p\}$.

63. Prove: Let X be a cofinite topological space. Then A' is closed for any subset A of X.

64. Consider the topological space $(\mathbf{R}, \mathcal{T})$ where $\mathcal{T}$ consists of $\mathbf{R}$, $\emptyset$ and all open infinite intervals $E_a = (a, \infty)$, $a \in \mathbf{R}$. Find the derived set of: (i) the interval $[4, 10)$; (ii) $\mathbf{Z}$, the set of integers.

65. Let $\mathcal{T}$ be the topology on $\mathbf{R}^* = \mathbf{R} \cup \{p\}$ defined in Problem 59.
   (i) Determine the accumulation points of the following sets:
   (1) open interval (a, b), $a, b \in \mathbf{R}$ (2) infinite open interval (a, ∞), $a \in \mathbf{R}$ (3) $\mathbf{R}$.
   (ii) Determine those subsets of $\mathbf{R}^*$ which have p as a limit point.

66. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be topologies on a set X with $\mathcal{T}_1$ coarser than $\mathcal{T}_2$, i.e. $\mathcal{T}_1 \subset \mathcal{T}_2$.
   (i) Show that every $\mathcal{T}_2$-accumulation point of a subset A of X is also a $\mathcal{T}_1$-accumulation point.
   (ii) Construct an example in which the converse of (i) does not hold.

CLOSED SETS, CLOSURE OF A SET, DENSE SUBSETS

67. Construct a non-discrete topological space in which the closed sets are identical to the open sets.

68. Prove: $\overline{A \cap B} \subset \bar{A} \cap \bar{B}$. Construct an example in which equality does not hold.

69. Prove: $\bar{A} \setminus \bar{B} \subset \overline{(A \setminus B)}$. Construct an example in which equality does not hold.

70. Prove: If A is open, then $A \cap \bar{B} \subset \overline{A \cap B}$.

71. Prove: Let A be a dense subset of $(X, \mathcal{T})$, and let B be a non-empty open subset of X. Then $A \cap B \neq \emptyset$.

72. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be topologies on X with $\mathcal{T}_1$ coarser than $\mathcal{T}_2$. Show that the $\mathcal{T}_2$-closure of any subset A of X is contained in the $\mathcal{T}_1$-closure of A.

73. Show that every non-finite subset of an infinite cofinite space X is dense in X.

74. Show that every non-empty open subset of an indiscrete space X is dense in X.

INTERIOR, EXTERIOR, BOUNDARY

75. Let X be a discrete space and let $A \subset X$. Find (i) $\text{int}(A)$, (ii) $\text{ext}(A)$, and (iii) $\text{b}(A)$.

76. Prove: (i) $\text{b}(A) \subset A$ if and only if A is closed.
   (ii) $\text{b}(A) \cap A = \emptyset$ if and only if A is open.
   (iii) $\text{b}(A) = \emptyset$ if and only if A is both open and closed.

77. Prove: If $\bar{A} \cap \bar{B} = \emptyset$, then $\text{b}(A \cup B) = \text{b}(A) \cup \text{b}(B)$.

78. Prove: (i) $A^\circ \cap B^\circ = (A \cap B)^\circ$; (ii) $A^\circ \cup B^\circ \subset (A \cup B)^\circ$. Construct an example in which equality in (ii) does not hold.

79. Prove: $\text{b}(A^\circ) \subset \text{b}(A)$. Construct an example in which equality does not hold.

80. Show that $\text{int}(A) \cup \text{ext}(A)$ need not be dense in a space X. (It is true if $X = \mathbf{R}$.)

81. Prove: Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be topologies on X with $\mathcal{T}_1$ coarser than $\mathcal{T}_2$, i.e. $\mathcal{T}_1 \subset \mathcal{T}_2$, and let $A \subset X$. Then:
   (i) The $\mathcal{T}_1$-interior of A is a subset of the $\mathcal{T}_2$-interior of A.
   (ii) The $\mathcal{T}_2$-boundary of A is a subset of the $\mathcal{T}_1$-boundary of A.

NEIGHBORHOODS, NEIGHBORHOOD SYSTEMS

82. Let X be a cofinite topological space. Show that every neighborhood of a point $p \in X$ is an open set.

83. Let X be an indiscrete space. Determine the neighborhood system $\mathcal{N}_p$ of any point $p \in X$.

84. Show that if $\mathcal{N}_p$ is finite, then $\bigcap\{N : N \in \mathcal{N}_p\}$ belongs to $\mathcal{N}_p$.

SUBSPACES, RELATIVE TOPOLOGIES

85. Show that every subspace of a discrete space is also discrete.

86. Show that every subspace of an indiscrete space is indiscrete.

87. Let $(Y, \mathcal{T}_Y)$ be a subspace of $(X, \mathcal{T})$. Show that $E \subset Y$ is $\mathcal{T}_Y$-closed if and only if $E = Y \cap F$, where F is a $\mathcal{T}$-closed subset of X.

88. Let $(A, \mathcal{T}_A)$ be a subspace of $(X, \mathcal{T})$. Prove that $\mathcal{T}_A$ consists of the members of $\mathcal{T}$ contained in A, i.e. $\mathcal{T}_A = \{G : G \subset A, G \in \mathcal{T}\}$, if and only if A is a $\mathcal{T}$-open subset of X.

89. Let $(Y, \mathcal{T}_Y)$ be a subspace of $(X, \mathcal{T})$. For any subset A of Y, let $\bar{A}$ and A° be the closure and interior of A with respect to $\mathcal{T}$ and let $(\bar{A})_Y$ and $(A^\circ)_Y$ be the closure and interior of A with respect to $\mathcal{T}_Y$. Prove (i) $(\bar{A})_Y = \bar{A} \cap Y$, (ii) $A^\circ = (A^\circ)_Y \cap Y^\circ$.

90. Let A, B and C be subsets of a topological space X with $C \subset A \cup B$. If A, B and $A \cup B$ are given the relative topologies, prove that C is open with respect to $A \cup B$ if and only if $C \cap A$ is open with respect to A and $C \cap B$ is open with respect to B.

EQUIVALENT DEFINITIONS OF TOPOLOGIES

91. Prove Theorem 5.11: Let X be a non-empty set and let there be assigned to each point $p \in X$ a class $\mathcal{A}_p$ of subsets of X satisfying the following axioms:

[A₁] $\mathcal{A}_p$ is not empty and p belongs to each member of $\mathcal{A}_p$.
[A₂] The intersection of any two members of $\mathcal{A}_p$ belongs to $\mathcal{A}_p$.
[A₃] Every superset of a member of $\mathcal{A}_p$ belongs to $\mathcal{A}_p$.
[A₄] Each member $N \in \mathcal{A}_p$ is a superset of a member $G \in \mathcal{A}_p$ such that $G \in \mathcal{A}_g$ for every $g \in G$.

Then there exists one and only one topology $\mathcal{T}$ on X such that $\mathcal{A}_p$ is the $\mathcal{T}$-neighborhood system of the point $p \in X$.

92. Prove Theorem 5.12: Let X be a non-empty set and let $k : \mathcal{P}(X) \to \mathcal{P}(X)$ satisfy the following Kuratowski Closure Axioms:

$$[\mathbf{K}_1]\ k(\emptyset) = \emptyset, \quad [\mathbf{K}_2]\ A \subset k(A), \quad [\mathbf{K}_3]\ k(A \cup B) = k(A) \cup k(B), \quad [\mathbf{K}_4]\ k(k(A)) = k(A)$$

Then there exists one and only one topology $\mathcal{T}$ on X such that $k(A)$ will be the $\mathcal{T}$-closure of $A \subset X$.

93. Prove: Let X be a non-empty set and let $i : \mathcal{P}(X) \to \mathcal{P}(X)$ satisfy the following properties:

$$\text{(i) } i(X) = X, \quad \text{(ii) } i(A) \subset A, \quad \text{(iii) } i(A \cup B) = i(A) \cup i(B), \quad \text{(iv) } i(i(A)) = i(A)$$

Then there exists one and only one topology $\mathcal{T}$ on X such that $i(A)$ will be the $\mathcal{T}$-interior of $A \subset X$.

94. Prove: Let X be a non-empty set and let $\mathcal{F}$ be a class of subsets of X satisfying the following properties:

(i) X and $\emptyset$ belong to $\mathcal{F}$.
(ii) The intersection of any number of members of $\mathcal{F}$ belongs to $\mathcal{F}$.
(iii) The union of any two members of $\mathcal{F}$ belongs to $\mathcal{F}$.

Then there exists one and only one topology $\mathcal{T}$ on X such that the members of $\mathcal{F}$ are precisely the $\mathcal{T}$-closed subsets of X.

95. Let a neighborhood of a real number $p \in \mathbf{R}$ be any set containing p and containing all the rational numbers of some open interval (a, b) where $a < p < b$.

(i) Show that these neighborhoods actually satisfy the neighborhood axioms and hence define a topology on the real line $\mathbf{R}$.
(ii) Show that any set of irrational numbers does not contain any accumulation points.
(iii) Show that any sequence of irrational numbers, such as $\langle \pi/2, \pi/3, \pi/4, \ldots \rangle$, does not converge.

Answers to Supplementary Problems

53. $\{X, \emptyset\}$, $\{X, \{a\}, \emptyset\}$, $\{X, \{b\}, \emptyset\}$ and $\{X, \{a\}, \{b\}, \emptyset\}$.

56. (ii) Discrete topology.

64. (i) $(-\infty, 10]$ (ii) $\mathbf{R}$

65. (i): (1) $[a, b]$, (2) $[a, \infty) \cup \{p\}$, (3) $\mathbf{R}^*$. (ii) Unbounded subsets of $\mathbf{R}$.

67. $X = \{a, b, c\}$, $\mathcal{T} = \{X, \emptyset, \{a, b\}, \{c\}\}$

75. (i) A, (ii) A^c, (iii) $\emptyset$

80. Let $X = \{a, b\}$ be an indiscrete space and let $A = \{a\}$.

Chapter 7

Continuity and Topological Equivalence

CONTINUOUS FUNCTIONS

Let $(X, \mathcal{T})$ and $(Y, \mathcal{T}^*)$ be topological spaces. A function f from X into Y is *continuous relative to* $\mathcal{T}$ and $\mathcal{T}^*$, or $\mathcal{T}$-$\mathcal{T}^*$ *continuous*, or simply *continuous*, iff the inverse image $f^{-1}[H]$ of every $\mathcal{T}^*$-open subset H of Y is a $\mathcal{T}$-open subset of X, that is, iff

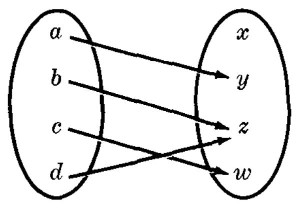
$$H \in \mathcal{T}^* \quad \text{implies} \quad f^{-1}[H] \in \mathcal{T}$$

We shall write $f: (X, \mathcal{T}) \to (Y, \mathcal{T}^*)$ for a function from X into Y when it is convenient to indicate the topologies involved.

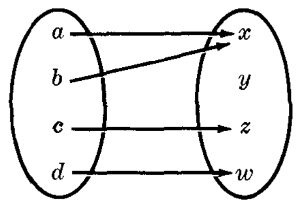
Example 1.1: Consider the following topologies on $X = \{a, b, c, d\}$ and $Y = \{x, y, z, w\}$ respectively:

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{a, b\}, \{a, b, c\}\}, \quad \mathcal{T}^* = \{Y, \emptyset, \{x\}, \{y\}, \{x, y\}, \{y, z, w\}\}$$

Also consider the functions $f: X \to Y$ and $g: X \to Y$ defined by the diagrams below:



f



g

The function f is continuous since the inverse of each member of the topology $\mathcal{T}^*$ on Y is a member of the topology $\mathcal{T}$ on X. The function g is not continuous since $\{y, z, w\} \in \mathcal{T}^*$, i.e. is an open subset of Y, but its inverse image $g^{-1}[\{y, z, w\}] = \{c, d\}$ is not an open subset of X, i.e. does not belong to $\mathcal{T}$.

Example 1.2: Consider any discrete space $(X, \mathcal{D})$ and any topological space $(Y, \mathcal{T})$. Then every function $f: X \to Y$ is $\mathcal{D}$-$\mathcal{T}$ continuous. For if H is any open subset of Y, its inverse $f^{-1}[H]$ is an open subset of X since every subset of a discrete space is open.

Example 1.3: Let $f: X \to Y$ where X and Y are topological spaces, and let $\mathcal{B}$ be a base for the topology on Y. Suppose for each member $B \in \mathcal{B}$, $f^{-1}[B]$ is an open subset of X; then f is a continuous function. For let H be an open subset of Y; then $H = \cup_i B_i$, a union of members of $\mathcal{B}$. But

$$f^{-1}[H] = f^{-1}[\cup_i B_i] = \cup_i f^{-1}[B_i]$$

and each $f^{-1}[B_i]$ is open by hypothesis; hence $f^{-1}[H]$ is the union of open sets and is therefore open. Accordingly, f is continuous.

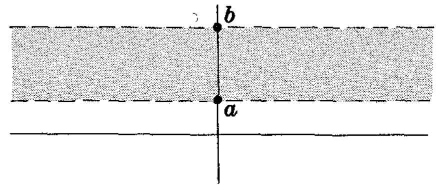
We formally state the result of the preceding example.

Proposition 7.1: A function $f: X \to Y$ is continuous iff the inverse of each member of a base $\mathcal{B}$ for Y is an open subset of X.

This proposition can in fact be strengthened as follows:

Theorem 7.2: Let $\mathcal{S}$ be a subbase for a topological space Y. Then a function $f: X \to Y$ is continuous iff the inverse of each member of $\mathcal{S}$ is an open subset of X.

Example 1.4: The projection mappings from the plane $\mathbf{R}^2$ into the line $\mathbf{R}$ are both continuous relative to the usual topologies. Consider, for example, the projection $\pi: \mathbf{R}^2 \to \mathbf{R}$ defined by $\pi(\langle x, y\rangle) = y$. Then the inverse of any open interval (a, b) is an infinite open strip as illustrated below:



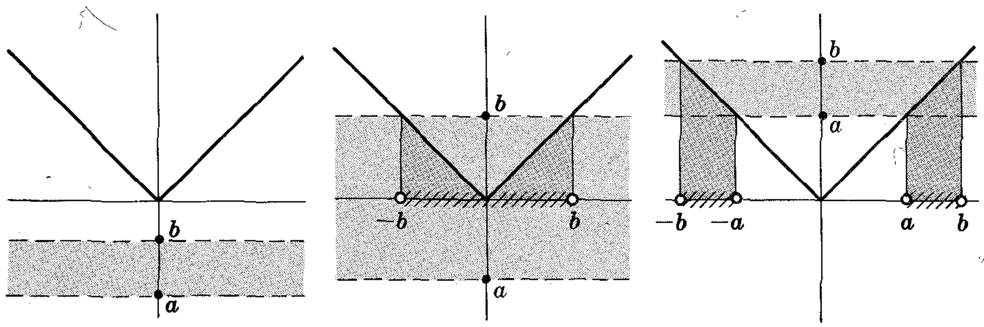
$\pi^{-1}[(a, b)]$ is shaded

Hence by Proposition 7.1, the inverse of every open subset of $\mathbf{R}$ is open in $\mathbf{R}^2$, i.e. π is continuous.

Example 1.5: The absolute value function f on $\mathbf{R}$, i.e. $f(x) = |x|$ for $x \in \mathbf{R}$, is continuous. For if $A = (a, b)$ is an open interval in $\mathbf{R}$, then

$$f^{-1}[A] = \begin{cases} \emptyset & \text{if } a < b \leq 0 \\ (-b, b) & \text{if } a < 0 < b \\ (-b, -a) \cup (a, b) & \text{if } 0 \leq a < b \end{cases}$$

as illustrated below. In each case $f^{-1}[A]$ is open; hence f is continuous.



$f^{-1}[A] = \emptyset$ $\qquad$ $f^{-1}[A] = (-b, b)$ $\qquad$ $f^{-1}[A] = (-b, -a) \cup (a, b)$

Continuous functions can be characterized by their behavior with respect to closed sets, as follows:

Theorem 7.3: A function $f: X \to Y$ is continuous if and only if the inverse image of every closed subset of Y is a closed subset of X.

CONTINUOUS FUNCTIONS AND ARBITRARY CLOSENESS

Let X be a topological space. A point $p \in X$ is said to be *arbitrarily close* to a set $A \subset X$ if

either (i) $p \in A$ or (ii) p is an accumulation point of A

Recall that $\bar{A} = A \cup A'$; so the closure of A consists precisely of those points in X which are arbitrarily close to A. Recall also that $\bar{A} = A^\circ \cup \mathrm{b}(A)$; hence p is arbitrarily close to A if p is either an interior or a boundary point of A.

Continuous functions can also be characterized as those functions which *preserve arbitrary closeness,* namely,

Theorem 7.4: A function $f: X \to Y$ is continuous if and only if, for any $p \in X$ and any $A \subset X$,

$$p \text{ arbitrarily close to } A \quad \Rightarrow \quad f(p) \text{ arbitrarily close to } f[A]$$

or

$$p \in \bar{A} \quad \Rightarrow \quad f(p) \in \overline{f[A]}$$

or

$$f[\bar{A}] \subset \overline{f[A]}$$

CONTINUITY AT A POINT

Continuity as we have defined it is a *global* property, that is, it restricts the way in which a function behaves on the entire set X. There also exists a corresponding local concept of *continuity at a point.*

A function $f: X \to Y$ is continuous at $p \in X$ iff the inverse image $f^{-1}[H]$ of every open set $H \subset Y$ containing $f(p)$ is a superset of an open set $G \subset X$ containing p or, equivalently, iff the inverse image of every neighborhood of $f(p)$ is a neighborhood of p, i.e.,

$$N \in \mathcal{N}_{f(p)} \quad \Rightarrow \quad f^{-1}[N] \in \mathcal{N}_p$$

Notice that, with respect to the usual topology on the real line $\mathbf{R}$, this definition coincides with the $\epsilon - \delta$ definition of continuity at a point for functions $f: \mathbf{R} \to \mathbf{R}$. In fact, the relationship between local and global continuity for functions $f: \mathbf{R} \to \mathbf{R}$ holds true in general; namely,

Theorem 7.5: Let X and Y be topological spaces. Then a function $f: X \to Y$ is continuous if and only if it is continuous at every point of X.

SEQUENTIAL CONTINUITY AT A POINT

A function $f: X \to Y$ is *sequentially continuous* at a point $p \in X$ iff for every sequence $\langle a_n \rangle$ in X converging to p, the sequence $\langle f(a_n) \rangle$ in Y converges to $f(p)$, i.e.,

$$a_n \to p \quad \text{implies} \quad f(a_n) \to f(p)$$

Sequential continuity and continuity at a point are related as follows:

Proposition 7.6: If a function $f: X \to Y$ is continuous at $p \in X$, then it is sequentially continuous at p.

Remark: The converse of the previous proposition is not true. Consider, for example, the topology $\mathcal{T}$ on the real line $\mathbf{R}$ consisting of $\emptyset$ and the complements of countable sets. Recall (see Example 7.3 of Chapter 5) that a sequence $\langle a_n \rangle$ converges to p if and only if it has the form

$$\langle a_1, a_2, \ldots, a_{n_0}, p, p, p, \ldots \rangle$$

Then for any function $f: (\mathbf{R}, \mathcal{T}) \to (X, \mathcal{T}^*)$,

$$\langle f(a_n) \rangle = \langle f(a_1), \ldots, f(a_{n_0}), f(p), f(p), f(p), \ldots \rangle$$

converges to $f(p)$. In other words, every function on $(\mathbf{R}, \mathcal{T})$ is sequentially continuous. On the other hand, the function $f(\mathbf{R}, \mathcal{T}) \to (\mathbf{R}, \mathcal{U})$ defined by $f(x) = x$, i.e. the identity function, is not $\mathcal{T}$-$\mathcal{U}$ continuous since $f^{-1}[(0,1)] = (0,1)$ is not a $\mathcal{T}$-open subset of $\mathbf{R}$.

OPEN AND CLOSED FUNCTIONS

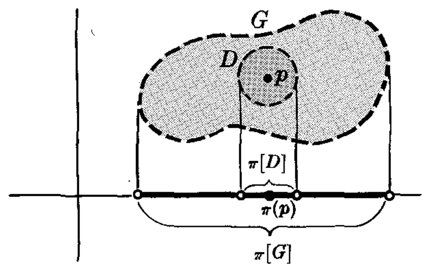
A continuous function has the property that the *inverse* image of every open set is open and the *inverse* image of every closed set is closed. It is natural then to ask about the following types of functions:

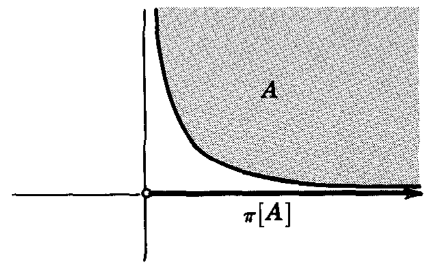
(1) A function $f: X \to Y$ is called an *open* (or *interior*) *function* if the image of every open set is open.

(2) A function $g: X \to Y$ is called a *closed function* if the image of every closed set is closed.

In general, functions which are open need not be closed and vice versa. In fact, the function in our first example is open and continuous but not closed.

Example 2.1: Consider the projection mapping $\pi : \mathbf{R}^2 \to \mathbf{R}$ of the plane $\mathbf{R}^2$ into the x-axis, i.e. $\pi(\langle x, y\rangle) = x$. Observe that the projection $\pi[D]$ of any open disc $D \subset \mathbf{R}^2$ is an open interval. Hence any point $\pi(p)$ in the image $\pi[G]$ of an open set $G \subset \mathbf{R}^2$ belongs to an open interval contained in $\pi[G]$, or $\pi[G]$ is open. Accordingly, π is an open function. On the other hand, π is not a closed function, for the set $A = \{\langle x, y\rangle : xy \geqq 1,\ x > 0\}$ is closed, but its projection $\pi[A] = (0, \infty)$ is not closed. (See diagrams below.)





HOMEOMORPHIC SPACES

A topological space $(X, \mathcal{T})$ is, as we have seen, a set X together with a distinguished class $\mathcal{T}$ of subsets of X, satisfying certain axioms. Between any two such spaces $(X, \mathcal{T})$ and $(Y, \mathcal{T}^*)$ there are many functions $f : X \to Y$. We choose to discuss continuous, or open, or closed functions rather than arbitrary functions since it is these functions which preserve some aspect of the structure of the spaces $(X, \mathcal{T})$ and $(Y, \mathcal{T}^*)$.

Now suppose there is some bijective (i.e. one-one and onto) mapping $f : X \to Y$. Then f induces a bijective function $f : \mathcal{P}(X) \to \mathcal{P}(Y)$ from the power set of X, i.e. the class of subsets of X, into the power set of Y. If this induced function also takes $\mathcal{T}$ onto $\mathcal{T}^*$, i.e. defines a one-to-one correspondence between the open sets in X and the open sets in Y, then the spaces $(X, \mathcal{T})$ and $(Y, \mathcal{T}^*)$ are identical from the topological point of view. Specifically:

Definition: Two topological spaces X and Y are called *homeomorphic* or *topologically equivalent* if there exists a bijective (i.e. one-one, onto) function $f : X \to Y$ such that f and f^{-1} are continuous. The function f is called a *homeomorphism.*

A function f is called *bicontinuous* or *topological* if f is open and continuous. Thus $f : X \to Y$ is a homeomorphism iff f is bicontinuous and bijective.

Example 3.1: Let $X = (-1, 1)$. The function $f : X \to \mathbf{R}$ defined by $f(x) = \tan \frac{1}{2}\pi x$ is one-one, onto and continuous. Furthermore, the inverse function f^{-1} is also continuous. Hence the real line $\mathbf{R}$ and the open interval $(-1, 1)$ are homeomorphic.

Example 3.2: Let X and Y be discrete spaces. Then, as seen in Example 1.2, all functions from one to the other are continuous. Hence X and Y are homeomorphic iff there exists a one-one, onto function from one to the other, i.e. iff they are cardinally equivalent.

Proposition 7.7: The relation in any collection of topological spaces defined by "X is homeomorphic to Y" is an equivalence relation.

Thus, by the Fundamental Theorem on Equivalence Relations, any collection of topological spaces can be partitioned into classes of topologically equivalent spaces.

TOPOLOGICAL PROPERTIES

A property P of sets is called *topological* or a *topological invariant* if whenever a topological space $(X, \mathcal{T})$ has P then every space homeomorphic to $(X, \mathcal{T})$ also has P.

Example 4.1: As seen in Example 3.1, the real line $\mathbf{R}$ is homeomorphic to the open interval $X = (-1, 1)$. Hence *length* is not a topological property since X and $\mathbf{R}$ have different lengths, and *boundedness* is not a topological property since X is bounded but $\mathbf{R}$ is not.

Example 4.2: Let X be the set of positive real numbers, i.e. $X = (0, \infty)$. The function $f: X \to X$ defined by $f(x) = 1/x$ is a homeomorphism from X onto X. Observe that the sequence

$$\langle a_n \rangle = \langle 1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots \rangle$$

corresponds, under the homeomorphism, to the sequence

$$\langle f(a_n) \rangle = \langle 1, 2, 3, \ldots \rangle$$

The sequence $\langle a_n \rangle$ is a Cauchy sequence; the sequence $\langle f(a_n) \rangle$ is not. Hence the property of being a Cauchy sequence is not topological.

Most of topology is an investigation of the consequences of certain topological properties as *compactness* and *connectedness*. In fact, formally topology is the study of topological invariants. In the next example, connectedness is defined and is shown to be a topological property.

Example 4.3: A topological space $(X, \mathcal{T})$ is *disconnected* iff X is the union of two open, non-empty, disjoint subsets, i.e.

$$X = G \cup H \quad \text{where} \quad G, H \in \mathcal{T}, \; G \cap H = \emptyset \quad \text{but} \quad G, H \neq \emptyset$$

If $f: X \to Y$ is a homeomorphism then $X = G \cup H$ if and only if $Y = f[G] \cup f[H]$ and so Y is disconnected if and only if X is.

The space $(X, \mathcal{T})$ is *connected* iff it is not disconnected.

TOPOLOGIES INDUCED BY FUNCTIONS

Let $\{(Y_i, \mathcal{T}_i)\}$ be any collection of topological spaces and for each Y_i let there be given a function $f_i: X \to Y_i$ defined on some arbitrary non-empty set X. We want to investigate those topologies on X with respect to which all the functions f_i are continuous. Recall that f_i is continuous relative to some topology on X provided the inverse image of each open subset of Y_i is an open subset of X. Thus we consider the following class of subsets of X:

$$\mathcal{S} = \bigcup_i \{f_i^{-1}[H] : H \in \mathcal{T}_i\}$$

That is, $\mathcal{S}$ consists of the inverse image of each open subset of every space Y_i. The topology $\mathcal{T}$ on X generated by $\mathcal{S}$ is called the topology *induced* (or *generated*) by the functions f_i. The main properties of $\mathcal{T}$ are listed in the next theorem.

Theorem 7.8: (i) All the functions f_i are continuous relative to $\mathcal{T}$.

(ii) $\mathcal{T}$ is the intersection of all the topologies on X with respect to which the functions f_i are continuous.

(iii) $\mathcal{T}$ is the smallest, i.e. coarsest, topology on X with respect to which the functions f_i are continuous.

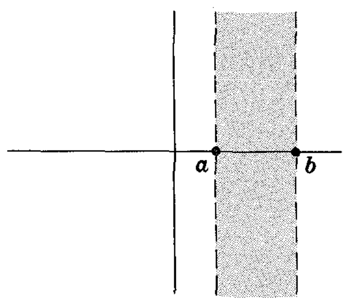
(iv) $\mathcal{S}$ is a subbase for the topology $\mathcal{T}$.

We shall call $\mathcal{S}$ the *defining subbase* for the topology induced by the functions f_i, i.e. the coarsest topology on X with respect to which the functions f_i are continuous.

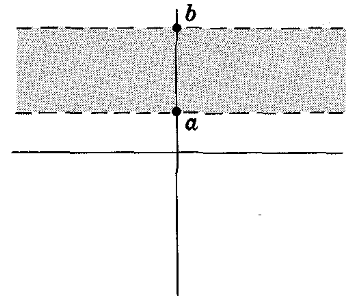
Example 5.1: Let π_1 and π_2 be the projections of the plane $\mathbf{R}^2$ into $\mathbf{R}$, i.e.,

$$\pi_1(\langle x, y \rangle) = x \quad \text{and} \quad \pi_2(\langle x, y \rangle) = y$$

Observe, as illustrated below, that the inverse image of an open interval (a, b) in $\mathbf{R}$ is an infinite open strip in $\mathbf{R}^2$.



$\pi_1^{-1}[(a,b)]$



$\pi_2^{-1}[(a,b)]$

Recall that these infinite open strips form a subbase for the usual topology on $\mathbf{R}^2$. Accordingly, the usual topology on $\mathbf{R}^2$ is the smallest topology on $\mathbf{R}^2$ with respect to which the projections π_1 and π_2 are continuous.

Solved Problems

CONTINUOUS FUNCTIONS

1. Prove: Let $f: X \to Y$ be a constant function, say $f(x) = p \in Y$, for every $x \in X$. Then f is continuous relative to any topology $\mathcal{T}$ on X and any topology $\mathcal{T}^*$ on Y.

Solution:

We need to show that the inverse image of any $\mathcal{T}^*$-open subset of Y is a $\mathcal{T}$-open subset of X. Let $H \in \mathcal{T}^*$. Now $f(x) = p$ for all $x \in X$, so

$$f^{-1}[H] \quad = \quad \begin{cases} X & \text{if } p \in H \\ \emptyset & \text{if } p \notin H \end{cases}$$

In either case $f^{-1}[H]$ is an open subset of X since X and $\emptyset$ belong to every topology $\mathcal{T}$ on X.

2. Prove: Let $f: X \to Y$ be any function. If $(Y, \mathcal{J})$ is an indiscrete space, then $f: (X, \mathcal{T}) \to (Y, \mathcal{J})$ is continuous for any $\mathcal{T}$.

Solution:

We want to show that the inverse image of every open subset of Y is an open subset of X. Since $(Y, \mathcal{J})$ is an indiscrete space, Y and $\emptyset$ are the only open subsets of Y. But

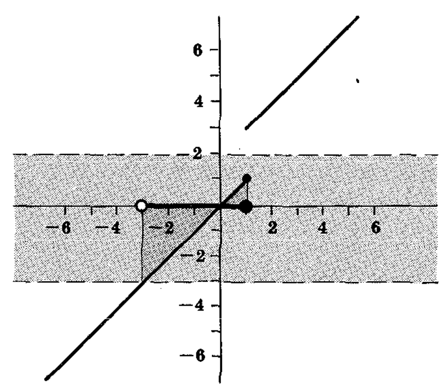
$$f^{-1}[Y] = X, \quad f^{-1}[\emptyset] = \emptyset$$

and X and $\emptyset$ belong to any topology $\mathcal{T}$ on X. Hence f is continuous for any $\mathcal{T}$.

3. Let $\mathcal{U}$ be the usual topology on the real line $\mathbf{R}$ and let $\mathcal{T}$ be the upper limit topology on $\mathbf{R}$ which is generated by the open-closed intervals $(a, b]$. Furthermore, let $f: \mathbf{R} \to \mathbf{R}$ be defined by

$$f(x) \quad = \quad \begin{cases} x & \text{if } x \leq 1 \\ x + 2 & \text{if } x > 1 \end{cases}$$

(See diagram on the right.)



(i) Show that f is not $\mathcal{U}$-$\mathcal{U}$ continuous.

(ii) Show that f is $\mathcal{T}$-$\mathcal{T}$ continuous.

Solution:

(i) Let $A = (-3, 2)$. Then $f^{-1}[A] = (-3, 1]$. Now $A \in \mathcal{U}$ but $f^{-1}[A] \notin \mathcal{U}$, so f is not $\mathcal{U}$-$\mathcal{U}$ continuous.

(ii) Let $A = (a, b]$. Then:

$$f^{-1}[A] = \begin{cases} (a, b] & \text{if } a < b \leq 1 \\ (a, 1] & \text{if } a < 1 < b \leq 3 \\ (a, b-2] & \text{if } a < 1 < 3 < b \\ \emptyset & \text{if } 1 \leq a < b \leq 3 \\ (1, b-2] & \text{if } 1 \leq a < 3 < b \\ (a-2, b-2] & \text{if } 3 \leq a < b \end{cases}$$

In each case, $f^{-1}[A]$ is a $\mathcal{T}$-open set. Hence f is $\mathcal{T}$-$\mathcal{T}$ continuous.

4. Suppose a function $f: (X, \mathcal{T}_1) \to (Y, \mathcal{T}_2)$ is not $\mathcal{T}_1$-$\mathcal{T}_2$ continuous. Show that if $\mathcal{T}_1^*$ is a topology on X coarser than $\mathcal{T}_1$ and if $\mathcal{T}_2^*$ is a topology on Y finer than $\mathcal{T}_2$, i.e. $\mathcal{T}_1^* \subset \mathcal{T}_1$ and $\mathcal{T}_2 \subset \mathcal{T}_2^*$, then f is also not $\mathcal{T}_1^*$-$\mathcal{T}_2^*$ continuous.

Solution:

Since $f: (X, \mathcal{T}_1) \to (Y, \mathcal{T}_2)$ is not continuous,

$$\exists G \in \mathcal{T}_2 \quad \text{for which} \quad f^{-1}[G] \notin \mathcal{T}_1$$

Now, $\mathcal{T}_1^* \subset \mathcal{T}_1$ and $\mathcal{T}_2 \subset \mathcal{T}_2^*$. Hence $G \in \mathcal{T}_2$ implies $G \in \mathcal{T}_2^*$, and $f^{-1}[G] \notin \mathcal{T}_1$ implies $f^{-1}[G] \notin \mathcal{T}_1^*$. Thus f is not continuous with respect to $\mathcal{T}_1^*$ and $\mathcal{T}_2^*$.

5. Show that the identity function $i: (X, \mathcal{T}) \to (X, \mathcal{T}^*)$ is continuous if and only if $\mathcal{T}$ is finer than $\mathcal{T}^*$, i.e. $\mathcal{T}^* \subset \mathcal{T}$.

Solution:

By definition, i is $\mathcal{T}$-$\mathcal{T}^*$ continuous if and only if

$$G \in \mathcal{T}^* \quad \Rightarrow \quad i^{-1}[G] \in \mathcal{T}$$

But $i^{-1}[G] = G$, so i is $\mathcal{T}$-$\mathcal{T}^*$ continuous, if and only if

$$G \in \mathcal{T}^* \quad \Rightarrow \quad G \in \mathcal{T}$$

that is, $\mathcal{T}^* \subset \mathcal{T}$.

6. Prove Theorem 7.2: Let $f: (X, \mathcal{T}) \to (Y, \mathcal{T}^*)$, and let $\mathcal{S}$ be a subbase for the topology $\mathcal{T}^*$ on Y. Then f is continuous if and only if the inverse of every member of the subbase $\mathcal{S}$ is an open subset of X, i.e. $f^{-1}[S] \in \mathcal{T}$ for every $S \in \mathcal{S}$.

Solution:

Suppose $f^{-1}[S] \in \mathcal{T}$ for every $S \in \mathcal{S}$. We want to show that f is continuous, i.e. $G \in \mathcal{T}^*$ implies $f^{-1}[G] \in \mathcal{T}$. Let $G \in \mathcal{T}^*$. By definition of subbase,

$$G = \cup_i (S_{i_1} \cap \cdots \cap S_{i n_i}) \quad \text{where} \quad S_{i_k} \in \mathcal{S}$$

Hence,

$$f^{-1}[G] = f^{-1}[\cup_i (S_{i_1} \cap \cdots \cap S_{i n_i})] = \cup_i f^{-1}[S_{i_1} \cap \cdots \cap S_{i n_i}]$$
$$= \cup_i (f^{-1}[S_{i_1}] \cap \cdots \cap f^{-1}[S_{i n_i}])$$

But $S_{i_k} \in \mathcal{S}$ implies $f^{-1}[S_{i_k}] \in \mathcal{T}$. Hence $f^{-1}[G] \in \mathcal{T}$ since it is the union of finite intersections of open sets. Accordingly, f is continuous.

On the other hand, if f is continuous then the inverse of all open sets, including the members of $\mathcal{S}$ are open.

7. Let f be a function from a topological space X into the unit interval $[0,1]$. Show that if $f^{-1}[(a,1]]$ and $f^{-1}[[0,b)]$ are open subsets of X for all $0 < a, b < 1$, then f is continuous.

Solution:

Recall that the intervals $(a,1]$ and $[0,b)$ form a subbase for the unit interval $I = [0,1]$. Hence f is continuous by the preceding problem, i.e. by Theorem 7.2.

8. Prove: Let the functions $f: X \to Y$ and $g: Y \to Z$ be continuous. Then the composition function $g \circ f: X \to Z$ is also continuous.

Solution:

Let G be an open subset of Z. Then $g^{-1}[G]$ is open in Y since g is continuous. But f is also continuous, so $f^{-1}[g^{-1}[G]]$ is open in X. Now

$$(g \circ f)^{-1}[G] \quad = \quad f^{-1}[g^{-1}[G]]$$

Thus $(g \circ f)^{-1}[G]$ is open in X for every open subset G of Z, or, $g \circ f$ is continuous.

9. Prove: Let $\{\mathcal{T}_i\}$ be a collection of topologies on a set X. If a function $f: X \to Y$ is continuous with respect to each $\mathcal{T}_i$, then f is continuous with respect to the intersection topology $\mathcal{T} = \cap_i \mathcal{T}_i$.

Solution:

Let G be an open subset of Y. Then, by hypothesis, $f^{-1}[G]$ belongs to each $\mathcal{T}_i$. Hence $f^{-1}[G]$ belongs to the intersection, i.e. $f^{-1}[G] \in \cap_i \mathcal{T}_i = \mathcal{T}$, and so f is continuous with respect to $\mathcal{T}$.

10. Prove Theorem 7.3: A function $f: X \to Y$ is continuous if and only if the inverse image of every closed subset of Y is a closed subset of X.

Solution:

Suppose $f: X \to Y$ is continuous, and let F be a closed subset of Y. Then F^c is open, and so $f^{-1}[F^c]$ is open in X. But $f^{-1}[F^c] = (f^{-1}[F])^c$; therefore $f^{-1}[F]$ is closed.

Conversely, assume F closed in Y implies $f^{-1}[F]$ closed in X. Let G be an open subset of Y. Then G^c is closed in Y, and so $f^{-1}[G^c] = (f^{-1}[G])^c$ is closed in X. Accordingly, $f^{-1}[G]$ is open and therefore f is continuous.

11. Prove Theorem 7.4: A function $f: X \to Y$ is continuous if and only if, for every subset $A \subset X$, $f[\bar{A}] \subset \overline{f[A]}$.

Solution:

Suppose $f: X \to Y$ is continuous. Now $f[A] \subset \overline{f[A]}$, so

$$A \subset f^{-1}[f[A]] \subset f^{-1}[\overline{f[A]}]$$

But $\overline{f[A]}$ is closed, and so $f^{-1}[\overline{f[A]}]$ is also closed; hence

$$A \subset \bar{A} \subset f^{-1}[\overline{f[A]}]$$

and therefore

$$f[\bar{A}] \subset \overline{f[A]} = f[f^{-1}[\overline{f[A]}]]$$

Conversely, assume $f[\bar{A}] \subset \overline{f[A]}$ for any $A \subset X$, and let F be a closed subset of Y. Set $A = f^{-1}[F]$; we wish to show that A is also closed or, equivalently, that $\bar{A} = A$. Now

$$f[\bar{A}] = f[\overline{f^{-1}[F]}] \subset \overline{f[f^{-1}[F]]} = \bar{F} = F$$

Hence

$$\bar{A} \subset f^{-1}[f[\bar{A}]] \subset f^{-1}[F] = A$$

But $A \subset \bar{A}$, so $\bar{A} = A$ and f is continuous.

12. Prove: Let $f: (X, \mathcal{T}) \to (Y, \mathcal{T}^*)$ be continuous. Then $f_A: (A, \mathcal{T}_A) \to (Y, \mathcal{T}^*)$ is continuous, where $A \subset X$ and f_A is the restriction of f to A.

Solution:

Observe that $f_A^{-1}[G] = A \cap f^{-1}[G]$ for any $G \subset Y$.

Let $G \in \mathcal{T}^*$. Then $f^{-1}[G] \in \mathcal{T}$, and so $A \cap f^{-1}[G] \in \mathcal{T}_A$ by definition of the induced topology. Thus $A \cap f^{-1}[G] = f_A^{-1}[G] \in \mathcal{T}_A$, so f_A is continuous.

CONTINUITY AT A POINT

13. Under what conditions will a function $f: X \to Y$ not be continuous at a point $p \in X$?

Solution:

A function $f: X \to Y$ is continuous at $p \in X$ iff, for every open set $H \subset Y$ containing $f(p)$, $f^{-1}[H]$ is a superset of an open set containing p. Hence f is not continuous at $p \in X$ if there exists at least one open set $H \subset Y$ containing $f(p)$ such that $f^{-1}[H]$ does not contain an open set containing p.

Equivalently, $f: X \to Y$ is not continuous at $p \in X$ iff $\exists$ a neighborhood N of $f(p)$ such that $f^{-1}[N]$ is not a neighborhood of p.

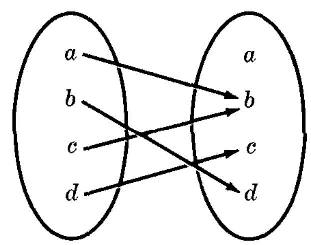
14. Consider the following topology defined on $X = \{a, b, c, d\}$:

$$\mathcal{T} = \{X, \emptyset, \{a\}, \{b\}, \{a, b\}, \{b, c, d\}\}$$

Let the function $f: X \to X$ be defined by the adjoining diagram.

(i) Show that f is not continuous at c.

(ii) Show that f is continuous at d.



Solution:

(i) Observe that $\{a, b\}$ is an open set containing $f(c) = b$ and that $f^{-1}[\{a, b\}] = \{a, c\}$. Hence f is not continuous at c since there exists no open set containing c which is contained in $\{a, c\}$.

(ii) The only open sets containing $f(d) = c$ are $\{b, c, d\}$ and X. Note that $f^{-1}[\{b, c, d\}] = X$ and $f^{-1}[X] = X$. Hence f is continuous at d since the inverse of each open set containing $f(d)$ is an open set containing d.

15. Suppose a singleton set $\{p\}$ is an open subset of a topological space X. Show that for any topological space Y and any function $f: X \to Y$, f is continuous at $p \in X$.

Solution:

Let $H \subset Y$ be an open set containing $f(p)$. But

$$f(p) \in H \quad \Rightarrow \quad p \in f^{-1}[H] \quad \Rightarrow \quad \{p\} \subset f^{-1}[H]$$

Hence f is continuous at p.

16. Prove: If $f: X \to Y$ is continuous at $p \in X$, then the restriction of f to a subset containing p is also continuous at p. More precisely, let A be a subset of a topological space $(X, \mathcal{T})$ such that $p \in A \subset X$, and let $f_A: A \to Y$ denote the restriction of $f: X \to Y$ to A. Then if f is $\mathcal{T}$-continuous at p, f_A will be $\mathcal{T}_A$-continuous at p where $\mathcal{T}_A$ is the relative topology on A.

Solution:

Let $H \subset Y$ be an open set containing $f(p)$. Since f is continuous at p,

$$\exists G \in \mathcal{T} \quad \text{such that} \quad p \in G \subset f^{-1}[H]$$

and so

$$p \in A \cap G \subset A \cap f^{-1}[H] = f_A^{-1}[H]$$

But, by definition of the induced topology, $A \cap G \in \mathcal{T}_A$; hence f_A is $\mathcal{T}_A$-continuous at p.

17. Prove Theorem 7.5: Let X and Y be topological spaces. Then a function $f: X \to Y$ is continuous if and only if it is continuous at every point $p \in X$.

Solution:

Assume f is continuous, and let $H \subset Y$ be an open set containing $f(p)$. But then $p \in f^{-1}[H]$, and $f^{-1}[H]$ is open. Hence f is continuous at p.

Now suppose f is continuous at every point $p \in X$, and let $H \subset Y$ be open. For every $p \in f^{-1}[H]$, there exists an open set $G_p \subset X$ such that $p \in G_p \subset f^{-1}[H]$. Hence $f^{-1}[H] = \bigcup\{G_p : p \in f^{-1}[H]\}$ a union of open sets. Accordingly, $f^{-1}[H]$ is open and so f is continuous.

18. Prove Proposition 7.6: If a function $f: X \to Y$ is continuous at $p \in X$, then it is sequentially continuous at p, i.e. $a_n \to p \Rightarrow f(a_n) \to f(p)$.

Solution:

We need to show that any neighborhood N of $f(p)$ contains almost all the terms of the sequence $\langle f(a_1), f(a_2), \ldots \rangle$.

Let N be a neighborhood of $f(p)$. By hypothesis, f is continuous at p; hence $M = f^{-1}[N]$ is a neighborhood of p. If the sequence $\langle a_n \rangle$ converges to p, then M contains almost all the terms of the sequence $\langle a_1, a_2, \ldots \rangle$, i.e. $a_n \in M$ for almost all $n \in \mathbf{N}$. But

$$a_n \in M \quad \Rightarrow \quad f(a_n) \in f[M] = f[f^{-1}[N]] = N$$

Hence $f(a_n) \in N$ for almost all $n \in \mathbf{N}$, and so the sequence $\langle f(a_n) \rangle$ converges to $f(p)$. Accordingly, f is sequentially continuous at p.

OPEN AND CLOSED FUNCTIONS, HOMEOMORPHISMS

19. Give an example of a real function $f: \mathbf{R} \to \mathbf{R}$ such that f is continuous and closed, but not open.

Solution:

Let f be a constant function, say $f(x) = 1$ for all $x \in \mathbf{R}$. Then $f[A] = \{1\}$ for any $A \subset \mathbf{R}$. Hence f is a closed function and is not an open function. Furthermore, f is continuous.

20. Let the real function $f: \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x^2$. Show that f is not open.

Solution:

Let $A = (-1, 1)$, an open set. Note that $f[A] = [0, 1)$, which is not open; hence f is not an open function.

21. Let $\mathcal{B}$ be a base for a topological space X. Show that if $f: X \to Y$ has the property that $f[B]$ is open for every $B \in \mathcal{B}$, then f is an open function.

Solution:

We want to show that the image of every open subset of X is open in Y. Let $G \subset X$ be open. By definition of a base, $G = \cup_i B_i$ where $B_i \in \mathcal{B}$. Now $f[G] = f[\cup_i B_i] = \cup_i f[B_i]$. By hypothesis, each $f[B_i]$ is open in Y and so $f[G]$, a union of open sets, is also open in Y; hence f is an open function.

22. Show that the closed interval $A = [a, b]$ is homeomorphic to the closed unit interval $I = [0, 1]$.

Solution:

The linear function $f: I \to A$ defined by $f(x) = (b - a)x + a$ is one-one, onto and bicontinuous. Hence f is a homeomorphism.

23. Show that area is not a topological property.

Solution:

The open disc $D = \{\langle r, \theta \rangle : r < 1\}$ with radius 1 is homeomorphic to the open disc $D^* = \{\langle r, \theta \rangle : r < 2\}$ with radius 2. In fact, the function $f: D \to D^*$ defined by $f(\langle r, \theta \rangle) = \langle 2r, \theta \rangle$ is a homeomorphism. Here $\langle r, \theta \rangle$ denotes the polar coordinates of a point in the plane $\mathbf{R}^2$.

24. Let $f: (X, \mathcal{T}) \to (Y, \mathcal{T}^*)$ be one-one and open, let $A \subset X$, and let $f[A] = B$. Show that the function $f_A: (A, \mathcal{T}_A) \to (B, \mathcal{T}^*_B)$ is also one-one and open. Here f_A denotes the restriction of f to A, and $\mathcal{T}_A$ and $\mathcal{T}^*_B$ are the relative topologies.

Solution:

If f is one-one, then every restriction of f is also one-one; hence we need only show that f_A is open.

Let $H \subset A$ be $\mathcal{T}_A$-open. Then by definition of the relative topology, $H = A \cap G$ where $G \in \mathcal{T}$. Since f is one-one, $f[A \cap G] = f[A] \cap f[G]$, and so

$$f_A[H] = f[H] = f[A \cap G] = f[A] \cap f[G] = B \cap f[G]$$

Since f is open and $G \in \mathcal{T}$, $f[G] \in \mathcal{T}^*$. Thus $B \cap f[G] \in \mathcal{T}^*_B$ and so $f_A : (A, \mathcal{T}_A) \to (B, \mathcal{T}^*_B)$ is open.

25. Let $f : (X, \mathcal{T}) \to (Y, \mathcal{T}^*)$ be a homeomorphism and let $(A, \mathcal{T}_A)$ be any subspace of $(X, \mathcal{T})$. Show that $f_A : (A, \mathcal{T}_A) \to (B, \mathcal{T}^*_B)$ is also a homeomorphism where f_A is the restriction of f to A, $f[A] = B$, and $\mathcal{T}^*_B$ is the relative topology on B.

Solution:

Since f is one-one and onto, $f_A : A \to B$, where $B = f[A]$, is also one-one and onto. Hence we need only show that f_A is bicontinuous, i.e. open and continuous. By the preceding problem f_A is open. Furthermore, the restriction of any continuous function is also continuous; hence $f_A : (A, \mathcal{T}_A) \to (B, \mathcal{T}^*_B)$ is a homeomorphism.

26. Show that any interval $A = (a, b)$ is connected as a subspace of the real line **R**. (See Example 4.3 for the definition of connectedness.)

Solution:

Suppose A is not connected. Then $\exists$ open sets $G, H \subset \mathbf{R}$ such that $A \cap G$ and $A \cap H$ are non-empty, disjoint and satisfy $(A \cap G) \cup (A \cap H) = A$. Define the function $f : A \to \mathbf{R}$ by

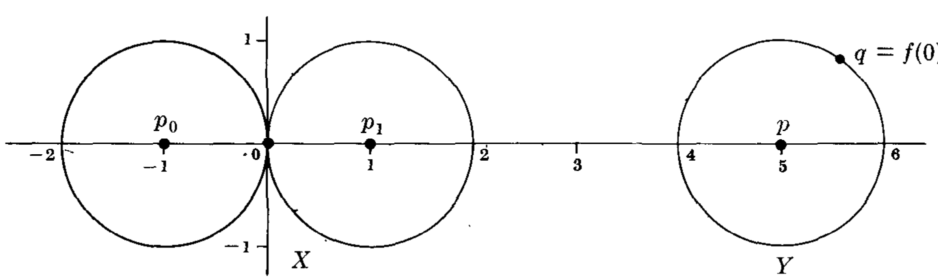
$$f(x) = \begin{cases} 1 & \text{if } x \in A \cap G \\ 0 & \text{if } x \in A \cap H \end{cases}$$

Then f is continuous, for the inverse of any open set is either $A \cap G$, $A \cap H$, $\emptyset$ or A and so is open. But then the intermediate value theorem applies, so $\exists x_0 \in A$ for which $f(x_0) = \frac{1}{2}$. But this is impossible, so A is connected.

27. Show that the following subsets of the plane $\mathbf{R}^2$ are not homeomorphic, where the topologies are the relativized usual topologies:

$$X = \{x : d(x, p_0) = 1 \text{ or } d(x, p_1) = 1;\ p_0 = \langle 0, -1 \rangle,\ p_1 = \langle 0, 1 \rangle\}$$

$$Y = \{x : d(x, p) = 1,\ p = \langle 0, 5 \rangle\}$$



Solution:

Suppose there exists a homeomorphism $f : X \to Y$; let $q = f(0)$, $X^* = X \setminus \{0\}$, and $Y^* = Y \setminus \{q\}$. Then $f : X^* \to Y^*$ is also a homeomorphism with respect to the relative topologies (see Problem 25).

We show that Y^* is connected. For if $q = \langle 5 + \cos \theta_0, \sin \theta_0 \rangle$, then the function

$$g : (0, 2\pi) \to Y^* \quad \text{defined by} \quad g(\theta) = \langle 5 + \cos(\theta_0 + \theta), \sin(\theta_0 + \theta) \rangle$$

is a homeomorphism. But the interval $(0, 2\pi)$ is connected, so Y^* is also connected.

On the other hand, X^* is not connected; for the sets

$$G = \{\langle x, y \rangle : x > 0\} \quad \text{and} \quad H = \{\langle x, y \rangle : x < 0\}$$

are both open in $\mathbf{R}^2$, so $G^* = X^* \cap G$ and $H^* = X^* \cap H$ are open subsets of X^*. Furthermore, G^* and H^* are non-empty, disjoint and satisfy $G^* \cup H^* = X^*$. Since connectedness is a topological property, X^* is not homeomorphic to Y^* and therefore there can exist no such function f.

TOPOLOGIES INDUCED BY FUNCTIONS

28. Let $\{f_i : X \to (Y_i, \mathcal{T}_i)\}$ be a collection of constant functions from an arbitrary set X into the topological spaces $(Y_i, \mathcal{T}_i)$. Determine the coarsest topology on X with respect to which the functions f_i are continuous.

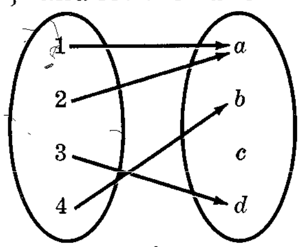
Solution:

Recall (see Problem 1) that a constant function $f: X \to Y$ is continuous with respect to every topology on X. Hence all the constant functions f_i are continuous with respect to the indiscrete topology $\{X, \emptyset\}$ on X. Since the indiscrete topology $\{X, \emptyset\}$ on X is the coarsest topology on X, it is also the coarsest topology on X with respect to which the constant functions are continuous.

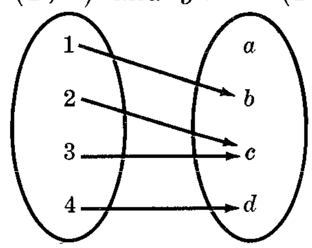
29. Consider the following topology on $Y = \{a, b, c, d\}$:

$$\mathcal{T} = \{Y, \emptyset, \{c\}, \{a,b,c\}, \{c,d\}\}$$

Let $X = \{1,2,3,4\}$ and let the functions $f: X \to (Y, \mathcal{T})$ and $g: X \to (Y, \mathcal{T})$ be defined by



f



g

Find the defining subbase $\mathcal{S}$ for the topology $\mathcal{T}^*$ on X induced by f and g, i.e. the coarsest topology with respect to which f and g are continuous.

Solution:

Recall that $$\mathcal{S} = \{f^{-1}[H] : H \in \mathcal{T}\} \cup \{g^{-1}[H] : H \in \mathcal{T}\}$$

that is, $\mathcal{S}$ consists of the inverses under f and g of the open subsets of Y. Hence

$$\mathcal{S} = \{X, \emptyset, \{1,2,4\}, \{3\}, \{2,3\}, \{1,2,3\}, \{2,3,4\}\}$$

30. Let $\mathcal{T}$ be the topology on the real line $\mathbf{R}$ generated by the closed-open intervals $[a, b)$, and let $\mathcal{T}^*$ be the topology on $\mathbf{R}$ induced by the collection of all linear functions

$$f: \mathbf{R} \to (\mathbf{R}, \mathcal{T}) \quad \text{defined by} \quad f(x) = ax + b, \quad a, b \in \mathbf{R}$$

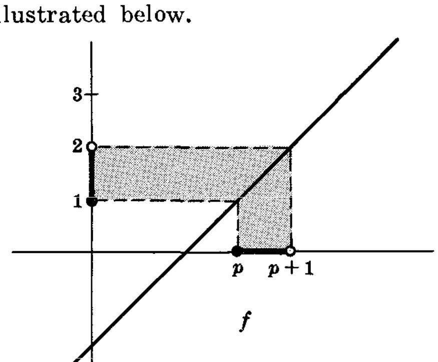
Show that $\mathcal{T}^*$ is the discrete topology on $\mathbf{R}$.

Solution:

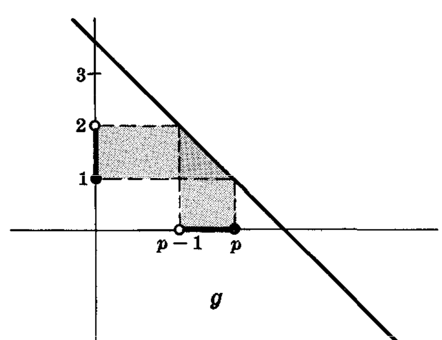
We want to show that, for every $p \in \mathbf{R}$, the singleton set $\{p\}$ is a $\mathcal{T}^*$-open set. Consider the $\mathcal{T}$-open set $A = [1, 2)$ and the functions $f: \mathbf{R} \to (\mathbf{R}, \mathcal{T})$ and $g: \mathbf{R} \to (\mathbf{R}, \mathcal{T})$ defined by

$$f(x) = x - p + 1 \quad \text{and} \quad g(x) = -x - p + 1$$

and illustrated below.



f



g

Now $A \in \mathcal{T}$ implies

$$f^{-1}[A] = [p, p+1) \quad \text{and} \quad g^{-1}[A] = (p-1, p]$$

belong to the defining subbase $\mathcal{S}$ for the topology $\mathcal{T}^*$. Hence the intersection

$$(p-1, p] \cap [p, p+1) = \{p\}$$

belongs to $\mathcal{T}^*$, and so $\mathcal{T}^*$ is the discrete topology on $\mathbf{R}$.

31. Prove Theorem 7.9: Let $\{f_i : X \to (Y_i, \mathcal{T}_i)\}$ be a collection of functions defined on an arbitrary non-empty set X, let

$$\mathcal{S} = \bigcup_i \{f_i^{-1}[H] : H \in \mathcal{T}_i\}$$

and let $\mathcal{T}$ be the topology on X generated by $\mathcal{S}$. Then:

(i) All the functions f_i are continuous relative to $\mathcal{T}$.

(ii) If $\mathcal{T}^*$ is the intersection of all topologies on X with respect to which the functions f_i are continuous, then $\mathcal{T} = \mathcal{T}^*$.

(iii) $\mathcal{T}$ is the coarsest topology on X with respect to which the functions f_i are continuous.

(iv) $\mathcal{S}$ is a subbase for $\mathcal{T}$.

Solution:

(i) For any function $f_i : (X, \mathcal{T}) \to (Y_i, \mathcal{T}_i)$, if $H \in \mathcal{T}_i$ then $f_i^{-1}[H] \in \mathcal{S} \subset \mathcal{T}$. Hence all the f_i are continuous with respect to $\mathcal{T}$.

(ii) By Problem 9, all the functions f_i are also continuous with respect to $\mathcal{T}^*$; hence $\mathcal{S} \subset \mathcal{T}^*$ and, since $\mathcal{T}$ is the topology generated by $\mathcal{S}$, $\mathcal{T} \subset \mathcal{T}^*$. On the other hand, $\mathcal{T}$ is one of the topologies with respect to which the f_i are continuous; hence $\mathcal{T}^* \subset \mathcal{T}$ and so $\mathcal{T} = \mathcal{T}^*$.

(iii) Follows from (ii).

(iv) Follows from the fact that any class of sets is a subbase of the topology it generates.

Supplementary Problems

CONTINUOUS FUNCTIONS

32. Prove that $f : X \to Y$ is continuous if and only if $f^{-1}[A^\circ] \subset (f^{-1}[A])^\circ$ for every $A \subset X$.

33. Let X and Y be topological spaces with $X = E \cup F$. Let $f : E \to Y$ and $g : F \to Y$, with $f = g$ on $E \cap F$, be continuous with respect to the relative topologies. Note that $h = f \cup g$ is a function from X into Y. (i) Show, by an example, that h need not be continuous. (ii) Prove: If E and F are both open, then h is continuous. (iii) Prove: If E and F are both closed, then h is continuous.

34. Let $f : X \to Y$ be continuous. Show that $f : X \to f[X]$ is also continuous where $f[X]$ has the relative topology.

35. Let X be a topological space and let $\chi_A : X \to \mathbf{R}$ be the characteristic function for some subset A of X. Show that χ_A is continuous at $p \in X$, if and only if p is not an element of the boundary of A. (Recall $\chi_A(x) = 1$ if $x \in A$, and $\chi_A(x) = 0$ if $x \in A^c$.)

36. Consider the real line $\mathbf{R}$ with the usual topology. Show that if every function $f : X \to \mathbf{R}$ is continuous, then X is a discrete space.

OPEN AND CLOSED FUNCTIONS

37. Let $f : (X, \mathcal{T}) \to (Y, \mathcal{T}^*)$. Prove the following:

(i) f is closed if and only if $\overline{f[A]} \subset f[\bar{A}]$ for every $A \subset X$;

(ii) f is open if and only if $f[A^\circ] \subset (f[A])^\circ$ for every $A \subset X$.

38. Show that the function $f: (0, \infty) \to [-1, 1]$ defined by $f(x) = \sin(1/x)$ is continuous, but neither open nor closed, where $(0, \infty)$ and $[-1, 1]$ have the relativized usual topologies.

39. Prove: Let $f: (X, \mathcal{T}) \to (Y, \mathcal{T}^*)$ be open and onto, and let $\mathcal{B}$ be a base for $\mathcal{T}$. Then $\{f[B] : B \in \mathcal{B}\}$ is a base for $\mathcal{T}^*$.

40. Give an example of a function $f: X \to Y$ and a subset $A \subset X$ such that f is open but f_A, the restriction of f to A, is not open.

HOMEOMORPHISMS, TOPOLOGICAL PROPERTIES

41. Let $f: X \to Y$ and $g: Y \to Z$ be continuous. Show that if $g \circ f: X \to Z$ is a homeomorphism, then g one-one (or f onto) implies that f and g are homeomorphisms.

42. Prove that each of the following is a topological property: (i) accumulation point, (ii) interior, (iii) boundary, (iv) density, and (v) neighborhood.

43. Prove: Let $f: X \to Y$ be a homeomorphism and let $A \subset X$ have the property that $A \cap A' = \emptyset$. Then $f[A] \cap (f[A])' = \emptyset$. (A subset $A \subset X$ having the property $A \cap A' = \emptyset$ is called *isolated*. The property of being isolated is thus a topological property.)

TOPOLOGIES INDUCED BY FUNCTIONS

44. Consider the following topology on $Y = \{a, b, c, d\}$: $\mathcal{T} = \{Y, \emptyset, \{a, b\}, \{c, d\}\}$. Let $X = \{1, 2, 3, 4, 5\}$ and let $f: X \to Y$ and $g: X \to Y$ be as follows:

$$f = \{\langle 1, a\rangle, \langle 2, a\rangle, \langle 3, b\rangle, \langle 4, b\rangle, \langle 5, d\rangle\}, \qquad g = \{\langle 1, c\rangle, \langle 2, b\rangle, \langle 3, d\rangle, \langle 4, a\rangle, \langle 5, c\rangle\}$$

Find the defining subbase for the topology on X induced by f and g.

45. Let $f: X \to (Y, \mathcal{T}^*)$. Show that if $\mathcal{S}$ is the defining subbase for the topology $\mathcal{T}$ induced by the one function f, then $\mathcal{S} = \mathcal{T}$.

46. Prove: Let $\{f_i : X \to (Y_i, \mathcal{T}_i)\}$ be a collection of functions defined on an arbitrary set X, and let $\mathcal{S}_i$ be a subbase for the topology $\mathcal{T}_i$ on Y_i. Then the class $\mathcal{S}^* = \bigcup_i \{f_i^{-1}[S] : S \in \mathcal{S}_i\}$ has the following properties: (i) $\mathcal{S}^*$ is a subclass of the defining subbase $\mathcal{S}$ of the topology $\mathcal{T}$ on X induced by the functions f_i; (ii) $\mathcal{S}^*$ is also a subbase for $\mathcal{T}$.

47. Show that the coarsest topology on the real line $\mathbf{R}$ with respect to which the linear functions

$$f: \mathbf{R} \to (\mathbf{R}, \mathcal{U}) \qquad \text{defined by} \qquad f(x) = ax + b, \quad a, b \in \mathbf{R}$$

are continuous is also the usual topology $\mathcal{U}$.

Answers to Supplementary Problems

33. (i) Let $X = (0, 2)$ and let $E = (0, 1)$ and $F = [1, 2)$. Then $f(x) = 1$ and $g(x) = 2$ are each continuous, but $h = f \cup g$ is not continuous.

44. $\{X, \emptyset, \{1, 2, 3, 4\}, \{5\}, \{2, 4\}, \{1, 3, 5\}\}$

45. *Hint.* Show that $\mathcal{S}$ is a topology.

Chapter 8

Metric and Normed Spaces

METRICS

Let X be a non-empty set. A real-valued function d defined on $X \times X$, i.e. ordered pairs of elements in X, is called a *metric* or *distance function* on X iff it satisfies, for every $a, b, c \in X$, the following axioms:

[M₁] $d(a, b) \geq 0$ and $d(a, a) = 0$.

[M₂] (Symmetry) $d(a, b) = d(b, a)$.

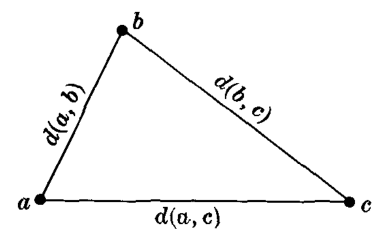
[M₃] (Triangle Inequality) $d(a, c) \leq d(a, b) + d(b, c)$.

[M₄] If $a \neq b$, then $d(a, b) > 0$.

The real number $d(a, b)$ is called the *distance* from a to b.

Observe that **[M₁]** states that the distance from any point to another is never negative, and that the distance from a point to itself is zero. The axiom **[M₂]** states that the distance from a point a to a point b is the same as the distance from b to a; hence we speak of the distance *between* a and b.

[M₃] is called the Triangle Inequality because if a, b and c are points in the plane $\mathbf{R}^2$ as illustrated on the right, then **[M₃]** states that the length $d(a, c)$ of one side of the triangle is less than or equal to the sum $d(a, b) + d(b, c)$ of the lengths of the other two sides of the triangle. The last axiom **[M₄]** states that the distance between two distinct points is positive.



We now give some examples of metrics. That they actually satisfy the required axioms will be verified later.

Example 1.1: The function d defined by $d(a, b) = |a - b|$, where a and b are real numbers, is a metric and called the *usual* metric on the real line $\mathbf{R}$. Furthermore, the function d defined by

$$d(p, q) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

where $p = \langle a_1, a_2 \rangle$ and $q = \langle b_1, b_2 \rangle$ are points in the plane $\mathbf{R}^2$, is a metric and called the *usual* metric on $\mathbf{R}^2$. We shall assume these metrics on $\mathbf{R}$ and $\mathbf{R}^2$, respectively, unless otherwise specified.

Example 1.2: Let X be any non-empty set and let d be the function defined by

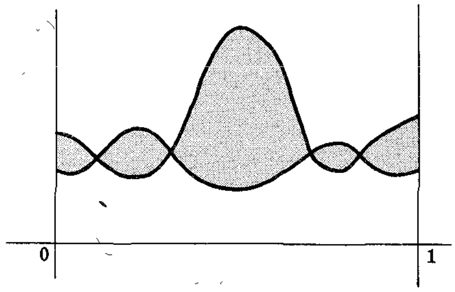
$$d(a, b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{if } a \neq b \end{cases}$$

Then d is a metric on X. This distance function d is usually called the *trivial* metric on X.

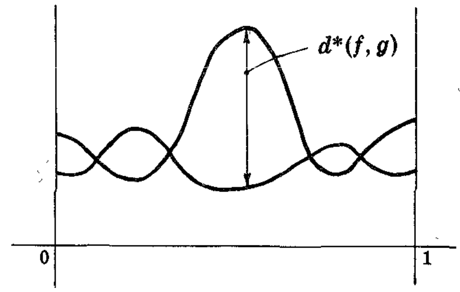
Example 1.3: Let $\mathcal{C}[0, 1]$ denote the class of continuous functions on the closed unit interval $[0, 1]$. A metric is defined on the class $\mathcal{C}[0, 1]$ as follows:

$$d(f, g) = \int_0^1 |f(x) - g(x)|\, dx$$

Here $d(f,g)$ is precisely the area of the region which lies between the functions as illustrated below.



$d(f,g)$ is shaded



Example 1.4: Again let $C[0,1]$ denote the collection of continuous functions on $[0,1]$. Another metric is defined on $C[0,1]$ as follows:

$$d^*(f,g) \;=\; \sup\{|f(x)-g(x)| : x \in [0,1]\}$$

Here $d^*(f,g)$ is precisely the greatest vertical gap between the functions as illustrated above.

Example 1.5: Let $p = \langle a_1, a_2\rangle$ and $q = \langle b_1, b_2\rangle$ be arbitrary points in the plane $\mathbf{R}^2$, i.e. ordered pairs of real numbers. The functions d_1 and d_2 defined by

$$d_1(p,q) \;=\; \max(|a_1-b_1|, |a_2-b_2|), \qquad d_2(p,q) \;=\; |a_1-b_1| + |a_2-b_2|$$

are distinct metrics on $\mathbf{R}^2$.

A function ρ satisfying **[M₁]**, **[M₂]** and **[M₃]**, i.e. not necessarily **[M₄]**, is called a *pseudometric*. Many of the results for metrics are also true for pseudometrics.

DISTANCE BETWEEN SETS, DIAMETERS

Let d be a metric on a set X. The *distance* between a point $p \in X$ and a non-empty subset A of X is denoted and defined by

$$d(p,A) \;=\; \inf\{d(p,a) : a \in A\}$$

i.e. the greatest lower bound of the distances from p to points of A. The *distance* between two non-empty subsets A and B of X is denoted and defined by

$$d(A,B) \;=\; \inf\{d(a,b) : a \in A,\ b \in B\}$$

i.e. the greatest lower bound of the distances from points in A to points in B.

The *diameter* of a non-empty subset A of X is denoted and defined by

$$d(A) \;=\; \sup\{d(a,a') : a, a' \in A\}$$

i.e. the least upper bound of the distances between points in A. If the diameter of A is finite, i.e. $d(A) < \infty$, then A is said to be *bounded*; if not, i.e. $d(A) = \infty$, then A is said to be *unbounded*.

Example 2.1: Let d be the trivial metric on a non-empty set X. Then for $p \in X$ and $A, B \subset X$,

$$d(p,A) \;=\; \begin{cases} 1 & \text{if } p \notin A \\ 0 & \text{if } p \in A \end{cases}, \qquad d(A,B) \;=\; \begin{cases} 1 & \text{if } A \cap B = \emptyset \\ 0 & \text{if } A \cap B \neq \emptyset \end{cases}$$

Example 2.2: Consider the following intervals on the real line $\mathbf{R}$: $A = [0,1)$, $B = (1,2]$.

If d denotes the usual metric on $\mathbf{R}$, then $d(A,B) = 0$. On the other hand, if d^* denotes the trivial metric on $\mathbf{R}$, then $d^*(A,B) = 1$ since A and B are disjoint.

The next proposition clearly follows from the above definitions:

Proposition 8.1: Let A and B be non-empty subsets of X and let $p \in X$. Then:

(i) $d(p,A)$, $d(A,B)$ and $d(A)$ are non-negative real numbers.

(ii) If $p \in A$, then $d(p,A) = 0$.

(iii) If $A \cap B$ is non-empty, then $d(A, B) = 0$.

(iv) If A is finite, then $d(A) < \infty$, i.e. A is bounded.

Observe that the converses of (ii), (iii) and (iv) are not true.

For the empty set $\emptyset$, the following conventions are adopted:

$$d(p, \emptyset) = \infty, \qquad d(A, \emptyset) = d(\emptyset, A) = \infty, \qquad d(\emptyset) = -\infty$$

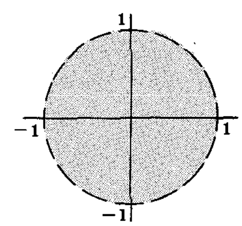
OPEN SPHERES

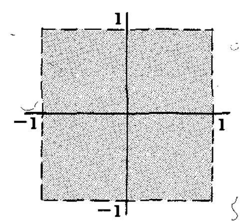
Let d be a metric on a set X. For any point $p \in X$ and any real number $\delta > 0$, we shall let $S_d(p, \delta)$ or simply $S(p, \delta)$ denote the set of points within a distance of δ from p:

$$S(p, \delta) = \{x : d(p, x) < \delta\}$$

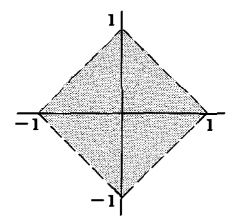
We call $S(p, \delta)$ the *open sphere*, or simply *sphere*, with center p and radius δ. It is also called a *spherical neighborhood* or *ball*.

Example 3.1: Consider the point $p = \langle 0, 0\rangle$ in the plane $\mathbf{R}^2$, and the real number $\delta = 1$. If d is the usual metric on $\mathbf{R}^2$, then $S_d(p, \delta)$ is the open unit disc illustrated on the right. On the other hand, if d_1 and d_2 are the metrics on $\mathbf{R}^2$ which are defined in Example 1.5, then $S_{d_1}(p, \delta)$ and $S_{d_2}(p, \delta)$ are the subsets of $\mathbf{R}^2$ which are illustrated below.





$S_{d_1}(p, \delta)$ is shaded



$S_{d_2}(p, \delta)$ is shaded

Example 3.2: Let d denote the trivial metric on some set X, and let $p \in X$. Recall that the distance between p and every other point in X is exactly 1. Hence

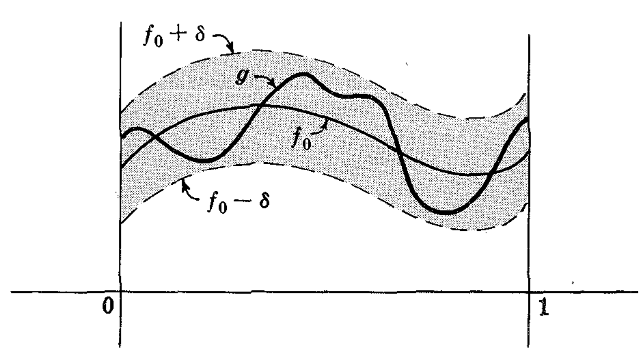
$$S(p, \delta) = \begin{cases} X & \text{if } \delta > 1 \\ \{p\} & \text{if } \delta \leq 1 \end{cases}$$

Example 3.3: Let d be the usual metric on the real line $\mathbf{R}$, i.e. $d(a, b) = |a - b|$. Then the open sphere $S(p, \delta)$ is the open interval $(p - \delta, p + \delta)$.

Example 3.4: Let d be the metric on the collection $\mathcal{C}[0,1]$ of all continuous functions on $[0,1]$ defined by

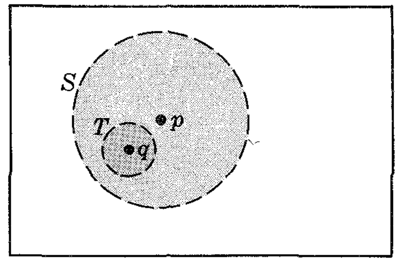
$$d(f, g) = \sup\{|f(x) - g(x)| : x \in [0,1]\}$$

(see Example 1.4). Given $\delta > 0$ and a function $f_0 \in \mathcal{C}[0,1]$, then the open sphere $S(f_0, \delta)$ consists of all continuous functions g which lie in the area bounded by $f_0 - \delta$ and $f_0 + \delta$, as indicated in the diagram below:



One important property of open spheres in metric spaces is given in the next lemma.

Lemma 8.2: Let S be an open sphere with center p and radius δ. Then for every point $q \in S$ there exists an open sphere T centered at q such that T is contained in S. (See the adjacent Venn diagram.)



METRIC TOPOLOGIES, METRIC SPACES

In general, the intersection of two open spheres need not be an open sphere. However, we will show that every point in the intersection of two open spheres does belong to an open sphere contained in the intersection. Namely,

Lemma 8.3: Let S_1 and S_2 be open spheres and let $p \in S_1 \cap S_2$. Then there exists an open sphere S_p with center p such that $p \in S_p \subset S_1 \cap S_2$.

Hence by virtue of Theorem 6.1 we have

Theorem 8.4: The class of open spheres in a set X with metric d is a base for a topology on X.

Definition: Let d be a metric on a non-empty set X. The topology $\mathcal{T}$ on X generated by the class of open spheres in X is called the *metric topology* (or, the topology *induced* by the metric d). Furthermore, the set X together with the topology $\mathcal{T}$ induced by the metric d is called a *metric space* and is denoted by (X, d).

Thus a metric space is a topological space in which the topology is induced by a metric. Accordingly, all concepts defined for topological spaces are also defined for metric spaces. For example, we can speak about open sets, closed sets, neighborhoods, accumulation points, closure, etc., for metric spaces.

Example 4.1: If d is the usual metric on the real line $\mathbf{R}$, i.e. $d(a, b) = |a - b|$, then the open spheres in $\mathbf{R}$ are precisely the finite open intervals. Hence the usual metric on $\mathbf{R}$ induces the usual topology on $\mathbf{R}$. Similarly, the usual metric on the plane $\mathbf{R}^2$ induces the usual topology on $\mathbf{R}^2$.

Example 4.2: Let d be the trivial metric on some set X. Note that for any $p \in X$, $S(p, \frac{1}{2}) = \{p\}$. Hence every singleton set is open and so every set is open. In other words, the trivial metric on X induces the discrete topology on X.

Example 4.3: Let (X, d) be a metric space and let Y be a non-empty subset of X. The restriction of the function d to the points in the subset Y, also denoted by d, is a metric on Y. We call (Y, d) a *metric subspace* of (X, d). In fact, (Y, d) is a subspace of (X, d), i.e. has the relative topology.

Frequently the same symbol, say X, is used to denote both a metric space and the underlying set on which the metric is defined.

PROPERTIES OF METRIC TOPOLOGIES

Since the topology of a metric space X is derived from a metric, one would correctly expect that the topological properties of X are related to the distance properties of X. For example,

Theorem 8.5: Let p be a point in a metric space X. Then the countable class of open spheres, $\{S(p, 1), S(p, \frac{1}{2}), S(p, \frac{1}{3}), \ldots\}$ is a local base at p.

Theorem 8.6: The closure $\bar{A}$ of a subset A of a metric space X is the set of points whose distance from A is zero, i.e. $\bar{A} = \{x : d(x, A) = 0\}$.

Observe that axiom [$\mathbf{M_4}$] implies that the only point with zero distance from a singleton set $\{p\}$ is the point p itself, i.e.,

$$d(x, \{p\}) = 0 \quad \text{implies} \quad x = p$$

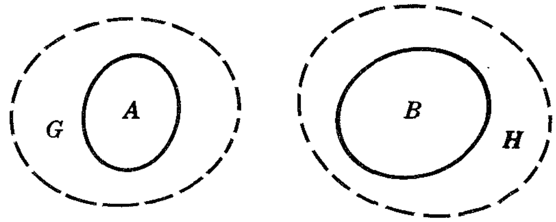
Hence by the preceding theorem, singleton sets $\{p\}$ in a metric space are closed. Accordingly, finite unions of singleton sets, i.e. finite sets, are also closed. We state this result formally:

Corollary 8.7: In a metric space X all finite sets are closed.

Thus we see that a metric space X possesses certain topological properties which do not hold for topological spaces in general.

Next follows an important "separation" property of metric spaces.

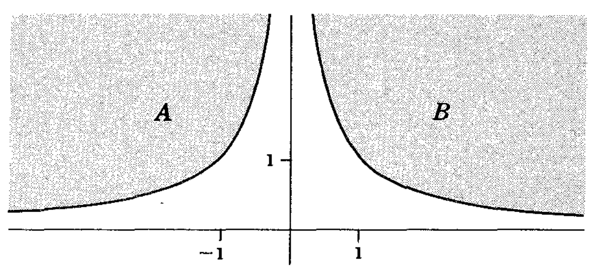
Theorem 8.8 (Separation Axiom): Let A and B be closed disjoint subsets of a metric space X. Then there exist disjoint open sets G and H such that $A \subset G$ and $B \subset H$. (See Venn diagram below.)



One might suspect from the above theorem that the distance between two disjoint closed sets is greater than zero. The next example shows that this is not true.

Example 5.1: Consider the following sets in the plane $\mathbf{R}^2$ which are illustrated below:

$$A = \{(x,y) : xy \geq -1,\ x < 0\}, \qquad B = \{(x,y) : xy \geq 1,\ x > 0\}$$

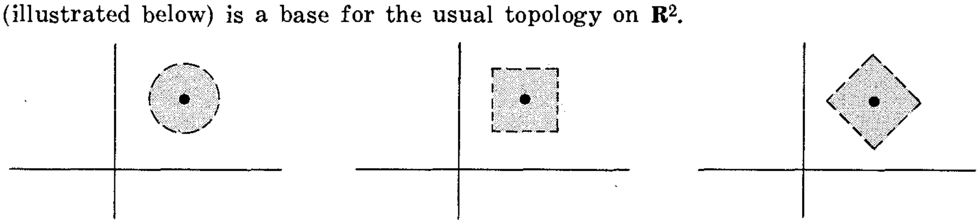


Observe that A and B are both closed and disjoint. However, $d(A,B) = 0$.

EQUIVALENT METRICS

Two metrics d and d^* on a set X are said to be equivalent iff they induce the same topology on X, i.e. iff the d-open spheres and the d^*-open spheres in X are bases for the same topology on X.

Example 6.1: The usual metric d and the metrics d_1 and d_2, defined in Example 1.5, all induce the usual topology on the plane $\mathbf{R}^2$, since the class of open spheres of each metric (illustrated below) is a base for the usual topology on $\mathbf{R}^2$.



$S_d(p, \delta)$ $\qquad$ $S_{d_1}(p, \delta)$ $\qquad$ $S_{d_2}(p, \delta)$

Hence the metrics are equivalent.

Example 6.2: Consider the metric d on a non-empty set X defined by

$$d(a, b) = \begin{cases} 2 & \text{if } a \neq b \\ 0 & \text{if } a = b \end{cases}$$

Observe that $s_d(p, 1) = \{p\}$; so singleton sets are open and d induces the discrete topology on X. Accordingly, d is equivalent to the trivial metric on X which also induces the discrete topology.

The next proposition clearly follows from the above definition.

Proposition 8.9: The relation "d is equivalent to d^*" is an equivalence relation in any collection of metrics on a set X.

METRIZATION PROBLEM

Given any topological space $(X, \mathcal{T})$, it is natural to ask whether or not there exists a metric d on X which induces the topology $\mathcal{T}$. The topological space $(X, \mathcal{T})$ is said to be *metrizable* if such a metric exists.

Example 7.1: Every discrete space $(X, \mathcal{D})$ is metrizable since the trivial metric on X induces the discrete topology $\mathcal{D}$.

Example 7.2: Consider the topological space $(\mathbf{R}, \mathcal{U})$, the real line $\mathbf{R}$ with the usual topology $\mathcal{U}$. Observe that $(\mathbf{R}, \mathcal{U})$ is metrizable since the usual metric on $\mathbf{R}$ induces the usual topology on $\mathbf{R}$. Similarly, the plane $\mathbf{R}^2$ with the usual topology is metrizable.

Example 7.3: An indiscrete space $(X, \mathcal{I})$ where X consists of more than one point is not metrizable. For X and $\emptyset$ are the only closed sets in an indiscrete space $(X, \mathcal{I})$. But by Corollary 8.7 all finite sets in a metric space are closed. Hence X and $\emptyset$ cannot be the only closed sets in a topology on X induced by a metric. Accordingly, $(X, \mathcal{I})$ is not metrizable.

The *metrization problem* in topology consists of finding necessary and sufficient topological conditions for a topological space to be metrizable. An important partial solution to this problem was given in 1924 by Urysohn as a result of his celebrated Urysohn's Lemma. It was not until 1950 that a complete solution to this problem was given independently by a number of mathematicians. We will prove Urysohn's results later. The complete solution to the metrization problem is beyond the scope of this text and the reader is referred to the classical text of Kelley, *General Topology*.

ISOMETRIC METRIC SPACES

A metric space (X, d) is *isometric* to a metric space (Y, e) iff there exists a one-one, onto function $f: X \to Y$ which preserves distances, i.e. for all $p, q \in X$,

$$d(p, q) = e(f(p), f(q))$$

Observe that the relation "(X, d) is isometric to (Y, e)" is an equivalence relation in any collection of metric spaces. Furthermore,

Theorem 8.10: If the metric space (X, d) is isometric to (Y, e), then (X, d) is also homeomorphic to (Y, e).

The next example shows that the converse of the above theorem is not true, i.e. two metric spaces can be homeomorphic but not isometric.

Example 8.1: Let d be the trivial metric on a set X and let e be the metric on a set Y defined by

$$e(a, b) = \begin{cases} 2 & \text{if } a \neq b \\ 0 & \text{if } a = b \end{cases}$$

Assume that X and Y have the same cardinality greater than one. Then (X, d) and (Y, e) are not isometric since distances between points in each space are different. But both d and e induce the discrete topology, and two discrete spaces with the same cardinality are homeomorphic; so (X, d) and (Y, e) are homeomorphic.

EUCLIDEAN m-SPACE

Recall that $\mathbf{R}^m$ denotes the product set of m copies of the set $\mathbf{R}$ of real numbers, i.e. consists of all m-tuples $\langle a_1, a_2, \ldots, a_m \rangle$ of real numbers. The function d defined by

$$d(p,q) \quad = \quad \sqrt{(a_1-b_1)^2 + \cdots + (a_m-b_m)^2} \quad = \quad \sqrt{\sum_{i=1}^{m} (a_i-b_i)^2} \quad = \quad \sqrt{\sum_{i=1}^{m} |a_i-b_i|^2}$$

where $p = \langle a_1, \ldots, a_m \rangle$ and $q = \langle b_1, \ldots, b_m \rangle$, is a metric, called the *Euclidean metric* on $\mathbf{R}^m$. We assume this metric on $\mathbf{R}^m$ unless otherwise specified. The metric space $\mathbf{R}^m$ with the Euclidean metric is called *Euclidean m-space* and will also be denoted by E^m.

Theorem 8.11: Euclidean m-space is a metric space.

Observe that Euclidean 1-space is precisely the real line $\mathbf{R}$ with the usual metric, and Euclidean 2-space is the plane $\mathbf{R}^2$ with the usual metric.

HILBERT SPACE

The class of all infinite real sequences

$$\langle a_1, a_2, \ldots \rangle \quad \text{such that} \quad \sum_{n=1}^{\infty} a_n^2 < \infty$$

i.e. such that the series $a_1^2 + a_2^2 + \cdots$ converges, is denoted by $\mathbf{R}^\infty$.

Example 9.1: Consider the sequences

$$p = \langle 1, 1, 1, \ldots \rangle \quad \text{and} \quad q = \langle 1, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \ldots \rangle$$

Since $1^2 + 1^2 + \cdots$ does not converge, p is not a point in $\mathbf{R}^\infty$. On the other hand, the series $1^2 + (\frac{1}{2})^2 + (\frac{1}{4})^2 + \cdots$ does converge; hence q is a point in $\mathbf{R}^\infty$.

Now let $p = \langle a_n \rangle$ and $q = \langle b_n \rangle$ belong to $\mathbf{R}^\infty$. The function d defined by

$$d(p,q) \quad = \quad \sqrt{\sum_{n=1}^{\infty} |a_n - b_n|^2}$$

is a metric and called the *l_2-metric* on $\mathbf{R}^\infty$. We assume this metric on $\mathbf{R}^\infty$ unless otherwise specified. The metric space consisting of $\mathbf{R}^\infty$ with the l_2-metric is called *Hilbert space* or *l_2-space* and will also be denoted by $\mathbf{H}$. We formally state:

Theorem 8.12: Hilbert space (or l_2-space) is a metric space.

Example 9.2: Let $\mathbf{H}_m$ denote the subspace of Hilbert space $\mathbf{H}$ consisting of all sequences of the form

$$\langle a_1, a_2, \ldots, a_{m-1}, a_m, 0, 0, 0, \ldots \rangle$$

Observe that $\mathbf{H}_m$ is isometric and hence homeomorphic to Euclidean m-space by the natural identification

$$\langle a_1, \ldots, a_m \rangle \quad \leftrightarrow \quad \langle a_1, \ldots, a_m, 0, 0, \ldots \rangle$$

Hilbert space exhibits two phenomena (not occurring in Euclidean m-space) described in the examples below:

Example 9.3: Consider the sequence $\langle p_n \rangle$ of points in Hilbert space where $p_k = \langle a_{1k}, a_{2k}, \ldots \rangle$ is defined by $a_{ik} = \delta_{ik}$; i.e. $a_{ik} = 1$ if $i = k$, and $a_{ik} = 0$ if $i \neq k$. Observe, as illustrated below, that the projection $\langle \pi_i(p_n) \rangle$ of $\langle p_n \rangle$ into each coordinate space converges to zero:

$$\begin{array}{lcl}
p_1 & = & \langle 1, 0, 0, 0, \ldots \rangle \\
p_2 & = & \langle 0, 1, 0, 0, \ldots \rangle \\
p_3 & = & \langle 0, 0, 1, 0, \ldots \rangle \\
p_4 & = & \langle 0, 0, 0, 1, \ldots \rangle \\
\cdot & & \;\;\cdot \;\; \cdot \;\; \cdot \;\; \cdot \\
\cdot & & \;\;\cdot \;\; \cdot \;\; \cdot \;\; \cdot \\
\cdot & & \;\;\cdot \;\; \cdot \;\; \cdot \;\; \cdot \\
 & & \;\downarrow \; \downarrow \; \downarrow \; \downarrow \\
\mathbf{0} & = & \langle 0, 0, 0, 0, \ldots \rangle
\end{array}$$

But the sequence $\langle p_n \rangle$ does not converge to $\mathbf{0}$, since $d(p_k, \mathbf{0}) = 1$ for every $k \in \mathbf{N}$; in fact, $\langle p_n \rangle$ has no convergent subsequence.

Example 9.4: Let $\mathbf{H}^*$ denote the proper subspace of $\mathbf{H}$ which consists of all points in $\mathbf{H}$ whose first coordinate is zero. Observe that the function $f: \mathbf{H} \to \mathbf{H}^*$ defined by $f(\langle a_1, a_2, \ldots \rangle) = \langle 0, a_1, a_2, \ldots \rangle$ is one-one, onto and preserves distances. Hence Hilbert space is isometric to a proper subspace of itself.

CONVERGENCE AND CONTINUITY IN METRIC SPACES

The following definitions of convergence and continuity in metric spaces are frequently used. Observe their similarity to the usual $\epsilon - \delta$ definitions.

Definition: The sequence $\langle a_1, a_2, \ldots \rangle$ of points in a metric space (X, d) converges to $b \in X$ if for every $\epsilon > 0$ there exists a positive integer n_0 such that

$$n > n_0 \quad \text{implies} \quad d(a_n, b) < \epsilon$$

Definition: Let (X, d) and (Y, d^*) be metric spaces. A function f from X into Y is continuous at $p \in X$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$d(p, x) < \delta \quad \text{implies} \quad d^*(f(p), f(x)) < \epsilon$$

The above definitions are equivalent to the definitions of convergence and continuity (in the metric topology) which were given for topological spaces in general.

NORMED SPACES

Let $\mathbf{V}$ be a real linear vector space, that is, $\mathbf{V}$ under an operation of vector addition and of scalar multiplication by real numbers satisfies the axioms $[\mathbf{V}_1]$, $[\mathbf{V}_2]$ and $[\mathbf{V}_3]$ of Chapter 2, Page 22. A function which assigns to each vector $v \in \mathbf{V}$ the real number $||v||$ is a *norm* on $\mathbf{V}$ iff it satisfies, for all $v, w \in \mathbf{V}$ and $k \in \mathbf{R}$, the following axioms:

$[\mathbf{N}_1]$ $||v|| \geq 0$ and $||v|| = 0$ iff $v = 0$.

$[\mathbf{N}_2]$ $||v + w|| \leq ||v|| + ||w||$.

$[\mathbf{N}_3]$ $||kv|| = |k|\,||v||$.

A linear space $\mathbf{V}$ together with a norm is called a *normed linear vector space* or simply a *normed space*. The real number $||v||$ is called the *norm* of the vector v.

Theorem 8.13: Let $\mathbf{V}$ be a normed space. The function d defined by

$$d(v, w) = ||v - w||$$

where $v, w \in \mathbf{V}$, is a metric, called the *induced metric* on $\mathbf{V}$.

Thus every normed space with the induced metric is a metric space and hence is also a topological space.

Example 10.1: The product set $\mathbf{R}^m$ is a linear vector space with addition defined by

$$\langle a_1, \ldots, a_m \rangle + \langle b_1, \ldots, b_m \rangle = \langle a_1 + b_1, \ldots, a_m + b_m \rangle$$

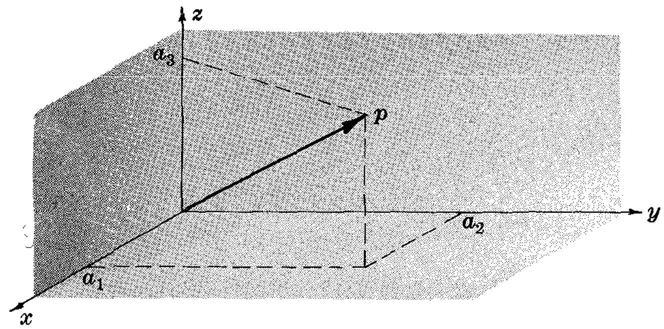
and scalar multiplication defined by

$$k \langle a_1, \ldots, a_m \rangle = \langle k a_1, \ldots, k a_m \rangle$$

The function on $\mathbf{R}^m$ defined by

$$|| \langle a_1, \ldots, a_m \rangle || = \sqrt{a_1^2 + \cdots + a_m^2} = \sqrt{\sum a_i^2} = \sqrt{\sum |a_i|^2}$$

is a norm and called the *Euclidean norm* on $\mathbf{R}^m$. Note that the Euclidean norm on $\mathbf{R}^m$ induces the Euclidean metric on $\mathbf{R}^m$. If $p = \langle a_1, a_2, a_3 \rangle$ is a point in $\mathbf{R}^3$, then $||p||$ corresponds precisely to the "length" of the arrow (or vector) from the origin to the point p as illustrated below.



Example 10.2: The following two functions are also norms on the linear space $\mathbf{R}^m$:

$$\|\langle a_1, \ldots, a_m\rangle\| = \max(|a_1|, |a_2|, \ldots, |a_m|)$$

$$\|\langle a_1, \ldots, a_m\rangle\| = |a_1| + |a_2| + \cdots + |a_m|$$

Let $\mathcal{F}(X, \mathbf{R})$ be the collection of all real-valued functions on a non-empty set X. Recall (see Theorem 2.9) that $\mathcal{F}(X, \mathbf{R})$ is a linear space with vector addition and scalar multiplication defined as follows:

$$(f+g)(x) \equiv f(x) + g(x) \quad \text{and} \quad (kf)(x) \equiv k\,f(x)$$

We shall frequently want to study classes of functions with certain other properties such as boundedness, continuity, etc. We shall use the following result from linear algebra:

Proposition 8.14: Let $\mathcal{A}(X, \mathbf{R})$ be a non-empty subcollection of $\mathcal{F}(X, \mathbf{R})$ satisfying the following two properties:

(i) If $f, g \in \mathcal{A}(X, \mathbf{R})$, then the sum $f + g \in \mathcal{A}(X, \mathbf{R})$.

(ii) If $f \in \mathcal{A}(X, \mathbf{R})$ and $k \in \mathbf{R}$, then the scalar multiple $kf \in \mathcal{A}(X, \mathbf{R})$.

Then $\mathcal{A}(X, \mathbf{R})$ is, itself, a linear vector space.

Example 10.3: The class $C[0,1]$ of all continuous real functions on the interval $I = [0,1]$ is a linear space since the sum and scalar multiples of continuous functions are continuous. The function on $C[0,1]$ defined by

$$\|f\| = \int_0^1 |f(x)|\,dx$$

is a norm which induces the metric on $C[0,1]$ defined in **Example 1.3**.

Example 10.4: The function on the linear space $C[0,1]$ defined by

$$\|f\| = \sup\{|f(x)| : x \in [0,1]\}$$

is also a norm. This norm induces the metric on $C[0,1]$ defined in Example 1.4.

Example 10.5: Let $\mathcal{B}(X, \mathbf{R})$ denote the subcollection of $\mathcal{F}(X, \mathbf{R})$ consisting of all bounded functions $f: X \to \mathbf{R}$. Then $\mathcal{B}(X, \mathbf{R})$ is a linear space since the sum and scalar multiples of bounded functions are also bounded. The function on $\mathcal{B}(X, \mathbf{R})$ defined by

$$\|f\| = \sup\{|f(x)| : x \in X\}$$

is a norm.

Example 10.6: We show later that the class $\mathbf{R}^\infty$ of all real sequences $\langle a_n\rangle$ such that $\sum |a_n|^2 < \infty$ is a linear space. The function on $\mathbf{R}^\infty$ defined by

$$\|\langle a_n\rangle\| = \sqrt{\sum |a_n|^2}$$

is a norm and called the l_2-norm on $\mathbf{R}^\infty$. Observe that this norm induces the l_2-metric in Hilbert space.

Solved Problems

METRICS

1. Show that in the definition of a metric the axiom $[\mathbf{M}_3]$ can be replaced by the following (weaker) axiom:

$[\mathbf{M}_3^*]$ If $a,b,c \in X$ are distinct then $d(a,c) \leq d(a,b) + d(b,c)$.

Solution:

Suppose $a = b$. Then

$$d(a,c) = d(b,c) = d(b,b) + d(b,c) \leq d(a,b) + d(b,c)$$

If $b = c$, the argument is similar. Lastly, suppose $a = c$; then

$$d(a,c) = 0 \leq d(a,b) + d(b,c)$$

Thus the Triangle Inequality follows from $[\mathbf{M}_1]$ if the points a, b and c are not all distinct.

2. Show that the trivial metric on a set X is a metric, i.e. that the function d defined by

$$d(a,b) = \begin{cases} 1 & \text{if } a \neq b \\ 0 & \text{if } a = b \end{cases}$$

satisfies $[\mathbf{M}_1]$, $[\mathbf{M}_2]$, $[\mathbf{M}_3^*]$ and $[\mathbf{M}_4]$.

Solution:

Let $a,b \in X$. Then $d(a,b) = 1$ or $d(a,b) = 0$. In either case, $d(a,b) \geq 0$. Also, if $a = b$ then, by definition of d, $d(a,b) = 0$. Hence d satisfies $[\mathbf{M}_1]$.

Let $a,b \in X$. If $a \neq b$, then $b \neq a$. Hence $d(a,b) = 1$ and $d(b,a) = 1$. Accordingly, $d(a,b) = d(b,a)$. On the other hand, if $a = b$ then $b = a$ and therefore $d(a,b) = 0 = d(b,a)$. Hence d satisfies $[\mathbf{M}_2]$.

Now let $a,b,c \in X$ be distinct points. Then $d(a,c) = 1$, $d(a,b) = 1$ and $d(b,c) = 1$. Hence

$$d(a,c) = 1 \leq 1 + 1 = d(a,b) + d(b,c)$$

and d satisfies $[\mathbf{M}_3^*]$.

Lastly, let $a,b \in X$ and $a \neq b$. Then $d(a,b) = 1$. Hence $d(a,b) \neq 0$, and d satisfies $[\mathbf{M}_4]$.

3. Let d be a metric on a non-empty set X. Show that the function e defined by

$$e(a,b) = \min(1, d(a,b))$$

where $a,b \in X$, is also a metric on X.

Solution:

Let $a,b \in X$. Since d is a metric, $d(a,b)$ is non-negative. Hence $e(a,b)$, which is either 1 or $d(a,b)$, is also non-negative. Furthermore, if $a = b$ then

$$e(a,b) = \min(1, d(a,b)) = \min(1,0) = 0$$

Hence e satisfies $[\mathbf{M}_1]$.

Now let $a,b \in X$. By definition $e(a,b) = d(a,b)$ or $e(a,b) = 1$. Suppose $e(a,b) = d(a,b)$; then $d(a,b) < 1$. Since d is a metric, $d(b,a) = d(a,b) < 1$. Consequently,

$$e(b,a) = d(b,a) = d(a,b) = e(a,b)$$

On the other hand, suppose $e(a,b) = 1$; then $d(a,b) \geq 1$. Hence $d(b,a) = d(a,b) \geq 1$. Consequently,

$$e(b,a) = 1 = e(a,b)$$

In either case e satisfies $[\mathbf{M}_2]$.

Now let $a,b,c \in X$. We want to prove the Triangle Inequality

$$e(a,c) \leq e(a,b) + e(b,c)$$

Observe that $e(a,c) = \min(1, d(a,c)) \leq 1$. Hence if $e(a,b) = 1$ or $e(b,c) = 1$, the Triangle Inequality holds. But if both $e(a,b) < 1$ and $e(b,c) < 1$, then $e(a,b) = d(a,b)$ and $e(b,c) = d(b,c)$. Accordingly,

$$e(a,c) = \min(1, d(a,c)) \leq d(a,c) \leq d(a,b) + d(b,c) = e(a,b) + e(b,c)$$

Thus in all cases the Triangle Inequality holds. Hence e satisfies $[\mathbf{M}_3]$.

Finally, let $a,b \in X$ and $a \neq b$. Then $d(a,b) \neq 0$. Hence $e(a,b) = \min(1, d(a,b))$ is also not zero. Thus e satisfies $[\mathbf{M}_4]$.

4. Let d be a metric on a non-empty set X. Show that the function e defined by

$$e(a,b) = \frac{d(a,b)}{1 + d(a,b)}$$

where $a, b \in X$, is also a metric on X.

Solution:

Since d is a metric, e clearly satisfies **[M$_1$]**, **[M$_2$]** and **[M$_4$]**. Hence we only need to show that e satisfies **[M$_3$]**, the Triangle Inequality. Let $a, b, c \in X$; then

$$\frac{d(a,b)}{1 + d(a,b) + d(b,c)} \leq \frac{d(a,b)}{1 + d(a,b)} = e(a,b)$$

and

$$\frac{d(b,c)}{1 + d(a,b) + d(b,c)} \leq \frac{d(b,c)}{1 + d(b,c)} = e(b,c)$$

Since d is a metric, $d(a,c) \leq d(a,b) + d(b,c)$. Hence

$$e(a,c) = \frac{d(a,c)}{1 + d(a,c)} \leq \frac{d(a,b) + d(b,c)}{1 + d(a,b) + d(b,c)}$$

$$= \frac{d(a,b)}{1 + d(a,b) + d(b,c)} + \frac{d(b,c)}{1 + d(a,b) + d(b,c)} \leq e(a,b) + e(b,c)$$

Thus e is a metric.

OPEN SPHERES

5. Prove Lemma 8.2: Let S be an open sphere with center p and radius δ, i.e. $S = S(p, \delta)$. Then for every point $q \in S$ there exists an open sphere T centered at q such that T is contained in S.

Solution:

Now $d(q,p) < \delta$ since $q \in S = S(p, \delta)$. Hence

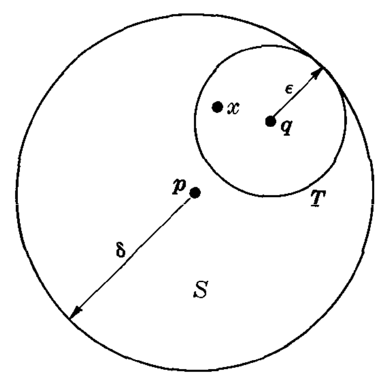
$$\epsilon = \delta - d(q,p) > 0$$

We claim that the open sphere $T = S(q, \epsilon)$, with center q and radius ϵ, is a subset of S.

Let $x \in T = S(q, \epsilon)$. Then $d(x,q) < \epsilon = \delta - d(q,p)$. So, by the Triangle Inequality,

$$d(x,p) \leq d(x,q) + d(q,p) < [\delta - d(q,p)] + d(q,p) = \delta$$

Thus $x \in S = S(p, \delta)$ since its distance from p is less than δ. So $x \in T$ implies $x \in S$, i.e. T is a subset of S (as indicated in the adjacent Venn diagram).



6. Let δ_1 and δ_2 be real numbers such that $0 < \delta_1 \leq \delta_2$. Show that the open sphere $S(p, \delta_1)$ is a subset of the open sphere $S(p, \delta_2)$.

Solution:

Let $x \in S(p, \delta_1)$. Then $d(x,p) < \delta_1 \leq \delta_2$. Hence $x \in S(p, \delta_2)$ and thus $S(p, \delta_1) \subset S(p, \delta_2)$.

7. Show that if S and T are open spheres with the same center, then one of them is a subset of the other.

Solution:

Say $S = S(p, \delta_1)$ and $T = S(p, \delta_2)$, i.e. S and T have the same center p with radii δ_1 and δ_2 respectively. But either $\delta_1 \leq \delta_2$ or $\delta_2 \leq \delta_1$. Hence by the preceding problem either $S \subset T$ or $T \subset S$.

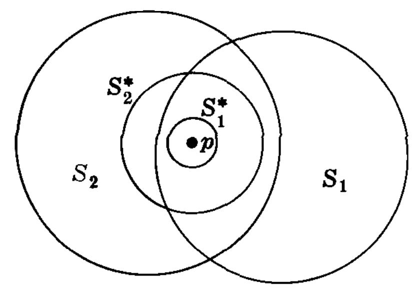
8. Prove Lemma 8.3: Let S_1 and S_2 be open spheres and let $p \in S_1 \cap S_2$. Then there exists an open sphere S_p with center p such that $p \in S_p \subset S_1 \cap S_2$.

Solution:

Since $p \in S_1$ and S_1 is an open sphere, there exists by Lemma 8.2 an open sphere S_1^* with center p such that $p \in S_1^* \subset S_1$. Similarly there exists an open sphere S_2^* with center p such that $p \in S_2^* \subset S_2$. Now S_1^* and S_2^* each has center p; so by Problem 7 one of them, say S_1^*, is contained in the other. Thus we have

$$p \in S_1^* \subset S_1 \qquad \text{and} \qquad p \in S_1^* \subset S_2^* \subset S_2$$

Accordingly, $p \in S_1^* \subset S_1 \cap S_2$. Hence we may take $S_p = S_1^*$. (See adjacent diagram.)



METRIC TOPOLOGIES

9. Prove: Let X be a metric space, and let $\mathcal{D}_p$ denote the class of open spheres with center $p \in X$. Then $\mathcal{D}_p$ is a local base at p.

Solution:

Let G be an open subset of X containing p. Since the open spheres in X form a base for the metric topology, $\exists$ an open sphere S such that $p \in S \subset G$. But by Lemma 8.2 $\exists$ an open sphere $S_p \in \mathcal{D}_p$, i.e. with center p, such that $p \in S_p \subset S \subset G$. Hence $\mathcal{D}_p$ is a local base at p.

10. Prove Theorem 8.5: Let X be a metric space. Then the countable class of open spheres

$$\mathcal{Z} = \{S(p,1),\ S(p,\tfrac{1}{2}),\ S(p,\tfrac{1}{3}),\ \ldots\}$$

with center $p \in X$, is a local base at p.

Solution:

Let G be an open subset of X containing p. By the preceding problem, $\exists$ an open sphere $S(p,\delta)$ with center p such that $p \in S(p,\delta) \subset G$. Since $\delta > 0$,

$$\exists n_0 \in \mathbf{N} \quad \text{such that} \quad 1/n_0 < \delta$$

Accordingly, $p \in S(p, 1/n_0) \subset S(p,\delta) \subset G$ where $S(p,1/n_0) \in \mathcal{Z}$. Hence $\mathcal{Z}$ is a local base at p.

11. Prove Theorem 8.6: The closure $\bar{A}$ of a subset A of a metric space X is the set of points whose distance from A is zero: $\bar{A} = \{x : d(x,A) = 0\}$.

Solution:

Suppose $d(p,A) = 0$. Then every open sphere with center p, and therefore every open set G containing p, also contains at least one point of A. Hence $p \in A$ or p is a limit point of A, and so $p \in \bar{A}$.

On the other hand, suppose $d(p,A) = \epsilon > 0$. Then the open sphere $S(p, \frac{1}{2}\epsilon)$ with center p contains no point of A. Hence p belongs to the exterior of A, and so $p \notin \bar{A}$. Accordingly, $\bar{A} = \{x : d(x,A) = 0\}$.

12. Show that a subset F of a metric space X is closed if and only if $\{x : d(x,F) = 0\} \subset F$.

Solution:

This follows directly from Problem 11 and the fact that a set is closed iff it is equal to its closure.

13. If F is a closed subset of a metric space X and $p \in X$ does not belong to F, i.e. $p \notin F$, then $d(p,F) \neq 0$.

Solution:

If $d(p,F) = 0$ and F is closed, then by Problem 12, $p \in F$. But by hypothesis $p \notin F$; so $d(p,F) \neq 0$.

14. Prove Theorem 8.8: Let A and B be closed disjoint subsets of a metric space X. Then there exist disjoint open sets G and H such that $A \subset G$ and $B \subset H$.

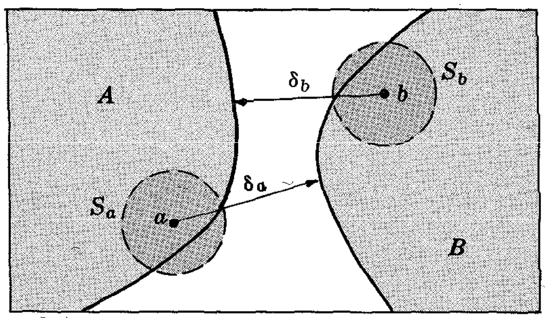
Solution:

If either A or B is empty, say $A = \emptyset$, then $\emptyset$ and X are open disjoint sets such that $A \subset \emptyset$ and $B \subset X$. Hence we may assume A and B are non-empty.

Let $a \in A$. Since A and B are disjoint, $a \notin B$. But B is closed; hence by the preceding problem, $d(a, B) = \delta_a > 0$. Similarly, if $b \in B$, then $d(b, A) = \delta_b > 0$. Set

$$S_a = S(a, \tfrac{1}{3}\delta_a) \quad \text{and} \quad S_b = S(b, \tfrac{1}{3}\delta_b)$$

so $a \in S_a$ and $b \in S_b$. (See the adjacent Venn diagram.)



We claim that the sets

$$G = \bigcup\{S_a : a \in A\} \quad \text{and} \quad H = \bigcup\{S_b : b \in B\}$$

satisfy the required conditions of the theorem. Now G and H are open since they are each the union of open spheres. Furthermore, $a \in S_a$ implies $A \subset G$, and $b \in S_b$ implies $B \subset H$. We must show that $G \cap H = \emptyset$.

Suppose $G \cap H \neq \emptyset$, say $p \in G \cap H$. Then

$$\exists\, a_0 \in A,\ b_0 \in B \quad \text{such that} \quad p \in S_{a_0},\ p \in S_{b_0}$$

Let $d(a_0, b_0) = \epsilon > 0$. Then $d(a_0, B) = \delta_{a_0} \leq \epsilon$ and $d(b_0, A) = \delta_{b_0} \leq \epsilon$. But $p \in S_{a_0}$ and $p \in S_{b_0}$, so

$$d(a_0, p) < \tfrac{1}{3}\delta_{a_0} \quad \text{and} \quad d(p, b_0) < \tfrac{1}{3}\delta_{b_0}$$

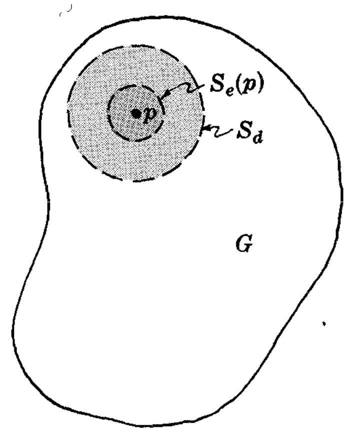
Therefore by the Triangle Inequality,

$$d(a_0, b_0) = \epsilon \leq d(a_0, p) + d(p, b_0) < \tfrac{1}{3}\delta_{a_0} + \tfrac{1}{3}\delta_{b_0} \leq \tfrac{1}{3}\epsilon + \tfrac{1}{3}\epsilon = \tfrac{2}{3}\epsilon$$

an impossibility. Hence G and H are disjoint and the theorem is true.

EQUIVALENT METRICS

15. Let d and e be metrics on a set X such that for each d-open sphere S_d with center $p \in X$ there exists an e-open sphere S_e with center p such that $S_e \subset S_d$. Show that the topology $\mathcal{T}_d$ induced by d is coarser (smaller) than the topology $\mathcal{T}_e$ induced by e, i.e. $\mathcal{T}_d \subset \mathcal{T}_e$.



Solution:

Let $G \in \mathcal{T}_d$. We want to show that G is also an e-open set. Let $p \in G$. Since G is d-open there exists a d-open sphere S_d with center p such that $p \in S_d \subset G$. By hypothesis, there exists an e-open sphere $S_e(p)$ with center p such that $p \in S_e(p) \subset S_d \subset G$. Accordingly, $G = \bigcup\{S_e(p) : p \in G\}$. Thus G is the union of e-open spheres, and so it is e-open. Hence $\mathcal{T}_d \subset \mathcal{T}_e$.

16. Let d and e be metrics on a set X such that for each d-open sphere S_d with center $p \in X$ there exists an e-open sphere S_e with center p such that $S_e \subset S_d$, and for each e-open sphere S_e^* with center $p \in X$ there exists a d-open sphere S_d^* such that $S_d^* \subset S_e^*$. Show that d and e are equivalent metrics, i.e. that they induce the same topology on X.

Solution:

By Problem 15, the topology $\mathcal{T}_d$ induced by d is coarser than the topology $\mathcal{T}_e$ induced by e, i.e. $\mathcal{T}_d \subset \mathcal{T}_e$. Also by Problem 15, $\mathcal{T}_e \subset \mathcal{T}_d$. Therefore $\mathcal{T}_d = \mathcal{T}_e$.

17. Show that the usual metric d on the plane $\mathbf{R}^2$ is equivalent to the metrics d_1 and d_2 on $\mathbf{R}^2$ defined in Example 1.5.

Solution:

Observe that we can inscribe a square in any circle as shown in Fig. (a) below, and we can inscribe a circle in a square as shown in Fig. (b). Now the points inside a circle form a d-open sphere and the points inside a square form a d_1-open sphere, so the metrics d and d_1 are equivalent by Problem 16.

Furthermore, we can inscribe a "diamond" in any circle as shown in Fig. (c), and we can inscribe a circle in any diamond as shown in Fig. (d). Since the points inside a "diamond" form a d_2-open sphere, the metrics d and d_2 are equivalent by Problem 16.

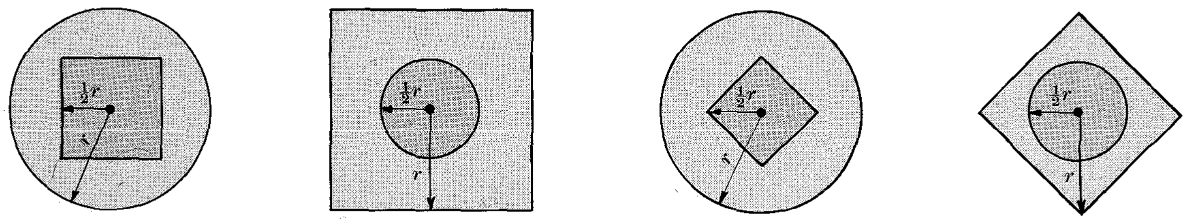


Fig. (a) **Fig.** (b) **Fig.** (c) **Fig.** (d)

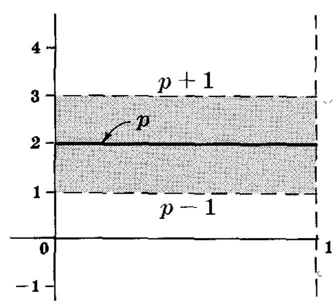
18. Let $\mathcal{C}[0,1]$ denote the collection of all real continuous functions defined on $I = [0,1]$. Consider the metrics d and e on $\mathcal{C}[0,1]$ defined by

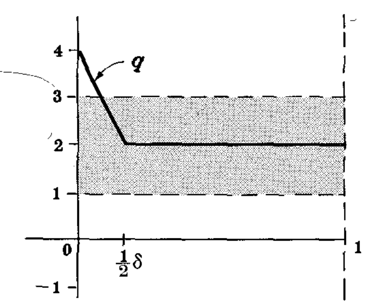
$$d(f,g) = \sup\{|f(x)-g(x)| : x \in I\}, \qquad e(f,g) = \int_0^1 |f(x)-g(x)|\,dx$$

(see Example 1.3 and Example 1.4). Show that the topology $\mathcal{T}_d$ induced by d is not coarser than the topology $\mathcal{T}_e$ induced by e, i.e. $\mathcal{T}_d \not\subset \mathcal{T}_e$.

Solution:

Let p be the constant function $p(x) = 2$ and let $\epsilon = 1$. Then the sphere $S_d(p,\epsilon)$ consists of all functions g for which g lies between the functions $p-1$ and $p+1$, i.e. such that $1 < g(x) < 3$ for all $x \in I$.





It is sufficient to show that $S_d(p,\epsilon)$ contains no e-open sphere with center p; i.e. for every $\delta > 0$, $S_e(p,\delta) \not\subset S_d(p,\epsilon)$. Let $\delta > 0$. Consider the function q consisting of the line segments between the points $(0,4)$ and $(\frac{1}{2}\delta, 2)$ and between $(\frac{1}{2}\delta, 2)$ and $(1,2)$, i.e. defined by

$$g(x) = \begin{cases} (-4x/\delta) + 4 & \text{if } 0 \leq x < \frac{1}{2}\delta \\ 2 & \text{if } \frac{1}{2}\delta \leq x \leq 1 \end{cases}$$

(see diagram above). Observe that the "area" between p and q is $\frac{1}{2}\delta$, i.e. $e(p,q) = \frac{1}{2}\delta$. Then $q \in S_e(p,\delta)$. But $d(p,q) = 2$; so $q \notin S_d(p,\epsilon)$. Thus $S_e(p,\delta) \not\subset S_d(p,\epsilon)$ for any $\delta > 0$. Hence $\mathcal{T}_d \not\subset \mathcal{T}_e$.

19. Let $\mathcal{C}[a,b]$ denote the collection of all continuous functions on a closed interval $X = [a,b]$. Consider the metrics d and e on $\mathcal{C}[a,b]$ defined by

$$d(f,g) = \sup\{|f(x)-g(x)| : x \in X\}, \qquad e(f,g) = \int_a^b |f(x)-g(x)|\,dx$$

Show that the topology $\mathcal{T}_e$ induced by e is coarser than the topology $\mathcal{T}_d$ induced by d, i.e. $\mathcal{T}_e \subset \mathcal{T}_d$.

Solution:

Let $S_e(p, \epsilon)$ be any e-open sphere in $C[a, b]$ with center $p \in C[a, b]$. Let $\delta = \epsilon/(b-a)$. In view of Problem 15 it is sufficient to show that $S_d(p, \delta)$, the d-open sphere with center p and radius δ, is a subset of $S_e(p, \epsilon)$, i.e. $S_d(p, \delta) \subset S_e(p, \epsilon)$.

Let $f \in S_d(p, \delta)$; then $\sup\{|p(x) - f(x)|\} < \delta = \epsilon/(b-a)$

Hence

$$e(p, f) = \int_a^b |p(x) - f(x)|\,dx \leq \int_a^b \sup\{|p(x) - f(x)|\}\,dx < \int_a^b \epsilon/(b-a)\,dx = \epsilon$$

So $f \in S_e(p, \epsilon)$ and therefore $S_d(p, \delta) \subset S_e(p, \epsilon)$.

NORMED SPACES

20. Prove Theorem 8.13: The function d defined by $d(v, w) = \|v - w\|$, where v and w are vectors in a normed space **V**, is a metric on **V**.

Solution:

Note that by **[N₁]**,

$$d(v, w) = \|v - w\| \geq 0 \quad \text{and} \quad d(v, v) = \|v - v\| = \|0\| = 0$$

Hence d satisfies **[M₁]**. Also, by **[N₃]**,

$$d(v, w) = \|v - w\| = \|(-1)(w - v)\| = |-1|\,\|w - v\| = \|w - v\| = d(w, v)$$

Hence d satisfies **[M₂]**. By **[N₂]**, $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in \mathbf{V}$. Accordingly if $a, b, c \in \mathbf{V}$, then substituting $v = a - b$ and $w = b - c$ we have

$$\|a - c\| = \|(a - b) + (b - c)\| = \|v + w\| \leq \|v\| + \|w\| = \|a - b\| + \|b - c\|$$

that is, $d(a, c) \leq d(a, b) + d(b, c)$. Hence d satisfies **[M₃]**.

Finally, if $v \neq w$ then $v - w \neq 0$; hence by **[N₁]**, $d(v, w) = \|v - w\| > 0$. Thus d satisfies **[M₄]**.

21. Prove the Cauchy-Schwarz Inequality: For any pair of points $p = \langle a_1, \ldots, a_m \rangle$ and $q = \langle b_1, \ldots, b_m \rangle$ in $\mathbf{R}^m$,

$$\sum_{i=1}^m |a_i b_i| \leq \|p\|\,\|q\| = \sqrt{\sum_{i=1}^m |a_i|^2}\,\sqrt{\sum_{i=1}^m |b_i|^2}$$

where $\|p\|$ is the Euclidean norm.

Solution:

If $p = 0$ or $q = 0$, then the inequality reduces to $0 \leq 0$ and is therefore true. So we need only consider the case in which $p \neq 0$ and $q \neq 0$, i.e. in which $\|p\| \neq 0$ and $\|q\| \neq 0$.

Now for any real numbers $x, y \in \mathbf{R}$, $0 \leq (x - y)^2 = x^2 - 2xy + y^2$ or, equivalently,

$$2xy \leq x^2 + y^2 \tag{1}$$

Since x and y are arbitrary real numbers, we can let $x = |a_i|/\|p\|$ and $y = |b_i|/\|q\|$ in (1). So, for any i,

$$2\,\frac{|a_i|}{\|p\|}\,\frac{|b_i|}{\|q\|} \leq \frac{|a_i|^2}{\|p\|^2} + \frac{|b_i|^2}{\|q\|^2} \tag{2}$$

But by definition of the Euclidean norm, $\sum |a_i|^2 = \|p\|^2$ and $\sum |b_i|^2 = \|q\|^2$. So summing (2) with respect to i and using $|a_i b_i| = |a_i|\,|b_i|$, we have

$$2\,\frac{\sum_{i=1}^m |a_i b_i|}{\|p\|\,\|q\|} \leq \frac{\sum_{i=1}^m |a_i|^2}{\|p\|^2} + \frac{\sum_{i=1}^m |b_i|^2}{\|q\|^2} = \frac{\|p\|^2}{\|p\|^2} + \frac{\|q\|^2}{\|q\|^2} = 2$$

that is,

$$\frac{\sum_{i=1}^m |a_i b_i|}{\|p\|\,\|q\|} \leq 1$$

Multiplying both sides by $\|p\|\,\|q\|$ gives us the required inequality.

22. Prove Minkowski's Inequality: For any pair of points $p = \langle a_1, \ldots, a_m \rangle$ and $q = \langle b_1, \ldots, b_m \rangle$ in $\mathbf{R}^m$,

$$\|p+q\| \leq \|p\| + \|q\| \quad \text{i.e.} \quad \sqrt{\sum |a_i + b_i|^2} \leq \sqrt{\sum |a_i|^2} + \sqrt{\sum |b_i|^2}$$

Solution:

If $\|p+q\| = 0$, the inequality clearly holds. Hence we need only consider the case in which $\|p+q\| \neq 0$.

Observe that, for real numbers $a_i, b_i \in \mathbf{R}$, we have $|a_i + b_i| \leq |a_i| + |b_i|$. Hence

$$\begin{aligned} \|p+q\|^2 &= \sum |a_i+b_i|^2 = \sum |a_i+b_i|\,|a_i+b_i| \\ &\leq \sum |a_i+b_i|\,(|a_i| + |b_i|) \\ &= \sum |a_i+b_i|\,|a_i| + \sum |a_i+b_i|\,|b_i| \end{aligned}$$

But by the Cauchy-Schwarz Inequality,

$$\sum |a_i+b_i|\,|a_i| \leq \|p+q\|\,\|p\| \quad \text{and} \quad \sum |a_i+b_i|\,|b_i| \leq \|p+q\|\,\|q\|$$

Then
$$\|p+q\|^2 \leq \|p+q\|\,\|p\| + \|p+q\|\,\|q\| = \|p+q\|\,(\|p\| + \|q\|)$$

Since we are considering the case $\|p+q\| \neq 0$, we can divide by $\|p+q\|$; this yields the required inequality.

23. Prove that the Euclidean norm,

$$\|p\| = \sqrt{\sum |a_i|^2} \quad \text{where} \quad p = \langle a_1, \ldots, a_m \rangle \in \mathbf{R}^m$$

satisfies the required axioms $[\mathbf{N}_1]$, $[\mathbf{N}_2]$ and $[\mathbf{N}_3]$.

Solution:

Now $[\mathbf{N}_1]$ follows from properties of the real numbers, and $[\mathbf{N}_2]$ is Minkowski's Inequality which was proven in the preceding problem. Hence we only need to show that $[\mathbf{N}_3]$ holds. But for any vector $p = \langle a_1, \ldots, a_m \rangle$ and any real number $k \in \mathbf{R}$,

$$\begin{aligned} \|kp\| &= \|k\langle a_1, \ldots, a_m\rangle\| = \|\langle ka_1, \ldots, ka_m\rangle\| \\ &= \sqrt{\sum |ka_i|^2} = \sqrt{\sum |k|^2\,|a_i|^2} = \sqrt{|k|^2 \sum |a_i|^2} \\ &= \sqrt{|k|^2}\sqrt{\sum |a_i|^2} = |k|\sqrt{\sum |a_i|^2} = |k|\,\|p\| \end{aligned}$$

Hence $[\mathbf{N}_3]$ also holds.

24. Prove Theorem 8.11: Euclidean m-space is a metric space, i.e. the Euclidean metric on $\mathbf{R}^m$ satisfies the axioms $[\mathbf{M}_1]$ to $[\mathbf{M}_4]$.

Solution:

Use Problem 23 and the fact that the Euclidean metric on $\mathbf{R}^m$ is induced by the Euclidean norm on $\mathbf{R}^m$.

25. Let $\langle a_1, a_2, \ldots \rangle$ be a convergent sequence of real numbers with the property that $a_n \leq b$ for all $n \in \mathbf{N}$. Show that $\lim a_n \leq b$.

Solution:

Suppose $\lim a_n = a > b$ and set $\epsilon = a - b > 0$. Since $a_n \to a$,

$$\exists n_0 \in \mathbf{N} \quad \text{such that} \quad a - a_{n_0} \leq |a - a_{n_0}| < \epsilon = a - b$$

Thus $-a_{n_0} < -b$ and therefore $b < a_{n_0}$, which contradicts the hypothesis. Accordingly, $\lim a_n \leq b$.

26. Prove Minkowski's inequality for infinite sums: If $\langle a_n \rangle, \langle b_n \rangle \in \mathbf{R}^\infty$, then

$$\|\langle a_n + b_n \rangle\| \leq \|\langle a_n \rangle\| + \|\langle b_n \rangle\| \quad \text{i.e.} \quad \sqrt{\sum_{n=1}^{\infty} |a_n + b_n|^2} \leq \sqrt{\sum_{n=1}^{\infty} |a_n|^2} + \sqrt{\sum_{n=1}^{\infty} |b_n|^2}$$

Solution:

By Minkowski's inequality for finite sums,

$$\sqrt{\sum_{n=1}^{m} |a_n + b_n|^2} \leq \sqrt{\sum_{n=1}^{m} |a_n|^2} + \sqrt{\sum_{n=1}^{m} |a_n|^2} \leq \sqrt{\sum_{n=1}^{\infty} |a_n|^2} + \sqrt{\sum_{n=1}^{\infty} |b_n|^2}$$

Since the above is true for every $m \in \mathbf{N}$, by the preceding problem it is also true in the limit.

27. Show that the l_2-norm on $\mathbf{R}^\infty$, i.e. $||\langle a_n \rangle|| = \sqrt{\sum |a_n|^2}$, satisfies the required axioms $[\mathbf{N}_1]$, $[\mathbf{N}_2]$ and $[\mathbf{N}_3]$.

Solution:

This is similar to the proof in Problem 23 that the Euclidean norm satisfies the axioms $[\mathbf{N}_1]$, $[\mathbf{N}_2]$ and $[\mathbf{N}_3]$.

28. Prove Theorem 8.12: Hilbert space (or l_2-space) is a metric space.

Solution:

Use Problem 27 and the fact that the l_2-metric on $\mathbf{R}^\infty$ is induced by the l_2-norm.

29. Let a and b be real numbers with the property that $a \leq b + \epsilon$ for every $\epsilon > 0$. Show that $a \leq b$.

Solution:

Suppose $a > b$. Then $a = b + \delta$ where $\delta > 0$. Set $\epsilon = \frac{1}{2}\delta$. Now $a > b + \frac{1}{2}\delta = b + \epsilon$ where $\epsilon > 0$. But this contradicts the hypothesis; so $a \leq b$.

30. Let $I = [0, 1]$. Show that the following is a norm on $\mathcal{C}[0, 1]$: $||f|| = \sup\{|f(x)|\}$.

Solution:

Recall that a real continuous function on a closed interval is bounded; so $||f||$ is well-defined. Since $|f(x)| \geq 0$ for every $x \in I$, $||f|| \geq 0$; also $||f|| = 0$ iff $|f(x)| = 0$ for every $x \in I$, i.e. iff $f = 0$. Thus $[\mathbf{N}_1]$ is satisfied.

Let $\epsilon > 0$. Then $\exists x_0 \in I$ such that

$$\begin{aligned} ||f + g|| &= \sup\{|f(x) + g(x)|\} \leq |f(x_0) + g(x_0)| + \epsilon \\ &\leq |f(x_0)| + |g(x_0)| + \epsilon \\ &\leq \sup\{|f(x)|\} + \sup\{|g(x)|\} + \epsilon \\ &= ||f|| + ||g|| + \epsilon \end{aligned}$$

Hence by Problem 29, $||f + g|| \leq ||f|| + ||g||$ and $[\mathbf{N}_2]$ is satisfied.

Now let $k \in \mathbf{R}$. Then

$$\begin{aligned} ||kf|| &= \sup\{|(kf)(x)|\} = \sup\{|k\,f(x)|\} = \sup\{|k|\,|f(x)|\} \\ &= |k| \sup\{|f(x)|\} = |k|\,||f|| \end{aligned}$$

and $[\mathbf{N}_3]$ is satisfied.

Supplementary Problems

METRICS

31. Let $\mathcal{B}(X, Y)$ be the collection of all bounded functions from an arbitrary set X into a metric space (Y, d). Show that the function e is a metric on $\mathcal{B}(X, Y)$:

$$e(f, g) = \sup\{d(f(x), g(x)) : x \in X\}$$

32. Let $d_1, \ldots, d_m$ be metrics on $X_1, \ldots, X_m$ respectively. Show that the following functions are metrics on the product set $X = \prod_i X_i$:

$$d(p, q) = \max\{d_1(a_1, b_1), \ldots, d_m(a_m, b_m)\}, \qquad e(p, q) = d_1(a_1, b_1) + \cdots + d_m(a_m, b_m)$$

Here, $p = \langle a_1, \ldots, a_m \rangle$, $q = \langle b_1, \ldots, b_m \rangle \in X = \prod_i X_i$.

33. Let $\mathbf{R}^* = \mathbf{R} \cup \{\infty, -\infty\}$ be the extended real line and let $f: \mathbf{R}^* \to [-1, 1]$ be defined by $f(x) = x/(1+|x|)$ if $x \in \mathbf{R}$, $f(\infty) = 1$ and $f(-\infty) = -1$. Show that the following function is a metric on $\mathbf{R}^*$: $d(x, y) = |f(x) - f(y)|$.

34. Let $\mathbf{R}^+$ denote the non-negative real numbers, and let $f: \mathbf{R}^+ \to \mathbf{R}^+$ be a continuous function such that (i) $f(0) = 0$, (ii) $f(x+y) \leq f(x) + f(y)$, and (iii) $x < y$ implies $f(x) < f(y)$. Show that if d is a metric on any set X then the composition function $f \circ d$ is also a metric on X.

35. Let ρ be a pseudometric on some set X. Let $\sim$ be the relation in X defined by

$$a \sim b \quad \text{iff} \quad \rho(a, b) = 0$$

(i) Show that $\sim$ is an equivalence relation in X.

(ii) Show that the following function is a metric on the quotient set $X/\!\sim \; = \{[a] : a \in X\}$: $d([a],[b]) = \rho(a, b)$. Here $[a]$ denotes the equivalence class of $a \in X$.

36. Let $\mathcal{R}[0, 1]$ denote the collection of (Riemann) integrable functions on $[0, 1]$. Show that the following function is a pseudometric on $\mathcal{R}[0, 1]$:

$$\rho(f, g) \quad = \quad \int_0^1 |f(x) - g(x)| \, dx$$

Also show by a counterexample that ρ is not a metric.

37. Show that a function d is a metric on a set X iff it satisfies the following two conditions: (i) $d(a, b) = 0$ iff $a = b$; (ii) $d(a, c) \leq d(a, b) + d(c, b)$.

DISTANCES BETWEEN SETS, DIAMETERS

38. Give an example of two closed subsets A and B of the real line $\mathbf{R}$ such that

$$d(A, B) = 0 \quad \text{but} \quad A \cap B = \emptyset$$

39. Let d be a metric on X. Show that for any subsets $A, B \subset X$: (i) $d(A \cup B) \leq d(A) + d(B) + d(A, B)$ and (ii) $d(\bar{A}) = d(A)$.

40. Let d be a metric on X and let A be any arbitrary subset of X. Show that the function $f: X \to \mathbf{R}$ defined by $f(x) = d(x, A)$ is continuous.

41. Consider the function $d: \mathbf{R}^2 \to \mathbf{R}$ defined by $d(\langle a, b\rangle) = |a - b|$ (i.e. the usual metric on $\mathbf{R}$). Show that d is continuous with respect to the usual topologies on the line $\mathbf{R}$ and the plane $\mathbf{R}^2$.

42. Let A be any subset of a metric space X. Show that $d(A) = d(\bar{A})$.

METRIC TOPOLOGIES

43. Let (A, d) be a metric subspace of (X, d). Show that (A, d) is also a topological subspace of (X, d), i.e. the restriction of d to A induces the relative topology on A.

44. Prove: If the topological space $(X, \mathcal{T})$ is homeomorphic to a metric space (Y, d), then $(X, \mathcal{T})$ is metrizable.

45. Prove Theorem 8.10: If (X, d) is isometric to (Y, e), then (X, d) is also homeomorphic to (Y, e).

46. Give an example to show that the closure of an open sphere

$$S(p, \delta) \quad = \quad \{x : d(p, x) < \delta\}$$

need not be the "closed sphere"

$$\bar{S}(p, \delta) \quad = \quad \{x : d(p, x) \leq \delta\}$$

47. Show that a closed sphere $\bar{S}(p, \delta) = \{x : d(p, x) \leq \delta\}$ is closed.

48. Prove: The sequence $\langle a_1, a_2, \ldots \rangle$ converges to the point p in a metric space X if and only if the sequence of real numbers $\langle d(a_1, p), d(a_2, p), \ldots \rangle$ converges to $0 \in \mathbf{R}$, i.e. $\lim a_n = p$ iff $\lim d(a_n, p) = 0$.

49. Prove: If $\lim a_n = p$ and $\lim b_n = q$ in a metric space X, then the sequence of real numbers $\langle d(a_1, b_1), d(a_2, b_2), \ldots \rangle$ converges to $d(p, q) \in \mathbf{R}$, i.e. $\lim d(a_n, b_n) = d(\lim a_n, \lim b_n)$.

EQUIVALENT METRICS

50. Let d be a metric on X. Show that the following metric is equivalent to d: $e(a, b) = \min\{1, d(a, b)\}$.

51. Let d be a metric on X. Show that the following metric is equivalent to d: $e(a, b) = \dfrac{d(a, b)}{1 + d(a, b)}$.

52. Let d and e be metrics on X. Suppose $\exists k, k' \in \mathbf{R}$ such that, for every $a, b \in X$,

$$d(a, b) \leq k\, e(a, b) \quad \text{and} \quad e(a, b) \leq k'\, d(a, b)$$

Show that d and e are equivalent metrics.

EUCLIDEAN m-SPACE, HILBERT SPACE

53. Let $p_1 = \langle a_{11}, a_{12}, \ldots, a_{1m} \rangle$, $p_2 = \langle a_{21}, a_{22}, \ldots, a_{2m} \rangle$, $\ldots$ be points in Euclidean m-space. Show that $p_n \to q = \langle b_1, b_2, \ldots, b_m \rangle$ if and only if, for $k = 1, \ldots, m$, $\langle a_{1k}, a_{2k}, a_{3k}, \ldots \rangle$ converges to b_k; i.e. the projection $\langle \pi_k(p_n) \rangle$ converges to $\pi_k(q)$ in each coordinate space.

54. Show that if G is an open subset of Hilbert Space $\mathbf{H}$, then $\exists p = \langle a_n \rangle \in G$ such that $a_1 \neq 0$.

55. Let H^* denote the proper subspace of Hilbert Space $\mathbf{H}$ which consists of all points in $\mathbf{H}$ whose first coordinate is zero. (i) Show that H^* is closed. (ii) Show that H^* is nowhere dense in $\mathbf{H}$, i.e. $\operatorname{int}(\overline{H^*}) = \emptyset$.

56. Let $p_1 = \langle a_{11}, a_{12}, \ldots \rangle$, $p_2 = \langle a_{21}, a_{22}, \ldots \rangle$, $\ldots$ be points in $\mathbf{R}^\infty$ and suppose that the sequence of real numbers $\langle \pi_k(p_n) \rangle = \langle a_{1k}, a_{2k}, a_{3k}, \ldots \rangle$ converge to $b_k \in \mathbf{R}$ for every $k \in \mathbf{N}$.

(i) Show that $q = \langle b_1, b_2, \ldots \rangle$ belongs to $\mathbf{R}^\infty$.

(ii) Show that the sequence $\langle p_1, p_2, \ldots \rangle$ converges to q.

HILBERT CUBE

57. The set $\mathbf{I}$ of all real sequences $\langle a_1, a_2, \ldots \rangle$ such that $0 \leq a_n \leq \frac{1}{n}$, for every $n \in \mathbf{N}$, is called the *Hilbert cube*.

(i) Show that $\mathbf{I}$ is a subset of $\mathbf{R}^\infty$.

(ii) Show that $\mathbf{I}$ is a closed and bounded subset of $\mathbf{R}^\infty$.

NORMED SPACES

58. Let $\mathcal{B}(X, \mathbf{R})$ denote the class of all real bounded functions $f: X \to \mathbf{R}$ defined on some non-empty set X. Show that the following is a norm on $\mathcal{B}(X, \mathbf{R})$: $||f|| = \sup\{|f(x)| : x \in X\}$.

59. Two norms, $||\cdots||_1$ and $||\cdots||_2$, on a linear space X are equivalent iff they induce equivalent metrics on X, i.e. iff they determine the same topology on X. Show that $||\cdots||_1$ is equivalent to $||\cdots||_2$ if and only if $\exists a_1, a_2, b_1, b_2 \in \mathbf{R}$ such that, for all $x \in X$,

$$a_1 ||x||_1 < ||x||_2 < b_1 ||x||_1 \quad \text{and} \quad a_2 ||x||_2 < ||x||_1 < b_2 ||x||_2$$

60. Let $||\cdots||$ be the Euclidean norm and let d be the induced Euclidean metric on the plane $\mathbf{R}^2$. Consider the function e defined by

$$e(p,q) \quad = \quad \begin{cases} ||p|| + ||q|| & \text{if } ||p|| \neq ||q|| \\ d(p,q) & \text{if } ||p|| = ||q|| \end{cases}$$

(i) Show that e is a metric on $\mathbf{R}^2$.

(ii) Describe an open sphere in the metric space $(\mathbf{R}^2, e)$.

61. Show that the following is a norm on $C[0,1]$: $\quad ||f|| \;=\; \int_0^1 |f(x)|\,dx.$

62. Let X be a normed space. Show that the function $f: X \to R$ defined by $f(x) = ||x||$ is continuous.

Answers to Supplementary Problems

36. The function $f:[0,1] \to \mathbf{R}$ defined by

$$f(x) \quad = \quad \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } 0 < x \leq 1 \end{cases}$$

is (Riemann) integrable, i.e. belongs to $\mathcal{R}[0,1]$. The zero function $g:[0,1] \to \mathbf{R}$, i.e. $g(x) = 0$ for all $x \in [0,1]$, also belongs to $\mathcal{R}[0,1]$. But $\rho(f,g) = 0$ and $f \neq g$. Hence ρ is not a metric as it does not satisfy $[\mathbf{M_4}]$.

38. Let $A = \{2, 3, 4, 5, \ldots\}$ and $B = \{2\tfrac{1}{2}, 3\tfrac{1}{3}, 4\tfrac{1}{4}, \ldots\}$.

46. Let d be the trivial metric on a set X containing more than one point. Then, for any $p \in X$,

$$S(p,1) \quad = \quad \{x : d(p,x) < 1\} \quad = \quad \{p\}$$

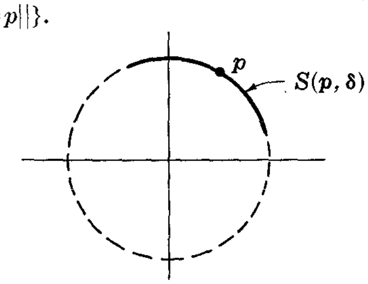
$$\bar{S}(p,1) \quad = \quad \{x : d(p,x) \leq 1\} \quad = \quad X$$

But d induces the discrete topology on X, and so every subset of X is both open and closed. Thus

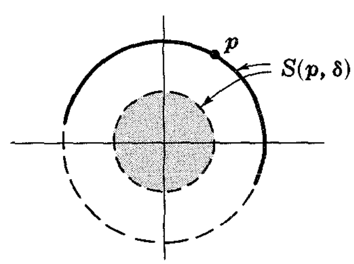
$$\overline{S(p,1)} \;=\; \overline{\{p\}} \;=\; \{p\} \;\neq\; \bar{S}(p,1)$$

58. *Hint.* Proof is similar to that of Problem 30.

60. (ii) If $||p|| \geq \delta$, then $S(p,\delta)$ is an arc of the circle $\{x : ||x|| = ||p||\}$. If $||p|| < \delta$, then $S(p,\delta)$ consists of the points interior to the circle $\{x : ||x|| = \delta - ||p||\}$ and the points on an arc of the circle $\{x : ||x|| = ||p||\}$.



$||p|| \geq \delta$



$||p|| < \delta$

61. $$||f+g|| \;=\; \int_0^1 |f(x) + g(x)|\,dx \;\leq\; \int_0^1 (|f(x)| + |g(x)|)\,dx$$

$$= \; \int_0^1 |f(x)|\,dx + \int_0^1 |g(x)|\,dx \;=\; ||f|| + ||g||$$