THEORETICAL COMPUTER SCIENCE 1

ONLY STUDY GUIDE FOR

COS1501

Author
T HÖRNE

Co-author
D BECKER

Critical reader
A E DU PREEZ

Editor
Carina Potgieter

Graphic designer
Ella Viljoen

Photographs
IlzeBotha (082 772 2482)
Shutterstock

SCHOOL OF COMPUTING
UNIVERSITY OF SOUTH AFRICA
PRETORIA

COS1501/1/2014-2020

ACKNOWLEDGEMENTS

In previous years many lecturers played a role in the development of the COS1501 (previously COS101S) study material. Our sincere thanks go to Willem Labuschagne, Martha Pistorius, Biffie Viljoen, Ruth de Villiers and Louise Leenen – some of whom are no longer lecturers in the School of Computing. We also thank Jeanetta du Preez for creating the mind map in study unit 1, the diagram for the worked example in section 1.2.2, and figures 5.1, 5.2 and 5.3 in study unit 5.

This study guide for COS1501 was reviewed and approved by the following team:

ROLE	**NAME**
Director, School of Computing	Sheryl Buckley
Educational advice	Hentie Wilson
Chair of Committee	Hentie Wilson
Academic field specialists	Sihem Belabbes
	Biffie Viljoen (study units 9 and 10)
Layout format	Hentie Wilson
Page layout	School of Computing
Printed and published by the	University of South Africa, Muckleneuk, Pretoria

CONTENTS

Introduction

1. What is discrete mathematics?

"Many problems of science deal with quantities so large that it is natural to assume that they are dense, continuously distributed, and that all real numbers can be used to measure them; centuries of development of "continuous mathematics" have given us extremely powerful tools for handling problems of this kind. Other problems are so small that we can deal with all the possible cases by hand. These are truly "finite" and defined problems. Some of the most important problems, however, fall in between: not big enough to assume density, continuity, etc., but not small enough to allow us to consider all cases. These intermediate problems are, for the most part, the problems with which discrete mathematics deals." (Roberts, 2001: 3743). Discrete mathematics became one of the fastest growing fields of modern mathematics because many of the physical and biology sciences problems such as time, mass, velocity, involve very large quantities. However, many problems fall in the middle ground and for these the tools of DISCRETE MATHEMATICS are especially relevant (according to Roberts, 2001: 3743, referring to the work by the mathematician and philosopher, Kemeny) With the help of powerful computers we are now able to replace computations done manually with computations done by computer.

> *"Mathematics is about solving problems. Mathematics explains patterns. Mathematics is a set of statements deduced logically from axioms and definitions. Mathematics uses abstraction to model the real world." (Ensley & Crawley, 2006)*

This module deals with the sub-field of the domain of discrete (not continuous) mathematics which is relevant to computing. In the module we shall be using tools such as

- set theory,
- relations and functions,
- vector and matrix manipulations,
- mathematical proofs, and
- propositional logic.

Rational-number concepts are among the most complex and important mathematical ideas children encounter during their secondary school years. We can consider the importance of this from a variety of perspectives, for example:

(a) *From a practical perspective*, our ability to deal effectively with rational number concepts vastly improves our understanding and our ability to handle situations and problems in the real world.
(b) *From a psychological perspective*, rational numbers provide a rich space within which we can develop and expand the mental structures necessary for our continued intellectual development.
(c) *From a mathematical perspective*, our understanding of rational numbers provides the foundation upon which our ability to do elementary algebraic operations can later be based.

Activity 1: Play ***Pre-knowledge and indigenous knowledge***

Have some fun by playing any of the following games with a friend or family member while understanding that mathematics forms their basis:

* Eeny-meeny-miney-moe

* One potato, 2 potato, 3 potato, four, 5 potato, 6 potato, 7 potato, MORE (on "MORE" the person is OUT)

* Draw an outline of an envelope (a square) with a cross over it without lifting your pen

* Grid game / noughts and crosses (3×3 or 4×4 grids)

What other games such as these did you play as a child? Describe them in a paragraph. This is great knowledge from your childhood and, if you wish, you can share it with others in the module's online discussion forum.

2. The purpose of the module

On completing this module, you will be able to critically apply the fundamental knowledge and skills of discrete mathematics. The module forms part of the theoretical foundation of a Computer Science major. This background is relevant to computing fields such as relational databases, the development of provably correct programs, and the analysis of algorithms that will contribute to the development of computing in Southern Africa, Africa, or globally. The module will support further studies and applications in the computing discipline.

3. Outcomes of the module

- *Specific outcome 1:* Think in an abstract way to construct logical arguments, using a variety of mathematical tools.
- *Specific outcome 2:* Construct proofs in a clear and concise way using mathematical reasoning techniques.
- *Specific outcome 3:* Demonstrate knowledge and understanding of the definitions, laws and operations of set theory.
- *Specific outcome 4:* Synthesise and critically analyse relations, functions and binary sets that are represented as sets containing ordered pairs.
- *Specific outcome 5:* Perform operations on vectors and matrices.

4. Syllabus

The syllabus topics include:

- Number sets
- Set theory
- Relations and functions
- Binary operations
- The fundamentals of logic

The following knowledge (declarative, causal, procedural, and contextual) is embedded in the module, and these "big ideas" will be assessed directly or indirectly through the assessment of the specific outcomes against the assessment criteria:

- How are mathematical writing skills different from everyday English?
- Which mathematical writing skills can be used as tools to solve logic problems?
- How does (the underpinning skill of) abstract reasoning (abstract structures) support problem-solving skills?
- How is the abstraction of facts and properties used to reason about events in the real world?
- How does set theory underpin the solving of everyday problems?
- How are operations on vectors and matrices applied in order to construct different ways of storing and listing numbered information in computing?
- How are the properties of relations and functions used to reason about events in the real world?
- In what way does academic rigour lead to disciplined and logical reasoning?
- How do you think, explore, write, and discuss in order to make connections between different abstract and concrete mathematical concepts?
- What mathematics underpins our play in puzzles, games of chance, chess, and card games?

5. How to study this module

Many textbooks or study guides provide us with detailed text that we have to read with comprehension and insight to make our own sense of the concepts and to practise the skills and internalise the values. Sometimes we may find science textbooks or study guides hard to read, until we start to apply the following three techniques, which often help at the start of a new study journey (you might find others elsewhere).

Activity 2: Independent study

(i) If at all possible, make your **study space** a separate space to support your independence. A crucial phase in the process of understanding and learning mathematics by problem solving is to articulate your ideas about mathematics, both orally (to hear yourself speak) and in writing. For better understanding it often helps to recite to yourself the materials that you read in your texts.

Train your brain to **think maths** at a certain time and in a certain place. Eventually it will take you no longer than 10 minutes per day to get in a maths mood. Not only will you save the time and emotional energy you once needed to psych yourself up to do maths, it will also help you remember more of what you are studying.

Use the "association learning concept": Attempt, as closely as possible, to study the **same subject at the same time in the same place every day**. You will find that, after a very short while, when you get to that time and place, you will automatically be in the subject "groove".

(ii) **Get help early** if there is something in the study material that you do not understand. Lagging or losing time is similar to committing academic suicide. Maths requires a sequential learning process, so if you fall behind, it will be difficult to catch up. Each topic builds on the previous one. It would be like going to a Spanish class without learning the current set of

vocabulary words – the lecturer/tutor would be talking to you using the new vocabulary, but you would not understand what is being said.

(iii) After studying, reinforce **the pleasure of studying** by doing something fun, such as watching television or going to a party. Experts have established that the positive reinforcement of behaviour (such as studying) will increase its frequency and duration.

Different kinds of documents hold information in different places and in different ways; these documents have different depths and breadths of coverage. By understanding the layout of the material you are reading, you can extract useful information much more effectively. You should also use the most appropriate reading strategy for each different document.

Where you only need the shallowest knowledge of the subject, you can ***scan*** material: read only the headings, introductions and summaries (like when you page through a magazine). If you need a moderate level of information on a subject, then you can ***skim*** the text: read the introductions and summaries in detail. You may also ***speed-read*** the contents, picking out and understanding key words and concepts, and paying attention to diagrams and graphs. Only when you need detailed knowledge of a subject is it worth ***studying*** the text. So, skim the material to get an overview of the subject and to get an understanding of its structure, so that you can fit the detail gained from a full, receptive reading of the material into that structure. We look at each of these reading methods in more detail in the activities that follow.

The way one should read a maths textbook or study guide is different from the traditional way students are taught to read textbooks in high school or college. Often, students are taught to read quickly or skim the material and, if they do not understand a word, they are supposed to keep on reading. Instructors of other courses want students to continue to read, so that they can pick up the unknown words and their meanings from the context. NOTE: This reading technique may work with your other modules, but using it in your maths course will probably leave you totally confused. If you skip some of the major concepts or words printed in bold or italics, you will not understand the concepts and will not be able to do the activities. In a mathematical subject, it might take you half an hour to read and understand just one page. If you do not understand everything in a section, you should go back, so make a note in the margin.

BEWARE: If you have a reading problem, it would be wise to spend time on a developmental reading course to increase your understanding of academic English (the language of this course) before taking any maths course, especially if you are studying in a language which is not your mother tongue, in which reading and writing are more difficult. This also applies if you are not used to studying, or if you have not established good study habits.

Scan-Question-Read-Recall-Review (SQ3R) is a good study technique for getting a deep understanding of a text. To some extent, this technique could also be used for this particular module.

- ***Scan*** the contents, introduction, chapter introductions and chapter summaries to pick up a shallow overview of the text. Form an opinion of whether it will be of any help. If it does not give you the information you want, discard it.
- ***Question*** while you ***skim*** the page, and make a note of any questions on the subject that come to mind, or of anything that particularly interests you. Perhaps skim the document again to see if anything stands out. These questions can be

considered study goals – understanding the answers can help you to structure the information in your own mind.

- ***(Study) Read*** the document while taking notes. Read through individual sections in detail, taking care to understand all the points that are relevant. In the case of some texts this reading may be very slow. This will particularly be the case if there is a lot of dense and complicated information. While you are reading, it can help to take notes in a mind map or concept map format. A concept map is a drawing that shows key ideas and relevant relationships (see activity 1 of each study unit).
- ***Recall*** after you have read appropriate sections of the document. Run through it in your mind several times. Isolate the core facts or the essential processes behind the subject, and then see how other information fits around them.
- ***Review*** the detail once you have run through the whole exercise above. This review can be done by rereading the document, by expanding your notes, or by discussing the material with your fellow students (peers) or with your tutor. A particularly effective method of reviewing information is to have to explain it to someone else in a group.

(From the webpage, Mindtools, for more details http://www.mindtools.com/pages/article/newISS_02.htm).

Activity 3: Scan-read *Study technique*

Quickly page through and SCAN the **headings with the paragraphs of description below them** (ignoring the Examples and Problems at first). This allows you to form a rough idea of the contents.

READ the **contents** page, and then go back and forth. Ask QUESTIONS such as:

- Which sections are important?
- Which sections are relevant to the different assignments you are given in this module?

While making this cursory survey, ask yourself:

- What are the key terms or themes? Stop when you identify a key term, read carefully what is said about it, and mark it in the study guide so that you can find it easily later on.
- What is "hidden" in the guide? There may be helpful **index** pages, or pages with summaries of formulas, special symbols and notations. In this guide, for example, there is a whole list of very important symbols that are key to your understanding. (What is interesting – and what makes it difficult – is that these symbols and notations often differ, depending on the conventions of the region in which they are used. We will alert you to such differences in the course of the study guide.)

Scanning is a technique you often use when looking up a word in the telephone book or dictionary, or when looking for the answer to an assignment question, where you search for key words or ideas. In most cases, you know what you're looking for, so you're concentrating on finding a particular answer. Scanning involves moving your eyes quickly down the page, seeking specific words and phrases. Scanning is also used when you first find a resource, to determine whether it will answer your questions. Once you've scanned a document, you might go back and skim-read it.

When scanning, look for the author's use of organisers (such as numbers, letters, steps, or the words "first", "second", or "next"). Look for words that are **bold**-faced, in *italics*, or in a different font size, style, or colour.

Research shows that people have more difficulty when reading off a computer screen than when reading off paper. Although they can read and comprehend at the same rate as when reading from paper, skimming on the computer is much slower than on paper.

Activity 4: Skim-read *Study technique*

CIRCLE the new words that you do not understand (using a pencil) and quickly identify the main ideas of the text. When you read the newspaper, you're probably not reading it word by word. Instead, you're skimming the text. Skimming is done at a speed three to four times faster than normal reading. People often skim when they have lots of material to read in a limited amount of time.

If you are reading large amounts of difficult mathematics or technical words, it may be useful to compile a **GLOSSARY** (a word list) during this skimming exercise. Keep this beside you as you read. It could also be useful to note down further explanations of the key concepts in your own words, and refer to them when necessary. (To help you, we have included the beginnings of a glossary as activity 2 in each study unit, to which you can add your own words.)

BEWARE: Over-underlining is a common fault of students; only the key words in a paragraph should be underlined. It should be done in ink or with a felt-tip highlighter, and it should be done only after you have finished the first part of your reading.

Research has shown that it is not how much time you study that is important; all that counts is how well you study during a given time. In fact, in at least one survey, students who studied more than 35 hours a week came out with poorer grades than those who studied less. Still, you will have to study at least 2 hours every day of the week to be successful in this module. Do not underestimate the volume of the module – it is a 120 study-hour module, and you will earn 12 credits when completing it.

There are many strategies that can be used when skimming. Some people read the first and last paragraphs using headings and summaries when they proceed through a document. You might only read the title, subtitles, subheadings, and illustrations. As a start, consider reading the first sentence of each paragraph. This technique is useful when you're seeking specific information, rather than reading for comprehension. Skimming works well to find specific information. It can also be used to get an overview of the information contained within graphs, tables, and charts. Take care not to skip these, but to find out what these graphics say and note down what you think they say.

Skim the section or chapter with the aim of starting a mind map. Look for items and concepts while reading the information in the section or unit in a more evaluative way.

Then, reflect on interrelationships between the identified concepts. The QUESTION now is:

- What are the interrelationships?
- What are the meaning and the purpose of these?

Visualisation is important and you are certainly going to start writing down key concepts. When you are reading a document in detail, it often helps if you highlight and underline important points/concepts and annotate as you go along. This emphasises information, and helps you to review important points later. Doing this

also helps to keep your mind focused on the material and stops it from wandering. It is best to make notes as you go along. Creating concept maps or using the study system suggested here are effective ways of studying.

Activity 5: Study-read *Study technique*

The next stage is to go back and start over, study-reading the paragraphs and making short notes (for example, on the DEFINITIONS, FORMULAS or the ACTIVITIES) in your mind map.

Study-reading is done carefully, thoroughly and thoughtfully. The key terms and concepts you have pinpointed have to be linked up, and for this the mind map and summaries are important. Summaries and mind maps also fix the knowledge more firmly in your mind. Pause while reading, consolidate what you remember, and consider how new information fits in with what you already know. We want to broaden your perspective and outlook, help you to identify problems and help you to resolve them in a new way in the context of teaching and learning.

Deeper reflection is where you expand the structure of the mind map, working towards a holistic picture. (Later, as you work through the prescribed activities of the section or chapter, keep returning to the mind map to fill in the detail.) Reflect on the value and meaning or categories, concepts, reasons, variables, formulas, and key terms.

Concentrate on including text in **bold** and *italic* type, boxes, tables and illustrations, summaries, and introductions. The objectives (or a bold introduction to a chapter) are very important for this overview, so use it to ensure you have all the key items in your map.

Read in such a manner that you will be able to make a summary of the contents of the chapter. Take care to look for the linking paragraphs that precede sub-sections, such as another definition that might extend the argument or indicate further exclusions or unique applications. The map will give you an overview of the story-line which you are going to study in detail.

Make up a colour and sign system for highlighting text and notes, such as

- red for main ideas
- blue for dates and numbers
- yellow for supporting facts
- circles, boxes, stars and checks in the margins to make reviewing easy

Start your own glossary of the words and concepts you do not know, and use a scientific dictionary to find the meanings of the words in a scientific context.

Watch for linking words such as “therefore” and “in essence”, which tell you what is being summarised.

Always record examples (usually indicated by “for example”). In fact, in subjects such as maths, your notes should focus mainly on your lecturer/tutor’s examples.

Take your time now to make a mind map of the whole guide (just the outline). Look out for these main ideas:

- Background: Study units 1, 2
- Set theory: Study units 3, 4
- Relations and functions: Study units 5, 6, 7
- Binary operations: Study unit 8
- Introduction to logic: Study units 9, 10

Now, consult your tutorial letter 101, and indicate on your mind map how the assignments will cover the sections above.

Activity 6: Work through a computer-aided instruction tutorial

You should receive a CD with an interactive computer-aided instruction (CAI) tutorial named "Relations". Information about this tutorial, including the navigation thereof, is provided in tutorial letter 101. This supplementary study aid covers the topics "sets" and "relations". It deals with basic sets and the main properties of relations such as reflexivity, irreflexivity, symmetry, antisymmetry and transitivity. It also explores the properties of different types of relation. You can work interactively through theory, examples and exercises. These concepts are discussed in study units 3 to 6 of the study guide. During the course of these study units we remind you to do specific parts of the tutorial by including the following picture in an activity: 

Activity 7: Recall examples and review activities

The next stage is to go back and start over. Yes, read it again (because it is through repetition that you can get new insight), and then go to the EXAMPLES.

Write out the example question in your notebook, and make sure that you understand the underlying concepts. Then write out your arguments, without looking at the solution. When you're done, compare your solution to the model solution. At this point, take care to understand that messy is good and that the authors of the guide have taken out the messy thinking to give you neat solutions – which is not necessarily how mathematicians work. If you cannot follow some argument, make a note of what you think and what your question is, and go on. Sometimes insight comes later when you come back.

Whenever you get to an activity, complete the activity in full, either in your notebook or on loose pages inserted and grouped together in the plastic folders of your file. Supplement this with your own notes.

Doing activities can be frustrating, or rewarding. Many students jump right into the activities, become frustrated, and stop studying. These students usually go directly to the maths problems and start working on them, without any preparation. When they get stuck on a problem, they read the solution. Then, they either try to work the problem backward to understand the steps in the problem, or they just copy down the answer. Other students go to the solutions and simply copy the steps. After getting stuck several times, these students will inevitably quit doing the activities. Doing the activities becomes a frustrating experience, and they may even quit working on the module altogether.

When doing activities, write down every step of the problem. Even if you can do the step in your head, write it down anyway. This will increase the amount of activity

time, but will be worthwhile. Doing every step is an easy way to memorise and understand the material. Another advantage is that when you rework the problems you did incorrectly, it is easy to review each step to find the mistake. In the long run, doing every step of the activity will save you time and frustration.

Understand the reasons for each step of the problem and check your answers. Do not get into the bad habit of memorising how to do problems without knowing the reasons for each step. Many students are smart enough to memorise the procedures required to complete a set of problems. However, when similar problems are presented in an exam, the student cannot solve them. To avoid this dilemma, keep reminding yourself about the rules, laws, or properties used to solve any given problem.

You should check the answers to your activities and to your assignments. Make a point of checking the answers to the self-assessment exercises in this study guide.

- First, check your answer by estimating the correct answer. Example: If you are multiplying 2.234 by 5.102 the answer should by a little over 10. Remember to estimate that 2 times 5 is 10.
- You can sometimes check your answers by substituting the answer back into the equation. The more you hone a skill, the faster you will become. This is very important, because increasing your reading and answer checking speed can help you to do quick checks to avoid careless errors in your assignments.

If you do not understand how to solve a problem, then

- review the material in this study guide that relates to the problem
- review the notes in the study unit text that relate to the problem
- review any similar problems, diagrams, examples or rules that explain the misunderstood material
- refer to a maths textbook, solutions guide, math computer program or DVD, or the internet to obtain a better understanding of the material
- call a friend
- skip the problem and contact your tutor or lecturer for help

Always finish your activity by successfully completing problems. Even if you get stuck, go back and successfully complete previous problems before quitting. You should end your activity assignment with feelings of success.

After finishing your activity, recall or write down the most important concepts you have learned. Recalling this information makes it easier to master these new concepts.

Once you know the correct reason for going from one step to another in solving a maths problem, you can answer any problem which requires a specific technique. Students who simply memorise how to solve problems instead of understanding the reasons for correctly working the steps will eventually fail their maths course.

In each study unit of this guide, there is a number of ***self-assessment exercises*** designed to

- assess the progress you have made towards the chapter objectives – allowing you to determine your own level of competence and what you still have to do to reach the requisite standard
- reinforce and expand the knowledge and insights you have derived from the chapter

Note: The ***solutions*** to the ***self-assessment exercises*** in this study guide are provided in tutorial letter 102. Evaluate your own solutions against these solutions.

Activity 8: Reflect with others

Link up with a group of students (called a peer-group), or with your tutorial group, if possible. Discuss your misunderstandings and insights with the group.

You can also participate in the discussion forum for this module on *my*Unisa.

Replace your previous misunderstandings: write down your new insights.

Activity 9: Celebrate your feelings of frustration, elation, inquiry

Learning to do something new (including mathematics) often involves a lot of effort, discomfort, excitement, or frustration. What feelings do you currently have when you think of mathematics?

These feelings cannot be avoided, so welcome them – they are symptoms of brain activity, which indicate that new connections between neurons are being made. After all, everything worthwhile takes effort. Studying mathematics is much like learning to play a new musical instrument or a sport: practice makes perfect. There is no royal road, just the hard road.

> *"There is no royal road to geometry", said Aristotle to his student, Alexander, when he had difficulty with mathematics.*

Do some maths activity every day. Work frequently and regularly. Make notes of your feelings as you go along. You can be successful!

6. Acknowledgements

Our sincere thanks go to Willem Labuschagne and Martha Pistorius for the valuable work they have done on the study material this study guide is based on.

Thank you as well to Ms Hentie Wilson, who assisted us with the design and layout of the guide and compiled the introduction.

We also thank Ms Jeanetta du Preez for compiling the mind map in study unit 1, and for providing the diagrams in sections 1.2.2 and 5.1.

Glossary of symbols

You have to thoroughly understand and be able to recognise at first sight the following "mathematical vocabulary". This takes time and focus. You can include the "vocabulary" in your mind maps and lists, with the equivalent term.

- *p*, *q*, *r,* etc. are used for propositions that have truth values.
- ¬ denotes "it is not the case that".
- ∧ denotes "and".
- ∨ denotes "or" meaning "either ..., or ..., or both".
- ∈ denotes "is a member of ...".
- ∀ denotes "for all".
- ∃ denotes "there exists (at least one)".
- → denotes "if ... then ...".
- ↔ denotes "if and only if" (iff).
- *a* mod *b* = *r* means that *r* is the remainder when a is divided by b.
- $\mathbb{Z}$ is a funny Z that denotes the set of integers {...,−3, −2, −1, 0, 1, 2, 3, ...}.
- $\mathbb{Z}^+$ which is the set of positive integers {1, 2, 3, 4, ...}.
- $\mathbb{Z}^{\geq}$ also denotes $\mathbb{N}$, which is the set of non-negative integers (natural numbers) {0, 1, 2, 3, 4, ...}.
- $\mathbb{Q}$ is a funny Q that denotes the set of all rational numbers, i.e. ratios of integers, defined by the set {a/b | a ∈ $\mathbb{Z}$, b ∈ $\mathbb{Z}$ and b ≠ 0}.
- $\mathbb{Q}^+$ denotes the set of positive rational numbers.
- $\mathbb{Q}^{\geq}$ denotes the set of non-negative rational numbers.
- $\mathbb{R}$ is a funny R that denotes the set of all real numbers.
- $\mathbb{R}^+$ denotes the set of positive real numbers.
- $\mathbb{R}^{\geq}$ denotes the set of non-negative real numbers.
- |m| is the absolute value of m.
- ⊆ denotes "is a subset of ...".
- ⊂ denotes "is a proper subset of ...".
- ∅ denotes "the empty set".
- ∩ denotes "intersection" (A ∩ B is a set that contains those elements common to both set A and to set B).
- ∪ denotes "union" (A ∪ B is a set that contains those elements in set A or set B).
- If A and B are sets, then A − B denotes the difference between A and B (the set that contains those elements in A that are not in B).

- If A is a set, A′ denotes the complement of A (the set that contains those elements of some universal set U that are not in A, i.e. U – A = A′).
- If A and B are sets, A + B denotes the symmetric difference between A and B (the set that contains those elements in either A or B, but not in both).
- If A is a set, |A| or n(A) denotes the number of elements in A.
- (a, b) denotes the ordered pair with a as first and b as second co-ordinate.
- If A and B are sets, A × B denotes the Cartesian product of A and B.
- (x_1, x_2, ..., x_n) denotes the ordered n-tuple with first co-ordinate x_1, second co-ordinate x_2, and so on.
- $\mathcal{P}$ is a funny P that denotes "the power set of ...".
- If R is a relation on some set A, then R is a relation from A to A.
- If R is a relation, then R^{-1} denotes the inverse relation of R.
- S ○ R denotes the composition of relations S and R. (We may write S ○ R or R; S.

  R; S reminds us that, in a certain sense, R is followed by S.)
- f: A → B should be read "f is a function from set A to set B".
- [x] denotes the equivalence class of x with regard to some previously specified equivalence relation.
- If f: X × X → X, then f is called a binary operation on X.
- *, □, as well as the more familiar symbols ·, +, −, and × are often used to denote binary operations.
- n! denotes the product 1·2·3·...·n. (n factorial).

Study unit 1 The development of number systems: $\mathbb{Z}^+$, $\mathbb{Z}^{\geq}$ and $\mathbb{Z}$

Key questions for this study unit

- *What are the differences amongst the three number systems $\mathbb{Z}^+$, $\mathbb{Z}^{\geq}$ (also referred to as $\mathbb{N}$) and $\mathbb{Z}$?*
- *Why did the invention of zero advance our understanding and ability regarding number systems?*
- *Why is a minus (−) times a minus a plus (+)? $- \times - = +$*
- *Why is a plus times a minus a minus? $+ \times - = -$*
- *What is the usefulness of properties such as commutativity?*
- *What is meant by the concepts "multiplicative and additive identities"; "additive inverse"; "absolute value"; "prime number"; and " n factorial (n!)"?*

1.1 Introduction to the study unit

In this study unit we review very briefly what the different classifications of the types of number we use are, and how these numbers can be re-written by scientists to make difficult and advanced mathematics possible. We also discuss the characteristics of numbers, which tell us what make them useful.

In this first study unit we are going to help you acquire three different skills/study habits which we think might help you in the future and might help you to master material in a shorter space of time.
These skills are
- getting an overview,
- the deep remembering of terms (linking it to your home language), and
- pronouncing these new terms correctly.

Activity 1-1: Overview ***Study skill***

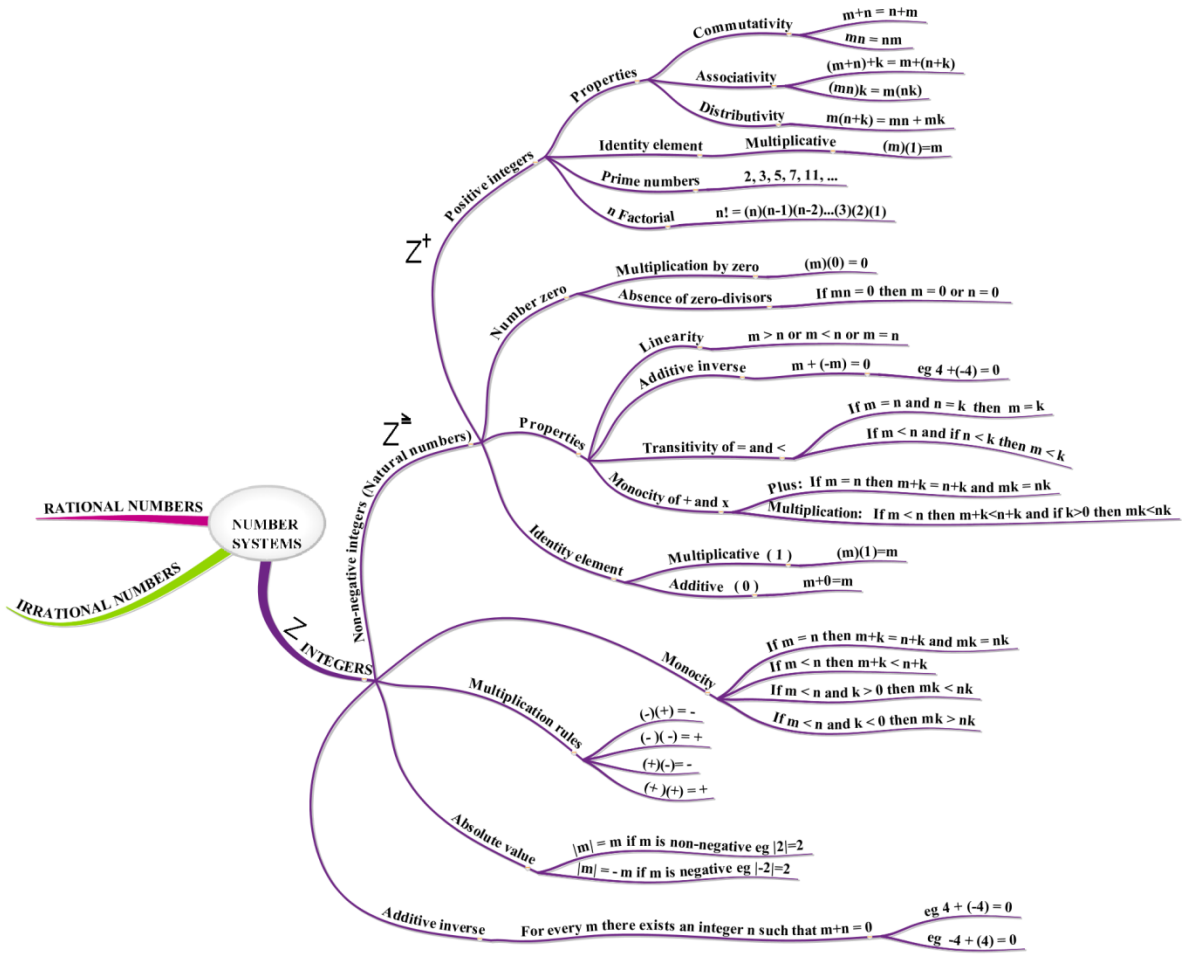
Draw a mind map of the different sections/headings you will deal with in this study session. Read section 4 (Syllabus) and section 5 (How to study this module) of the introduction to this study guide again, then page through this study unit with the purpose of completing the map.

Check your map against ours for concepts. It should include integers, positive integers and non-negative integers (natural numbers). It should also include the properties "commutativity", "associativity", "distributivity", and so on, the concepts "absolute value", "prime number" and "n factorial", and the multiplication rules.

Did you add colour to your map? Your map is your own drawing and should not look like those of other people; in fact, yours should be unique. Still, a mind map should contain key concepts that can be checked or correlated for correctness.

The concepts in the sections focus on the knowledge we expect you to have as underpinning ("building blocks") for this module and further modules. Without a clear understanding of these concepts you will find it difficult to identify and correctly complete tasks. Please take some time to ensure that you understand these concepts.

A mind map for the number systems $\mathbb{Z}$, $\mathbb{Z}^{\geq}$, and $\mathbb{Z}^+$ (or $\mathbb{N}$):



You can compare your mind map with the above map and then fill in possible missing concepts in your map if needed. After studying the next study unit you can also include the concepts "rational numbers" and "irrational numbers" with their properties in an extended mind map.

Activity 1-2: Concepts ***Conceptual skill***

We will list concepts at the beginning of each study unit to introduce and summarise concepts or terms. You may use this space to test your own knowledge (write in pencil) and then correct your understanding afterwards (erase and write the correct description). Your understanding can be deepened by also jotting down the term as you know it in your home language.

English term	Description	Term in your home language
Positive integers		
Non-negative integers		
Integers		
Property: Commutativity		
Property: Associativity		
Property: Distributivity		
Multiplicative identity		
Additive identity		
Multiplication by zero		
Additive inverse		
Absolute value		
Prime number		
n factorial (n!)		

Activity 1-3: Pronunciation ***Communication skill***

If you are uncertain about the pronunciation of some of the terms in the concept list, it will be a good idea to consult a science dictionary that will give you the pronunciation of scientific words. Investing in such a dictionary, or accessing one on the internet, will be very helpful if you are not sure about the pronunciation of a specific word.

1.2 Positive integers: $\mathbb{Z}^+$

In this section we look at the number system $\mathbb{Z}^+$ by which we mean the set of *all positive integers* namely the numbers 1, 2, 3, … and so on. We can indicate a *set* by writing down curly brackets and then writing the *members* belonging to the set inside the brackets. The members (or elements) of a set are separated by commas. We can denote the set of positive integers as follows:
$\mathbb{Z}^+ = \{1, 2, 3, \ldots\}$. (In study unit 3 we will learn more about sets.)

These are the numbers we would use to count sheep, for example. Historically, before discovering numbers such as zero and the negative integers, people worked with these numbers. A child, for example, learns to use these numbers before learning to use zero or fractions.

You all know how to add together or multiply positive integers. In this section we look at a number of *properties* of the addition and multiplication of positive integers. The principles ("commutativity", "associativity" and "distributivity") will be familiar to you; we will only add some terminology that you will often encounter in your Computing modules.

1.2.1 Commutative property

The following rules specify the commutative property of addition and multiplication for all positive integers within a mathematical context.

Property: Commutativity

For all positive integers m and n, addition and multiplication is commutative.

This means that

(a) $m + n = n + m$ (i.e. addition is commutative)

(b) $(m)(n) = (n)(m)$ (i.e. multiplication is commutative)

How do we know that addition and multiplication are commutative? Can I (as a student) trust and believe you (the lecturer)?

I would reply "As your lecturer, I cannot at this stage prove my claim that addition and multiplication are commutative, because this is outside the scope of this fundamental module. The proof is usually discussed at Honours level in any good set theory course, so you have to trust me at this stage when I say that these properties hold. But my claim is reasonable, because we can substitute specific values for m and n, and through examples show that it is a truthful statement."

We illustrate commutativity of addition by looking at an example.

Example

Let's take m = 3 and n = 6, for instance,

then $m + n = 3 + 6$ and
$n + m = 6 + 3$,
i.e. $3 + 6 = 9$ and
also $6 + 3 = 9$.

So for m = 3 and n = 6 we get $3 + 6 = 6 + 3 = 9$.

Similarly, it is also true for multiplication:

$(m)(n) = (3)(6) = 18$, and
$(n)(m) = (6)(3) = 18$.

So for m = 3 and n = 6 we get $(3)(6) = (6)(3) = 18$.

You might also ask, "Why do you write the product of m times n as (m)(n)? Can't we write $m \times n$ or mn or $m \cdot n$?"

We can write (m)(n) or mn or $m \cdot n$. I would tend to avoid $m \times n$, since the $\times$ symbol is sometimes mistaken for the letter x, and also because we later use $\times$ for a very specific sort of product, namely the Cartesian product of sets.

Given that addition is commutative, it should not surprise us that multiplication is also commutative, because the multiplication of positive integers is simply *repeated addition*. That is, a product such as (6)(3) might be viewed as the sum 6 + 6 + 6.

One can employ many examples to illustrate the commutativity property.

In stating the commutative properties, we used the letters "m" and "n". These symbols are *variables*; they are not the names of specific things. The idea is that for *every* positive integer whose name is substituted by m and *every* positive integer whose name is substituted by n, we will get a true statement of the form

m + n = n + m (Let m = 113 and n = 25 then 113 + 25 = 25 + 113.)

or of the form

mn = nm (In our example, (113)(25) = (25)(113).)

The variables thus serve to include a great number of specific statements in one brief, packaged statement.

Note that we wrote mn and not (m)(n) in the discussion. We do not need to use the brackets, because it is clear what mn means. However, the moment we substitute the variables with numbers such as 113 and 25, the brackets are essential. We cannot write 11325 when we actually mean (113)(25).

When you come across a statement involving variables, it's a good idea to mentally substitute specific values for the variables, just to keep the feeling of being in control.

1.2.2 Associative property

The next property we consider is the associativity of the addition and multiplication of positive integers.

Property: Associativity

For all positive integers m, n and k, addition and multiplication are associative.

This means that

(a) m+(n+k) = (m+n)+k (i.e. addition is associative)

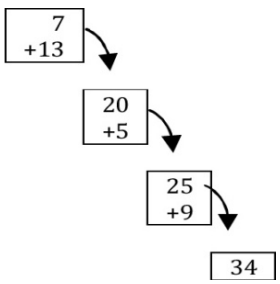
(b) (m)(nk) = (mn)(k) (i.e. multiplication is associative)

Bookkeepers often have to add up long lists of numbers. Some like to start at the top of the list and work downwards. Others like to start at the bottom of the list and work upwards. Do you know why they get the same answers?

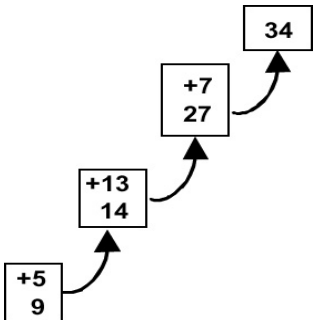
WORKED EXAMPLE

Bookkeepers use the principles of the **commutativity** and **associativity** of addition. Let's illustrate with an example of what they often need to do. Consider the list of numbers that have to be added: 7, 13, 5, 9.

Working downwards we get: ((7 + 13) + 5) + 9, where the brackets show that we added 7 and 13 first, then added 5 to the result, and then 9 to that result:



Working upwards we would have: ((9 + 5) + 13) + 7, where the brackets show that we first added 9 and 5, then 13 to the result, and then 7 to that result:



Now we look at the rules that we applied:

((7 + 13) + 5) + 9
= 9 + ((7 + 13) + 5) by commutativity
= 9 + (5 + (7 + 13)) by commutativity
= 9 + (5 + (13 + 7)) by commutativity
= 9 + ((5 + 13) + 7) by associativity
= (9 + (5 + 13)) + 7 by associativity
= ((9 + 5) + 13) + 7 by associativity

1.2.3 Distributive property

Another interesting property of the addition and multiplication of positive numbers is the property of distributivity.

Property: Distributivity

For all positive integers m, n and k, we say that multiplication is distributive over addition. This means that

m(n + k) = mn + mk, and since multiplication is commutative,

(n + k)m = m(n + k)
= mn + mk by distributivity
= nm + km by commutativity

What makes the distributive property important? Let's look at examples.

Examples

Let x be a variable representing some positive integer.

In order to write $18x^5 + 12x^4 + 3x^3$
as $3x^3(6x^2 + 4x + 1)$
we take out the *common factor*, namely $3x^3$. This means that we are relying on the distributive property. Note that we have used distributivity over more than two terms.

We can read the general definition of distributivity in a different way:
"mn + mk is equal to m(n + k)".

Furthermore, if we read m(n + k) = mn + mk from left to right, it tells us how to multiply out expressions.

Suppose we want to simplify (3x + 2)(x + 4).

One can read (3x + 2) as playing the role of m, x playing the role of n, and 4 playing the role of k:

$$\begin{aligned}&(3x + 2)(x + 4)\\ &= (3x + 2)x + (3x + 2)4 \quad \text{by distributivity}\\ &= 3x^2 + 2x + 12x + 8 \quad \text{by distributivity}\\ &= 3x^2 + 14x + 8 \quad \text{by distributivity}\end{aligned}$$

Activity 1-4:

Do you think that addition is distributive over multiplication, i.e. that m + (nk) = (m+n)(m+k)? Substitute a few values for m, n and k to see whether you get the same answer for the left-hand side and the right-hand side of the equation.

1.2.4 Multiplicative identity

Does the number 1 play a special role in some operations on numbers?

Property: Multiplicative identity

There exists a positive integer, namely 1 (the multiplicative identity element), that has the property that for every positive integer m, m·1 = m.

Of course, by commutativity we also have 1·m = m·1. Here, we call 1 the *identity* in respect of multiplication, because any number multiplied by 1 is identical to the given number.

"Isn't this a very useless property?" one might ask. "Everyone knows it, after all." Well, this property is used very often as a trick to simplify some complicated expression or to get it into some required form.

At this stage it is a bit difficult to give you convincing examples of this trick, since it requires further mathematical skills. For instance, in a calculus course, one uses this trick to show that the root of x, i.e. $\sqrt{x}$, is differentiable. But you cannot at this stage be expected to follow examples like that, so let's make a deal. You keep your eyes open from now on for situations in which the trick is used. We'll do the same and try to draw your attention to them.

Different notation: It is important that you should be aware of the fact that you will find a variety of notations (or names) for the same concept (such as the positive integers) in the literature. We call the positive integers $\mathbb{Z}^+$, but some books call it P. So remember, whenever you pick up a mathematical textbook, you will have to make an effort to figure out exactly what notation it uses.

1.3 Non-negative integers: $\mathbb{Z}^{\geq}$

With $\mathbb{Z}^{\geq}$ (or $\mathbb{N}$) we mean the set of *all non-negative integers (natural numbers)*, namely the numbers 0, 1, 2, 3, 4, ... and so on. We can write $\mathbb{Z}^{\geq}$ = {0, 1, 2, 3, ...}.

The only difference between $\mathbb{Z}^{\geq}$ and $\mathbb{Z}^+$ is that $\mathbb{Z}^{\geq}$ has as member the **additional number zero**.

Note: In some books the letter $\mathbb{N}$ (the set of natural numbers) is reserved for the positive integers {1, 2, 3, …}. However, in this study guide we stick to the convention of using $\mathbb{N}$ for the set of natural numbers {0, 1, 2, 3, …}.

Strange as it might seem, it took thousands of years before mankind conceived the idea of having a number such as zero. Of what use is it?

Scenario

Suppose we want to buy sheep, but farmer Mokoena has lost all his sheep in the drought. If we don't know about zero, we could write:
Farmer Mokoena has no sheep.

If we have the number zero at our disposal, we could write:
Farmer Mokoena has 0 sheep.

It was not obvious that zero was necessary while people were mainly interested in counting the number of sheep. It is only when we want to rate things according to *a scale* that we begin to get the idea that there is a difference between "zero" and "nothing".

Example

Suppose we think of the marks a lecturer awards for an exam. One particular student, let's call him Mr Jones, decided to do no work at all. This might result in Mr Jones being awarded a mark of zero! But Mr Singh, who is not enrolled for the module, would get nothing. In other words, he will have no mark for the module at all, so we would just delete his name off our class list in reality. Mathematically these situations are not the same; 0 is not the same as "nothing".

The invention of zero by the Hindus in the seventh or eighth century AD in India made the first successful form of positional notation possible. This means that the position of a digit within a number is important. This concept is worth spending a little time on, since it may well be the most powerful justification for the extension of $\mathbb{Z}^+$ to $\mathbb{Z}^{\geq}$. So, let's think about notation for numbers. Strictly speaking, in ordinary language one could get away with names for numbers: names such as "one thousand eight hundred and sixty", "thirteen" and "seventy-two". But this makes arithmetic difficult. If the ordinary man in the street is to learn to add and multiply easily, what we need are abbreviated names for numbers which, in some way or other, make arithmetic easy.

Well, we could consider the simplest notation of all, consisting of vertical strokes, which works as follows:

1	for the number one
11	for two
111	for three
1111	for four
11111	for five
.	
.	
.	
11111111111111111111	for twenty-three, and so on.

Those of you who are familiar with binary numbers will realise that these vertical strokes do not represent binary notation.

Addition becomes very easy with this notation. To add three and four, for example, one just takes the bunch of strokes representing three, i.e. 111, and puts it with the bunch representing four, i.e. 1111, to get 1111111, which then represents the sum of three and four. When it comes to bigger numbers though, it is not only tedious to write down all the strokes, but it is also very easy to make a mistake and put in one extra stroke, or one stroke too few. Also, there is no way to indicate zero in this system, except by the *absence* of a stroke. This tends to result in the idea that zero is nothing, and we have discussed the weakness of this idea in our example of Mr Jones and Mr Singh.

The early civilizations of Sumer and Babylonia started to get a handle on the problem of representing large numbers by devising a positional system of notation in which the place (position) occupied by a symbol determined its value. Hence we talk about a place-value system.

Activity 1-5: Place value

Determine how a place-value system works.

First, you can choose any positive integer greater than 1. Call whichever number you choose the *base* of your system (or *radix*, meaning root).

The Arabs chose ten, which we still use today. In contrast, some ancient tribes in South America chose four, while the Mayan tribe chose twenty, and the Babylonians sixty.

Next you need symbols for the numbers smaller than your base. Call these symbols *digits*. Now you can represent a number as a row of digits, in which the rightmost digit represents a number of units, the next digit to its left represents a number of groups which each has as many things in it as your base, and so on. The further to the left you go, the greater the value of the digit. We will get back to this a little later.

Following the invention of zero by the Hindus and the Arabs, we ended up with several excellent systems of notation. Nowadays the man in the street uses the decimal system.

Let's see how it works in the familiar decimal system (of which the base is 10).

Activity 1-6: Place values in the decimal system

We have as digits the familiar symbols 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9.

(a) How do we represent one hundred and thirteen in this number system?

(b) How would you multiply (186)(10) in the decimal system?

(a) This is easy for us, since we simply write 113. When we look at this closely, and say the number out loud, we see that the *position* of each digit within the number indicates the place value of each digit, so 113 can be regarded as
(1 times $10^2 = 100$) plus (1 times 10^1) plus 3, or $(1)(10^2) + (1)(10^1) + 3$.

Now, if one doesn't have the number zero represented by some digit, then the following problem arises. How does one indicate the difference between numbers such as 300, 30 and 3? The symbol 0 pushes the digit 3 to the left in 30 and in 300 and so increases its value. Also how would you write the number "three hundred and five" in the decimal system? This can be written as $(3)(10^2) + (0)(10^1) + 5$.

(b) The decimal system has made it simple to multiply a hundred and eighty-six by ten, i.e. (186)(10): you just put a zero to the right of 186, to get 1860.

Addition and multiplication are not made more difficult by the inclusion of 0.

These operations remain:

- commutative and
- associative;
- the distributive property is retained, and
- 1 is still the multiplicative identity.

However, we gain two important new properties.

1.3.1 The existence of an additive identity

We call 0 the *identity* with respect to addition because, when adding zero, the number we start with is identical to the end result.

Property: Additive identity

There exists a non-negative integer, namely 0 (the additive identity element), that has the property that, for every non-negative integer m,
$m + 0 = m$.

Of course, by commutativity we also have $0 + m = m + 0 = m$ for any non-negative integer m.

For example, let m = 2, then

$$0 + 2 = 2 + 0 = 2.$$

Sometimes this property is used when we want to solve equations. We see how this is done in examples provided later in this study unit.

1.3.2 Multiplication by zero

Zero is a special number. Why do we say this?

Definition: Multiplication by zero

There is only one way to get a product equal to zero, and that is to have zero as one of the factors, i.e.

(a) for every non-negative integer m, $m \cdot 0 = 0$, and

(b) if m and n are non-negative integers such that $m \cdot n = 0$ then either m = 0 or n = 0, or both.

We can use (b) above when we use factorisation to solve some quadratic equation such as $x^2 + 2x + 1 = 0$. An example illustrating how this definition is applied is provided later in this study unit. If you have trouble remembering how factorisation works, we have included a number of self-assessment exercises at the end of this study unit. Try to do the exercises by yourself then evaluate your answers by studying the solutions to the self-assessment exercises as provided in tutorial letter 102.

1.4 Integers: $\mathbb{Z}$

In this section we look at the symbol, $\mathbb{Z}$, that we use to represent the set of *all integers* in both directions (negative and non-negative), namely the numbers ... , –3, –2, –1, 0, 1, 2, … . We can write $\mathbb{Z}$ = {... , –3, –2, –1, 0, 1, 2, …}.

Just as it took a long time for zero to be invented, so the invention of the negative integers had to wait until it became a necessity within the society of the day.

In bookkeeping, scribes found it easier to keep track of debits and credits if, instead of entering figures into two different columns, the debits were indicated by putting a little hyphen in front of the number representing the size of the debit. So, for instance, a debit of 113 Δ (where Δ represents some currency) could be written as –113. Apart from making things convenient for scribes, the invention of negative numbers was necessary to clear up some other mathematical difficulties.

Consider the following silly argument, which can be modified to demonstrate that any two numbers are equal (which is nonsense, of course).

Silly argument: Are any two numbers equal?

The argument is based on the idea that if both sides of an equation represent the same number, then the square roots also represent the same number.

Let x be 5 and y be 3.
Then $x + y = 2z$. (where z is 4)

Multiply each side by $(x – y)$. (We can do this because $x – y \neq 0$.)
This gives $(x – y)(x + y) = (x – y)(2z)$
or $x^2 – y^2 = 2xz – 2yz$.

Add $y^2 + z^2 - 2xz$ to each side, to get
$x^2 – 2xz + z^2 = y^2 – 2yz + z^2$

This means that each side is a perfect square.
Thus, $(x – z)^2 = (y – z)^2$

Now, take the square root of each side
$x – z = y – z$
Add z to each side, then
$x = y$
i.e. $5 = 3$ (Oops, this is interesting!)

Activity 1-7: Fixing the silly argument with numbers

Can you spot what is wrong with the argument used above?

Well, remember that we call p a *square root* of q if $p^2 = q$.

Let's consider the number 4.

We know that $(-2)^2 = (+2)^2 = 4$.
This means that if $p^2 = 4$ then $p^2 = (\pm 2)^2$
i.e. $p = +2$ or -2.

The important thing to remember here is that usually when we talk about taking "the square root", we take the *positive* root. *The symbol* $\sqrt{\ }$ *in fact means "the positive square root".*

So from $(x - z)^2 = (y - z)^2$ we should get a positive object on each side. But recall from our "silly argument" that y was 3 and z was 4. Clearly $y - z$ is not positive.

This helps us to see the flaw in the argument given above: you have to realise that there are things such as negative numbers and that in our reasoning we have gone to a negative number when in fact the symbol $\sqrt{\ }$ *required a non-negative number. A modern mathematician will not have any trouble fixing the argument by using absolute values (a topic we will deal with soon).*

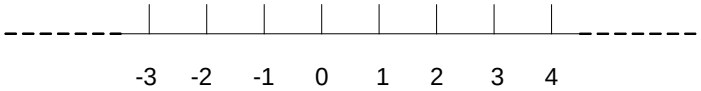
Many uses were found for the expanded system of integers. Just think again of rating things according to a scale.

Some scales, such as marks for an exam, can conveniently contain a range from some minimum to some maximum. Other scales have, rather than a minimum or a maximum, some point in the middle which is of importance. Think for instance of temperature measured in centigrade, with a mid-point indicating the temperature at which water freezes.

°C .... –50 .... 0 .... 50 100

From such a point in the middle one can move in two directions, one of which could be called the positive direction and the other the negative direction. For example, a business might choose as mid-point its break-even point at which it has covered its expenses but not made any profit, and then choose increasing net profit as the positive direction and increasing net cost as the negative direction.

Addition and multiplication were not made much more complicated by the inclusion of negative integers. Addition can be performed quite mechanically by drawing a *number line* on which the names of integers are written at some fixed unit of distance apart. A number line can be represented as follows:

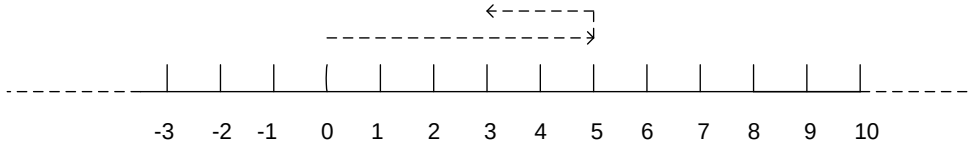


You might now want to know whether the concept of a number line has any use at all. Let's use an example to illustrate that it is, in fact, a very useful concept.

Example: Addition

Let zero be the starting point. To find the sum x + y, one moves x steps in the positive or negative direction (depending, of course, on whether x is positive or negative) and then, from where you stop, y steps in the appropriate direction (again depending on whether y is positive or negative). The place where you stop gives the sum x + y.

We look at an example: 5 + (−2)



Clearly 5 + (−2) = (−2) + 5 = 3.
In the case when two positive numbers are added (e.g. 3 + 4), the movement is only in a positive direction.

Multiplication is a little more complicated. There are four different possibilities for a product (x)(y).

Example: Multiplication

Firstly, the simplest case is when both x and y are non-negative. Then, of course, we know that the product is non-negative, because both x and y are also members of $\mathbb{Z}^{\geq}$.

A second simple case is that of a product (x)(y) in which x is negative but y is not negative. Then
xy = x + x + … + x (y times), for example,
(−2)(3) = (−2) + (−2) + (−2) = −6,
so we can find the product easily by using the principle of repeated addition.

Thirdly, remember that a product (x)(y) can also be written as xy. Now suppose we need to find xy when y is negative but x is not. Can we say "xy = yx by commutativity, and so xy = y + y + ... + y (x times)"?

No, we can't. If we don't yet know what xy is, then we can't be sure that the multiplication of integers is commutative.

One has to be careful not to use a more advanced fact to prove one of the simpler facts on which it rests; doing so would create a vicious circle.

The fourth case is just as tricky.

What happens when we have to find the product xy, when both x and y are negative?

The usual way to calculate the product in tricky cases is given by the general rule:

Fact 1: If x is negative, then x = –a for some positive integer a. Otherwise x = a.

Fact 2: If y is negative, then y = –b for some positive integer b. Otherwise y = b.

Step 1: Find ab.

Step 2: If one of x and y is negative, then xy = –(ab).
Otherwise, xy = ab.

Let's consider the following rules:

RULES: $(+)(-) = -$; $(-)(+) = -$; $(-)(-) = +$

A plus times a minus is a minus, a minus times a plus is a minus, and a minus times a minus is a plus.

How can we justify the rules? Let's do it by looking at a concrete situation.

Worked example

Suppose you deposit money into, or withdraw money from your bank account every day. Let's denote an increase of, say, R110 in your account by the positive integer 110, and a decrease of R110 by the negative integer –110. Let's indicate days in the future by a positive number, and days in the past by a negative number (e.g. three days from now is indicated by 3, while three days ago is indicated by –3).

The following four cases all fit our general rule for multiplication:

* If you increase the amount in your account by R5 every day for the next 3 days, your balance should change by R15. Our mathematical representation gives (5)(3) = 15 using the rule that a plus times a plus is a plus. So our rule fits the situation.

* If you increase the amount in your account daily by R5, then three days ago your balance would have been R15 less. Our mathematical representation gives 5(–3) = –15 using the rule that a plus times a minus is a minus. So our rule fits the situation.

* If you withdraw R5 from your bank account daily, then within 3 days your balance will change by (–5)3 = –15, i.e. it will decrease by R15. So a minus times a plus is a minus. (We see that 5(–3) = (–5)3 = –15.)

* If you withdraw R5 from your bank account daily, then three days ago your balance would have been different by (–5)(–3) = 15, i.e. it would have been R15 more than the present amount. So a minus times a minus is a plus.

Activity 1.8: Addition and multiplication of integers

Investigate whether the addition and multiplication properties hold for integers.

Let's summarise the properties of addition and multiplication of integers:

- Both addition and multiplication are commutative and also associative.
- Multiplication is distributive over addition.
- There is an identity element with respect to multiplication, namely 1.
- There is an identity element with respect to addition, namely 0.
- A product is zero, if and only if, at least one of the factors is zero.

There is one additional property which we consider in the next section.

1.5 The additive inverse, absolute values and prime numbers

Definition: Additive inverse

For every integer x there exists an integer, i.e. an integer denoted by –x, such that

x + (–x) = 0.

Don't fall into the trap of thinking that – x always denotes a negative number. See and read the hyphen (the minus sign) as an abbreviation for the phrase "– x is the additive inverse of ".

So if x = –2, for instance,
then –x is the additive inverse of –2,
i.e. –x = –(–2) = 2.

We can use the "additive identity element" and an "additive inverse" when we want to change an equation to get it into some standard form.

Example: Additive inverse

The idea of an additive inverse is simple. If you think of x as representing a certain number of steps in either the positive or the negative direction, then –x represents an equal number of steps in the opposite direction.

For instance, given that x + 3 = 0, we can add –3 to both sides to get

x + 3 + (–3) = 0 + (–3)
i.e. x + 0 = –3 (0 is the additive identity element)
i.e. x = –3

There is no need to speak of "subtraction". It is just convenient for us to write, for example, 4 + (–1) as 4 – 1.

An additive inverse can play a role in the solving of equations of the form $p^2 = q$. Let's look at the following example, and also remember that we call p a *square root* of q if $p^2 = q$.

Example

Consider the technique for solving quadratic equations known as "completing the square". It is based on the fact that an equation, such as $x^2 = 9$ can easily be solved by taking the square root on each side.

We know that $(–3)^2 = (3)^2 = 9$, so we use this to solve $x^2 = 9$:

If $x^2 = 9$, then
$x = 3$ or $x = –3$.
So 3 and –3 are square roots of 9.

Given a more complicated equation, we can follow the same reasoning as in the previous example.

Example

Let's consider $x^2 - 2x = 15$.

We can try to get the left-hand side into the form (something)^2. To do this, we use the fact that adding 0 (the additive identity element) to one side doesn't change a thing, and we write 0 as
$1^2 - 1^2 = 1 - 1$.

So we get

$x^2 - 2x + (1 - 1)$	$= 15$	(adding 0 to the left-hand side)
i. e. $x^2 - 2x + 1$	$= 16$	(adding 1 to both sides)
i. e. $(x - 1)^2$	$= 16$	(completing the square)
	$= (\pm 4)^2$	

Therefore $x - 1 = 4$ or $x - 1 = -4$
i. e. $x = 5$ or $x = -3$

In addition to involving the "additive identity element" and an "additive inverse" when solving quadratic equations, we can also apply the "multiplication by zero" definition (as defined earlier in this study unit).

For instance

Suppose we want to solve $x^2 - 2x - 3 = 0$.

We can factorise the left-hand side
to get $(x - 3)(x + 1) = 0$.

Now, we know that whenever a product of an equation is equal to zero (as in the equation above), at least one of the factors must be zero, so we know that

	$x - 3 = 0$	or	$x + 1 = 0$
i.e.	$x - 3 + 3 = 0 + 3$	or	$x + 1 + (-1) = 0 + (-1)$
So	$x + 0 = 0 + 3$	or	$x + 0 = 0 - 1$
i.e.	$x = 3$	or	$x = -1$.

From these examples we see that zero is a very special and unique integer, and an additive inverse also comes in handy sometimes.

There is another interesting thing about $\mathbb{Z}$ when we consider the concept of size. When we discussed the addition of integers, we formed a mental picture of $\mathbb{Z}$ as a number line, with the members of $\mathbb{Z}^+$ extending towards the right and their additive inverses extending towards the left. This mental picture suggests a way to *order* the integers.

We say that x is less than y (abbreviated by $x < y$) if x lies to the left of y on the number line. This amounts to saying that $x < y$ if $y + (-x)$ is positive.

Activity 1-9: Number line

Draw a number line (as given above) to show your understanding of the concept $-5 < 2$.

Is the number 2 to the right of the zero on your line? Is the negative number to the left of the zero?

There is another way in which we use the concept of size. In ordinary life we speak not only of big credits but also of big debits, of big profits and of big losses, and so on. So, an "absolute value" is useful for large numbers.

Activity 1-10: Communicating big negative numbers

How can you communicate the idea that a negative number (such as $-113\,000$) is very big?

We can do this by defining the *absolute value* of a number. This tells us how far from zero the number sits on the number line, without taking into account the direction. Let's look at the definition.

Definition: Absolute value

For any integer x, the absolute value of x (denoted by |x|) is defined to be either:

x if x is non-negative, or

–x if x is negative.

This sounds more complicated than it really is. We can look at some examples.

Examples

|2| makes sense since 2 lies two steps to the right of 0, and
|–2| also makes sense since the absolute value does not care in which direction one has to step.

In terms of <, we saw that 2 was greater than –5. In terms of the absolute values, the situation is reversed, since –5 is more steps away from 0 than 2 is, so

$|2| < |-5|$.

We can apply the concept of absolute value to fix the *invalid argument* involving square roots which we used earlier to show that 5 = 3. The crucial step there was taking the square root on each side of

$(x - z)^2 = (y - z)^2$ to get
$(x - z) = (y - z)$

In order to *fix the argument*, one can choose the positive or the negative root on each side by the following reasoning:

Since $\sqrt{w^2} = |w|$ we can go from
$(x - z)^2 = (y - z)^2$
to $|x - z| = |y - z|$.
Now since x = 5, y = 3 and z = 4, we know that x – z is positive,
so that $|x - z| = x - z$,
whereas y – z is negative.

So, to get rid of the absolute value signs we must go to its *additive inverse*, i.e. $|y - z| = -(y - z)$.

This gives $-(y - z)$
$= -y - (-z)$
$= -y + z$
$= z - y$
$= 4 - 3$
$= 1$

Also: $x - z$
$= 5 - 4$
$= 1$

Now $(x - z)^2 = |x - z| = x - z = 1$, and
$(y - z)^2 = |y - z| = -(y - z) = 1$.

This time we get a valid answer!

Now we can move on to another concept. Remember the concept of prime numbers from high school maths? Refer to the following definition, in case you have forgotten.

Definition: Prime number

A positive integer p greater than 1 is defined to be a *prime number* if its only factors are 1 and p.

The list of prime numbers thus includes the numbers 2, 3, 5, 7, 11, 13, 17, 19, and so on. We see that the number 7 is prime because its only factors are 1 and 7. On the other hand, a number such as 4 is not prime since it can be factored as $4 = 2 \cdot 2$ where 2 is greater than 1.

Before we conclude this study unit, we would like to introduce you to the concept of a factorial, if you are not already familiar with it. This is a concept you will often encounter in your further studies.

Definition: n factorial (n!)

If n is any positive number, then n factorial, denoted by n!, is calculated as follows:

$n! = n(n-1)(n-2)\ldots(4)(3)(2)(1)$

To find the value of 6!, for example, we need to do the following calculation:
$6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$.

Activity 1-11: Self-assessment exercises *Application skills*

1. Factorise the following expressions (as revision of your school maths):

(a) $x^2 + 6x + 9$
(b) $x^2 - x - 2$
(c) $x^2 - 5x + 6$
(d) $x^2 + 4x - 12$

2. Solve $x^2 - 4x + 4 = 0$ by factorising.

3. Complete the square to solve $x^2 - 4x = 12$.

4. Is 21 a prime number?

5. What is the value of 5! (5 factorial)?

Note: The solutions to the self-assessment exercises are provided in tutorial letter 102.

1.6 The nine laws for $\mathbb{Z}^{\geq}$

To conclude, we give the nine laws for $\mathbb{Z}^{\geq}$:

Law 1 (commutativity):
For all non-negative integers m and n,
$m + n = n + m$ and $mn = nm$.

Law 2 (associativity):
For all non-negative integers m, n and k,
$m + (n + k) = (m + n) + k$ and $m(nk) = (mn)k$.

Law 3 (distributivity):
For all non-negative integers m, n and k,
$m(n+k) = (mn) + (mk)$.

Law 4 (existence of a multiplicative identity element):
For all non-negative integers m,
$m \cdot 1 = m$.

Law 5 (linearity):
For all non-negative integers m and n, exactly one of the following statements are true:
$m < n$, $m = n$, $m > n$.

Law 6 (monotonicity of + and × respectively*):*
For all non-negative integers m, n and k,
if $m = n$, then $m + k = n + k$ and $mk = nk$;
if $m < n$, then $m + k < n + k$; and
if $k > 0$, $mk < nk$.

Law 7 (transitivity of = and < respectively):
For all non-negative integers m, n and k,
if $m = n$ and $n = k$, then $m = k$, and
if $m < n$ and $n < k$, then $m < k$.

Law 8 (existence of an additive identity element):
For all non-negative integers m,
$m + 0 = m$.

Law 9 (absence of zero-divisors):
For all non-negative integers m and n,
$mn = 0$ if and only if $m = 0$ or $n = 0$.

What about Z?

All the laws listed above hold for $\mathbb{Z}$, except for the monotonicity law, which looks slightly different for $\mathbb{Z}$:

Law 6 (monotonicity):
For all integers m, n and k,
if m = n, then m + k = n + k and mk = nk;
if m < n, then m + k < n + k;
if k > 0, then mk < nk; and
if k < 0, then mk > nk (negative numbers must also be taken into account).

$\mathbb{Z}$ has one law that $\mathbb{Z}^{\geq}$ does not have:

Law 10 (existence of additive inverses):
For every integer m there exists an integer n such that
m + n = 0.

1.7 In summary of the study unit

In this study unit you ensured that you can answer the following basic questions:

- What are the differences amongst the three number systems: $\mathbb{Z}^+$, $\mathbb{Z}^{\geq}$ (also referred to as $\mathbb{N}$) and $\mathbb{Z}$?
- Why does zero play an important role in some number systems?
- Why is a minus (–) times a minus equal to a plus, i.e. (–)(–) = +?
- Why is a minus (–) times a plus equal to a minus, i.e. (–)(+) = –?
- What is the usefulness of properties such as commutativity?

It is important to understand the properties of the different number systems very well. We understand that you might not have had this knowledge as background, so hopefully you have corrected some misunderstandings you might have had. If so, this is good and you know that you can go on to the next section.

In the next study session we will concentrate on the rational numbers and the real numbers.

NOTES

Study unit 2 Rational and real numbers: $\mathbb{Q}$ and $\mathbb{R}$

Key questions for this study unit

- *What is meant by the concepts "rational number"; "irrational number"; "reductio ad absurdum"; "repeating decimal"; "quadratic formula"; and "the Theorem of Pythagoras"?*
- *Why can we not divide by zero?*
- *Where do the real numbers fit in?*
- *What is the meaning of the concept "proof by contradiction"?*

2.1 Introduction to this study unit

This study unit follows on study unit 1. In study unit 1 we discussed positive integers, non-negative integers and the integer number system. Here we shall look at the rational numbers, describe what an irrational number is, and finally consider real numbers. These number systems do not form the complete set of all numbers. We are, for example, not considering complex numbers.

Activity 2-1: Overview **Study skill**

Draw a mind map of the different sections/headings you will deal with in this study session. Then page through the study unit with the purpose of completing the map.

Your mind map should include the concepts of rational numbers, irrational numbers, repeating decimals, a quadratic formula and the Theorem of Pythagoras. The concepts in study units 1 and 2 focus on the knowledge we expect you to have as underpinning ("building blocks") for this module and further modules.

Activity 2-2: Concepts **Conceptual skill**

Use this space to test your own knowledge (write in pencil) and then correct your understanding afterwards (erase and write the correct description). Your understanding can be deepened by also jotting down the term as you know it in your home language.

English term	Description	Term in your home language
Rational number		
Irrational number		
Real numbers		
Multiplicative inverse		
Theorem of Pythagoras		
General form of odd and even integers		
Expressing 0 as a ratio		

2.2 The rational numbers: $\mathbb{Q}$

So far the numbers at our disposal do not allow us to solve simple equations such as $2x = 1$. Let's see how such an equation could arise in practice.

Example

A cook uses 1 bag of flour to bake 2 loaves of bread. How much flour does she need to bake a single loaf?

Well, if x represents the amount of flour needed for one loaf, then what do we know about x? Certainly, we know that two times x is one whole bag of flour, because we get 2 loaves of bread from 1 bag of flour. We see that x is the number by which 2 must be multiplied to give 1, and so we can write this as $2x = 1$.

Furthermore we know that x is not an integer. To see this, we note firstly that there are no integers between 0 and 1. Then we see that x must be greater than 0. Why? Because $(2)(0) = 0$. This is less than the required result, so zero is too small.

Next we see that x must be less than 1, since $(2)(1) = 2$, which is bigger than the required result, so one is too big.

So, for example, the x above could be written as $1/2$. Now it is clear that we have to multiply 2 by ½ to get 1.

We therefore have to expand $\mathbb{Z}$ by including new numbers. We want all numbers of the form p/q, where you should regard the notation p/q as an abbreviation for *the number by which q must be multiplied to give p*.

But there is a very important exception: We do not want q to be zero, i.e. we do not want to allow division by zero. Why not? There are two possible reasons.

- Suppose p ≠ 0 and q = 0. Then p/q would mean *the number by which q must be multiplied to give p*. But no such number exists, since q is zero, and zero multiplied by anything must give zero, whereas p is not zero.
- Suppose p = q = 0. Then p/q = 0/0 would mean *the number by which q must be multiplied in order to give p*. Now the problem is that any number will do, since 0 multiplied by any number is zero. And this is no good, for we would never know what number is being talked about when someone refers to 0/0.

A famous mathematician called De Morgan, who lived nearly two centuries ago, showed some of the disastrous things that would happen if we tried to make it possible to divide by zero.

One would be able to prove that any number is equal to zero (which, of course, is nonsense). The nonsense proof works as follows:

Suppose $x = a$.
It is possible to show that $x = 0$?
Multiply by x: $x^2 = ax$
Subtract a^2: $x^2 - a^2 = ax - a^2$
i.e. $(x + a)(x - a) = a(x - a)$
Divide by $x - a$ *(using the statement above* $x = a$, *so* $x - a = 0$*):* $x + a = a$
Subtract a: $x = 0$ *QED*

*Note: We indicate the end of a proof by QED, which is an acronym of the Latin phrase "*quod erat demonstrandum*", which means "that which was to be demonstrated".*

It is wiser to accept that division by zero can never be allowed!

Definition: Rational numbers

We expand $\mathbb{Z}$ to a new number system $\mathbb{Q}$ which is the set of all numbers of the form p/q where p and q are integers and q is not zero. We call such numbers *rational numbers*.

Of course, every integer can be written in the form p/q just by giving it a denominator equal to 1, i.e. by making q one (1).

How does one add and multiply rational numbers? To multiply is very easy. For instance

$$\left(\frac{1}{2}\right)\left(\frac{3}{5}\right) = \frac{1 \cdot 3}{2 \cdot 5} = \frac{3}{10}.$$

We simply multiply the tops (called *numerators*) with each other and multiply the bottoms (called *denominators*) with each other.

So, the general rule is:

(a/b)(c/d) = ac/bd

Addition is not so simple. In order to add, say, 1/2 and 3/5 we have to find a *common denominator*, which in this case is 10.

Then

$$\frac{1}{2}+\frac{3}{5}=\frac{5+6}{10}=\frac{11}{10}.$$

A hidden part of this reasoning involves the idea of *equivalent fractions*. You know that

1/2 = 5/10 = 10/20 = 30/60, etc.

This is because 1 remains the multiplicative identity for $\mathbb{Q}$, and 1 can be written as any fraction in which the numerator and denominator are the same, for instance as 5/5 (the number by which the 5 at the bottom must be multiplied to give the 5 at the top). So 5/10 and 10/20 and 30/60 are all equal to 1/2, since they are all obtained from 1/2 by multiplying by 1, with 1 written in an appropriate form.

To express 1/2 as an equivalent fraction with a denominator of 10, multiply by 5/5. Similarly, to express 3/5 as an equivalent fraction with 10 as denominator, multiply by 2/2.

How does one find common denominators? There are two methods:

The quick method just multiplies the denominators of the given fractions. For example, in the case of 1/2 + 3/5 this method gives 10 as the common denominator. The drawback of this method is that it does not always give the least common denominator, although for the purposes of this module, it's not a problem.

The more elegant method involves using things called prime factors to build the least common denominator. We don't require you to use this method, so we will ignore it for this module.

Let's summarise the properties of addition and multiplication in $\mathbb{Q}$. As you probably expect,

- we have commutativity and associativity for both addition and multiplication,
- the distributive property holds,
- 0 and 1 are identities for addition and multiplication respectively,
- products are zero if and only if at least one of the factors is equal to zero, and
- additive inverses exist for all rational numbers.

There is one further property.

Definition: Multiplicative inverses

For every non-zero rational number x there exists a rational number called the *multiplicative inverse,* denoted by 1/x which is such that (x)(1/x) = 1.

The idea of a multiplicative inverse is simple. If you think about the notation 1/x for a moment, you will recall that it means *the number by which x must be multiplied to give 1*.

Given x, the number 1/x is called the *reciprocal* of x (in older books). This fits the following idea: Think of every rational number as having a numerator and a denominator. Now the multiplicative inverse of any number can be found by simply turning it upside down, i.e. writing the numerator as the denominator and the denominator as the numerator.

We use this property whenever we want to solve linear equations of the form ax = b. Multiplying both sides by 1/a gives the solution x = b/a. Remember that this cannot be done if a = 0.

There is no need to speak of division. To say that one is dividing x by y is the same as to say that one is multiplying x by 1/y.

2.3 The 11th law of $\mathbb{Q}$

Similar laws hold for $\mathbb{Q}$ as for $\mathbb{Z}$, but set $\mathbb{Q}$ has one law that the set $\mathbb{Z}$ does not have:

Law 11 (the existence of multiplicative inverses):
For every non-zero rational number x there exists a rational number y such that xy = 1

2.4 The real numbers: $\mathbb{R}$

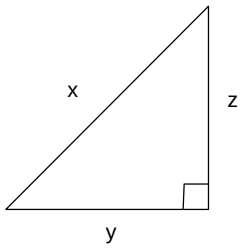
Suppose a ladder leaning against a wall reaches 1 meter high when its foot is 1 meter away from the wall. How long is the ladder?

The *Theorem of Pythagoras* can help us solve the problem, because the situation can be represented by a right-angled triangle in which the length of the ladder is given by the unknown x.

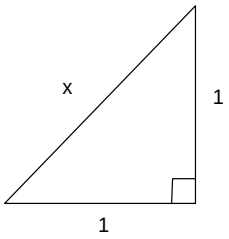
In general, given a right-angled triangle with sides of lengths x, y and z as shown below, Pythagoras' theorem tells us that

$$x^2 = y^2 + z^2$$

i.e. *the square on the hypotenuse is the sum of the squares on the other two sides.*



So if x, which indicates the hypotenuse, represents the length of the ladder, and y and z are both equal to 1,
we have $x^2 = 1^2 + 1^2$
i.e. $x = \sqrt{2}$



This does not tell us very much yet.
How big is $\sqrt{2}$? What number is it?
Well, it turns out that $\sqrt{2}$ is not any of the numbers we've met so far, i.e. $\sqrt{2}$ is not a rational number. This is not quite as easy to show as it was to show that $1/2$ is not an integer. But it is worth looking at the proof.

Theorem

There is no rational number whose square equals 2.

Proof

We will use a technique known as *reductio ad absurdum*, also called *proof by contradiction* (to be discussed in greater detail in subsequent study units).

Suppose there is some rational number x whose square equals 2, i.e.
$x^2 = 2$ for some $x = p/q$, with $p, q \in \mathbb{Z}$ and $q \neq 0$. (We may write x in the form p/q because x is a rational number.)

Any fraction can be expressed in *lowest terms* by cancelling any common factors shared by the numerator and denominator. So we may assume that p/q is a fraction in lowest terms.

By the next argument we show that both p and q are even, i.e. both p and q are multiples of 2 and hence have a common factor, namely 2, which is a direct contradiction of what we have just noted.

If $(p/q)^2 = 2$
then $p^2 = 2q^2$.

This means p^2 is an even integer (since it is a multiple of 2).

Note: *Even integers can be written in the form 2k for some integer k.*

We will now show that q^2 is also an even integer.

Firstly, we claim that the square of an odd integer is odd.

Note: *Odd integers can be written in the form 2k + 1 for some integer k.*

Squaring 2k + 1
gives $(2k + 1)^2 = 4k^2 + 4k + 1$
$= 2(2k^2 + 2k) + 1$
$= 2m + 1$

where m is the integer $2k^2 + 2k$. Hence $(2k + 1)^2$ is odd.

This means that p by itself cannot be odd (remember that we showed that p^2 is an even integer), and p must therefore be even.

Returning to the proof of the theorem discussed above, we have now shown that p = 2k for some k because it is even, so that

$p^2 = 4k^2 = 2q^2$
i.e. $q^2 = 2k^2$
and thus q^2 is even, and ultimately q is also even.
Hence both p and q are even.

Remember that we assumed p and q to have no common terms, hence we derived a *contradiction*. Something is wrong somewhere. Every step in our argument can be justified, except the assumption that $x^2 = 2$ for some x = p/q, and therefore we conclude that the initial assumption is false. That is, no rational number x, written in the form p/q, can be found such that $x^2 = 2$.

QED

Activity 2-3: The general form of an even or odd integer

Take note of the general form of an even or odd integer provided in the previous proof. What is the general form of an integer that is a multiple of 3?

Now we have some idea of what $\sqrt{2}$ is not. It is not any of the rational numbers. In order to solve simple equations such as $x^2 = 2$, we need to expand $\mathbb{Q}$ by including what we call irrational numbers, such as $\sqrt{2}$.

Activity 2-4: Addition and multiplication involving irrational numbers

Suppose $a \neq 0$ is rational and b is irrational. How could one demonstrate that a + b and ab are both irrational?

We can get a clearer picture of what *irrationals* are like, if we use decimal notation. Just as a non-negative integer such as thirty could be written as

$$30 \text{ signifying } (3 \cdot 10) + (0 \cdot 1),$$

any other number can be expressed in decimal notation, for instance $-3/8$ could be written as

$$-0.375 \text{ signifying}$$

$$-[(0 \cdot 1) + (3 \cdot 1/10) + (7 \cdot 1/100) + (5 \cdot 1/1000)]$$

i.e. $-\left[\frac{3}{10} + \frac{7}{100} + \frac{5}{1000}\right]$.

Some decimals cannot be expressed in such a simple form, because the fractional part goes on forever, for example

$$1/3 = 0.3333 \ldots$$

and $5/7 = 0.7142857142857 \ldots$
and $\sqrt{2} = 1.4142 \ldots$

Of course you know how to express 1/3 and 5/7 as decimals: just divide the bottom into the top. But you may be wondering how we got the decimal expansion of $\sqrt{2}$.

We know that $1 < \sqrt{2} < 2$, i.e. that $\sqrt{2}$ lies between 1 and 2. How do we know this? Well, $1^2 = 1$ is too small, and $2^2 = 4$ is too big, because $(\sqrt{2})^2 = 2$.

Now then, let's take as our first approximation of $\sqrt{2}$ the number 1.5. Since $(1.5)^2 = 2.25$, we see that 1.5 is too big, i.e. $\sqrt{2} < 1.5$. What about 1.4 as an approximation? But $(1.4)^2 = 1.96$ so 1.4 is too small, i.e. $1.4 < \sqrt{2}$. Matters have improved. Earlier we knew only that $1 < \sqrt{2} < 2$. Now we know that $1.4 < \sqrt{2} < 1.5$.

So let's look at the approximation 1.45 for $\sqrt{2}$. But $(1.45)^2 = 2.1025$ which is too big. So we look at 1.44, and so forth. Eventually we will get to the point where we can show that

$$1.4142 < \sqrt{2} < 1.4143.$$

Clearly we can go on to calculate $\sqrt{2}$ to an arbitrary number of positions after the comma.

Let's return to the point we made earlier, namely that *some decimal expansions never terminate.* There is a subtle difference between the decimal expansions of the rationals 1/3 and 5/7, on the one hand, and those of the irrationals, such as $\sqrt{2}$ on the other. In the case of the rationals, the expansions *repeat* a digit (3 in 0.333 ...) or a group of digits (714285 in 0.7142857142857 ...), and so we call these expansions *repeating decimals*.

In the case of $\sqrt{2}$, it can be proved that the decimal expansion *never* even begins to repeat.

Definition: Real numbers

The expanded number system that consists of the **combination** of the rational plus the irrational numbers is called $\mathbb{R}$, i.e. the set of the *real numbers*.

In general, irrational numbers correspond to non-repeating decimals, while rationals correspond to repeating decimals (often the digit that repeats is 0, in which case we may speak of a terminating decimal). This makes it easy for us to construct new irrationals. Just form a non-repeating decimal, e.g. 0.10110111011110 ...

What about arithmetic in $\mathbb{R}$? Well, addition and multiplication have exactly the same useful properties as addition and multiplication in $\mathbb{Q}$, but it is much harder to describe exactly what is going on. You see, to handle infinite decimals, one really needs the concept of a limit, which you will encounter only if you take a Calculus module. Not having the concept at our disposal, we will not investigate the arithmetic of real numbers any further.

There are also, of course, the *complex number system*, but a discussion of this topic is beyond the scope of this module.

Activity 2-5: Expressing zero as a ratio

Express 0 as a ratio. Why does the rule of the multiplicative inverse exclude 0?

There are many ways to express 0 as a ratio, for instance 0/1 (since 0/1 represents the number by which 1 must be multiplied to give 0 and we know that number is 0), or 0/2 (since 0/2 represents the number by which 2 must be multiplied to give 0, and we know that multiplying 2 by 0 will do the trick), and so on.

In general, 0 may be written in the form 0/b, where b is any non-zero integer.

The reason the multiplicative inverse is defined for non-zero rationals only, is that a multiplicative inverse for 0 would have to be a number that, when multiplied by 0, gives 1. But no such number exists; in the previous study unit we have already learned that any number multiplied by 0 will be 0.

Activity 2-6: n/n is equal to 1

Why is a ratio of the form n/n equal to 1?

Simple. The ratio n/n represents the number by which the denominator n must be multiplied to give the numerator n. By which number should n be multiplied to give n back again? The multiplicative identity, 1, of course.

Activity 2-7: Integers as fractions

Does it make sense to call 113 a fraction?

People use the word "fraction" ambiguously, that is, we cannot be sure what they mean. Often they have in mind "a part of the whole", in other words a number between 0 and 1. Certainly 113 is not a fraction in this sense, but in the context of the present section, any number that can be written as a ratio may be called a fraction, and since 113 may be written as 113/1, we are allowed to refer to 113 as a fraction.

To avoid confusion between the two uses of the word "fraction", we call a number between 0 and 1 a *proper fraction* and a number such as 113/1 an *improper fraction*.

Activity 2-8: Self-assessment exercises Application skills

1. Define the words "even" and "odd" for positive integers.

For the following questions, substantiate your answers with proofs or counterexamples.

2. Is it the case that $m+ (n \cdot k) = (m + n) (m + k)$ for all positive integers m, n, and k?
3. Are there any even prime numbers besides 2?
4. If m and n are even, is m + n even?
5. If m and n are odd, is $m \cdot n$ odd?

2.5 In summary of the study unit

In this study unit you ensured that you can answer the following questions:

- How do we express the concept of a rational number?
- What is the major difference between a rational and an irrational number?
- How is a real number defined?
- Why can't we divide by zero?
- Which property holds for set $\mathbb{Q}$ but not for set $\mathbb{Z}$?

The next study unit is an introduction to set theory.

NOTES

NOTES

Study unit 3 Sets

Key questions for this study unit

- *What is meant by the following concepts: "set", "list notation", "set-builder notation", "element", "subset", "proper subset", "set equality", "union", "intersection", "difference", "symmetric difference", "complement", "set equality"?*
- *How would you describe sets correctly using list notation and set-builder notation?*
- *How would you construct new sets from old ones by forming subsets, unions, intersections, complements, differences and symmetric differences?*
- *What does it mean if we say that two sets are disjoint?*
- *How is the cardinality of a set defined?*

3.1 Introduction to this study unit

The previous study unit covered study material that is very important to understand so that the remaining units in this study guide make sense. In this study unit, we start with the serious stuff and introduce you to set theory. Set theory is just an important-sounding word that refers to discussions on topics that have to do with sets. It is essential for your computing studies that you understand the concepts we cover in this study unit.

Activity 3-1: Overview ***Study skill***

Draw a mind map of the different sections/headings you will deal with in this study session. Then page through the study unit with the purpose of completing the map.

Your map should include the concepts "set", "elements", "set membership", "universal set", "empty set", "subset", "set union", "set intersection", "set complement", "set difference", "symmetric difference" and "power set".

Activity 3-2: Concepts ***Conceptual skill***

Test your own knowledge (write in pencil) and then correct your understanding afterwards (erase and write the correct description). Often a young language may not have all the terms in a discipline; can you think of some examples?

English term	Description	Term in your home language
Set		
List notation		
Set-builder notation		
Element/member		
Set equality		
Universal set		
Empty set		
Subset/proper subset		
Set union		
Set intersection		
Set difference / relative complement		
Set complement		
Disjoint sets		
Symmetric difference		
Set cardinality		
Power set		

3.2 Why do set theory?

Firstly, we need to explain what a set is. As a simple example, think of a set as a bag. Not a real bag made of plastic or paper, but an imaginary one. Inside the bag you will find *distinct objects* of the set. These objects can also be called *members* or *elements* of the set.

We do set theory for a variety of reasons.

- We eventually want to arrive at the concepts "relation" and "function"; and
- we want to learn how to prove theorems, and set theory is a convenient source of reasonably simple proofs on which to practise.

3.3 How do we talk about sets?

Suppose we want to tell you something interesting about the set of all positive integers less than 5. Then, of course, we can refer to it as "the set of all positive integers less than 5". But this appears rather clumsy.

Another possibility is to use *list notation*. This involves writing down *curly* brackets (to represent the set or bag) and then writing the names of all the members of the set inside the curly brackets, separated by commas. This gives, in the case of our example, {1, 2, 3, 4}. This is a pretty neat and clear description of the set of all positive integers less than 5, not so? (List notation is often referred to as the *roster method*.)

A third possibility is to give the set a name and subsequently, in the discussion, to use the name. For instance, we might begin a discussion by saying "Let A be the set of all positive integers less than 5". For the rest of our discussion you will know that whenever we talk of the set A, then we mean the set of all positive integers less than 5. We use this trick often, particularly in referring to the sets $\mathbb{Z}^+$, $\mathbb{Z}^{\geq}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$.

The fourth and last possibility is to use *set-builder notation*. At first sight it is more complicated than, say, list notation. But it compensates for that by being more useful. You see, list notation is severely limited; we cannot write down all the elements of a set when the set we want to describe has lots of elements.

Example

Suppose we want to talk about some positive integers using list notation. We can't possibly write down the names of all the elements.

All we can do is to list the names of the first few and then put down what is called an *ellipsis*, i.e. three little dots that stand for "and so on":

{3, 5, 7, ...}

But notice how the dots introduce an element of vagueness. After all, if we look at the set {3, 5, 7, ...}, we are unsure whether we're talking about the set of odd integers greater than 1 or about the set of prime numbers greater than 2. The latter set differs from the former, as we can see by listing a few more elements:

{3, 5, 7, 9, 11, 13, 15, 17, 19, 21, ...}

is not the same as

{3, 5, 7, 11, 13, 17, 19, 23, ...}.

We'll be discussing the concept "is the same" presently.

Set-builder notation avoids having to use dots, by using a *variable* instead. We use set-builder notation to write $\mathbb{Z}^+$ as

{ x | x is a positive integer}.

Read this in the following manner:

{ x | x is a positive integer}
↑ ↑ ↑ ↑
The set ... of all x's ... such that ... x is a positive integer.

The letter x is called a variable, because it is not the name of a specific number. In mathematics the word "variable" means "place-holder", because one gets a specific statement if one replaces the variable with the name of something.

Examples

From "x > 3" we get the specific statement "5 > 3" if we push the name 5 into the space kept open by the variable x.

Another example: Suppose we want to talk about the set of all positive integers less than 5. Using set-builder notation, we can write

{x | x is a positive integer less than 5}.

Abbreviations: Set membership

Suppose we want to say "3 is a *member* of $\mathbb{Z}$". Then it will be convenient if we can abbreviate the phrase "is a member of". The symbol we usually use for this abbreviation is "$\in$". So, "3 is a member of $\mathbb{Z}$" can be written as "$3 \in \mathbb{Z}$". We can also say "3 is an *element* of $\mathbb{Z}$".

It could be that some element, say –2, is not a member of the set $\mathbb{Z}^+$, then we say "–2 is *not a member* of $\mathbb{Z}^+$", and we may write "$-2 \notin \mathbb{Z}^+$".

Similarly, in order to say "2 is a member of {0, 1, 2}", we may write "$2 \in \{0, 1, 2\}$", and to say "3 is not a member of {0, 1, 2}", we may write "$3 \notin \{0, 1, 2\}$".

The symbol "$\in$" is a streamlined version of the letter epsilon (ε) which is the Greek version of the English letter e, and e is the first letter of the word "element".

Sets should be well-defined. In general, a set can be described in different ways, and we illustrate this by the following examples:

Examples

The set of even non-negative integers less than 10 can be described in (at least) three ways:

{0, 2, 4, 6, 8},
{x | x is an even non-negative integer less than 10}, and
{x | $x \in \mathbb{Z}^{\geq}$, x is an even integer less than or equal to 8}.

These three descriptions might look different, but clearly they refer to the same collection of things. We indicate this by writing

{0, 2, 4, 6, 8} = {x | x is an even non-negative integer less than 10}
= {x | $x \in \mathbb{Z}^{\geq}$, x is an even integer less than or equal to 8}.

Another example: Suppose we want to talk about the set of all negative integers greater than –5. We can describe this set by using
list notation: {–4, –3, –2, –1}, or
set-builder notation: {x | x is a negative integer greater than –5}.

We may write

{–4, –3, –2, –1} = {x | x is a negative integer greater than –5}
= {y | $y \in \mathbb{Z}, -4 \leq y < 0$}.

There are also other alternatives that can describe this set.

Here the "=" stands for "is the same set as" or, if you prefer, "is equal to".
Note that "x is an even non-negative integer less than 10", and "x is a negative integer greater than –5" are referred to as property descriptions.

Note: Different variables can be used when defining a set,
e.g. {y | y ∈ ℤ, –4 ≤ y < 0} = {z | z ∈ ℤ, –4 ≤ z < 0}, and other alternatives can also define this set.

But how do we check whether or not two sets are equal? We can't always expect it to be obvious.

Example

Consider, on the one hand
{x | x is a real number and $1 < x < 2$}, and on the other hand
{x | x is a real number and $x^2 - 3x + 2 < 0$}.
It is certainly not obvious whether or not these descriptions refer to the same set.

The test for equality can be done using the following principle: The important thing about a set is the elements which are inside the set, just as the important thing about a shopping bag is the groceries inside. So it makes sense to regard two *sets* as *equal if they have precisely the same elements*. So, for example {3, 4} = {4, 3}.

Although the order in which the elements are listed differs, exactly the same numbers are in {3, 4} as are in {4, 3}, so these sets are equal.

It is important to note from this example that *the order in which the elements are listed is not significant*.

You might wonder whether there are some instances in which the order is important, but we will discuss this in the next study unit.

Another example: {5, 7} = {5, 5, 7}.

Although the number 5 is listed twice in the right-hand set, it does not tell us anything we don't already know. The numbers which are in {5, 7} are 5 and 7, and these are exactly the numbers that are in {5, 5, 7}. So these sets are equal.

This brings up another important point: Elements may live in more than one set. A repetition does not change the elements of a set.

Since repeatedly listing the name of an element is just a waste of energy, we will not write descriptions such as {2, 3, 2} or {1, 2, 3, 3, 3}. We often have to test whether or not two sets are equal, and it is usually not enough to convince ourselves that they are equal – we also have to convince everyone else, and so we have to write out a proof. This is discussed further in the next study unit.

Note: Sets do not need to have only numbers as elements. We may actually throw different objects of our choice into a "bag". Let's look at a few examples:

We can compile sets such as {a, b, c} (a set with some letters as elements), {1, 2, a, b} (a set with different kinds of element), or {Thabo, Amy, Hanifa} (a set with names of people as elements).

Another example: "boxA" can be considered to be a set with elements "black pen", "red pen", "pencil", "rubber$_1$" and "rubber$_2$". More elements can be included in this set – can you think of more items to put in "boxA"? A set should always be well-defined, thus *all* the distinct elements in "boxA" should be listed:

boxA = {black pen, red pen, pencil, rubber$_1$, rubber$_2$}

We looked at sets that have some elements, but do all sets have elements?

We call a set with *no elements* an *empty set*, denoted by $\emptyset$. Using list notation to indicate the empty set: { }. (Our imaginary bag has *no* elements.)

If we want to use set-builder notation to describe this set we can consider some set that does not have any elements. We know, for example, that there does not exist an integer that is, at the same time, positive and less than 1. If we use set builder notation, one possibility is $\{x < 1 \mid x \in \mathbb{Z}^+\}$.

There are numerous acceptable alternatives:
the set of all rational numbers greater than 2 and less that −2,
the set of all odd integers divisible by 2 without a remainder,
the set of all negative integers greater than 0, and so on.
These descriptions have one thing in common: they all give an entrance requirement that is impossible to fulfil, thereby ensuring that the set has no members.

You might ask: Is it possible for $\emptyset$ to be an element of some set? (We can put $\emptyset$ into our imaginary bag: {$\emptyset$}.)

Examples

Consider the set {1}. This set only has one element namely 1.

The set {$\emptyset$, 1, {1} } has three elements namely $\emptyset$, 1 and {1}.

As we have previously seen, we can include any element of our choice in some set. All the elements are separated by commas.

Does it bother you that ∅ and {1} could be elements of a set? Later in this study unit we will see that a special kind of set namely a "powerset" can be formed with ∅ and some other sets as members.

Activity 3-3: Self-assessment exercises *Application skills*

1. Ambiguous (or indistinct/unclear) descriptions of sets are given below. In each case, describe the set more concisely, firstly using list notation and then using set-builder notation.
   (a) The set of all even non-negative integers less than 10.
   (b) The set of all odd negative integers greater than –13.
   (c) The set of all positive integers less than 1.
   (d) The set of all real numbers greater than 2.

2. The following are descriptions of sets, given in list or in set-builder notation. In each case give an unambiguous description in English.
   (a) $\{-1, 0, 1\}$
   (b) $\{x \in \mathbb{R} \mid 0 < x < 1\}$
   (c) $\{0\}$
   (d) $\{\mathbb{Z}\}$

In Activity 6 of the introductory unit of this study guide and in tutorial letter 101 we discuss the interactive computer-aided instruction (CAI) tutorial that is available on a CD which you should have received. This tutorial will help you to understand the concept "sets".

Activity 3-4: CAI tutorial

You can now do the "sets" part of the CAI tutorial.

3.4 How to build new sets from old ones

Note: In the next study unit we introduce Venn diagrams illustrating definitions in this study unit.

When we want to talk about some *subset*, we need to place this subset in the context of some *universal* set. A *universal* set is simply the collection of all things of the kind we want to talk about. If not otherwise specified, we usually name a universal set "U".

Now let's talk about *subsets*. Suppose we take the set of integers $\mathbb{Z}$ to be our *universal set*, then we can form some set {1, 2} from $\mathbb{Z}$. What did we do to form {1, 2}? Well, by singling out the elements 1 and 2, we have in a certain sense thrown away all the other members of $\mathbb{Z}$.

The simplest way to build a new set is to throw away some of the elements of an old set. It's rather like having a bag of sweets, pulling one out, and eating it. Something has changed; the bag of sweets is not the same as it was.

Definition: Subset

If A and B are sets from a universal set U, we say that A is a subset of B if and only if every element of A is also an element of B.

We may abbreviate "A is a subset of B" by writing "$A \subseteq B$".

We can abbreviate "if and only if" by writing "iff", so we can write "A is a subset of B iff every element of A is also an element of B". Only in the next study unit will we see what is exactly meant by "iff".

Note: One can throw away none, one, or more elements from some set B, then the resulting set, let's call it A, is a *subset* of B, i.e. $A \subseteq B$.

Just as we write "$x \leq y$" to say that x is less than or equal to y (or x is not greater than y), so "$A \subseteq B$" means that A has no elements which do not also belong to B, i.e. every element of A is also an element of B.

Examples

Suppose that B = {1, 2, 3}. Throwing away the element 2 gives the subset A = {1, 3}. So A is a subset of B, i.e. $A \subseteq B$.
(Each element of A is an element of B.)

When we throw away no element of B, then $B \subseteq B$, or it could happen that we throw away all the elements of B, then $\{\} \subseteq B$.

Another example: Let C = **{** ∅, {∅} **}** with ∅ and {∅} being the elements of C. (The element {∅} is highlighted so that one can remember that this is **one** element of C.)

We can form subsets of C by throwing away some elements from C:

Throwing away both elements of C gives the subset **{ }**;
throwing away the element ∅ gives the subset **{** {∅} **}**;
throwing away the element {∅} gives the subset **{∅}**; and finally
of course, $C \subseteq C$.

We have formed all the subsets of C namely **{ }**, **{∅}**, **{** {∅} **}** and C.

It is interesting to note that
{ } $\subseteq$ **{ }**; **{ }** $\subseteq$ **{∅}**; **{ }** $\subseteq$ **{** {∅} **}**; **{ }** $\subseteq$ C; and
{∅} $\subseteq$ C; **{** {∅} **}** $\subseteq$ C; but
{∅} is **not** a subset of **{** {∅} **}** since
the only element of **{∅}** is ∅ and the only element of **{** {∅} **}** is {∅},
so the element of **{∅}** is not an element of **{** {∅} **}**.

Note: We can consider *any* set and throw away *all* its elements, then we are left with the subset { }. This means that { } is a subset of *any* set.

Definition: Proper subset

If C and D are sets from a universal set U, and if every element of C is an element of D, but D has some element(s) that is/are not in C, (i.e. C has less elements than D), we call C a proper subset of D.
We may abbreviate "C is a proper subset of D" by writing $C \subset D$.

In order to form a proper subset A of B, we have to throw away at least one element from B to form A. This means that if A is a proper subset of B, then B has one or more elements that is/are in B but that is/are not in A.

When we write "$A \subseteq B$" we mean that A is a proper subset of B, or A is equal to B, so even if A is a proper subset of B we may write "$A \subseteq B$". We usually just talk about "subsets" unless we specifically want to mention a "proper subset".

Another way to build new sets is to form the union of two sets. This is rather like going to school on Monday with your History and English books in your bag, on Tuesday with your English and Maths books, and on Wednesday with your History books, your English books and your Maths books as well. If you think of your bag on Monday as set A, your bag on Tuesday as set B, and your bag on Wednesday as set C, then set C is the union of sets A and B.

Definition: Set union

The *union* of sets A and B is denoted by $A \cup B$, and is the set of all those elements which belong to A or to B (or to both). More briefly,
$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$.

This definition allows us to say
"if $x \in A \cup B$, then $x \in A$ or $x \in B$", and
"if $x \in A$ or $x \in B$, then $x \in A \cup B$".

In the next study unit we will see that these two statements can be combined:
"$x \in A \cup B$ iff $x \in A$ or $x \in B$".

When we say "x is an element of A or B", we mean "either x is an element of A, or x is an element of B, or x is an element of both A and B", i.e we mean that x is an element of A or B in an *inclusive* sense.

Example

Let $A = \{1, 2\}$ and $B = \{0, 1\}$,
then $A \cup B = \{0, 1, 2\}$,
i.e. the set of those elements that belong to A or to B.

A third way to construct sets is to form the intersection of given sets. This is somewhat like going to school on Monday and Tuesday with the same books as before, and on Wednesday taking along only your English books. Set C now consists of precisely the books common to set A and set B.

Definition: Set intersection

The *intersection* of sets A and B is denoted by $A \cap B$, and is the set of all those elements which belong to both A and B at the same time. More briefly, $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.

Now we can write
"if $x \in A \cap B$, then $x \in A$ and $x \in B$", and
"if $x \in A$ and $x \in B$, then $x \in A \cap B$".

We can combine these two statements:
"$x \in A \cap B$ iff $x \in A$ and $x \in B$".

Example

Let $A = \{1, 2\}$ and $B = \{0, 1\}$,
then $A \cap B = \{1\}$,
i.e. the set of those elements that belong to both A and B.

A fourth way to construct a set is to form the complement of one set relative to another. This is a bit like taking the same books to school on Monday and on Tuesday as previously, and then on Wednesday taking only your History books. Set C now only has as elements those books in set A that did not also appear in set B, so in a way it's as if you subtracted set B from set A.

Definition: Set difference

The *complement of B relative to A* is denoted by $A - B$, and is the set of all those elements of A which do not belong to B. More briefly,
$A - B = \{x \mid x \in A \text{ and } x \notin B\}$.
This is also referred to as the *difference* between sets A and B.

We can write
"if $x \in A - B$, then $x \in A$ and $x \notin B$", and
"if $x \in A$ and $x \notin B$, then $x \in A - B$".

These two statements can be combined:
"$x \in A - B$ iff $x \in A$ and $x \notin B$".

Example

Let $A = \{1, 2\}$ and $B = \{0, 1\}$,
then $A - B = \{2\}$,
i.e. the set of those elements that belong to A but not to B.

There is a special case of the relative complement construction that will be of importance to us. But first, some background material.

It often happens that all the sets involved in a particular discussion are subsets of a single larger set. For instance, if a problem involves the sets H = {Java, C^{++}, $C^{\#}$}, I = {Haskell, Prolog}, and J = {COBOL, Visual Basic}, then these are all subsets of U = {x | x is a programming language}. This means that {Java, C^{++}, $C^{\#}$, Haskell, Prolog, COBOL, Visual Basic} $\subseteq$ U.

Since all the elements of sets H, I and J are also elements of U, U can be called a *universal set* for this particular problem.
Now, if U is a universal set and B is a subset of U, then U – B is called the *complement* of B for short (i.e. we may omit the phrase "relative to U") and is denoted more briefly by B′. So B′ = {x | x ∉ B}.

Definition: Set complement

Let A be a subset of a universal set U. Then the *complement* of A is defined as the set of all elements that belong to U but do not belong to A, i.e. A′ = {x | x ∈ U and x ∉ A} (or A′ = {x | x ∉ A}).

Example

Let U = {0, 1, 2, 3} and A = {0, 1}, then A′ = U – A = {2, 3},
i.e. the set of all those elements that belong to U but not to A.

Notation: You will find that different books refer to the complement of B in different ways, for example by writing B^C or ~B. So always make sure what notation a particular author is using.

Definition: Symmetric set difference

The *symmetric difference* between two sets A and B, written as A + B, is defined as the set of all elements that belong to A or to B but not to both:
A + B = { x | x ∈ A or x ∈ B, but not both.}

Now we can write
"if x ∈ A + B, then x ∈ A or x ∈ B, but not both", and
"if x ∈ A or x ∈ B, but not both, then x ∈ A + B".

We combine these two statements:
"x ∈ A + B iff x ∈ A or x ∈ B, but not both".

Example

Let A = {0, 1, 2, 3} and B = {0, 1, 3, 4}, then A + B = {2, 4},
i.e. the set of those elements that belong to A or to B, but not to both.

Isn't A + B exactly the same as A ∪ B? Let us investigate by looking at an example.

Activity 3-5: Symmetric difference between sets

Use an example to investigate whether or not A + B and A ∪ B have the same result.

Let A = {3, 4} and B = {4, 5, 6} with U = {1, 2, 3, 4, 5, 6, 7, 8}, then

A ∪ B = {3, 4} ∪ {4, 5, 6} = {3, 4, 5, 6} and
A + B = {3, 4} + {4, 5, 6} = {3, 5, 6}.

In this example we see that A ∪ B ≠ A + B.

It is clear that the results from these two operations are not the same. This is due to the fact that, as we have mentioned before, in the English language, there are two different ways in which one can interpret the word "or".

The definition of $A \cup B$ uses the *inclusive* meaning of "or", that is, $\{x \in U \mid x \in A \text{ or } x \in B, \textit{or both}\}$.

The definition of A + B on the other hand, uses the *exclusive* meaning of "or" to be $\{x \in U \mid$ either $x \in A$ or $x \in B$, *but not both*$\}$. The members common to both A and B are not included in A + B. This means that $A + B = (A \cup B) - (A \cap B)$.

Our example also shows that
$(A \cup B) - (A \cap B) = \{3, 4, 5, 6\} - \{4\} = \{3, 5, 6\} = A + B$.

Is it possible that $C \cup D = C + D$ for some sets C and D? Yes, whenever $C \cap D = \emptyset$, then $C \cup D = C + D$. It is also true that whenever $C \cup D = C + D$, then $C \cap D = \emptyset$.

An example can illustrate this: Suppose $A = \{2, 4\}$ and $B = \{5, 6\}$, then we have $A \cup B = A + B = \{2, 4, 5, 6\}$. (Note that A and B have no common elements.)

Definition: Set disjointness

Two sets A and B are called *disjoint* if they have no elements in common, i.e. there is not a single element that live in both A and B, i.e. $A \cap B = \emptyset$.

Example

Let $A = \{1, 2, 3, 4\}$ and $B = \{5, 6\}$ with $U = \{1, 2, 3, 4, 5, 6\}$.

Then we can say that A and B are disjoint. There are no elements that belong to both A and B, i.e. $A \cap B = \emptyset$.

Sometimes, it could be important for us to know how many elements some set has.

Definition: Set cardinality

Let A be a set with k distinct elements that can be counted. The *number* of *elements* in A is called the *cardinality* of the set. The cardinality of A is denoted by n(A), and n(A) = k. We can use the notation n(A) or |A|.

Note: In this study guide we normally use the notation |A|.

Example

Let $A = \{1, 2, 3, 4\}$ and $B = \{5, 6\}$ be subsets of $U = \{1, 2, 3, 4, 5, 6\}$,
then $|A| = 4$, $|B| = 2$ and $|U| = 6$,
i.e. A has four, B has two, and U has six elements.

Note that the cardinality of $\emptyset$ is 0 since the set $\emptyset = \{\ \}$ has no elements.

Before we conclude this study unit, we introduce the very important concept of a *power set.*

Definition: Power set

Given a set A with n distinct elements, the *power set of A*, denoted by $\mathcal{P}(A)$, is the set that has as its members *all* the subsets of A.

The cardinality of $\mathcal{P}(A)$ is 2^n, i.e. $|\mathcal{P}(A)| = 2^n$.

When we determine the power set of some set B (i.e. $\mathcal{P}(B)$), we first have to determine all the subsets of B. All these subsets of B are the elements of $\mathcal{P}(B)$. We illustrate this with an example:

Example

Let B = {1, 2, 3}. Which sets are all subsets of B?

Let us list them: ∅, {1}, {2}, {3}, {1, 2}, {1, 3}, {2, 3} and {1, 2, 3}. These are all the *members* of $\mathcal{P}(B)$.

The power set of B, i.e. $\mathcal{P}(B)$ is therefore

$\mathcal{P}(B) = \{ \emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\} \}$.

One can throw away *all* the elements of *any* set, then the subset $\{ \} = \emptyset$ is formed. This means that ∅ is a *subset of any set*. When a power set is formed, ∅ will always be included as a *member* of the power set.

We now want to look at some *subsets* of $\mathcal{P}(B)$. Note that every member of a *subset of $\mathcal{P}(B)$* is also a member of $\mathcal{P}(B)$.

Some subsets of $\mathcal{P}(B)$:

$\{\emptyset\} \subseteq \mathcal{P}(B)$, since {∅} has only one member namely ∅, and ∅ is also a member of $\mathcal{P}(B)$.

$\{ \emptyset, \{1\}, \{2\} \} \subseteq \mathcal{P}(B)$, since every member of { ∅, {1}, {2} } (i.e. ∅, {1} and {2}) are also members of $\mathcal{P}(B)$.

$\{ \{1\}, \{1, 2, 3\} \} \subseteq \mathcal{P}(B)$, since very member of { {1}, {1, 2, 3} } (i.e. {1} and {1, 2, 3}) are also members of $\mathcal{P}(B)$.

$\{ \} \subseteq \mathcal{P}(B)$. The subset { } is formed by throwing away all the members of $\mathcal{P}(B)$.

We also have that $\mathcal{P}(B) \subseteq \mathcal{P}(B)$.

Can you determine the other subsets of $\mathcal{P}(B)$?

Activity 3-6: Self-assessment exercises ***Application skills***

1. Let a universal set be U = {1, 2, 3, 4, 5} and let A = {1, 2, 3} and B = {3, 4, 5}. Determine the required sets.
   (a) $A \cup B$ and $B \cup A$
   (b) $A \cap B$ and $B \cap A$

(c) $A - B$ and $B - A$
(d) $A + B$ and $B + A$

2. Let a universal set be $U = \{a, e, i, o, u\}$ and let $A = \{i, o, u\}$ and $B = \{a, e, o, u\}$. Determine the following sets:
   (a) A' and $(A')'$
   (b) B' and $(B')'$
   (c) $A \cup B$ and $(A \cup B)'$
   (d) $A' \cap B'$
   (e) $A \cap B$ and $(A \cap B)'$
   (f) $A' \cup B'$
   (g) $A - B$ and $B - A$
   (h) $A \cap B'$ and $B \cap A'$
   (i) $A + B$ and $B + A$

3. Let a universal set be $U = \{1, 2, 3, 4, 5\}$, and let $A = \{3\}$ and $B = \{\{3\}, 4, 5\}$. Determine $\mathcal{P}(A)$ and $\mathcal{P}(B)$.

4. Let a universal set be $U = \{a, e, i, o, u\}$ and let $A = \{i, o, u\}$ and $B = \{a, e, o, u\}$. Determine the following sets:
   (a) $\mathcal{P}(A)$ and $\mathcal{P}(B)$.
   (b) $\mathcal{P}(A \cap B)$ and $\mathcal{P}(A) \cap \mathcal{P}(B)$.
   (c) $\mathcal{P}(A')$ and $(\mathcal{P}(A))'$ (The complement of $\mathcal{P}(A)$ is taken relative to $\mathcal{P}(U)$.)
   (d) $\mathcal{P}(A) \cup \mathcal{P}(B)$ and $\mathcal{P}(A \cup B)$.

3.5 In summary of the study unit

In this study unit you ensured that you can answer the following questions on set theory:

- What does it mean if set A is a subset of set B?
- Why do we need a universal set?
- What does the empty set represent?
- Which elements does the union of two sets have?
- Which elements does the intersection of two sets have?
- What does the cardinality of a set mean?
- What does $a \in A$ represent?
- What does it mean if two sets are disjoint?
- What is the complement of a set with respect to the universal set?
- What is the difference between two sets?
- Which elements do the symmetric difference between two sets have?
- What is the power set of a given set?

In the following study unit we will learn more about Venn diagrams that can be used to give a graphical representation of a set or of operations involving sets. We also look at proofs where sets are involved.

Study unit 4 Proofs involving sets

Key questions for this study unit

- *How can we represent sets using Venn diagrams?*
- *How can we use Venn diagrams to investigate whether or not the theoretic identities of sets hold?*
- *How do we prove the theoretic identities of sets using the phrase "if and only if"?*
- *How do we handle the proof of a set identity involving specific sets?*
- *How do we establish the falsity of a universal assertion by providing a counterexample?*
- *What is useful about the Inclusion-exclusion principle?*

Activity 4-1: Overview *Study skill*

Draw a mind map of the different sections/headings that you will deal with in this study session. Then page through the study unit with the purpose of completing the map.

Your map should include the concepts of a Venn diagram, different kinds of proof involving sets, a counterexample in the context of sets and the inclusion-exclusion principle.

Activity 4-2: Concepts *Conceptual skill*

Test your own knowledge (write in pencil) and then correct your understanding afterwards (erase and write the correct description). Often a young language may not have all the terms in a discipline; can you think of some examples?

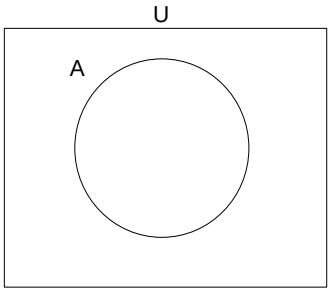
English term	Description	Term in your home language
Venn diagram		
iff		
Set equality		
Set identity		
Counterexample		
Inclusion-exclusion principle		

4.1 Venn diagrams

One can draw pictures of sets. If the drawing is done as described below, we call the picture a *Venn diagram*.

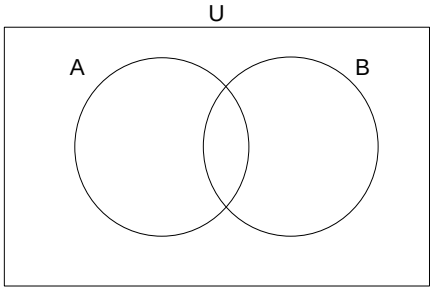
Let's begin with the simplest case: a single set A, which is a subset of a universal set U (for example, A might be $\mathbb{Z}^+$ and U might be $\mathbb{R}$).

Draw a rectangle to represent the "bag" U, called the *universal set,* and draw a circle inside the rectangle to represent A. The diagram shows that all the elements of A are also elements of U, but there might be elements of U that fall outside the boundaries of the set A:

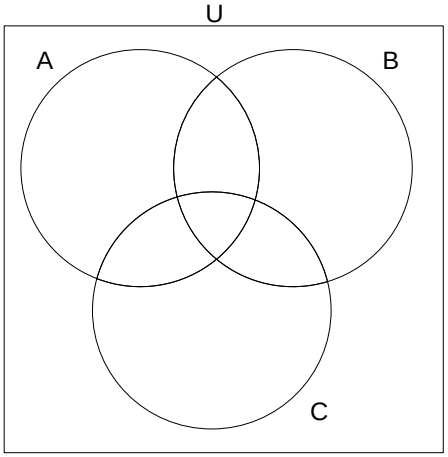


Venn diagrams start to get interesting when we perform constructions involving two or three subsets of U.

Suppose we are given two sets A and B and a universal set U. As before, we represent each of the sets by means of a circle inside the rectangle, as shown in the following figure:

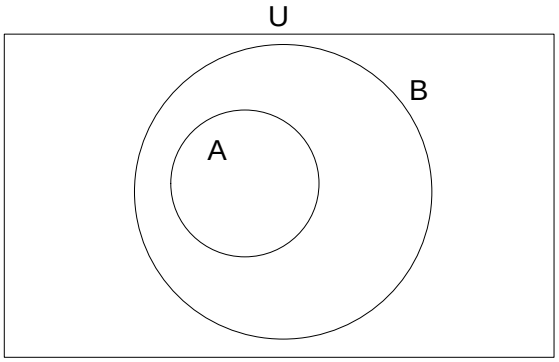


We could also have a Venn diagram with three sets A, B and C, as shown in the following figure:



Let us recap on the definition for "subset": For all sets A and B, A is a subset of B, i.e. $A \subseteq B$, iff every element of A is also an element of B.

In the following Venn diagram we see that every element of A is also an element of B, and every element of A or B is also an element of U:



Note: B could have some elements that are not in A, and U could have some elements that are not in A or B.

Example

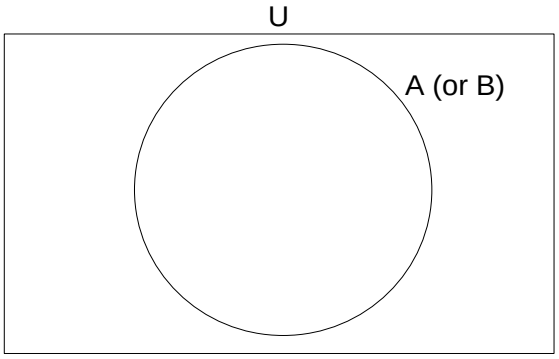
Let A = {♣, [♣]} and B = {♣, [♣], ♠}. Every element in A, namely ♣ and [♣], is also an element of B. This means that $A \subseteq B$.

As we have seen in the previous study unit, it is possible that two sets can be regarded as being equal if they have the same distinct elements. We provide a formal definition:

Definition: Set equality

For any sets A and B, if $A \subseteq B$ and $B \subseteq A$, then every element of A is also an element of B, and every element of B is also an element of A, i.e. A = B.

In terms of Venn diagrams we can look at it in this way:

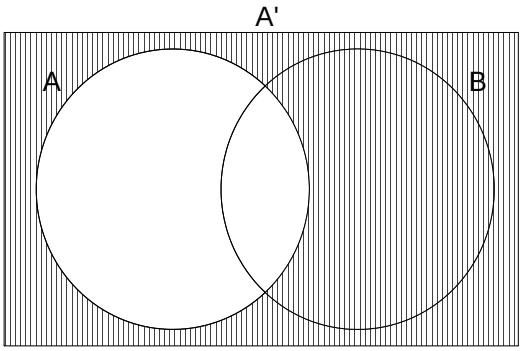


Note: In all the Venn diagrams that follow, the shaded area represents the result of the particular set operation and the whole rectangle represents the set "U".

We provide the specific definition of each set operation as it was provided in the previous study unit. Relate the shaded area in each Venn diagram to the relevant definition.

For example: If x is an element of $A \cap B$, then x is an element of both A and B. In a Venn diagram representing $A \cap B$, the shaded area shows the area where the elements live that belong to ***both*** *A and B.*

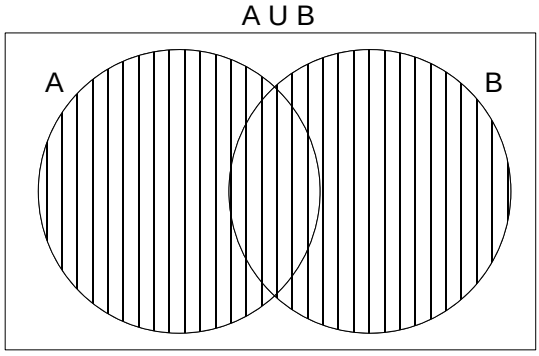
Set complement:
Set $A' = \{x \mid x \notin A\}$, which is the complement of set A, is represented as follows in a Venn diagram:



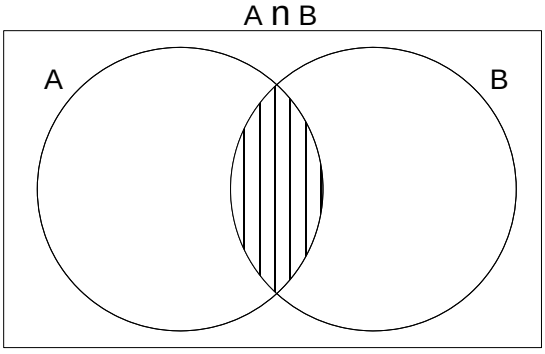
As can be seen from the Venn diagram for A′, no element x lives inside A (i.e. $x \notin A$).

We can now illustrate a number of operations on sets by shading the appropriate areas in Venn diagrams.

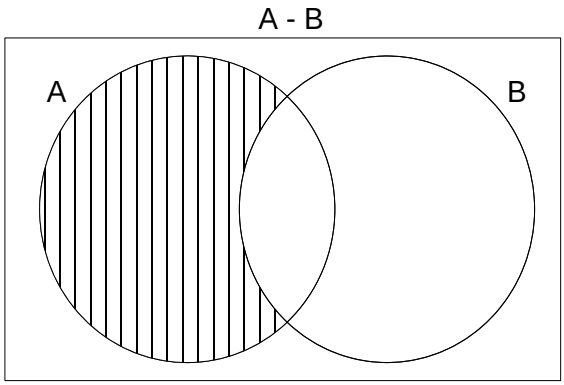
Union:
The union of any sets A and B, i.e. $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$, is depicted by the shaded area:



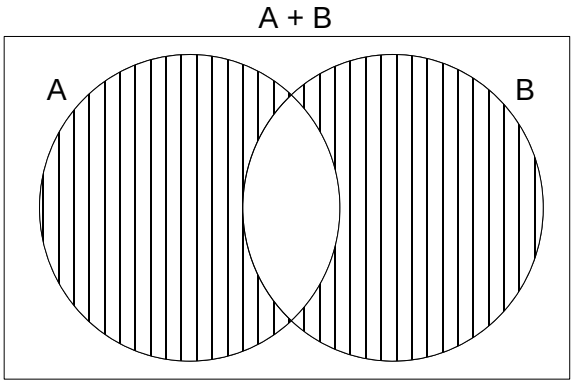
Intersection:
The intersection of any sets A and B, i.e. $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$, is depicted by the shaded area:



Set difference:
The complement of set B relative to set A, i.e. $A - B = \{x \mid x \in A \text{ and } x \notin B\}$, is depicted by the shaded area:



Symmetric difference:
For sets A and B, the symmetric difference,
i.e. $A + B = \{x \mid x \in A \text{ or } x \in B, \text{ but not both}\}$, is depicted by the shaded area:



Let's draw a Venn diagram for a more complex expression before the self-assessment exercises can be attempted.

Activity 4-3: Venn diagrams

Let $A, B, C \subseteq U$. Draw the Venn diagram for $[(A \cup B) - (A \cap B)] \cup C$.

How do we go about doing this? By drawing the Venn diagram for $[(A \cup B) - (A \cap B)] \cup C$ in stages, one considerably reduces one's chances of making unnecessary mistakes.

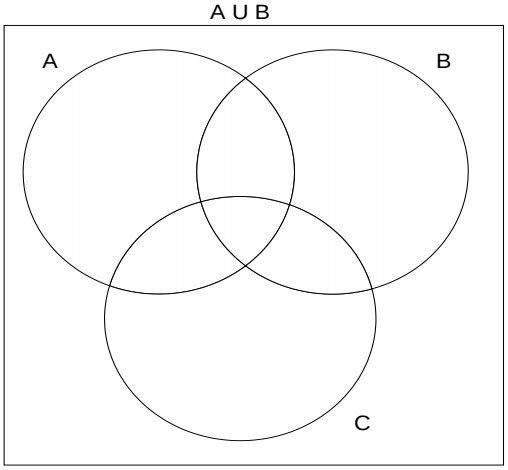
First draw the Venn diagram for $(A \cup B) - (A \cap B)$ in stages:
Draw the diagrams for $(A \cup B)$ and $(A \cap B)$ separately, and then draw the diagram for $(A \cup B) - (A \cap B)$.

(Compare the Venn diagram of $(A \cup B) - (A \cap B)$ with that of $A + B$ which we saw previously in this study unit. Clearly, these two operations give the same result!)

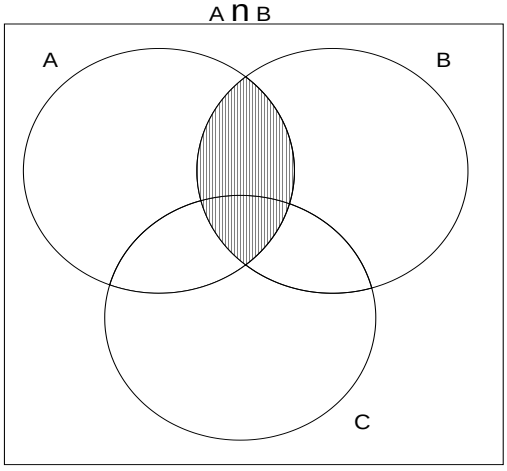
Finally the Venn diagram for $[(A \cup B) - (A \cap B)] \cup C$ can be drawn.

Note: Include the universal set U and *all three* sets A, B and C in *each* diagram.

Step 1: $A \cup B$ is depicted by the shaded area:

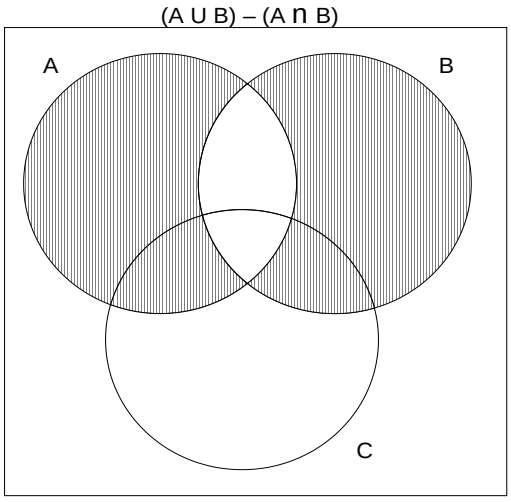


Step 2: $A \cap B$ is depicted by the shaded area:

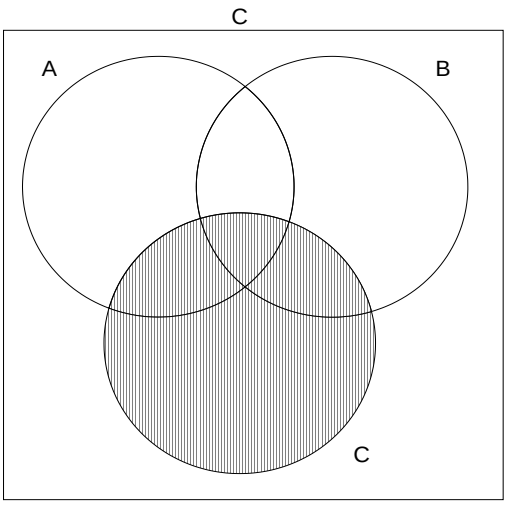


Consider the shaded areas in the above diagrams. The shaded area for $(A \cup B) - (A \cap B)$ can be derived from the following: the shaded area $A \cap B$ should be removed from the shaded area $A \cup B$ as we see in step 3. (The elements of $(A \cup B) - (A \cap B)$ live in $(A \cup B)$ but not in $(A \cap B)$.)

Step 3: $(A \cup B) - (A \cap B)$ is depicted by the shaded area:

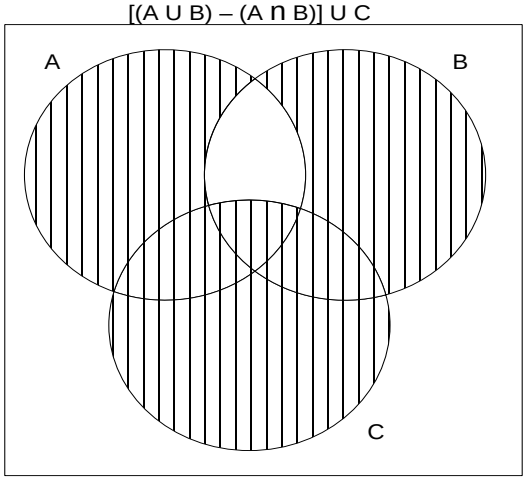


Step 4: C is depicted by the shaded area:



Consider the shaded areas in the above diagrams. The shaded area for $[(A \cup B) - (A \cap B)] \cup C$ can be derived from the following: add the shaded area of C to that of $[(A \cup B) - (A \cap B)]$ as we see in step 5. (The elements of $[(A \cup B) - (A \cap B)] \cup C$ live in $(A \cup B) - (A \cap B)$ or in C, or in both.)

Finally, step 5: $[(A \cup B) - (A \cap B)] \cup C$ is depicted by the shaded area:



Activity 4-4: Self-assessment exercises ***Application skills***

1. Using a Venn diagram for two subsets, X and Y of U, show, by shading the appropriate region, what each of the following sets looks like:

(a) $(X \cup Y)'$
(b) $X' \cap Y'$
(c) $(X \cap Y)'$
(d) $X' \cup Y'$

2. Using a Venn diagram for three subsets X, Y and Z of U, show, by shading the appropriate region, what each of the following sets looks like (do these step by step):

(a) $X - (Y \cup Z)$
(b) $(X - Y) \cup (X - Z)$
(c) $X \cap (Y - Z)$
(d) $(X \cap Y) - (X \cap Z)$
(e) $X \cap (Y + Z)$
(f) $(X \cap Y) + (X \cap Z)$

Note : The solutions to the self-assessment exercises are provided in tutorial letter 102.

4.2 Proofs

We first look at an example. Let A = {1, 2} and B = {2, 4} with U = {1, 2, 4}. Then $A \cup B$ = {1, 2, 4}. Similarly $B \cup A$ = {1, 2, 4}. In other words, $A \cup B = B \cup A$ for the given sets.

If we think a little bit about unions, we realise that the order of A and B does not play a role in the definition of union. And so it ought to be true for all possible choices of sets A and B that $A \cup B = B \cup A$.

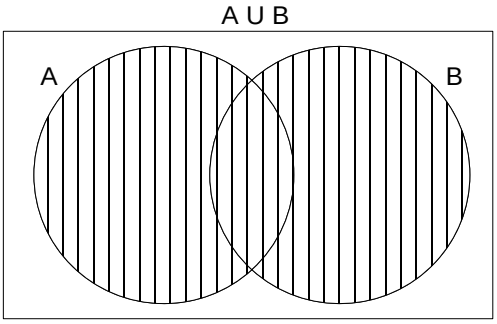
Now if only we could *prove* that for all sets A and B, $A \cup B = B \cup A$, then it would not be necessary to test whether {2, 7} ∪ {5, 9} is the same set as {5, 9} ∪ {2, 7}, and whether {3, 4, 5} ∪ {1} is equal to {1} ∪ {3, 4, 5}, and so forth.

If we could prove that $A \cup B = B \cup A$ for all subsets A and B of a universal set U, then we could say that
$A \cup B = B \cup A$ is an *identity.* But how can we establish the identity
$A \cup B = B \cup A$ by a strong and convincing argument?

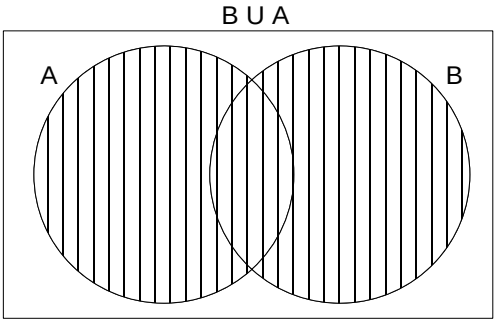
Well, one can investigate whether or not $A \cup B = B \cup A$ is an *identity* by first drawing a Venn diagram shaded to represent $A \cup B$. Then draw a Venn diagram shaded to represent $B \cup A$. If the shaded portions are the same, it shows that $A \cup B = B \cup A$ for any sets A and B.

Let's consider $A \cup B = B \cup A$ again:

Left-hand side: Draw the Venn diagram of $A \cup B$ by first shading A and then B.



Right-hand side: Draw the Venn diagram of $B \cup A$ by first shading B and then A.



By using Venn diagrams, we have shown that $A \cup B = B \cup A$ for any sets A and B. Perhaps our "proof" is not very convincing, because it relies on our willingness to accept that Venn diagrams give an accurate picture of sets. *The study of more rigorous proofs is called logic* and we will give you just a taste of logic later in this module.

There are some limitations to the use of Venn diagrams in proofs. One has already been mentioned, namely that some people might claim that the proofs are not rigorous enough, because there are assumptions built into the technique that are not obvious – such as the assumption that Venn diagrams give a good picture of sets. This objection is not our main problem, since we do not, on this level need to be very formal and rigorous.

However, Venn diagrams become difficult to draw when there are more than three subsets involved, but we will not investigate these further.

Activity 4-5: Self-assessment exercises *Application skills*

Use Venn diagrams to determine, for all subsets X, Y and W of a universal set U, whether or not the following equations hold:

(a) $X - (Y \cap W) = (X - Y) \cup (X - W)$
(b) $X \cap (Y \cap W) = (X \cap Y) \cap W$
(c) $X \cap (Y \cup W) = (X \cap Y) \cup (X \cap W)$
(d) $(X')' = X$

When looking at a Venn diagram, one is inclined to forget that some (or all) of the sets represented in the picture may be empty. We will therefore also practise *writing out proofs in words*. We begin by practising on simple examples for which we have already drawn Venn diagrams.

Example

Let us prove that, for all subsets A and B of U, $A \cup B = B \cup A$.

In our proof we have to show that the set $A \cup B$ has exactly the same elements as the set $B \cup A$.

There is a standard technique for doing proofs of this kind. First we show that every element of the left-hand side set is also an element of the right-hand side set (i.e. we show that LHSset $\subseteq$ RHSset). Then we go in reverse and show that every element of the right-hand side set is also an element of the left-hand side set (i.e. RHSset $\subseteq$ LHSset).
If LHSset $\subseteq$ RHS set and RHSset $\subseteq$ LHSset, then we may conclude that LHSset = RHSset.

Let's begin by choosing any element x of $A \cup B$. Now $x \in A \cup B$, and this means, by the definition of "$\cup$", that $x \in A$ or $x \in B$. But the statement "$x \in A$ or $x \in B$" means the same as the statement "$x \in B$ or $x \in A$". Therefore we know that $x \in B$ or $x \in A$. By the definition of "$\cup$", this means that $x \in (B \cup A)$.

So we have shown that no matter which element x of $A \cup B$ we choose,
if $x \in (A \cup B)$ then $x \in B \cup A$,
so $A \cup B \subseteq B \cup A$.

(This ends the first half of the proof.)

For the *converse* part of the proof, begin by choosing any element x of $B \cup A$.

If $x \in B \cup A$
then $x \in B$ or $x \in A$
i.e. $x \in A$ or $x \in B$
i.e. $x \in (A \cup B)$.

Now we have shown that
if $x \in B \cup A$ then $x \in A \cup B$,
so $B \cup A \subseteq A \cup B$.

(This ends the second half of the proof.)

We conclude (from both halves of the proof) that $A \cup B = B \cup A$.

QED

Note that we have not tried to be particularly brief in the proof. Conciseness comes with practice. It is more important that the reader should be happy with each line of the proof before going on to the next line.

Let's look at a second example.

Example

Prove that $A \cup (B \cup C) = (A \cup B) \cup C$ for all sets A, B, $C \subseteq U$.

We first prove that $A \cup (B \cup C) \subseteq (A \cup B) \cup C$ and then that $(A \cup B) \cup C \subseteq A \cup (B \cup C)$:

Let $x \in A \cup (B \cup C)$.
Then $x \in A$ or $x \in (B \cup C)$
i.e. $x \in A$ or ($x \in B$ or $x \in C$)
i.e. ($x \in A$ or $x \in B$) or $x \in C$ (since surely the statements
"$x \in A$ or ($x \in B$ or $x \in C$)" and
"($x \in A$ or $x \in B$) or $x \in C$" mean the same)
i.e. $x \in A \cup B$ or $x \in C$
i.e. $x \in (A \cup B) \cup C$.
(This ends the first half of the proof.)

Conversely, let $x \in (A \cup B) \cup C$.

Then $x \in (A \cup B)$ or $x \in C$
i.e. ($x \in A$ or $x \in B$) or $x \in C$
i.e. $x \in A$ or ($x \in B$ or $x \in C$)
i.e. $x \in A$ or $x \in (B \cup C)$
i.e. $x \in A \cup (B \cup C)$.
(This ends the second half of the proof.)

We conclude (from both halves of the proof) that $A \cup (B \cup C) = (A \cup B) \cup C$.

At this stage you might want to know whether we can somehow shorten this type of proof. The arguments used in the examples above fall into two halves. The first half begins with the assumption "Let x belong to the left-hand set" and reasons are given step by step up to the conclusion "then x belongs to the right-hand set". The second half begins with the assumption "Let x belong to the right-hand set" and reasons are logically given up to the conclusion "then x also belongs to the left-hand set".

The second half is just a mirror image of the first.

In such cases, we use the phrase "if and only if", which we abbreviate as "iff". The "if" part represents the backward reasoning of the second half, while the "only if" part represents the forward reasoning of the first half. We will discuss this further in a subsequent study unit.

If we look back at our subset definition in the previous study unit, we can now see that "A is a subset of B "iff" every element of A is also an element of B" means that "A is a subset of B if every element of A is also an element of B" and "if A is a subset of B then every element of A is also an element of B".

4.3 Working with $(X \cap Y)'$ and $(X \cup Y)'$

If we turn to the Venn diagram of "complement" in a previous section of this study unit, we see that A′ is the set U − A, i.e. the set of all elements that belong to U but not to A.

In set-builder notation: $A' = \{x \in U \mid x \notin A\}$. We may also write this set as $\{x \mid x \notin A\}$ if everyone knows what our universal set is. Remember that the reader must always be kept informed about our choice of universal set.

When working with the complement, the symbol "$\notin$" can be used as follows: $x \in Y'$ iff $x \notin Y$.

However, when we use the notation "$\notin$" combined with "$\cup$" or "$\cap$", we get a few surprises!

Look at the following example:

$x \in (X \cup Y)'$
iff $x \notin (X \cup Y)$
iff $x \notin X$ and $x \notin Y$
iff $x \in X'$ and $x \in Y'$
iff $x \in X' \cap Y'$

This shows that $(X \cup Y)' = X' \cap Y'$.

If we look at the definition of "union" in a previous section of this study unit, we associate "union", i.e. "$\cup$", with the word "or". For example, if x resides in $A \cup B$ (i.e. $x \in (A \cup B)$), then x belongs to A *or* to B *or* to both of them.

The statement "$x \notin X \cup Y$" tells us that x resides outside $X \cup Y$, and under this condition, x can belong to neither X nor Y, i.e. x cannot reside in either of the two sets. This means that x resides outside X *and* x resides outside Y, i.e. $x \notin X$ and $x \notin Y$.

It helps to draw Venn diagrams for $A \cup B$ and $(A \cup B)'$ – then we can see it clearly.

Next consider the definition of "intersection", i.e. "$\cap$", as portrayed in a previous section of this study unit. For example, if x resides in $A \cap B$ (i.e. $x \in A \cap B$), then x belongs to A *and* to B. We associate "intersection" with the word "and". For the symbol "$\notin$", it is rather different.

Over to you – try to continue the following expansion …

$x \in (X \cap Y)'$
iff $x \notin X \cap Y$
.
.
.
Can you prove that $(X \cap Y)' = X' \cup Y'$? If you are not sure about the proof, draw the Venn diagrams and look at the definitions of "union", "intersection" and "complement" again.

Activity 4-6: Self-assessment exercises *Application skills*

Using *if and only if* statements, write out a proof for each of the following identities, where X, Y and W are arbitrary subsets of a universal set U:

(a) $(X')' = X$
(b) $X - (Y \cap W) = (X - Y) \cup (X - W)$
(c) $X \cap (Y \cap W) = (X \cap Y) \cap W$
(d) $X \cap (Y \cup W) = (X \cap Y) \cup (X \cap W)$

We can combine proof methods when investigating whether or not some given statement is always true. First we draw Venn diagrams for the sets in a given statement. If the statement appears to hold, i.e. the final Venn diagrams are equivalent, we give a formal proof. If not, we give a ***counterexample***.

Example

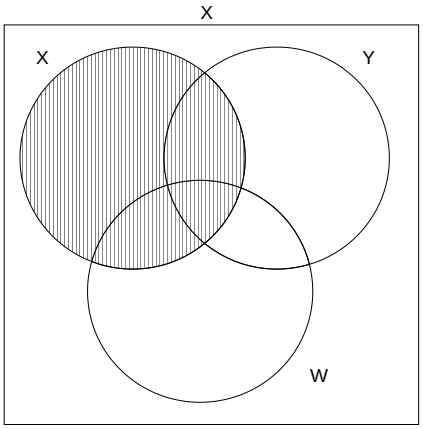
Use Venn diagrams to investigate whether or not, for all subsets X, Y, and W of U, $X - (Y \cup W) = (X - Y) \cap (X - W)$.

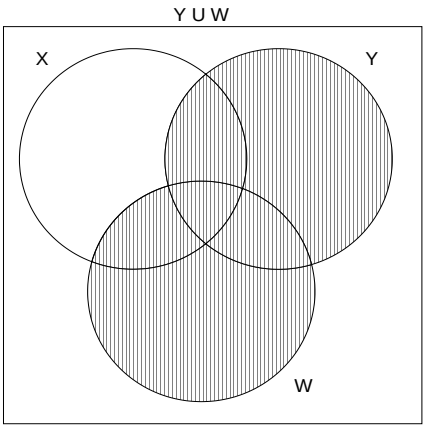
If it appears to be true, provide a proof; if not, provide a counterexample.

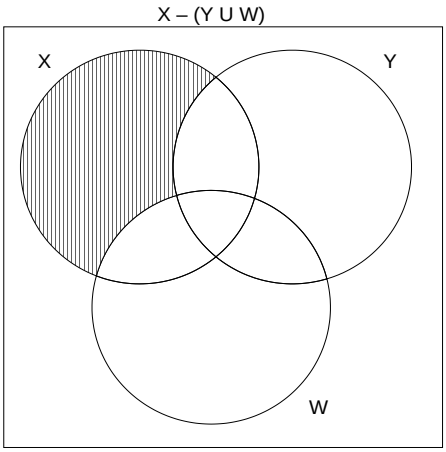
Solution:

We draw the Venn diagrams *step by step* as follows:

Left-hand side:

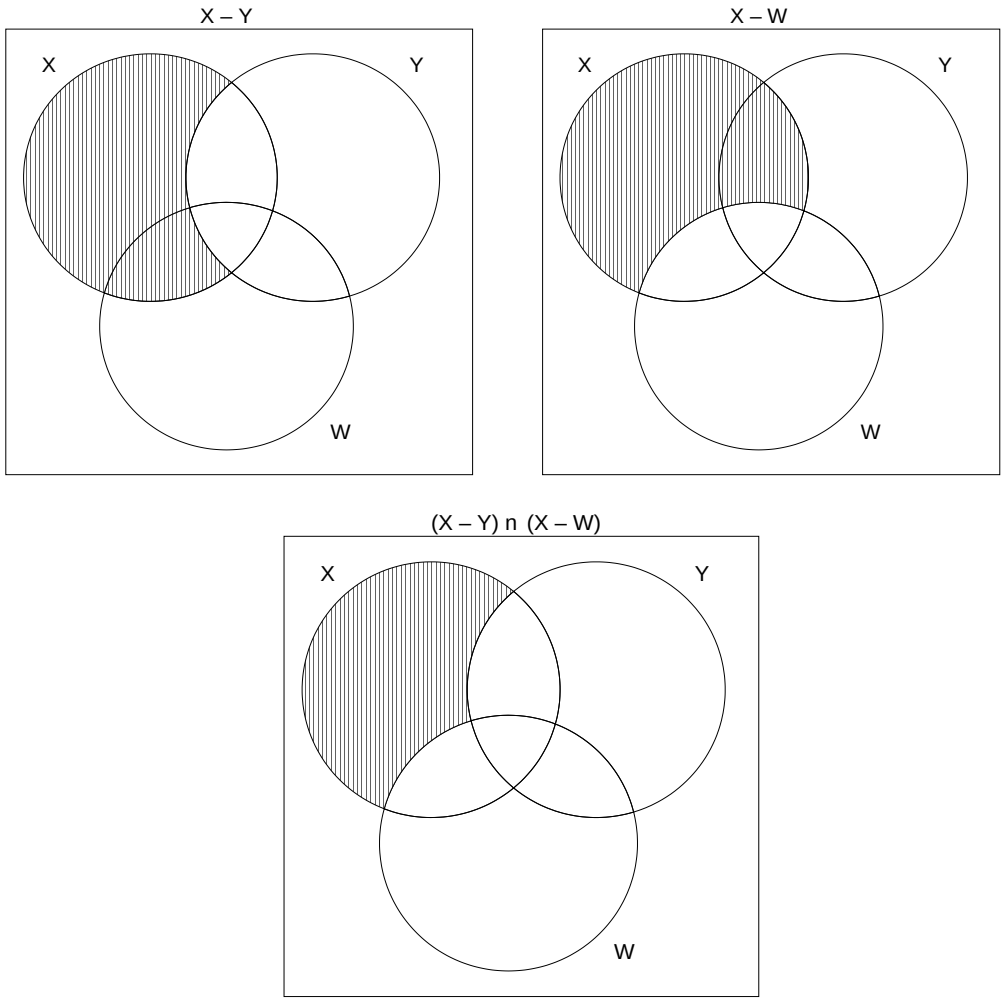






The shaded area represents the set of elements that live in X but not in $Y \cup W$.

Right-hand side:



The shaded area represents the set of elements that live in X – Y and in X – W.

Since the two resulting Venn diagrams are identical, it appears that the statement holds, so we provide a proof. (In this proof we apply the definitions of the set operations provided in the previous study unit.)

$x \in X - (Y \cup W)$
iff $x \in X$ and $x \notin (Y \cup W)$
iff $x \in X$ and $x \in (Y \cup W)'$
iff $x \in X$ and $(x \in Y'$ and $x \in W')$
iff $x \in X$ and $(x \notin Y$ and $x \notin W)$
iff $(x \in X$ and $x \notin Y)$ and $(x \in X$ and $x \notin W)$
iff $x \in (X - Y)$ and $x \in (X - W)$
iff $x \in (X - Y) \cap (X - W)$

Thus $X - (Y \cup W) = (X - Y) \cap (X - W)$ for all subsets X, Y and W of U.

So far we have only dealt with proofs of equality, and these involve proofs which go in two directions and which can be shortened using *iff*.

There are other kinds of proof that are important too. For instance, we might want to show that two sets are not necessarily equal.

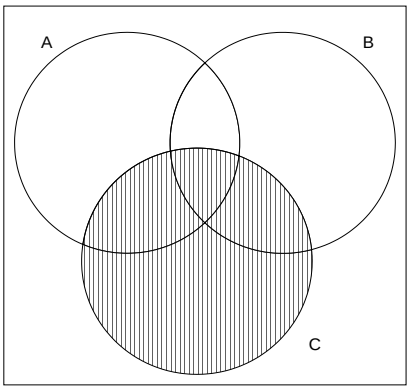
Activity 4-7: Set equality / inequality

Use Venn diagrams to investigate whether or not, for all A, B, C $\subseteq$ U,
$(A + B) \cup C = (A \cup C) + (B \cup C)$.

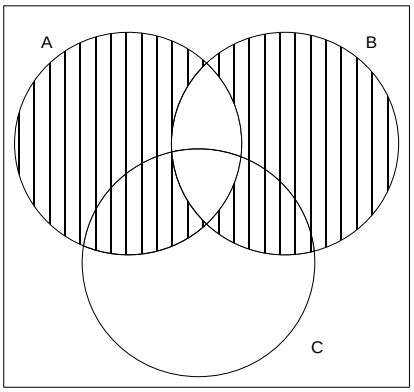
We first draw the Venn diagrams for $(A + B) \cup C$ and for $(A \cup C) + (B \cup C)$.

Left-hand side:

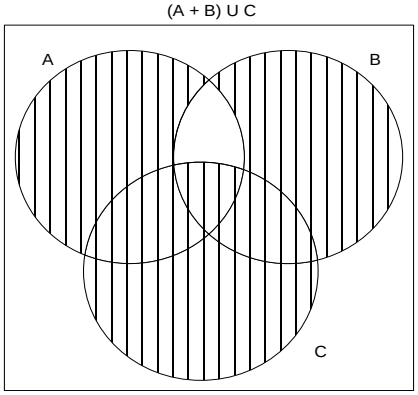
C



A + B

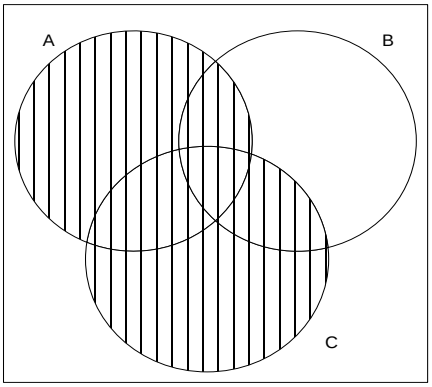


(A + B) U C

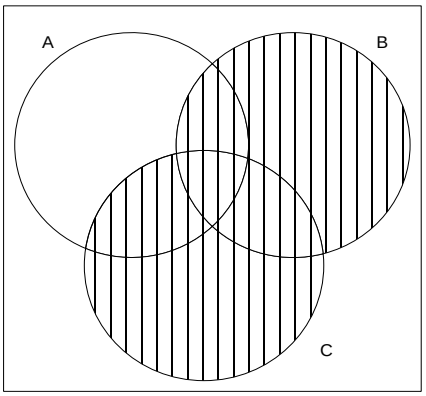


Right-hand side:

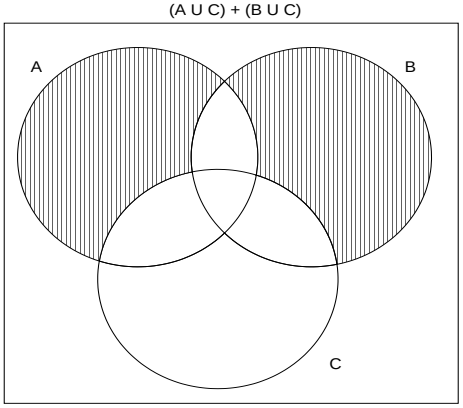
A U C



B U C



(A U C) + (B U C)



$(A + B) \cup C$ is the set with elements that live in $(A + B)$ or C, and
$(A \cup C) + (B \cup C)$ is the set with elements that live in
either $(A \cup C)$ or $(B \cup C)$, but not both.
In this case, it is clear that we have been trying to prove something that is not true, i.e. $(A + B) \cup C$ is *not* equal to $(A \cup C) + (B \cup C)$ for any sets A, B and C.

In order to prove that $(A + B) \cup C$ is not necessarily equal to $(A \cup C) + (B \cup C)$, we need to find a *counterexample*, that is, a choice of sets A, B and C such that $(A + B) \cup C \neq (A \cup C) + (B \cup C)$.

So what we want is a concrete example of sets which shows that the left-hand side is different from the right-hand side.

The sets we choose for the ***counterexample*** must be chosen in such a way that some element is present in a part of the respective diagrams where they *differ.* In the example we used above, the two *final diagrams differ in respect of C*, so we choose some *element that appears in C only*. (We choose $4 \in C$ but $4 \notin A$ and $4 \notin B$. We choose any other elements from U to live in sets A, B and C.)

Counterexample:

Let $A = \{1, 2\}$, $B = \{2, 3\}$, and $C = \{1, 4\}$, with
$U = \{1, 2, 3, 4\}$.

Left-hand side:
$(A + B) \cup C = \{1, 3\} \cup \{1, 4\} = \{1, 3, 4\}$ i.e. those elements that reside in $\{1, 3\}$ or $\{1, 4\}$ or both.

Right-hand side:
$(A \cup C) + (B \cup C) = \{1, 2, 4\} + \{1, 2, 3, 4\} = \{3\}$, i.e. an element that resides in either $\{1, 2, 4\}$ or $\{1, 2, 3, 4\}$ but not both.

This *counterexample* shows that the statement $(A + B) \cup C = (A \cup C) + (B \cup C)$ does not hold for all subsets A, B and C of U.

Perhaps this is a good moment to say more about the term "identity", which we referred to earlier in this study unit. You know that an equation such as ax = b has either no real solution (e.g. when a = 0 and b = 2), or a real solution (e.g. when a = 3 and b = 0), or else allows every real number to be a solution (e.g. when a = 0 and b = 0).

Definition: An identity

An equation which is satisfied by every possible value of the unknown(s) is called an *identity*.

On the previous pages we have been showing that certain equations such as

$A \cup B = B \cup A$

are identities, i.e. are satisfied by any subsets A and B of any universal set.

We have also shown that the equation $(A + B) \cup C = (A \cup C) + (B \cup C)$
is *not* an identity by finding values for the unknowns A, B, and C which do not satisfy the equation.

Another type of proof involves showing, not that one set is equal to another, but that one set is a subset of another set. This is easy, since it involves just the first half of an equality proof.

We give an example of such a proof. The example involves the very important concept of *power sets* which was defined in the previous study unit.

Example

$\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$, i.e. the union of power sets is a subset of the power set of the union.

(Refer to the definitions of "subset", "union" and "power set" in the previous study unit.)

It is important to note that, in the proof that follows, because X is an element of a power set, it represents a set.

Let $X \in \mathcal{P}(A) \cup \mathcal{P}(B)$

Then $X \in \mathcal{P}(A)$ or $X \in \mathcal{P}(B)$

i.e. $X \subseteq A$ or $X \subseteq B$

i.e. the members of X all live in A or all live in B, i.e. the members of X all live in $A \cup B$

i.e. $X \subseteq A \cup B$ (remember that X is a set)

i.e. $X \in \mathcal{P}(A \cup B)$.

Sometimes we might also want to show that one set is not necessarily a subset of another. As you may have suspected, we do this by finding a *counterexample*.

Activity 4-8: Self-assessment exercises *Application skills*

1. Is it the case that for all X, Y, $Z \subseteq U$, $X + (Y \cap Z) = (X + Y) \cap (X + Z)$?

2. Find examples of sets A and B such that $\mathcal{P}(A \cup B)$ is not a subset of $\mathcal{P}(A) \cup \mathcal{P}(B)$.

3. Is it the case that, for all X, $Y \subseteq U$, $\mathcal{P}(X) \cap \mathcal{P}(Y) = \mathcal{P}(X \cap Y)$? Justify your answer.

4. Use Venn diagrams to investigate whether or not, for all sets X, Y, $Z \subseteq U$, $X - (Y - Z) = (X - Y) - Z$. If the statement appears to hold, give a proof; if not, give a counterexample.

5. Use Venn diagrams to investigate whether or not, for all subsets A, B and C of U, $(A \cap B) + (C \cap A) = (A \cap B') \cup (B - C)$. If the statement appears to hold, give a proof; if not, give a counterexample.

Activity 4-9: CAI tutorial

Take note of the counterexamples provided in the tutorial.

A more detailed discussion of different types of proof is to be found in the study unit on logic.

4.4 The Inclusion-exclusion principle

As we saw in the previous study unit, it could sometimes be important for us to know how many elements some set has. We call the number of elements in a set the *cardinality* of the set. We indicate the cardinality of some set C by |C| or by n(C).

Examples

We determine the cardinality of some sets:

(a) $\emptyset$ The empty set has no elements.
The cardinality of set $\emptyset$: $|\emptyset| = 0$.

(b) {1, 2, 3} 1, 2, and 3 are the elements of {1, 2, 3}.
The cardinality of set {1, 2, 3}: $|\{1, 2, 3\}| = 3$.

(c) {$\emptyset$, {1, 2}} $\emptyset$ and {1, 2} are the elements of {$\emptyset$, {1, 2}}.
The cardinality of set {$\emptyset$, {1, 2}}: $|\{\emptyset, \{1, 2\}\}| = 2$.

Sometimes we want to determine how many elements live in some sets. The intersection of sets can be empty or some sets might have common members. The Inclusion-exclusion principle can be applied to determine the number of elements that live in some sets.

Theorem4.1: Inclusion-exclusion principle

For all finite sets X and Y, $|X \cup Y| = |X| + |Y| - |X \cap Y|$.

Proof

To calculate |X| + |Y|, count the elements of X and of Y and add the two numbers. The elements that belong to both X and Y will have been counted twice, so subtract $|X \cap Y|$.

We look at an example to illustrate this principle:

Example

Let X = {a, b, c, 1} and Y = {1, 2, 3}. It is clear that $X \cap Y = \{1\}$ thus $|X \cap Y| = 1$. We see that

|X| = 4 and |Y| = 3, thus $|X \cup Y| = |X| + |Y| - |X \cap Y| = 4 + 3 - 1 = 6$.

Check the answer: $X \cup Y = \{a, b, c, 1, 2, 3\}$. How many elements live in this set?

The theorem can be applied when X and Y have no common members. This leads to the sum rule:

Definition: Sum rule

If X and Y are disjoint sets (i.e. $X \cap Y = \emptyset$), and |X| = m and |Y| = n, then $|X \cup Y| = m + n$.

An example illustrates this rule:

Example

Let X = {a, b} and Y = {1, 2, 3}. It is clear that $X \cap Y = \{\ \}$. We see that |X| = 2 and |Y| = 3, thus $|X \cup Y| = 2 + 3 = 5$.

Returning to Venn diagrams for a moment, we illustrate the usefulness of these diagrams for extracting information regarding the size (or cardinality) of certain subsets, using the information we are given about a number of sets.

When dealing with the Inclusion-exclusion principle, we encounter two different types of question. In the first kind, we are told explicitly how many participants (or objects) is part of specific sets, and how many live in certain intersections of sets. We are then required to determine how many are part of a certain set or subset, to the *exclusion* of others.

In the second type of question *the unknown*, which we must determine, is the number of persons (or objects) that live in an intersection of sets, frequently the central intersection of three sets. So we call the unknown x, and carry on from there, taking special care to do the algebra accurately!

The *Inclusion-exclusion principle* is best illustrated by examples:

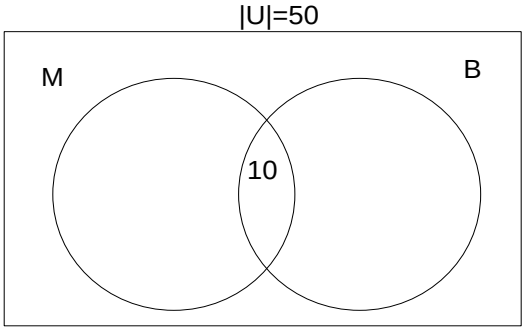
Example

In a group of 50 learners, 25 play Mastermind, 30 play basketball, and 10 play both.
(a) How many students play Mastermind or basketball (or both)?
(b) How many students do not play either Mastermind or basketball?

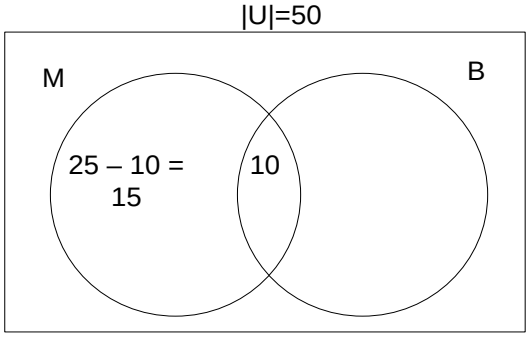
Let U be the set of all the students in the group, M the set of those playing Mastermind, and B the set of those playing basketball. Then $|U| = 50$, $|M| = 25$, $|B| = 30$, and $|M \cap B| = 10$.

Let's determine the number of students who play Mastermind or basketball by using Venn diagrams.

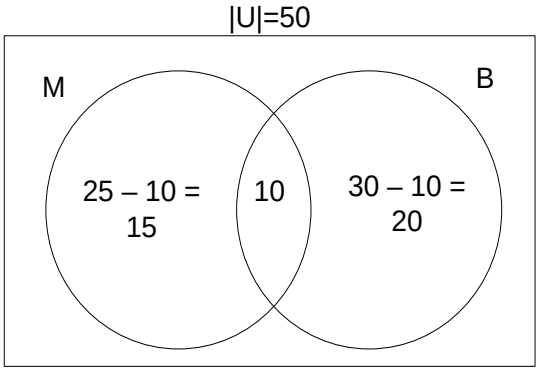
We use the above information to fill in the number of elements that live in each region, starting with the region in the middle. We know $|M \cap B| = 10$:



How many elements reside in M but not in $M \cap B$? Since $|M| = 25$ and $|M \cap B| = 10$, there are $25-10 = 15$ elements in this region.



Take the region on the right-hand side. How many elements live in B but not in $M \cap B$? Since $|B| = 30$ and $|M \cap B| = 10$, there are $30-10 = 20$ elements in this region.



Now all the regions have been filled in and we can answer the question:

(a) $|M \cup B| = 15 + 10 + 20 = 45$, i.e. 45 students play Mastermind or basketball.

Alternatively, by the Inclusion-exclusion principle, we get the same result: $|M \cup B| = |M| + |B| - |M \cap B| = 25 + 30 - 10 = 45$.

(b) We determine the number of students who do not play either Mastermind or basketball:

There are 50 students in total and 45 play Mastermind or basketball, so $|(M \cup B)'| = 50 - 45 = 5$, i.e. 5 do not play Mastermind or basketball.

Example

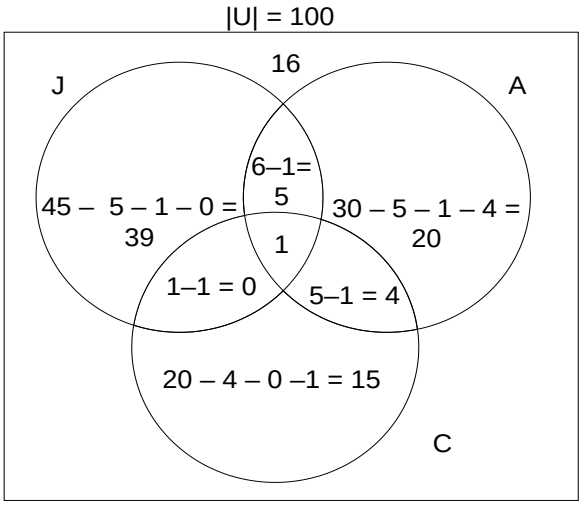
A questionnaire filled in by the 100 subscribers to Blue Scalpel Medical Insurance who submitted no claims during 2009 reveals that 45 jog regularly, 30 do aerobics regularly, 20 cycle regularly, 6 jog and do aerobics, 1 jogs and cycles, 5 do aerobics and cycle, and 1 jogs, cycles and does aerobics.

(a) How many of these healthy people do not participate regularly in any of the three activities mentioned?
(b) How many only jog?

We use the same Venn diagram to answer both questions. Let us firstly display the available information neatly. Let U be the set of subscribers who filled in the questionnaire, J the set of those that jog, A the set of those that do aerobics and C the set of those that cycle. Then

$|U| = 100$,
$|J| = 45$,
$|A| = 30$,
$|C| = 20$,
$|J \cap A| = 6$,
$|J \cap C| = 1$,
$|A \cap C| = 5$, and
$|J \cap A \cap C| = 1$.

Start by filling in 1 in the central intersection region (1 person participates in all three) then determine the other values by successive subtractions.
This gives the following diagram:



(a) How many of these healthy people do not participate regularly in any of the three activities mentioned?

The set $J \cup A \cup C$ turns out to have as members
$|J \cup A \cup C| = 39 + 5 + 20 + 4 + 15 + 1 = 84$ subscribers, therefore
$|(J \cup A \cup C)'| = |U| - |J \cup A \cup C| = 100 - 84 = 16$ do not participate regularly in any of the three activities mentioned.

(b) How many only jog?
From the diagram it is clear that 39 jog as their only exercise.

Activity 4-10: Self-assessment exercises (Inclusion-exclusion principle)

1. Suppose that of 1000 first-year students, 700 take Mathematics, 400 take Computer Science, and 800 take Mathematics or Computer Science.

   (a) How many take Mathematics and Computer Science?
   (b) How many students take Maths, but not Computer Science?
   (c) How many students do not take any of the two subjects?

2. A builder has a team of 64 construction workers. Of these, 45 are trained in the use of heavy machinery, i.e. cranes, bulldozers and backhoes. A total of 22 can operate cranes, 26 can operate backhoes, 4 can operate cranes and bulldozers, 6 can operate backhoes and bulldozers, 8 can operate cranes and backhoes, and 1 can operate all three kinds of machine. How many can operate bulldozers?

2. A large software company employs 22 software engineers for the design of systems. Of these engineers, 17 are well versed in the secrets of a formal method (FM), 9 can use the Unified Modelling Language (UML), and 9 are familiar with the use of entity-relationship (ER) diagrams. If 5 engineers can use both an FM and UML, 4 both an FM and ER diagrams, and 7 both UML and ER diagrams, answer the following questions:
   (a) How many engineers can use all 3 techniques, namely an FM, UML and ER diagrams?
   (b) How many engineers can use UML only?

4.5 Proofs on specific sets

We can prove that two specific sets are equal. Let's look at an example:

Example

Prove that
$\{w \in \mathbb{R} \mid w^2 - 3w + 2 < 0\} = \{z \in \mathbb{R} \mid 1 < z < 2\}$.

We should prove that each member of the left-hand side set belongs to the right-hand side set and conversely.

Proof

$x \in \{w \in \mathbb{R} \mid w^2 - 3w + 2 < 0\}$
iff $x \in \mathbb{R}$ and $x^2 - 3x + 2 < 0$
iff $x \in \mathbb{R}$ and $(x - 2)(x - 1) < 0$
iff $x \in \mathbb{R}$ and either $(x - 2 < 0$ and $x - 1 > 0)$ or $(x - 2 > 0$ and $x - 1 < 0)$
(since a minus times a plus gives a minus, or a plus times a minus gives a minus)
iff $x \in \mathbb{R}$ and either $(x < 2$ and $x > 1)$ or $(x > 2$ and $x < 1)$
(there are no real numbers that are simultaneously greater than 2 and less than 1)
iff $x \in \mathbb{R}$ and $1 < x < 2$
iff $x \in \{x \in \mathbb{R} \mid 1 < x < 2\}$ iff $x \in \{z \in \mathbb{R} \mid 1 < z < 2\}$

Note: Any variable can be used for a set description - it does not change the members in the set.

To go from the step $x \in \mathbb{R}$ and ((x < 2 and x > 1) or (x > 2 and x <1)) to the step $x \in \mathbb{R}$ and (1 < x < 2), we gave as justification that (x > 2 and x < 1) is false. But we must not forget that we are making use of "iff" and therefore we must also know why it is correct to go in the opposite direction. Given a true statement such as "The earth is round", we can add any other statement to it using "or", whether the added statement is true or not, and the resulting statement will still be true. For example, the statement "The earth is round or elephants are green" is true even though "elephants are green" is not true.

So we may go from $x \in \mathbb{R}$ and (1 < x < 2)
to the statement $x \in \mathbb{R}$ and ((x < 2 and x > 1) ***or (x > 2 and x < 1)****)*
even though "(x > 2 and x < 1)" is false.

Activity 4.11 Self-assessment exercises *Application skills*

Use the technique illustrated in the previous example to prove the following:

1. $\{y \in \mathbb{Z}^+ \mid y \text{ is an even prime number}\} = \{u \in \mathbb{Z}^+ \mid u^2 = 4\}$

2. $\mathcal{P}(\{0, 1\}) = \{\emptyset\} \cup \{\{0\}\} \cup \{\{1\}\} \cup \{\{0,1\}\}$

3. $\{x \in \mathbb{R} \mid x^2 + 6x + 5 < 0\} = \{x \in \mathbb{R} \mid -5 < x < -1\}$

4. $\{x \in \mathbb{Z} \mid x^2 - 5x + 4 < 0\} = \{x \in \mathbb{Z}^+ \mid x \text{ is a prime factor of } 6\}$

5. $\{x \in \mathbb{R} \mid x^2 + x - 2 > 0\} = \{x \in \mathbb{R} \mid x < -2 \text{ or } x > 1\}$

We conclude this study unit by completing a few additional activities designed around the topics covered here.

Activity 4-12: Self-assessment exercises ***Application skills***

1. Determine whether or not for V, W, Z $\subseteq$ U, if V $\subseteq$ W, then V $\cup$ Z $\subseteq$ W $\cup$ Z and V $\cap$ Z $\subseteq$ W $\cap$ Z. Provide either a proof or a counterexample, whichever is appropriate.

2. Is it the case that, for all subsets X, Y, W $\subseteq$ U, if X = Y and Y = W, then X = W, and if X $\subset$ Y and Y $\subset$ W, then X $\subset$ W? Justify your answer.

3. Is it the case that, for all subsets X of U, X $\cup$ $\emptyset$ = X? Justify your answer.

4. Is it true that, for all subsets V and W of U, V $\cap$ W = $\emptyset$ iff V = $\emptyset$ or W = $\emptyset$? Justify your answer. This claim has two parts:
(i) if V $\cap$ W = $\emptyset$ then V = $\emptyset$ or W = $\emptyset$, and (ii) if V = $\emptyset$ or W = $\emptyset$ then V $\cap$ W = $\emptyset$.
Both these parts must hold for the claim to be true.

5. Is it the case that for every subset X of U there exists a subset Y of U such that X $\cup$ Y = $\emptyset$? Justify your answer.

6. Is it the case that for every subset X of U there is some subset Y such that X $\cap$ Y = U? Justify your answer.

7. Using "*if and only if*"statements, prove the following:

   (a) X + Y = Y + X for all X, Y $\subseteq$ U.
   (b) X $\cap$ (Y + Z) = (X $\cap$ Y) + (X $\cap$ Z) for all X, Y, Z $\subseteq$ U.

4.6 In summary of the study unit

In this study unit you ensured that you can answer the following basic questions:

- How do you prove the theoretic identities of sets using Venn diagrams?
- How do you prove the theoretic identities of sets using the phrase "if and only if"?
- How do you establish the falsity of universal assertions by providing a counterexample?
- How do you handle proofs involving specific sets?
- What is the usefulness of the Inclusion-exclusion principle?

In the next study unit we will discuss the concept of a relation.

Study unit 5 Relations

Key questions for this study unit

- *What is meant by the term "ordered pair"?*
- *Do you understand the concepts of "a Cartesian product", "a relation", "a composition relation", and "an inverse relation"?*
- *Are you able to test whether a relation is: reflexive? irreflexive? symmetric? antisymmetric? transitive?*

Activity 5-1: *Study skill*

Draw a mind map of the different sections/headings you will deal with in this study session. Draw the map as you did the others, paging through the study unit with the purpose of completing the map.

Your map should include the concepts of an ordered pair, a Cartesian product, a relation, domain, codomain, range, reflexivity, irreflexive, symmetry, antisymmetry, transitivity, the composition of relations, inverse relations.

Activity 5-2: Concepts *Conceptual skill*

Test your own knowledge (write in pencil) and then correct your understanding afterwards (erase and write the correct description). Often a young language may not have all the terms in a discipline; can you think of some examples?

English term	Description	Term in your home language
Ordered pair		
Cartesian product		
Relation		
Codomain		
Domain		
Range		
Binary relation		
Reflexive relation		
Irreflexive relation		
Symmetric relation		
Antisymmetric relation		
Transitive relation		
Trichotomy		
Inverse relation		
Composition of relations		

5.1 Ordered pairs

As we saw earlier, {1, 2} = {2, 1}, so the order in which the elements of a set are listed does not matter. We now introduce a way to "encode" a new concept. This is done by using round brackets (parentheses) instead of curly brackets {braces} to get something that looks like this: (3, 0). The parentheses tells us that we are not talking about a set, but that we are dealing with an *ordered pair*, in which 3 is the first co-ordinate and 0 the second co-ordinate. Since the order is important, it follows that (3, 0) ≠ (0, 3).

The original use for this notation was to plot points on a graph in a plane. A Cartesian co-ordinate system in a plane consists of two perpendicular lines, one horizontal and one vertical, with a unit of measurement marked on each line. The horizontal line is usually called the *x-axis* and the vertical line the *y-axis*. The point where they meet is called *the origin*. The origin can be represented by the ordered pair (0, 0). Figure 5-1 illustrates this idea:

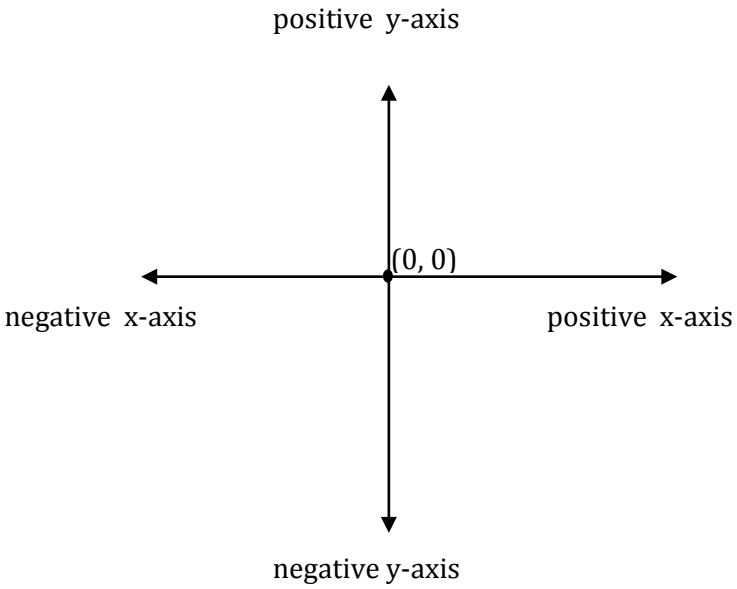


Figure 5-1

Now, each point in the plane occurs at an address provided by an ordered pair. Let's look at (4, −2) as an example. The first co-ordinate tells us to move, from our starting point at the *origin*, 4 steps in the positive direction along the x-axis. The second co-ordinate tells us that we should then move 2 steps in the negative direction along the y-axis.

Figure 5-2 shows us where (4, –2) is located in the plane:

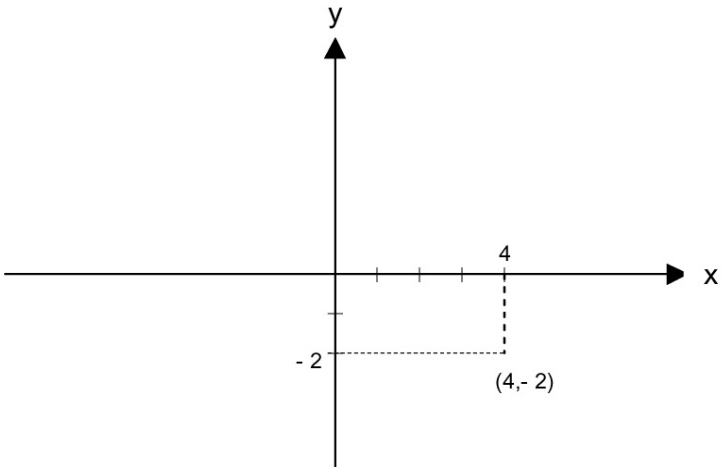


Figure 5-2

Apart from being able to represent points in a plane, ordered pairs have many other uses. All the uses are based on the fact that we can use the *fixed order* of the co-ordinates to provide information about some relationship. The resulting set of ordered pairs is then conveniently called a *relation*.

5.2 Relations

Let's look at an example. Consider the set {2, 3, 5}. We know that 2 < 3, 2 < 5 and 3 < 5. We can call this specific relationship "is less than" and we can represent it graphically by drawing arrows from 2 to 3, 2 to 5, and 3 to 5. This relationship is illustrated in Figure 5-3:

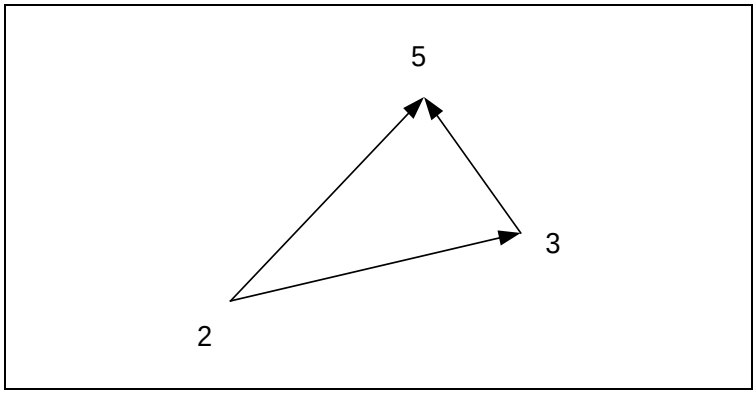


Figure 5-3

Applying what we have just learned about ordered pairs, we can use ordered pairs to represent the elements of this example of an *order relation*. In the place of the arrows, we can consider the ordered pairs (2, 3), (2, 5) and (3, 5). The set of these pairs, R = { (2, 3), (2, 5), (3, 5) }, completely captures the information in our picture. R is an order relation (which we could also call "<") on the set {2, 3, 5}.

Note: There are many other examples of order relations other than the "<" order relation.

As we will see in this study unit, if R is a relation on {2, 3, 5}, it means that R is a relation from {2, 3, 5} to {2, 3, 5}, i.e. the first and second co-ordinates come from the set {2, 3, 5}.

Instead of writing $(2, 3) \in R$ we could also use what is called *infix notation*, and write 2 R 3. Usually this specific relation is given the name < rather than R, so we write $2 < 3$ (using infix notation) rather than $(2, 3) \in <$.

Now it is important to note that any set of ordered pairs may be called a relation. It also often happens that a relation represents a *relationship* between one sort of thing and another. We can then say that all the first co-ordinates in the ordered pairs are of the same sort, whereas the second co-ordinates are all of a sort which may be different from the sort of the first co-ordinates.

Let's look at an example.

Example

In the relation
$R = \{(x_i , y_i) \mid x_i$ is the husband of y_i, $i = 1, 2, \ldots, n\}$,
the first co-ordinates are all men and the second co-ordinates are all women.
We describe the relation as the relation R from M to W, where
$M = \{x_1, x_2, x_3, \ldots x_n\}$ is a set of men and $W = \{y_1, y_2, y_3, \ldots y_n\}$ is a set of women.
Ordered pairs live in R, i.e. $(x_1, y_1) \in R$ (x_1 is the husband of y_1),
$(x_2, y_2) \in R$ (x_2 is the husband of y_2), and so on.

Note: We can write $(x_i , y_i) \in R$ or we could use *infix notation* to write $x_i \, R \, y_i$.

Another example

When a C^{++} compiler translates a source program into the machine language of the object program, it constructs a symbol table which contains the following sets:

S: the set of symbolic names such as variables, constants and types
A: the set of possible attributes for elements of S, such as integer, real number, Boolean, character, etc.
L: the set of locations (or addresses) in memory where the elements of S are stored.

The information in this table can be encoded into two relations:
- relation R_1 with first co-ordinates from S and second co-ordinates from A, and
- relation R_2 with first co-ordinates from S and second co-ordinates from L.

If you have not been exposed to programming, you need not be concerned if you do not understand the concept of a symbol table. Just imagine it as a table with three columns: column 1 contains a set of names, column 2 a set of attributes (or properties) of the names in column 1, and column 3 a set of locations where the names can be found.

Let's investigate more carefully this idea of having a relation from one set to another.

Definition: Cartesian product

For any sets A and B, the *Cartesian product* of A and B is denoted by A × B and is equal to the set {(x, y) | x ∈ A and y ∈ B}.

The Cartesian product A × B denotes a set of ordered pairs such that all the first co-ordinates are members of A and all the second co-ordinates are members of B. Each member of A is combined with each member of B in the set of ordered pairs A × B consists of.

A few examples should make this concept clear.

Examples

Suppose A = {2, 3, 4} and B = {5, 6}. Then

(a) $A \times B = \{ (2, 5), (2, 6), (3, 5), (3, 6), (4, 5), (4, 6) \}$
(b) $B \times A = \{ (5, 2), (5, 3), (5, 4), (6, 2), (6, 3), (6, 4) \}$
(c) $B^2 = B \times B = \{ (5, 5), (5, 6), (6, 5), (6, 6) \}$.
(d) $A^2 = A \times A = \{ (2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4), (4, 2), (4, 3), (4, 4)\}$.

Note that in example (c) we can also write $B \times B$ as B^2.

Although the two sets A and B might contain different types of things, they often contain exactly the same type of thing. In the example above, both A and B contain some natural numbers.

Definition: Relation

A subset of a Cartesian product C × D is called a *relation from C to D*.

Examples

With A = {2, 3, 4} and B = {6, 7}, the following are some relations from A to B:

(a) ∅
(b) { (3, 7) }
(c) { (2, 6), (2, 7) }
(d) { (2, 6), (3, 6), (4, 6) }
(e) $A \times B$.

Note: ∅ is a subset of A × B, so even though it has no ordered pairs as elements, it can also be called a relation from A to B. The relations in examples (a) to (e) are all subsets of A × B.

Activity 5-3: Examples of relations

If A = {0, 1, 2} and B = {1, 3, 5}, give two examples of relations from A to B.

A relation R from A to B means that R consists of a set of ordered pairs where the first co-ordinate is an element of A and the second co-ordinate is an element of B.

Therefore $R_1 = \{(0, 5)\}$ is a relation from A to B, and
$R_2 = \{(0, 1), (1, 3), (2, 5)\}$ is also a relation from A to B.

Note: The Cartesian product A × B can also be called a relation from A to B.

Activity 5-4: Self-assessment exercises *Application skills*

Let A = {1, 2, 3, 4}, B = {2, 5}, and C = {3, 4, 7}.
Write the following Cartesian products in list notation:

(a) $A \times B$
(b) $B \times A$
(c) $(A \cup B) \times C$
(d) $(A + B) \times B$

Activity 5-5: CAI tutorial

All the concepts in this study unit are covered in the CAI tutorial. Working through the relevant theory, examples and exercises will help you to understand these concepts.

If R is some relation from A to B, all the members of A and B do not necessarily appear as first and second co-ordinates respectively in the relation. This leads us to the following definitions:

Definitions: Domain, range and codomain of a relation

Suppose T is a relation from X to Y, then Y is called the *codomain* of T.

We define the *domain* of T (dom(T)) and the *range* of T (ran(T)) as the following subsets of X and Y respectively:

dom(T) = {x | for some $y \in Y$, $(x, y) \in T$}, i.e. the set of all elements that actually appear as first co-ordinates in the ordered pairs of T.

ran(T) = {y | for some $x \in X$, $(x, y) \in T$}, i.e. the set of all elements that actually appear as second co-ordinates in the ordered pairs of T.

Note: $dom(T) \subseteq X$. The domain of relation T, i.e. dom(T), is a *subset* of X, but it is not necessarily equal to X.

Furthermore, $ran(T) \subseteq Y$. The range of relation T, i.e. ran(T), is a *subset* of the codomain Y, but it is not necessarily equal to Y.

Note: One has to be a little careful when reading textbooks, because some authors use the word range to describe what we have called the codomain. But rest assured, we are following the most common practice.

Example

Let S = { (a, 1) }, { (b, 1), (a, 2) } be a relation from {a, b, c} to {1, 2, 3}.

Then $dom(S) = \{a, b\} \subseteq \{a, b, c\}$ and $ran(S) = \{1, 2\} \subseteq \{1, 2, 3\}$. The codomain is the set {1, 2, 3}.

Definition: Binary relation

If R is any subset of a Cartesian product $X \times Y$, then R is called a binary relation from X to Y (or "between X and Y"). A subset R of $X \times Y$ is called the rule for the relation.

If $R \subseteq X \times X$, we say that R is a binary relation on X.

*The members of a relation $R \subseteq X \times X$ are **ordered** pairs, so it is called a **binary** relation. We discuss n-ary relations in the next study unit.*

Examples

Let A = {2, 3, 4} and B = {a, b}.

Some examples of relations from A to B: $\emptyset$, { (2, a) }, { (2, a), (4, b) } and A × B = { (2, a), (2, b), (3, a), (3, b), (4, a), (4, b) }.

The codomain of all these relations is {a, b} = B; the domain and range of $\emptyset$ is the set $\emptyset$; the domain of { (2, a), (4, b) } is {2, 4} and its range is {a, b}; the domain of A × B is A and its range is B.

5.3 Properties of relations

We now look specifically at some relations R that are subsets of A × A for some set A. Such relations may, of course, be called relations from A to A, but it is more usual to refer to them as relations *on* A.

In this section we define some properties of relations.

Definition: Reflexive

A relation R on A (also written as $R \subseteq A \times A$) is called *reflexive* on A iff for every $x \in A$, we have $(x, x) \in R$.

The idea is that $R \subseteq A^2$ is reflexive on A if every element of A is related (in the context of R) *to itself*.

Examples

Let A = {2, 3, 5}. In order for some relation S to be reflexive on A, we should have { (2, 2), (3, 3), (5, 5) } $\subseteq$ S, which means that, among the elements of S, there should at least be the ordered pairs (2, 2), (3, 3) and (5, 5). Note that *each* element of A should be related to itself.

Now S = { (2, 2), (3, 3), (5, 5), (2, 3) } is an example of a reflexive relation on A.

Another example: Let R be the relation on $\mathbb{Z}$ defined by

$(x, y) \in R$ iff $x - y = 7k$ for some integer k.

(If $x - y = 7k$, it means that $x - y$ is a multiple of 7.)

Is R reflexive? I.e. is it true that $(x, x) \in R$ for all $x \in \mathbb{Z}$?

Yes, for all $x \in \mathbb{Z}$ we have $x - x = 0 = (7)(0)$ with k = 0

i.e. $(x, x) \in R$ and therefore R is reflexive on $\mathbb{Z}$.

Definition: Irreflexive

A relation $R \subseteq A \times A$ is called *irreflexive* iff there is *no* $x \in A$ such that $(x, x) \in R$. In other words, for any $x \in A$, $(x, x) \notin R$.

The idea is that $R \subseteq A^2$ is *irreflexive* iff there is *no* element of A related (in the context of R) *to itself*.

Examples

Let A = {2, 3, 5}. We look at properties of some relations on A:

R = {(3, 2), (2, 5), (3, 5)} is *irreflexive* since *no* member of A is related to itself, i.e. not one of the pairs (2, 2), (3, 3) and (5, 5) belong to R.

S = {(2, 2), (2, 5), (3, 5)} is neither reflexive nor irreflexive. Why do we say this? S is not reflexive since not all elements of A are related to itself, i.e. not all the pairs (2, 2), (3, 3) and (5, 5) belong to S. S is not irreflexive either since (2, 2) is an element of S. For irreflexivity not one of the ordered pairs (2, 2), (3, 3) and (5, 5) should belong to S.

Another example: Let R be the relation $\mathbb{Z}$ defined by $(x, y) \in R$ iff $x < y$.

Is R irreflexive? I.e. is it true that $(x, x) \notin R$ for all $x \in \mathbb{Z}$?

Yes, it can never be the case that $x < x$, therefore $(x, x) \notin R$ for all $x \in \mathbb{Z}$.

Definition: Symmetric

A relation $R \subseteq A \times A$ is *symmetric* iff R has the property that, for all $x, y \in A$, if $(x, y) \in R$, then $(y, x) \in R$.

The idea is that whenever an ordered pair (x, y) lives in R, then a pair with the *order of the elements reversed* (i.e. the mirror image pair) must also live in R.

Examples

Let B = {1, 2, 3}. We look at properties of some relations on B:

(a) $R_1 = \{ (1, 2), (2, 1), (1, 3), (3, 1) \}$ is symmetric and irreflexive.
(b) $R_2 = \{ (1, 1), (2, 2), (3, 3), (2, 3) \}$ is reflexive on B but not symmetric.
(c) $R_3 = \{ (1, 1), (2, 2), (3, 3), (2, 3), (3, 2) \}$ is reflexive on B and symmetric.
(d) $R_4 = \{ (1, 1), (2, 3), (3, 3) \}$ is neither reflexive on B, nor irreflexive, nor symmetric.

Another example: Let R be the relation on $\mathbb{Z}$ defined by
$(x, y) \in R$ iff $x - y = 7k$ for some integer k.

Is R symmetric? I.e. is it true that for all $x, y \in \mathbb{Z}$, if $(x, y) \in R$ then $(y, x) \in R$?

Yes. Assume $(x, y) \in R$ then $x - y = 7k$
i.e. $y - x = 7(-k)$ for some $-k \in \mathbb{Z}$, thus $(y, x) \in R$.

Definition: Antisymmetric

A relation $R \subseteq A \times A$ is *antisymmetric* iff R has the property that, for all $x, y \in A$, if $x \neq y$ and $(x, y) \in R$ then $(y, x) \notin R$.

We also look at an alternative definition:
A relation $R \subseteq A \times A$ is *antisymmetric* iff R has the property that, for all $x, y \in A$, if $(x, y) \in R$ and $(y, x) \in R$, then $x = y$.

The idea is that whenever $x \neq y$ and (x, y) lives in R, then a pair with the *order of the elements reversed* (i.e. the mirror image pair) may *not* live in R.

Examples

Let A = {a, b, c}.

Consider the relation P = {(a, b), (b, b), (b, c)} on A:
$a \neq b$ and $(a, b) \in P$, and also $b \neq c$ and $(b, c) \in P$, but the mirror image pairs namely (b, a) and (c, b) do not live in P. It follows that P is *antisymmetric*.

Consider another relation: Let $R \subseteq \mathbb{Z} \times \mathbb{Z}$ be defined by $(x, y) \in R$ iff $x \leq y$.
Is R antisymmetric?
I.e. is it true that for all $x, y \in \mathbb{Z}$, if $x \neq y$ and $(x, y) \in R$, then $(y, x) \notin R$?

Yes. Assume $x \neq y$ and $(x, y) \in R$, then $x \leq y$
i.e. it is not the case that $y \leq x$, thus $(y, x) \notin R$.

(Alternatively, assume $(x, y) \in R$ and $(y, x) \in R$, then $x \leq y$ and $y \leq x$ i.e. $x = y$.)

Activity 5-6: Symmetry, antisymmetry

One might easily come to the conclusion that *not symmetric* means the same as antisymmetric. Is this conclusion true? Use the relation R on
A = {1, 2, 3} where R = {(1, 2), (2, 1), (2, 3)} to test whether *not symmetric* means the same as antisymmetric.

R is not symmetric, since $(2, 3) \in R$ but $(3, 2) \notin R$. But R is *not* antisymmetric either, since (1, 2) and (2, 1) are both inside R, but $1 \neq 2$.

Definition: Transitive

A relation $R \subseteq A \times A$ is *transitive* iff R has the property that for all $x, y, z \in A$, whenever $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$.

Transitivity means that if x "is related to" y and y "is related to" z, then x "is related to" z with y acting as a sort of intermediary. One could say that, if the relation is transitive, and one can get from x to z in two steps, then one can also get from x to z in one step.

Examples

Let R = {(1, 1), (2, 2), (1, 2), (2, 1) } be a relation on A = {1, 2, 3}.

This relation is transitive: Consider the ordered pairs (1, **2**) and (**2**, 1):
if **2** acts as the "intermediary" then $(1, 1) \in R$.
If **1** acts as the "intermediary" for (2, **1**) and (**1**, 2) then $(2, 2) \in R$.
One should also consider pairs such as (1, **1**) and (**1**, 1): **1** acts as the "intermediary" so $(1, 1) \in R$. In this case the pair (1, 1) plays a triple role.
All the different possibilities should be investigated systematically.

Another example: Let R be the relation on $\mathbb{Z}$ defined by
$(x, y) \in R$ iff $x - y = 7k$ for some integer k.
Is R transitive? I.e. is it true for all $x, y, z \in \mathbb{Z}$ that if $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$?

Yes. Assume $(x, y) \in R$ and $(y, z) \in R$,
then $x - y = 7k$ ① and $y - z = 7m$ ② for some $k, m \in \mathbb{Z}$.
Add ① and ②, then $(x - y) + (y - z) = 7k + 7m$
i.e. $x - z = 7(k + m)$, which is a multiple of 7, therefore $(x, z) \in R$.

Many relationships in real life are transitive, for example "is heavier than". Others might give one the impression that they should be transitive, but are in fact not. In soccer, for instance, we might have "Liverpool beats Real Madrid", and "Real Madrid beats Arsenal", but this does not necessarily mean that "Liverpool beats Arsenal".

Definition: Trichotomy

A relation R on A satisfies the requirement for *trichotomy* iff, for every x and y chosen from A such that $x \neq y$, we have that x and y are comparable,
i.e. for all $x, y \in A$ such that $x \neq y$, x R y or y R x (i.e. $(x, y) \in R$ or $(y, x) \in R$).

This means that we should be able to *compare **any** two distinct elements* x and y belonging to A in terms of the relation. So, if we pick *any* distinct elements x and y from A, then either x is the first co-ordinate and y the second co-ordinate in an ordered pair, i.e. x R y, or x is the second and y the first co-ordinate in an ordered pair, i.e. y R x (it is possible that both x R y and y R x since "or" is used in the *inclusive* sense in this definition).
All distinct elements x and y that belong to A should be considered when investigating whether x R y or y R x. (It is possible that pairs of the form (x, x) could belong to some relation that satisfies trichotomy.)

Examples

Suppose S = { (1, 1), (3, 2), (2, 1), (3, 1), (3, 1) } is a relation on A = {1, 2, 3}, then S satisfies the requirement for trichotomy since *any* element of A is related to *each other element* of A that is different from itself. This means that *all* the *distinct* elements belonging to A are comparable:
$(3, 2), (2, 1), (3, 1) \in S$.

Another example: Suppose T = { (3, 2), (2, 1), (2, 3) } is a relation on A = {1, 2, 3}, then T does not satisfy trichotomy since not all distinct elements of A are comparable. We have $1, 3 \in A$ and $1 \neq 3$ but neither (1, 3) nor (3, 1) are elements of T.

Yet another example: Let $R \subseteq \mathbb{Z} \times \mathbb{Z}$ be defined by $(x, y) \in R$ iff $x \leq y$.

(A pair (x, y) belongs to R iff x is less than or equal to y. Instead of "R" we could just as well have named the relation "≤". For example, $1 \leq 2$, i.e. $(1, 2) \in R$ but $6 \nleq 2$ thus $(6, 2) \notin R$.)

If $x, y \in \mathbb{Z}$ and $x \neq y$ then either x lies to the left of y, or y lies to the left of x on the *number line* (mentioned in study unit 1), so either $(x, y) \in \leq$ or $(y, x) \in \leq$.

Thus ≤ satisfies trichotomy.

Activity 5-7: The properties of relations

Let A = {1, 2, 3}. Give an example of a relation on A that is reflexive on A and symmetric and transitive.

Let R be a relation on A. For R to be reflexive, all the elements of A need to be paired with themselves, i.e. (1, 1), (2, 2) and (3, 3) must live in R.

For R to be symmetric means that if $(x, y) \in R$, then $(y, x) \in R$.

Now suppose (1, 2) ∈ R and (2, 3) ∈ R, then we need to add (2, 1) and (3, 2) to ensure that R is symmetric.
At this stage { (1, 1), (2, 2), (3, 3), (1, 2), (2, 3), (2, 1), (3, 2) } ⊆ R.

Finally, we need to build in transitivity. We have (1, **2**) and (**2**, 3) in R, which means that (1, 3) needs to be added to R. We also have (1, **2**) and (**2**, 1), so (1, 1) needs to be in R, which is already the case. We have (3, **2**) and (**2**, 3), so (3, 3) must be included. Then we have (2, **1**) and (**1**, 3) (remember we added (1, 3)) so (2, 3) needs to be in R, which is true. Finally, we have (2, **3**) and (**3**, 2), so (2, 2) needs to in R, which is true.

We have added (1, 3), so we need to add (3, 1) to maintain symmetry.

Thus R = { (1, 1), (2, 2), (3, 3), (1, 2), (2, 3), (2, 1), (3, 2), (1, 3), (3, 1) }.

It is also possible to build new relations from old ones. There are various ways of doing this, two of which we will consider next. Firstly, we will look at the inverse of a relation, and secondly we will investigate how the composition of two relations is formed.

Definition: Inverse relation

Given a relation R with domain A and range B, the relation R^{-1} (read "R inverse") with domain B and range A is called the *inverse of R* and is defined such that

$(x, y) \in R$ if and only if $(y, x) \in R^{-1}$.

Note that this definition also tells us that the inverse of R^{-1} is R, so we could simply say that these two relations are inverses of each other.
(Adapted from *Ensley & Crawley,* 2006.)

The best way to understand this definition is to look at an example.

Example

Let X = {a, b, c} and R = { (a, b), (b, c), (a, c) }.

We get the inverse of a relation by switching the co-ordinates of the ordered pairs, so R^{-1} = { (b, a), (c, b), (c, a) }.

Definition: Relation composition

Given relations R from A to B and S from B to C, the composition of R followed by S (S ○ R or R; S) is the relation from A to C defined by

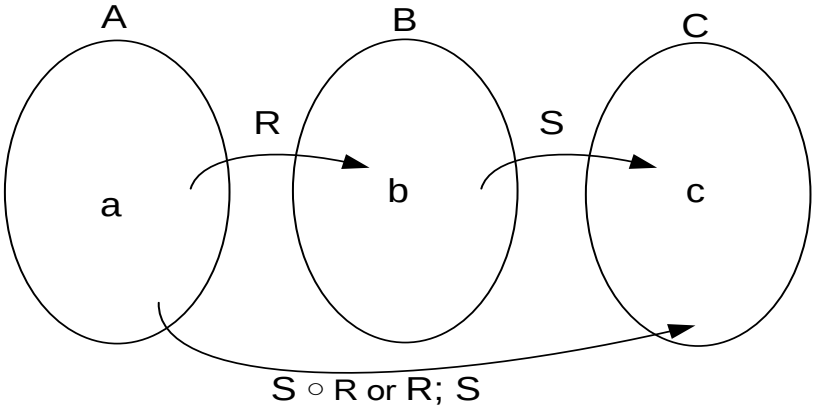
S ○ R = R; S = {(a, c) | there is some b ∈ B such that (a, b) ∈ R and (b, c) ∈ S}.

Using the notation R; S rather than S ○ R helps us to remember that R is followed by S.

It is worth spending some time thinking about this definition. If x is a first co-ordinate in some pair of R; S, can you see that x must be a member of A? This is the case because if x appears as a first co-ordinate in R; S, then there has to be some b ∈ B such that (x, b) ∈ R. Similarly, if y appears as a second co-

ordinate in S ○ R, then there has to be some b ∈ B such that (b, y) ∈ S, so y must be a member of C. Therefore the definition ensures that S ○ R ⊆ A × C.

The following diagram illustrates how S ○ R links element a of set A to element c of set C by using element b of set B as an intermediate point.



Examples

Let R = { (1, a), (2, b) } be a relation from {1, 2} to {a, b} and let
S = { (a, s), (b, s), (b, t) } be a relation from {a, b} to {s, t}.

Determine the relations S ○ R (i.e. R; S) and R ○ S (i.e. S; R). The relation S ○ R (or R; S) is the composition of **R followed by S**.

The first co-ordinates of S ○ R come from {1, 2} (dom(R)) and the second co-ordinates come from {s, t} (ran(S)).

If we want to determine S ○ R, i.e. R;S, then we first write down a pair of R, let's take (1, **a**), then we look for a pair in S that has as first co-ordinate an **a**. If we link (1, **a**) of R with (**a**, s) of S (**a** is the linking co-ordinate), then by the "relation composition" definition, (1, s) ∈ S ○ R.

No other member of S has "a" as first co-ordinate, so (1, a) in R cannot link with any other member in S.

Let's consider the other member of R, namely (2, b), and inspect S to see whether or not (2, b) can link with some members in S. We see that (2, **b**) can link with (**b**, s) and also with (**b**, t). By definition (2, s) and (2, t) are members of S ○ R.

We have looked at all possible pairs that can link, so
S ○ R = { (1, s), (2, s), (2, t) }.

Determine R ○ S: The relation R ○ S (or S; R) is the composition of S followed by R. We start with S, but there are no members in S that can link with members in R, so R ○ S = ∅.

More examples

In the following examples, we consider relations defined on X = {a, b, c}.

We can form the compositions R; R (i.e. R ○ R), R; R^{-1} (i.e. R^{-1} ○ R), R^{-1}; R (i.e. R ○ R^{-1}) and R^{-1}; R^{-1} (i.e. R^{-1} ○ R^{-1}):

(a) Let R = { (a, a), (b, c), (c, b) }, then R^{-1} = { (a, a), (c, b), (b, c) }.

To determine **R**; *R*, we start with the pair (a, **a**) of **R**, and then we look for a pair in *R* that has as first co-ordinate an **a**, and then see where it takes us.

Link (a, **a**) of **R** with (**a**, a) of *R*, then (a, a) ∈ **R**; *R*. Continuing in this vein gives R ○ R = { (a, a), (b, b), (c, c) }.

Since R^{-1} = R, we have R^{-1} ○ R = { (a, a), (b, b), (c, c) }.

Similarly R ○ R^{-1} = { (a, a), (b, b), (c, c) }.

And of course R^{-1} ○ R^{-1} = {(a, a), (b, b), (c, c)}.

(b) Let R = { (a, b), (b, c), (a, c) } then R^{-1} = { (b, a), (c, b), (c, a) }.

Only the pair (a, **b**) of **R** can link with a pair in *R*, namely (**b**, c) of *R*, thus

R; *R* = {(a, c)}.

R; R^{-1} = { (a, a), (b, a), (b, b), (a, b) }. We obtain the four ordered pairs in R; R^{-1} as follows:

(a, a): link (a, b) of R with (b, a) of R^{-1}, or link (a, c) of R with (c, a) of R^{-1};

(b, a): link (b, c) of R with (c, a) of R^{-1};

(b, b): link (b, c) of R with (c, b) of R^{-1}; and finally

(a, b): link (a, c) of R with (c, b) of R^{-1}.

Similarly, R ○ R^{-1}= { (b, b), (b, c), (c, c), (c, b) } and R^{-1} ○ R^{-1} = { (c, a) }.

Activity 5-8: Self-assessment exercises *Application skills*

1. Let P and R be relations on A = { 1, 2, 3, {1}, {2} }:
   P = { (1,{1}), (1, 2) } and R = { (1, {1}), (1, 3), (2, {1}), (2, {2}), ({1}, 3), ({2}, {1}) }.

   Justify your answers to the following questions.
   (a) Is R irreflexive?
   (b) Is R reflexive?
   (c) Is R symmetric?
   (d) Is R antisymmetric?
   (e) Is R transitive?
   (f) Does R satisfy the requirement for trichotomy?
   (g) Determine the relation R ○ R.

(h) Determine the relation $R \circ P$.

(i) Give the subset T of R where $(a, B) \in T$ iff $a \in B$.

2. Let A = {a, b}. For each of the specifications given below, find suitable examples of relations on $\mathcal{P}(A)$.

First of all let us write down $\mathcal{P}(A)$: $\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{a,b\}\}$,

and $\mathcal{P}(A) \times \mathcal{P}(A) = \{ (\emptyset, \emptyset), (\emptyset, \{a\}), (\emptyset, \{b\}), (\emptyset, \{a,b\}), (\{a\}, \emptyset), (\{a\}, \{a\}), (\{a\}, \{b\}), (\{a\}, \{a, b\}), (\{b\}, \emptyset), (\{b\}, \{a\}), (\{b\}, \{b\}), (\{b\}, \{a, b\}), (\{a, b\}, \emptyset), (\{a, b\}, \{a\}), (\{a, b\}, \{b\}), (\{a, b\},\{a,b\}) \}$.

(a) R is reflexive on $\mathcal{P}(A)$, symmetric, and transitive.

(b) R is reflexive on $\mathcal{P}(A)$ and symmetric, but not transitive.

(c) R is reflexive on $\mathcal{P}(A)$ and transitive, but is not symmetric and not antisymmetric.

(d) R is simultaneously symmetric and antisymmetric.

(e) R is irreflexive, antisymmetric and transitive.

3. Prove that if R is a relation on X, then R is transitive iff $R \circ R \subseteq R$.

5.4 In summary of the study unit

We can summarise the properties of the relations we considered in this unit:

Let R be a relation on A:
R is *reflexive* on A iff for all $x \in A$, we have $(x, x) \in R$.
R is *irreflexive* iff for any $x \in A$, $(x, x) \notin R$.
R is *symmetric* iff for all $x, y \in A$, if $(x, y) \in R$, then $(y, x) \in R$.
R is *antisymmetric* iff for all $x, y \in A$, if $x \neq y$ and $(x, y) \in R$, then $(y, x) \notin R$, or alternatively: for all $x, y \in A$, if $(x, y) \in R$ and $(y, x) \in R$, then $x = y$.
R is *transitive* iff for all $x, y, z \in A$, if $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$.

In this study unit you ensured that you can answer the following questions regarding set theory:

- What does the term ordered pair mean?
- What is meant by the term relation?
- How do we define the Cartesian product of two sets?
- What does it mean if a relation is reflexive on some set?
- What does it mean if a relation is irreflexive?
- What does it mean if a relation is symmetric?
- What does it mean if a relation is antisymmetric?
- What does it mean if a relation is transitive?
- What does the notation $R \subseteq A^2$ mean?
- When does a relation satisfy trichotomy?
- How does one form the inverse relation R^{-1} of a given relation R?
- How do we form the composition of two relations?

In the following study unit we will learn more about additional properties that relations might have and also study special kinds of relation.

Study unit 6 Special kinds of relation

Key questions for this study unit

- *Provide a list of the properties each kind of special relation has.*
- *What do we mean by the terms "order relation" and "n-ary relation"?*
- *What do we have to check if we want to determine whether or not a relation is a function?*
- *What do we mean by a partial or linear order?*
- *How do we approach a proof if we want to use "reductio ad absurdum" as a proof technique?*
- *What are the properties of an equivalence relation?*
- *What is meant by an equivalence class and a partition?*

Activity 6-1: Overview *Study Skill*

Draw a mind map of the different sections/headings you will deal with in this study session. Page through this study unit with the purpose of completing the map.

Your map should include the concepts of order relations, partial orders, trichotomy, total order, equivalence relations, equivalence classes, partitions, n-ary relations, functions, domain, range and codomain.

Activity 6-2: Concepts *Conceptual skill*

Test your own knowledge and then correct your understanding afterwards. How does your understanding deepen as you jot down the terms used in your home language?

English term	Description	Term in your home language
A weak partial order		
A strict partial order		
A weak total (or linear) order		
A strict (or linear) total order		
Equivalence relation		
Equivalence class		
Partition		
n-ary relation		
A functional relation		
A function		

6.1 Order relations

In the previous study unit we discussed relations, as well as special properties a relation may or may not have. In this study unit we shall look at special classes or kinds of relation. The properties a relation has, define the kind of relation it is classified as.

Activity 6-3: CAI tutorial

All the concepts in this study unit are covered in the CAI tutorial. Working through the relevant theory, examples and exercises will help you to understand these concepts. You can also re-visit concepts of study unit 5.

Familiar concepts from study unit 5 play a role in the following definition of a special kind of relation, namely a weak partial order:

Definition: Weak partial order

A relation R on a set A is called a *weak partial order* iff R is
reflexive on A,
antisymmetric, and
transitive.

We will later see that it is possible for some order relation to be irreflexive, antisymmetric and transitive. We call such a relation a "strict partial order".

Examples

Let A = { {a}, {a, b} }. A relation S on A is defined by (B, C)$\in$ S iff B $\subseteq$ C, i.e.
S = { ({a}, {a}), ({a}, {a, b}), ({a, b}, {a, b}) }. (Each first co-ordinate is a subset of the second co-ordinate.)

Prove that S is reflexive on A, antisymmetric and transitive (a weak partial order).

Reflexivity: Is it true that (B, B) $\in$ S for all B $\in$ A?

Yes, S is reflexive on A:
Each element of A is related to itself: ({a}, {a}) $\in$ S and ({a, b}, {a, b}) $\in$ S.

Antisymmetry: Is it true that for all B, C $\in$ A, if B $\neq$ C and (B, C) $\in$ S then (C, B) $\notin$ S?

Yes, S is antisymmetric:
We see that {a} $\neq$ {a, b} and ({a}, {a, b}) $\in$ S, but ({a, b}, {a}) $\notin$ S.

Transitivity: Is it true that for all B, C, D $\in$ A, if (B, C)$\in$ S and (C, D)$\in$ S, then (B, D) $\in$ S?

Yes, S is transitive:
We have ({a}, {a}$\}\in$ S and ({a}, {a}$\}\in$ S, then also ({a}, {a}$\}\in$ S (({a}, {a}) plays a triple role);
({a}, {a}$\}\in$ S and ({a}, {a, b}$\}\in$ S, then also ({a}, {a, b}$\}\in$ S;
({a}, {a, b}$\}\in$ S and ({a, b}, {a, b}$\}\in$ S, then also ({a}, {a, b}$\}\in$ S; and lastly
({a, b}, {a, b}$\}\in$ S and ({a, b}, {a, b}$\}\in$ S, then also ({a, b}, {a, b}$\}\in$ S.

Another example: The relation R on $\mathbb{Z}^+$ is defined by x R y iff x = k·y for some $k \in \mathbb{Z}^+$. (x is a multiple of y.) Is R a weak partial order?

Let us synthesize some ordered pairs that belong to R:
We see that (6, 2) ∈ R: x = 6 = (3)(2) = (k)(y) with k = 3. Some other elements of R are (6, 3), (35, 5) and (24, 4), and so on. These pairs meet the requirement that x = ky for some $k \in \mathbb{Z}^+$.

Investigate whether R is reflexive on $\mathbb{Z}^+$, antisymmetric and transitive:

Reflexivity:
For each $x \in \mathbb{Z}^+$ we have that x = 1·x with k = 1, so (x, x) ∈ R.

Therefore R is reflexive on $\mathbb{Z}^+$.

Antisymmetry: For all $x, y \in \mathbb{Z}^+$, if $x \neq y$ and (x, y) ∈ R, is (y, x) ∉ R?

Suppose $x \neq y$ and (x, y) ∈ R
then m·y = x for some $m \in \mathbb{Z}^+$

i.e. y = (1/m) · x, but $1/m \notin \mathbb{Z}^+$. (Remember, $x \neq y$.)
Hence (y, x) ∉ R.

Therefore R is antisymmetric.

Transitivity: For all $x, y, z \in \mathbb{Z}^+$ such that (x, y) ∈ R and (y, z) ∈ R. Does it follow that (x, z) ∈ R?

Suppose (x, y) ∈ R and (y, z) ∈ R
then x = m·y ① for some $m \in \mathbb{Z}^+$ and y = k·z ② for some $k \in \mathbb{Z}^+$.

Now we substitute ② into ①,
then x = (m·k)·z with $k \cdot m \in \mathbb{Z}^+$.
Hence (x, z) ∈ R.

Therefore R is transitive.

R is a weak partial order because it is reflexive, antisymmetric and transitive.

Activity 6-4: Weak partial orders

For each of the following relations, determine whether or not the relation is a weak partial order on the given set:

(a) Let A = {a, b, {a, b} }. S is the relation on A is defined by (c, B) ∈ S iff c ∈ B.
(b) Define $R \subseteq \mathbb{Z} \times \mathbb{Z}$ by x R y iff x + y is *even*.
(c) Define R on $\mathbb{Z} \times \mathbb{Z}$ by (a, b) R (c, d) if either a < c or else (a = c and b ≤ d).

It is possible that an order relation could be irreflexive, antisymmetric and transitive? Yes, we call it a strict partial order.

Definition: Strict partial order

A relation R on a set A is called a strict partial order iff R is
irreflexive,
antisymmetric, and
transitive.

Example

Let A = {1, 2, 3} and let S on A be the relation S = { (1, 2), (1, 3), (2, 3) }.
(Each first co-ordinate is less than the second co-ordinate.)

Is S is a strict partial order, i.e. is S irreflexive, antisymmetric and transitive?

Irreflexivity: Is it true that $(x, x) \notin S$ for any $x \in A$?

Yes, S is irreflexive:
No element of A is related to itself, so $(x, x) \notin S$ for any element of A.

Antisymmetry: Is it true that for all $x, y \in A$, if $x \neq y$ and $(x, y) \in S$ then $(y, x) \notin S$?

Yes, S is antisymmetric:
We see $1 \neq 2$ and $(1, 2) \in S$, but $(2, 1) \notin S$,
$1 \neq 3$ and $(1, 3) \in S$, but $(3, 1) \notin S$, and
$2 \neq 3$ and $(2, 3) \in S$, but $(3, 2) \notin S$.

Transitivity: Is it true that for all $x, y, z \in A$, if $(x, y) \in S$ and $(y, z) \in S$, then $(x, z) \in S$?

Yes, S is transitive:
We have $(1, 2) \in S$ and $(2, 3) \in S$, then also $(1, 3) \in S$.

It follows that S is a strict partial order because it is irreflexive, antisymmetric and transitive.

This relation S also satisfies the condition of trichotomy (as defined in the previous study unit). Any element of A is related to any other element of A that is different from itself. We have (1, 2), (1, 3) and (2, 3) $\in$ S.

Activity 6-5: Strict partial orders

For each of the following relations, determine whether or not the relation is a strict partial order on the given set:

(a) Let A = {a, {a}, {b}} and let S on A be the relation S = { (a, {a}), (a, {b})}.
(b) Define $R \subseteq (\mathbb{Z} \times \mathbb{Z}) \times (\mathbb{Z} \times \mathbb{Z})$ by (a, b) R (c, d) iff $a < c$.

Why do we use the word partial? The reason for this is that, if we choose any two elements from A, say x and y, they need not be related. The two most basic examples of partial orders are the usual ordering of numbers, on the one hand, and the subset relation on the other. A comparison of these relations illustrates that the "is less than or equal to" relation has a property (namely "trichotomy") that the subset relation does not have. We will presently give a name to these new kinds of relation with this property. The word "partial" indicates that some of the relations that we are talking about might not have this special property.

Let's look at the concept of power sets again. Consider the set
U = {1, 2}. Then $\mathcal{P}$(U) has as members the subsets ∅, {1}, {2} and {1, 2}.

Let R be a relation on $\mathcal{P}$(U) defined by R = {(A, B) | A is a subset of B}.

Then we can claim that ∅ R ∅, ∅ R {1}, ∅ R {2}, ∅ R {1, 2}, {1} R {1}, {1} R {1, 2}, {2} R {2}, {2} R {1, 2} and {1, 2} R {1, 2}. (We use infix notation.)

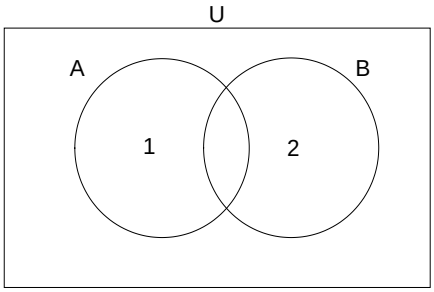
Activity 6-6: Notation

Replace the letter R used above by the symbol we usually use for the subset relation, namely ⊆. Then read the definition of R again.
E.g. {1} R {1, 2} can now be written as {1} ⊆ {1, 2}. It should now be clear what is meant by {1} R {1, 2}.

Can we say that any two subsets of U can be compared (*are comparable*) in terms of the subset relation? The answer is NO.

For instance, pick the subsets {1} and {2}. Not every element of {1} belongs to {2}, so we do not have {1} ⊆ {2}. Similarly, not every element of {2} belongs to {1}, so we don't have {2} ⊆ {1}. This means that it is not the case that if we pick any two elements from the subset relation on $\mathcal{P}$(U), that they are related, i.e. that they are comparable.

The situation can be illustrated by the following Venn diagram:



In this diagram A = {1}, B = {2} and U = {1, 2}.

Order relations are important in computer science because of our interest in sorting elements in a certain order. We usually want to sort lists of things that are ordered in such a way that *any two things are comparable* in terms of the order relation, i.e. given two different things in the list, we want to be sure that one of them comes before the other. So we might as well give a special name to the order relations we are most likely to use.

Definition: A total (or linear) order relation

A relation R on a set A is called a total (or linear) order if R is a partial order on A which satisfies the additional property that for all x, y ∈ A, either x R y or y R x, i.e. R satisfies the condition of trichotomy (as defined in the previous study unit).

A simple way to remember what this means is to think of any two members of the set A as being comparable (related).

Examples

The relation $\leq$ is a weak total order relation on $\mathbb{Z}$. (In Activity 6.6 you will determine that $\leq$ is reflexive, antisymmetric and transitive.)
We define the relation: $\leq = \{(m, n) | \ m \leq n$ with $m, n \in \mathbb{Z}\}$
i.e. for all $m, n \in \mathbb{Z}$, either $m \leq n$ or $n \leq m$.

The relation $\leq$ is a *weak total order relation* on $\mathbb{Z}$ because we can compare *any* two integers in terms of this *total order relation*. We see that either the first co-ordinate is less than the second, or they are equal, or the second co-ordinate is less than the first.

In an example that we encountered previously in this study unit, we proved that the relation

$S = \{ (1, 2), (1, 3), (2, 3) \}$ on $A = \{1, 2, 3\}$

is irreflexive, antisymmetric, transitive and satisfies trichotomy, thus we can say that S is a *strict total order relation* on A.

Let's consider some elements of $\mathbb{Z}$ and then compare them:

$0, -1$	then $-1 \leq 0$.
$1, 1$	then $1 \leq 1$.
$45, 113$	then $45 \leq 113$.
$-20, -250$	then $-250 \leq -20$ … and so on.

Any element in $\mathbb{Z}$ is either smaller or greater than, or equal to any other element in $\mathbb{Z}$.

We see that it is possible to differentiate between weak and strict total order relations. The difference being that a weak total order relation is reflexive, whereas a strict total order relation is irreflexive. This means that "$\leq$" is a weak total order and "$<$" is a strict total order.

For your own notes: A summary of the properties of order relations

Let's consider some relations on A:

A *weak partial order* is reflexive on A, antisymmetric and transitive,
a *strict partial order* is irreflexive, antisymmetric and transitive,
a *weak total (or linear) order* is reflexive on A, antisymmetric and transitive, and satisfies trichotomy, and
a *strict total (or linear) order* is irreflexive, antisymmetric and transitive, and satisfies trichotomy.

N.B.: Note the way in which we differentiate between a "weak" order and a "strict" order.

6.2 Some comments on proof strategies

When we consider relations (or functions, a topic which is studied later in this study unit) and we want to investigate whether or not they have certain properties, we must keep the following in mind:
One cannot use examples to prove general statements of the form:

"for all x, ...", or
"for all pairs (x, y), ..."

We need to use *abstract reasoning* to produce a *general proof.*

We have made the point that an example does not constitute a general proof. However, examples have a valuable role to play in helping us to get an intuitive "feel" for some problems.

Suppose we are investigating a relation T, where T is the relation on $\mathbb{Z}$ defined by the rule $(x, y) \in T$ iff $x - y = 3k$ for some $k \in \mathbb{Z}$.

The first thing you should do is to get a "gut feel" for T by determining some of its outputs. We know that the difference between x and y should be a multiple of 3, so we determine the outputs of T when the following ordered pairs are the inputs:

(10, 7)	where 10−7	= 3	= 3(1)
(80, 50)	where 80−50	= 30	= 3(10)
(28, 73)	where 28−73	= −45	= 3(−15)
(−9, −3)	where −9−(−3)	= − 6	= 3(−2)
(2, 2)	where 2 − 2	= 0	= 3(0)

Thinking about these pairs and outputs will help you as you consider the tests for the various properties, and will make it easier to decide on your approach to the proof – or, maybe to provide a counterexample.

Example

Say, for example, we want to investigate whether or not the previous relation T is reflexive on $\mathbb{Z}$, i.e. for all $x \in \mathbb{Z}$, is it the case that $(x, x) \in T$? We must provide a general proof:

$x - x = 0$
i.e. $x - x = 3 \cdot 0$ with $k = 0$,
so $(x, x) \in T$ for all $x \in \mathbb{Z}$.

Therefore T is reflexive on $\mathbb{Z}$.

We can, however, definitely use an example to show that a general statement is *false*. Find a case for which the statement does not hold and you have done it! This is called a ***counterexample***. (We used counterexamples in a previous study unit.)

Example

We use a counterexample to prove that the relation T in the previous example does not satisfy trichotomy:

Consider the elements 1, 5 $\in \mathbb{Z}$. Neither (1, 5) nor (5, 1) are elements of T (1 – 5 = –4 and 5 – 1 = 4, and neither –4 nor 4 can be written as a multiple of 3).

So T does not satisfy trichotomy.

Another approach to prove something is called the brute force approach. What we do in this type of proof is to write down all the elements we are working with, and show that whatever we want to prove holds for all the possible elements. The first exercise in the following activity is an example of such a proof.

Activity 6-7: Self-assessment exercises *Application skills*

1. Let X = {a, b, c}. Write down all strict partial orders on X. Which of them are linear?

Strict partial orders on X are irreflexive, antisymmetric and transitive. We write down a few examples:

{ (a, b) }, { (a, c) }, … write down the other 4 strict partial orders that have only one element;

{ (a, b), (a, c) }, { (a, b), (c, b) }, … write down the other 4 strict partial orders that have two elements; and

{ (a, b), (b, c), (a, c) }, … write down the other 5 strict partial orders that have three elements.

2. In each of the following cases, determine whether or not R is some sort of order relation on the given set X (weak partial, weak total, strict partial, or strict total). Justify your answer.

   (a) X = { ∅, {0}, {2} } and R = { (∅, {0}) , (∅, {2}) }.
   (b) X = { ∅, {∅}, {{∅}} } and R = $\subseteq$ (i.e. R is the relation of all ordered pairs where each first co-ordinate is a subset of the 2nd co-ordinate, and R $\subseteq$ X × X. For example, ∅ $\subseteq$ {∅}, so (∅, {∅}) $\in \subseteq$ or (∅, {∅}) $\in$ R.)
   (c) X = $\mathbb{Z}$ and R = $\leq$.
   (d) X = $\mathbb{Z}$ and R = >.
   (e) X = $\mathbb{Z}^+$ and R is defined by the requirement that x R y iff x divides into y with zero remainder, i.e. y = kx for some k $\in \mathbb{Z}^+$. This means that x is a *factor* of y and y is a *multiple* of x.
   Note: x R y is another way of saying (x, y) $\in$ R.

6.3 Equivalence relations

During your computing studies you will often come across the term "equivalence relation". This type of relation is used in many applications and what make them special are the properties that these relations have.

Definition: Equivalence relation

A relation R on a set A is called an equivalence relation if R is
reflexive on A,
symmetric, and
transitive.

Objects within a relation may have different characteristics. It is sometimes very useful to lump together objects that have one particular characteristic in common and ignore the other characteristics. This is where equivalence relations come in handy.

Examples

Suppose A is the set of all people that were living in the Republic of South Africa on the night of the last census. Let R be the relation on A defined by x R y if x slept under the same roof as y on that particular night. If we ignore the possibility of someone sleepwalking from one house to another, then R is an equivalence relation. This means that we divide the population not into humans with individual characteristics, but into households. The single property we use to separate objects into groups is the place where a person slept on that particular night.

Another example: Suppose students are evaluated for an assignment on a scale from A to E. For example, if a student has a score such that $0 \leq \text{score} \leq 20$, he gets an E. If a student has a score such that $21 \leq \text{score} \leq 40$, she gets a D, and so on. All the students who get an E will be in the same equivalence class although they had different individual scores. The same is true for all the other symbols (A to D). This relation divides the students into different equivalence classes and is an equivalence relation.

This leads us to a very important concept regarding equivalence relations: an equivalence relation R on some set A *partitions* A into *equivalence classes* in the following way:

Definition: Equivalence class

For each $x \in A$ we define the equivalence class [x] by
$[x] = \{y \mid y \in A \text{ and } x \, R \, y\}$.

Although it takes a bit of practice to feel comfortable with equivalence classes, you have been using them all your life without realising it. Let's illustrate this:

Consider the rational numbers (or fractions). You are probably used to thinking of rational numbers as quotients of integers a/b with b ≠ 0. But do you regard 1/3 as different from 2/6, or 2/5 as different from 20/50? No of course not.

Now, let A be the set of all quotients a/b where a, b $\in \mathbb{Z}$ and b ≠ 0.
Then A is not exactly the set of rational numbers, because we can find different quotients in A which represent the same rational number. We can define this relation R on A by a/b R c/d iff ad - bc = 0. (b ≠ 0 and d ≠ 0.)

Activity 6-8: Equivalence relation

By using the description of the relation R above, show that R is an equivalence relation on A.

Now all the quotients a/b, 2a/2b, 3a/3b, 4a/4b, … are equivalent to one another, in other words, each belongs to the equivalence class [a/b]. These equivalence classes are actually the rational numbers. When we write 1/3, we really mean [1/3], the equivalence class of 1/3, because when we write 1/3 we are talking simultaneously about all the equivalent fractions 1/3, 2/6, 4/12, and so on.

Activity 6-9: Equivalence class

Let R be the relation on $\mathbb{Z}$ defined by $(x, y) \in R$ iff $y - x$ is even.

Then $[x] = \{ y \mid y - x = 2k \text{ for some } k \in \mathbb{Z} \}$
$= \{ y \mid y = 2k + x \text{ for some } k \in \mathbb{Z} \}$

We substitute values for x until we no longer encounter new equivalence classes:

Let $x = 0$, then $[0] = \{ y \mid y = 2k + 0 \text{ for some } k \in \mathbb{Z} \}$
$= \{ ..., -8, -6, -4, -2, 0, 2, 4, 6, 8, ... \}$
This is the set of *even* integers.

We also have …[0] = [−2] = [2] = [4]…etc.

Let x = 1, then $[1] = \{ y: y = 2k + 1 \text{ for some } k \in \mathbb{Z} \}$
$= \{ ..., -7, -5, -3, -1, 1, 3, 5, 7, ... \}$
This is the set of *odd* integers.
We also have …[−1] = [1] = [−3] = [3]…etc.

Note: The *equivalence classes* [0] and [1] are the "*parts*" of the *partition* S of the set $\mathbb{Z}$ induced by the relation R.

S = { [0], [1] }. This shows that there are only two different (non-empty) equivalence classes, namely [0] and [1].

It is clear that $[0] \cap [1] = \emptyset$ and $[0] \cup [1] = \mathbb{Z}$.

A given equivalence class [x] can be described in many ways. To be precise, [x] = [y] for each $y \in [x]$. We say that we have made a choice of representative x or y, depending on whether we denote this one equivalence class by [x] or by [y]. When we deal with equivalence relations, it is conventional to pick as representative the smallest non-negative member of the equivalence class. E.g. if {2, 4, 6, …} is an equivalence class of some relation, we denote the class by [2].

Let R_3 be the relation $\mathbb{Z}$ that maps an integer x to the result of calculating x modulo 3. This means that x maps to the remainder of x/3. Since the remainder can only be 0, 1 or 2, we have three equivalence classes, namely [0], [1] and [2].

(For integers x, y ($y \neq 0$) and a, x modulo y = a means that a is the remainder when x divided by y.)

Activity 6-10: Self-assessment exercises ***Application skills***

1. Let X = {a, b, c}. Write down all the equivalence relations on X.

2. In each of the following cases, determine whether or not the given relation R on X is an equivalence relation. If it is, describe the equivalence class(es) of R. Justify your reasoning.
   (a) X = {a, b, c} and R = {(c, c), (b, b), (a, a)}
   (b) X = {a, b, c} and R = X × X
   (c) X = $\mathcal{P}$(Y) where Y = {1, 2, 3} and R consists of all pairs (C, D) such that C ∩ {2} = D ∩ {2}

3. Let R be the relation on $\mathbb{Z}$ such that (x, y) ∈ R iff x – y is a multiple of 4.

   (a) Do tests on R for all of the following properties: reflexivity, irreflexivity, symmetry, antisymmetry, transitivity and trichotomy.
   (b) Now say what kind of relation R is.
   (c) If R is an equivalence relation, give the equivalence classes of R, and show some members of each class.

4. Suppose $\mathbb{Q}^+$ is the set of all positive quotients n/m, where n, m ∈ $\mathbb{Z}^+$, i.e. $\mathbb{Q}^+$ is the set of positive rational numbers. Let R be the relation on $\mathbb{Q}^+$, defined by the rule (x, y) ∈ R iff y = (a·x) / b for some a, b ∈ $\mathbb{Z}^+$. Prove that R is an equivalence relation and show the equivalence classes of R.

5. Prove that if R is a relation on $\mathbb{Z}^+$, then R is symmetric iff R = R^{-1}.

We have seen that the basic thing an equivalence relation R on A does, is to split the set A into a bunch of subsets, each of which is an equivalence class. We formally state this.

Theorem 6.1

(i) If R is an equivalence relation on A, then x ∈ [x] for each x ∈ A, i.e. every member of A belongs to some equivalence class with respect to R.
(ii) If x R y, then [x] = [y], i.e. if two elements are equivalent with respect to R, then they belong to the same equivalence class.
(iii) If [x] = [y], then x R y, i.e. if x and y are different representatives of the same equivalence class, then x and y are equivalent with respect to R.
(iv) Either [x] = [y] or [x] ∩ [y] = ∅, i.e. different equivalence classes do not have any elements in common.

Proof

(i) R is reflexive on A, so (x, x) ∈ R for all x ∈ A, i.e. x ∈ [x] for all x ∈ A.

(ii) Suppose (x, y) ∈ R. We want to show that [x] = [y], so we have to show that the two sets, [x] and [y], are equal.

w ∈ [x]
iff x R w
iff w R x (by symmetry)
iff w R y (since x R y and R is transitive)
iff y R w (by symmetry)
iff w ∈ [y].
Hence [x] = [y].

(iii) Suppose [x] = [y]. We want to show that x R y.

$y \in [y]$ (by (i)),
i.e. $y \in [x]$ (since [x] = [y])
i.e. x R y.

(iv) Let us suppose that $[x] \neq [y]$. We will show that $[x] \cap [y] = \emptyset$. The method we use is known as *reductio ad absurdum*, or (in English) *reduction to an absurdity*. The idea is to assume temporarily the opposite of what we want to prove, and to derive a *contradiction*. Since contradictions cannot be tolerated, it would mean that something is wrong somewhere. But the only weak link in our argument is the temporary assumption we have made, so *we have to conclude that our assumption was wrong*.

Right, let's start. We assume that $[x] \neq [y]$. Our temporary assumption is: $[x] \cap [y] \neq \emptyset$.

A consequence of this assumption is that there is some $v \in A$ with $v \in [x]$ and $v \in [y]$. Then x R v and y R v. By symmetry, v R y. By transitivity, this means that x R y and thus that [x] = [y]. But this contradicts our original statement that $[x] \neq [y]$.

So we *reject the supposition* that $[x] \cap [y] \neq \emptyset$.

We have therefore shown the following:

There are two possibilities for [x] and [y]. Either [x] = [y] or $[x] \neq [y]$.
But if $[x] \neq [y]$, then $[x] \cap [y] = \emptyset$.
So either [x] = [y] or $[x] \cap [y] = \emptyset$

QED

The above theorem suggests a simple and fail-safe way to build equivalence relations. We'll get to it in a moment. But first, we reconsider a new term we have already referred to earlier in the context of equivalence relations.

When we split a set A into a bunch of non-empty subsets in such a way that each element of A lives in one of these subsets, and no two of these subsets have any elements in common, then we say that we have partitioned A. More formally:

Definition: Partitions

For a nonempty set A, a *partition of A* is a set $S = \{S_1, S_2, S_3, \ldots\}$. The members of S are subsets of A (each set S_i is called a *part* of S) such that

1. for all i, $S_i \neq \emptyset$ (that is, each *part* is nonempty),
2. for all i and j, if $S_i \neq S_j$, then $S_i \cap S_j = \emptyset$ (that is, different parts have nothing in common), and
3. $S_1 \cup S_2 \cup S_3 \cup \ldots = A$ (that is, every element in A is in some *part* S_i).

If S is a partition of A, then there exists an equivalence relation R on A whose equivalence classes are precisely the *parts* (elements) of S.

Examples

For example, let A = {a, 2, b}. Let's split A into two subsets {a, 2} and {b}. Then {{a, 2}, {3}} is a partition of A, because

1. the two subsets are not empty,
2. the intersection of {a, 2} and {b} is empty, and
3. the union of {a, 2} and {b} gives {a, 2, b} = A.

But suppose we throw the subsets {a} and {b} of A in a set, then the resulting set {{a}, {b}} is not a partition of A, because 2 has no part (subset) where it belongs; {a} $\cup$ {b} = {a, b} $\neq$ {a, 2, b}.

Activity 6-11: Partitions

Suppose we split A = {goat, pig, cow} into the subsets {goat, pig} and {pig, cow}. Is the set { {goat, pig}, {pig, cow} } a partition of A?

No, it is not. One of the conditions of a partition is that the intersection of its parts must be the empty set.

{goat, pig} $\cap$ {pig, cow} = {pig} $\neq \emptyset$. So {{goat, pig}, {pig, cow}} is not a partition of A.

A partition splits a set into subsets just like a fence subdivides a farm into different parts. Each animal on the farm reside in some part of the farm, and no animal can be in two areas at the same time.

Theorem 6.1 tells us that *every equivalence relation R on A partitions A*. What is of great interest to us is that one can go backwards: If A is partitioned, we can find an equivalence relation on A which has as its equivalence classes precisely the partitioning subsets.

Example

Consider the partition of {1, 2, 3} given by { {1, 2}, {3} }.

The subset {1, 2} tells us that [1] = {1, 2} = [2]. This means that a relation R has the pairs (1, 1), (2, 2), (1, 2) and (2, 1) as elements.

And the subset {3} tells us that [3] = {3}, i.e. (3, 3) $\in$ R.

So R = { (1, 1), (2, 2), (1, 2), (2, 1), (3, 3) }.

Of course, the fact that something works for one example doesn't mean it will always work. So let's try to prove that it always works.

Theorem 6.2

Suppose A is a nonempty set, and suppose further that we have a partition of A. Then the relation R defined on A by (x, y) $\in$ R iff x and y belong to the same partitioning subset, is an equivalence relation on A.

Proof

Let's see whether R is *reflexive* on A. Suppose x is an arbitrary member of A. Then certainly x is in the same partitioning subset of A as itself. So x R x.
Let's see whether R is *symmetric*. Suppose x R y, then x and y belong to the same partitioning subset of A. But we can just as well say that y and x belong to the same partitioning subset of A. So y R x.
Let's see whether R is *transitive.* Suppose x R y and y R z. The former implies that x and y live in the same partitioning subset. The latter implies that y and z live in the same partitioning subset. Since y can only live in one partitioning subset, it follows that x, y and z are all present in the same partitioning subset. So, more particularly, x and z live in the same partitioning subset, which means that x R z.

QED

Activity 6-12: Self-assessment exercises ***Application skills***

Determine whether P is a partition of X in each of the following cases. If it is, describe the corresponding equivalence relation.

(a) X = {1, 2, 3} and P = { $\emptyset$, {1}, {2, 3} }
(b) X = {1, 2, 3} and P = { {1}, {2}, {1, 3} }
(c) X = {1, 2, 3} and P = { {1,3}, {2} }
(d) X = {1, 2, 3} and P = { {1}, {2} }
(e) X = $\mathbb{Z}$ and P = {{0}, $\mathbb{Z}^+$, Neg} where Neg = {x | x $\in$ $\mathbb{Z}$ and x < 0}
(f) X = $\mathbb{Z}$ and P = { [0], [1], [2], [3], [4] } where
[0] = {x | x − 0 is divisible by 5 with zero remainder}
[1] = {x | x − 1 is divisible by 5 with zero remainder}
[2] = {x | x − 2 is divisible by 5 with zero remainder}
[3] = {x | x − 3 is divisible by 5 with zero remainder}
[4] = {x | x − 4 is divisible by 5 with zero remainder}.

Let us conclude this section by asking the following question: Why does one need to know about equivalence relations and partitions?

Firstly, you will apply this knowledge in the field of Boolean algebra, which you will encounter in other Computing modules.

Secondly, in later modules you will encounter things called finite state machines, and you will learn how to reduce the complexity and cost of such finite state machines by a minimisation process. The minimisation process involves defining a certain equivalence relation on the machine.

6.4 n-ary relations

Up to now we have worked with relations that had ordered pairs as members. Just as we can form ordered pairs, we can form ordered triples (x_1, x_2, x_3) in which x_1 is the first co-ordinate, x_2 the second co-ordinate and x_3 the third co-ordinate; ordered quadruples (4-tuples) (x_1, x_2, x_3, x_4) in which x_1 is the first co-ordinate, x_2 the second, x_3 the third, and x_4 the fourth; ordered quintuples (5-tuples) (x_1, x_2, x_3, x_4, x_5), and so on.

In general, if n is some positive integer greater than, or equal to 2, then (x_1, x_2, ..., x_n) is an ordered n-tuple in which x_1 is the first co-ordinate, x_2 the second, ..., and x_n the n^{th} and last co-ordinate.

Now, a set of ordered pairs (a relation), is actually a binary or 2-ary relation; the word binary tells us that the relation consists of ordered pairs, not triples or 6-tuples.

A set of ordered triples such as {(1, 0, 0), (0, 1, 2), (-1, -2, 5)} will be called a 3-ary or ternary relation. Similarly a 4-ary (quaternary) relation consists of ordered quadruples, and in general an n-ary relation consists of ordered n-tuples.

Just as a binary relation R could be viewed as a subset of a Cartesian product $A \times B$ (or $A \times A$), so an n-ary relation can be viewed as a subset of a Cartesian product $A_1 \times A_2 \times , \dots , \times A_n$, where by $A_1 \times A_2 \times , \dots , \times A_n$ is the set of all ordered n-tuples with first co-ordinates from A_1, second co-ordinates from A_2, ... , and n-th co-ordinates from A_n.

But why are these complications useful? All the important relations dealt with so far have been binary relations.

Worked example

Consider the student records of some university. They contain various different kinds of data. The data items of each type can be grouped quite naturally into a set. This might, for example, yield the following sets:

X_1 = {x | x is a valid student name}
X_2 = {x | x is a valid student number}
X_3 = {x | x is a valid date of birth}
X_4 = {x | x is either MALE or FEMALE} = {MALE, FEMALE}
X_5 = {x | x is a valid student home address}
X_6 = {x | x is a valid module}
X_7 = {x | x is a valid number of credits}
X_8 = {x | x is a valid value for *balance brought forward*}
X_9 = {x | x is a valid value for *paid to date*}
X_{10} = {x | x is a valid value for *amount owing*}

Suppose we choose a student from the set X_1 above, and then select, in the same order as the sets above, a data item associated with that student from each of the remaining sets. We end up with a 10-tuple (or just "a tuple") that looks something like this:

(Joe Johnson, 519-229-7, 14-FEB-1990, MALE, 21 GREEN ROAD IRENE 0002, COS1521, 159, R700-00, R2500-00, R350-00).

The set of all such 10-tuples forms a 10-ary relation
$R \subseteq X_1 \times X_2 \times X_3 \times X_4 \times X_5 \times X_6 \times X_7 \times X_8 \times X_9 \times X_{10}$.

Of course, for some applications not all items of data are relevant. We can, for instance, restrict ourselves to forming tuples from the sets X_1, X_4 and X_7, for example (Mary Wright, FEMALE, 113).

The set of all such 3-tuples forms a 3-ary relation $S \subseteq X_1 \times X_4 \times X_7$.

The relations that we form depend on the applications we have in mind for the data. But the point is that by forming appropriate n-ary relations, we can organise the storage of information in a computer in such a way that the data is easily accessible, the storage space is efficiently utilised, and the data can easily be modified. This approach to database organisation is called the *relational database model*. This will be discussed further in the modules on databases.

6.5 Functions

A relation from X to Y can be *functional*. What does this term mean?

Definition: Functional relation

Suppose R is a relation from X to Y (i.e. $R \subseteq X \times Y$). Then R is *functional* iff any element x in X that appears as a first co-ordinate in an ordered pair of R, does so *in exactly one ordered pair.*

That is, if $(x, y) \in R$ and $(x, z) \in R$, then $y = z$.

We can say that a relation is functional iff each first co-ordinate lives with only one second co-ordinate in an ordered pair of the relation. Let's look at some examples.

Examples

Suppose S is a relation from {1, 2, 3} to {a, b, c}, then S = { (1, a), (2, c) } is functional. We can claim this because 1 and 2 are members of {1, 2, 3}, and each one of the elements 1 and 2 appears as first co-ordinate in exactly one pair of S.

If we consider a relation R on $\mathbb{Z}$ defined by $(x, y) \in R$ iff $y = x + 5$, then we can determine whether or not this relation is functional:

Suppose $(x, y) \in R$ and $(x, z) \in R$. Is it the case that $y = z$?
Well, since $(x, y) \in R$, $y = x + 5$ and since $(x, z) \in R$, $z = x + 5$.
So $z = x + 5 = y$, which means that R is functional.

We have shown that each first co-ordinate lives with only one special second co-ordinate in an ordered pair of R.

There are many examples of relations from X to Y, whether they are functional or not, that might be very small subsets of X × Y. Some elements of X and of Y might not appear at all as co-ordinates of pairs in these relations. For example, if X = Y = $\mathbb{Z}$ and R = { (2, 3), (3, 4), (5, 100) }, then most integers do not appear as co-ordinates of the functional relation R. We should be able to refer to the subsets of X and Y whose elements play a role in such a relation. Let's recap on the definitions of domain, range and codomain that were provided in the previous study unit.

Definitions: Domain / Range / Codomain

Suppose R is a relation from X to Y, then the *domain* of R (dom(R)) is a subset of X, and the range of R (ran(R)) is a subset of the *codomain* of R where the codomain is the set Y.

We have $dom(R) = \{x \mid \text{for some } y \in Y, (x, y) \in R\}$ (i.e. the set of first co-ordinates) and

$ran(R) = \{y \mid \text{for some } x \in X, (x, y) \in R\}$ (i.e. the set of second co-ordinates).

This brings us to a very important point in this study unit, namely the concept of a function. A function is a special kind of relation.

Definition: Function

Suppose $R \subseteq A \times B$ is a binary relation (i.e. it involves two sets) from a set A to a set B. We may call R a *function* from A to B if every element of A appears exactly once as the first co-ordinate of an ordered pair in R (i.e. f is functional), and the domain of R is exactly the set A, i.e. dom(R) = A.

This function is denoted by $R: A \rightarrow B$, i.e. R is a function from A to B.

Examples

Suppose S = {(1, c)} is a relation from A = {1, 3} to B = {a, c}, then S is functional but *it is not a function* since dom(S) ≠ {1, 3}.

In a previous example we proved that the relation R on $\mathbb{Z}$ defined by $(x, y) \in R$ iff y = x + 5, is functional. Now we also determine the domain of R:

Dom(R)
= {x | for some $y \in \mathbb{Z}$, $(x, y) \in R$}
= {x | for some $y \in \mathbb{Z}$, y = x + 5}
= {x | x + 5 is an integer} ($y \in \mathbb{Z}$ and y = x + 5, thus $(x + 5) \in \mathbb{Z}$.)
= $\mathbb{Z}$.

R is *functional* and *dom(R) = $\mathbb{Z}$*, thus R is a *function*. We may write R: $\mathbb{Z} \rightarrow \mathbb{Z}$.

*Note: If it must be proved that the relation R in the above example is a function, then a general statement such as "every element of $\mathbb{Z}$ appears exactly once as the first co-ordinate of an ordered pair in R, and the domain of R is exactly the set $\mathbb{Z}$" does not constitute a proof. Rather: It should be **proved** that R is **functional** and dom(R) should be **determined** (by using the definition of "domain") to **prove** that **dom(R)** = **A**.*

We determine a few ordered pairs belonging to R: if x = 0 then x + 5 = 5, thus (0, 5) ∈ R; if x = 1 then x + 5 = 6, thus (1, 6) ∈ R; if x = −2 then x + 5 = 3, thus (−2, 3) ∈ R.

We usually give relations names such as R or S, but because functions are so important, we usually denote these as f, g or h, and so on. We use the notation f: X → Y to indicate that f is a function from X to Y.

Suppose f is a function from A to B. If a pair (a, b) belongs to f, then we know that b is the only element that can be the second co-ordinate next to the element a. This means we can use a new way to say that (a, b) belongs to f. We can write in words "b is the image of a under f", or "a maps to b", or we can write in symbols f(a) = b, i.e. "f of a is equal to b".

The equation f(a) = b means, of course, that f(a) is an alternative name we may use for that particular b which stands next to a in the ordered pair.

The usefulness of functions is based on the fact that we can think of a function as a rule that tells us how to get from A to B – if you're at the point a ∈ A, go to f(a) ∈ B. This works only because we are never uncertain about where to go – for each a ∈ A there is only one f(a) ∈ B, since "a" appears as first co-ordinate in only one ordered pair of f. Can you spot the application of this reasoning in the following function g?

Let A = {Mary, Thabo, Shawren} and B = {Memel, Polokwane, Durban, Ibhayi} and let g = {(Mary, Memel), (Thabo, Ibhayi), (Shawren, Durban)} be a function from A to B such that a person and the city where he/she resides are grouped together in an ordered pair.

Activity 6-13: Determine whether or not a given relation is a function

As we have seen in the definition of a function, there are two things we need to determine when we want to determine whether some relation is a function:

(i) Is the relation functional?
(ii) What is the domain of the relation?

Suppose we are given a relation R from A to B and we need to prove that R *is a function* from A to B, then we should prove that

R is *functional*, i.e. if $(x, y) \in R$ and $(x, z) \in R$, then $y = z$, and we should also prove that the *domain* of R, i.e. is the set
$dom(R) = \{x \mid \text{for some } y \in Y, (x, y) \in R\}$, is equal to A.

Note: Vague explanations will not do, so formal proofs should be provided.

Let's quickly recap: A function f from A to B is a binary relation with *domain A* and *codomain* B, with the property that for every $x \in A$, there is exactly one element $y \in B$ such that $(x, y) \in f$. (f is *functional*.)

We note that for *every* element x in A, there is *exactly one* element y in B such that $(x, y) \in f$. This means that *dom(f) must be equal to A*, i.e. each domain element must appear exactly once as first co-ordinate, and that $ran(f) \subseteq B$.

Now if we consider some *relation* from X to Y, it will not necessarily be true that every element in X appears as first co-ordinate in the relation, nor is it necessarily true that the elements in X are related to exactly one element in the codomain Y.

To illustrate: Let $S = \{(1, a), (1, c)\}$ be the relation from X to Y where $X = \{1, 2, 3\}$ and $Y = \{a, b, c\}$. It is clear that 1 is the only element in X that appears as first co-ordinate in ordered pairs of S (*each* element in X does *not* appear as first co-ordinate, i.e. $dom(S) \neq X$). It is also the case that 1 appears as first co-ordinate in more than one ordered pair, so 1 is related to *more than one* element in the codomain (so S is not functional). $dom(S) \neq X$ and S is not functional, thus S is not a function.

Let's investigate whether the following relation S is a function:

Example

Let $S = \{(4, a), (5, b), (6, a)\}$ be a relation from X to Y with $X = \{4, 5, 6\}$ and $Y = \{a, b, c\}$.

S is functional:
If $(x, y) \in S$ and $(x, z) \in S$, then
either $x = 4$ with $y = a = z$, or $x = 5$ with $y = b = z$, or $x = 6$ with $y = a = z$.
This means that for every $x \in X$, there is exactly one element $y \in Y$ such that $(x, y) \in S$.

We also see that $dom(S) = \{4, 5, 6\} = X$.

We have shown that S is functional and that $dom(S) = X$, thus S is a function.

On the other hand, $S^{-1} = \{(a, 4), (b, 5), (a, 6)\}$ is not functional, because it is possible to find pairs $(x, y) \in S^{-1}$ and $(x, z) \in S^{-1}$ such that $y \neq z$, namely the pairs with x = a, y = 4 and z = 6.

More examples

We look at some examples of functions that arise in Computing:

Since a function f: A → B may be thought of as something which transforms input (the elements a of A) into output (the elements b of B), one can regard a C++ compiler as a function that transforms a source program (the input) into its corresponding object program (the output).

In C++ a function called *trunc*, the truncation function, are often used and it is a function from $\mathbb{R}$ to $\mathbb{Z}$ which transforms a real number into an integer by deleting any fractional part, e.g. trunc (3.78) = 3, trunc (5) = 5, trunc (−7.22) = −7.

SOMETHING FOOLISH THAT MIGHT HELP YOU REMEMBER THE BASIC IDEA: Think of a function f: A → B as a monster that eats little woolly a-nimals that live in A and spits out each little woolly animal's b-ackbone to be b-uried in B.

Another example

Prove that f defined by $(x, y) \in f$ iff $y = 5x^2 + 3$ is a function on $\mathbb{R}$.

We investigate whether f is functional and determine whether dom(f) = $\mathbb{R}$:

Suppose $(x, y) \in f$ and $(x, z) \in f$. Is it the case that y = z?
Well, since $(x, y) \in f$, $y = 5x^2 + 3$ and since $(x, z) \in f$, $z = 5x^2 + 3$.
So $z = 5x^2 + 3 = y$, which means that f is functional. Furthermore,

Dom(f)
$= \{x \mid \text{for some } y \in \mathbb{R}, (x, y) \in f\}$
$= \{x \mid \text{for some } y \in \mathbb{R}, y = 5x^2 + 3\}$
$= \{x \mid 5x^2 + 3 \text{ is a real number}\}$
$= \mathbb{R}$.

We can now say that f is a *function* because we *proved* that f is *functional* and that *dom(f) = $\mathbb{R}$*.

Activity 6-14: Self-assessment exercises *Application skills*

1. Give 5 functions from A = {1, 2, 3, 4} to B = {a, b, c}.

2. Give all the functions from A = {a, b} to B = {5, 6, 7}.

3. Give 3 functions from A × A to B if A = {a, b} and B = {5, 6, 7}.

4. Let R be a relation on A = {1, 2, 3, {1}, {2}} defined by
   R = { (1, {1}), (1, 3), (2, {1}), (2, {2}), ({1}, 3), ({2}, {1}) }.

   (a) Is R a function from A to A?
   (b) Is ran(R) equal to the codomain of R?

5. Consider the set $\mathcal{P}(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$. Show that the relations f, g, and h described below are functional and have as domains $\mathcal{P}(A)$, $\mathcal{P}(A) \times \mathcal{P}(A)$, and $\mathcal{P}(A) \times \mathcal{P}(A)$ respectively:
   (a) Let $f = \{(x, y) \mid x, y \in \mathcal{P}(A) \text{ and } y = x'\}$.
   (b) Let $g = \{((u, v), y) \mid (u, v) \in \mathcal{P}(A) \times \mathcal{P}(A) \text{ and } y = u \cup v\}$.
   (c) Let $h = \{((u, v), y) \mid (u, v) \in \mathcal{P}(A) \times \mathcal{P}(A) \text{ and } y = u \cap v\}$.

6. For each of the following relations from X to Y, determine whether or not the relation may be regarded as a function from X to Y.
   (a) $X = Y = \mathbb{Z}$ and $R = \{(x, y) \mid y = x\}$.
   (b) $X = Y = \mathbb{Z}$ and $R = \{(x, y) \mid y = x + 1\}$.
   (c) $X = Y = \mathbb{Z}$ and $R = \{(x, y) \mid y = 3 - x\}$.
   (d) $X = Y = \mathbb{Z}$ and $R = \{(x, y) \mid y = \sqrt{x}\}$, where the notation $\sqrt{x}$ refers to the positive square root of x.
   (e) $X = Y = \mathbb{Z}$ and $R = \{(x, y) \mid y^2 = x\}$.
   (f) $X = Y = \mathbb{R}$ and $S = \{(x, y) \mid x^2 + y^2 = 1\}$.

7. Is the relation R on $\mathbb{Z}^+$, which consists of all pairs (x, y) such that $y = x - 1$, a function from $\mathbb{Z}^+$ to $\mathbb{Z}^+$?

8. Let A = {a, b, c}. Consider all the equivalence relations on A (see activity 6.9 (1)). How many relations are also functions from A to A? (We use brute force and then abstract reasoning in our answer. We recommend the latter.)

9. Let A = {a, b, c}.
   (In the answers to the questions that follow, we use brute force (refer to activity 6-6(1)) and also abstract reasoning in our proofs. We recommend the latter.)
   (a) How many weak partial orders on A (reflexive, antisymmetric and transitive) are also functions from A to A?
   (b) How many strict partial orders on A (irreflexive, antisymmetric and transitive relations) are also functions from A to A?

6.6 In summary of the study unit

In this study unit you ensured that you can answer the following questions on special kinds of relation:

- What does the term order relation mean?
- What is the difference between a partial order and a total (linear) order?
- What is the difference between a weak total order and a strict total order?
- What does the term trichotomy mean?
- What are n-ary relations?
- How is an equivalence relation defined?
- What is an equivalence class?
- What does the term partition mean?
- What does the term “functional” mean?
- When is a relation a function?

In the following study unit we will learn more about the *properties* of functions.

Study unit 7 More about functions

Key questions for this study unit

- *What is the usefulness of properties of functions such as "surjectivity", "injectivity" and "bijectivity"?*
- *How do we test whether a function has the properties listed above?*
- *How do we form the composition of two functions?*
- *How do we construct the inverse of a function?*

Activity 7-1: Overview *Study skill*

Draw a mind map of the different sections/headings you will deal with in this study session. Then page through the unit with the purpose of completing the map.

Your map should include the concepts of surjectivity, injectivity, an invertible function, the identity function, function composition, the composition of relations and bijectivity.

Activity 7-2: Concepts *Conceptual skill*

Test your own knowledge and then correct your understanding afterwards. How does your understanding deepen as you jot down the terms used in your home language?

English term	Description	Term in your home language
Surjective function		
Injective function		
Composition of relations		
Composition of functions		
Inverse relation		
Bijective function		
Invertible function		
Identity function		

7.1 Surjective functions

Suppose we have a function f: A → B. As you know, one may think of f as a rule that tells one how to get from A to B. A very sensible question to ask is the following: *Can every b ∈ B be reached from some a ∈ A?*

It's rather like having a set A of factories, a set B of markets, and a set f of roads connecting *each* factory in A with one of the markets in B. (In other words, f is a function from A to B and we interpret each ordered pair in f as saying "There is a road from my first co-ordinate to my second co-ordinate".) Because f is a function, *every* factory in A is connected to a market in B by *only one* road. It is important to know whether all the markets in B can be reached.

In order to discuss this question, we give a special name to the set of "markets" that can be reached.

Definition: The range of a function

Given a function f: A → B, the *range* or *image set* of f is the subset {f(x) | x ∈ A} of B, which may be denoted by either ran(f) or f[A] .

In other words, the *range* of a function f: A → B is the subset ran(f) of B consisting of those elements in B for which there exists an x ∈ A such that f(x) ∈ ran(f).

Does this definition remind you of the definition of the range of a relation that was provided in the previous study unit? If f is a function on A, the sets {f(x) | x ∈ A} and {y | for some x ∈ A, (x, y) ∈ f } both defines ran(f).

Examples

Suppose A = {1, 2, 3} and B = {5, 7, 9} and h: A → B is the function defined by the following: h(1) = 5, h(2) = 5 and h(3) = 7.

Then ran(h) = {5, 7}. If 1, 2 and 3 represent factories and 5, 7 and 9 represent markets, then the fact that ran(h) is not equal to codomain B means that not all the markets can be reached using the roads in h.

Let's look at another example: Suppose f: $\mathbb{Z} \to \mathbb{Z}$ is defined by y = 2x (the codomain is $\mathbb{Z}$). Then the range ran(f) (or f[$\mathbb{Z}$]) consists of all the even integers. Why do we say this?

ran(f) = {f(x) | x ∈ $\mathbb{Z}$} = {2x | x ∈ $\mathbb{Z}$}
= {y | for some x ∈ $\mathbb{Z}$, y = 2x} (y = 2x, i.e. x = y/2)
= {y | y/2 is an integer}
= {y | y is an even integer}

Note that if x ∈ $\mathbb{Z}$ then x = y/2 is also an integer. It follows that y/2 is an integer only if y is an even integer. So the range of f is *not equal to* the codomain $\mathbb{Z}$ because the odd integers are not included in the range.

If, for some reason, we want to prove that ran(f) ≠ $\mathbb{Z}$, we can provide a ***counterexample***: Choose y = 3, then there is no x ∈ $\mathbb{Z}$ such that 2x = 3 (i.e. x = 3/2 ∉ $\mathbb{Z}$) and so 3 ∉ ran(g). Thus ran(g) ≠ $\mathbb{Z}$.

Yet another example: What is the range of g: $\mathbb{Z} \to \mathbb{Z}$ defined by g(x) = x + 5?

ran(g) = {g(x) | x $\in \mathbb{Z}$}
= {x + 5 | x $\in \mathbb{Z}$}
= {y | y – 5 is an integer} (y = x + 5, i.e. x = y – 5)
= $\mathbb{Z}$

So for the function g the range is *equal to* the codomain.

Definition: Surjectivity

Given a function f: A → B, we say that f: A → B is *surjective* iff the range of f is equal to the codomain of f, i.e. f[A] = B.

Since the range of f is defined to be a *subset* of the codomain, we know that f[A] $\subseteq$ B.

For equality, we also require that B $\subseteq$ f[A]. So f: A → B is surjective iff for every b $\in$ B, we can find some a $\in$ A such that b = f(a).

Activity 7-3: Surjective functions

Explain *in your own words* what it means when a function is surjective.

Do you agree that g: $\mathbb{Z} \to \mathbb{Z}$ defined by g(x) = x + 5 is surjective because g[$\mathbb{Z}$] = $\mathbb{Z}$? (Determine g[$\mathbb{Z}$] as shown in a previous example.)

And do you agree that f: $\mathbb{Z} \to \mathbb{Z}$ defined by f(x) = 2x is *not* surjective because f[$\mathbb{Z}$] ≠ $\mathbb{Z}$? (Provide a counterexample as shown in a previous example.

People often say f maps ***onto*** B when they want to explain what it means when a function is ***surjective***.

An example of a surjective function would be the following:
Let A = {1, 2, 3} and B = {5, 7}.

Define f: A → B by
f(1) = 5, f(2) = 5, and f(3) = 7, then
f is the function {(1, 5), (2, 5), (3, 7)}.

Clearly f[A] = {5, 7} = B.

Note: The range of f, i.e. f[A], is the **set** {5, 7} - one cannot merely give 5, 7 as the range.

Activity 7-4: Self-assessment exercises *Application skills*

1. In each of the following cases, write down the possible surjective functions from X to Y. We will do the first one.

(a) X = {a, b} and Y = {c}.

To obtain a surjective function from X to Y, we must try to fill in the template { (a,), (b,) } in such a way that all the elements of Y are used. This can only be done as follows: g = { (a, c), (b, c) }.

(b) X = {a, b} and Y = {c, d}.
(c) X = {a, b} and Y = {c, d, e}.

2. Let f: $\mathbb{Z} \to \mathbb{Z}$ be defined by f(x) = x + 1.

(a) Determine f[$\mathbb{Z}$] (or ran(f)). (Do not give specific examples of elements in f[$\mathbb{Z}$].)
(b) Is f surjective? If f is not surjective, provide a counterexample to show why it is not surjective.

3. Let g: $\mathbb{Z} \to \mathbb{Z}$ be defined by g(x) = 4x + 8.

(a) Determine g[$\mathbb{Z}$] (or ran(g)). (Do not give specific examples.)
(b) Is g surjective? If g is not surjective, provide a counterexample to show why it is not surjective.

7.2 Injective functions

Suppose a company buys 5 new word processor packages. Nine typists are hoping to obtain one of the new word processor packages.

If we think of A as the set of word processors and of B as the set of typists, then an allocation of word processors to typists can be thought of as a function f : A $\to$ B, where for each a $\in$ A the image f(a) is the lucky typist who gets a word processor.

Suppose one of the typists is the favourite of the company director. Would it be a good idea for him to give her two of the word processors? I doubt it. Some unlucky typist who didn't get a word processor would be most upset. We see that it is important for f to map different word processors to different typists.

Another way to think of it is that each typist should get *at most* one word processor.

Let's formalise this property we want f to have.

Definition: Injectivity

A function f: A $\to$ B is *injective* iff f has the property that whenever $f(a_1) = f(a_2)$ then $a_1 = a_2$.

What would the formal version of this reasoning look like? If a typist gets a word processor, then the typist can be called $f(a_1)$, where a_1 is the word processor she gets. To make sure that she gets no more than a single word processor, we could require that whenever $f(a_1) = f(a_2)$ then $a_1 = a_2$, i.e. whenever a typist receives word processors a_1 and a_2, then it should be the case that a_1 and a_2 represent the same word processor.

Alternative definition:

A function f: A $\to$ B is *injective* iff f has the property that whenever $a_1 \neq a_2$ then $f(a_1) \neq f(a_2)$.

The definition captures the idea that different word processors must go to different typists. Instead, we could have captured the idea that each typist should get at most one word processor.

In the part on logic that follows in a later study unit, we will show you that the following two conditions:

(i) if $a_1 \neq a_2$ then $f(a_1) \neq f(a_2)$, and
(ii) if $(a_1) = f(a_2)$ then $a_1 = a_2$

are equivalent, since the one is what we call the contrapositive of the other.

*Instead of using the term "**injective**", people often say a function f is **one-to-one**.*

Examples

Let A = {1, 2, 3} and B = {6, 7, 8}.

The function f: A $\rightarrow$ B defined by
f (l) = 7,
f(2) = 6, and
f(3) = 8

is injective since different elements of the domain clearly go to different elements of the range.

However, the function h: A $\rightarrow$ B defined by
h(l) = 7,
h(2) = 8, and
h(3) = 8

is *not* injective, since h(2) = 8 = h(3) but 2 ≠ 3.

Consider our old friend, the function g: $\mathbb{Z} \rightarrow \mathbb{Z}$ defined by g(x) = x + 5. We determine whether or not g is injective.

Assume g(u)= g(v)
then u + 5 = v + 5
i.e. u = v.

Therefore f is injective.

Activity 7-5: Self-assessment exercises *Application skills*

In each of the following cases, write down the injective (one-to-one) functions from X to Y (if possible). We will do the first one.

(a) X = {2, 4} and Y = {1}.

We can build an injective function from X to Y by filling in the template { (2,), (4,) } in such a manner that different pairs contain different elements of Y. But this is not possible, because Y has only one member, and there are two pairs. So there is *no* injective function from X to Y in this case.

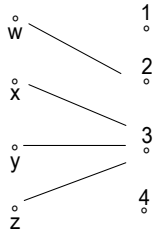
(b) X = {2, 4} and Y = {1, 3}.
(c) X = {2, 4} and Y = {1, 3, 5}.

Consider h: $\mathbb{Z} \rightarrow \mathbb{Z}$ defined by h(x) = 2x – 5. Determine whether or not h is injective.

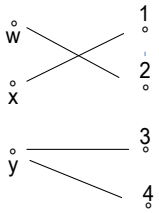
Examples

We look at some diagrams that show different correspondences:

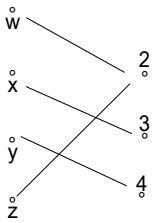
(a) a function



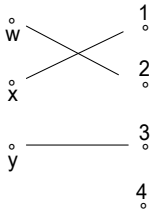
(b) not a function



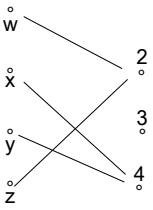
(c) surjective, not injective



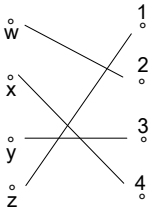
(d) injective, not surjective



(e) neither surjective, nor injective



(f) injective, and surjective



Activity 7-6: Self-assessment exercises

For each of the above diagrams, write down the corresponding relation or function it represents, then provide the reason(s) why the relation or function has the given property or properties.

7.3 The composition of relations / functions

Let's forget about functions for a moment and recap some definitions that are closely related to what we discuss in this section. So we have a little bit of repetition here.

Suppose we have a relation R from A to B, in other words $R \subseteq A \times B$, and a relation S from B to C, in other words $S \subseteq B \times C$.

In study unit 5 we have already seen that there is an important way to build a new relation from R and S, called *forming the composition* of R followed by S. This new relation is denoted by $S \circ R$ (pronounced *S little circle R*) or we can write R; S. According to the convention, although it is the composition of R followed by S, we write down $S \circ R$, i.e. we write down S first (i.e. on the left) and then R. (Remember, by using the notation R; S rather than $S \circ R$, it helps us to remember that R is followed by S.)

What exactly is $S \circ R$?

Definition: The composition of relations

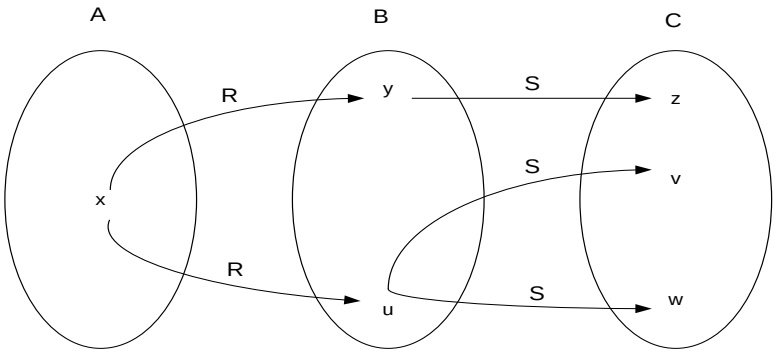
Given relations R from A to B and S from B to C, *the composition of R followed by S* is the relation from A to C defined by

$S \circ R = \{(a, c) \mid$ there is some $b \in B$ such that a R b and b S c$\}$.

It is worth spending some time thinking about this definition. If x is a first co-ordinate in some pair of $S \circ R$, is it clear to you that x must be a member of A? We can say this because if x appears as a first co-ordinate in $S \circ R$, then there has to be some $b \in B$ such that $(x, b) \in R$. Similarly, if y appears as a second co-ordinate in $S \circ R$ then there has to be some $b \in B$ such that $(b, y) \in S$, which means that y must be a member of C. So the definition ensures that $S \circ R \subseteq A \times C$.

What is the intuitive idea behind the composition definition? Let's think in terms of a concrete example: Suppose A is a set of people who work in Johannesburg, but live in Limpopo, and B is a set of busses taking people from Johannesburg to Polokwane. C is a set of taxi services between Polokwane and other towns in Limpopo. Think of each ordered pair (a, b) in $R \subseteq A \times B$ as representing the fact that "there is bus service for person a to town b". Suppose further that each ordered pair (b, c) in $S \subseteq B \times C$ represents "there is a regular taxi service between b and c". Then $S \circ R$ consists of pairs (a, c) which may be thought of as saying "there is an efficient transportation service for person a to town c".

The following diagram illustrates how $S \circ R$ links elements of A to elements of C by using elements of B as intermediate points.



Examples

Suppose A = {1, 2, 3}, B = {4, 5} and C = {6, 7, 8, 9}.
Let R = { (1, 4), (1, 5), (3, 5) } and S = { (4, 7), (5, 8) }.

Then $S \circ R$ = { (1, 7), (1, 8), (3, 8) }.

With the same sets A, B, and C, let f: A → B be { (1, 4), (2, 4), (3, 5) } and let g: B → C be { (4, 7), (5, 9) }.

Then $g \circ f$ = { (1, 7), (2, 7), (3, 9) }.

Did it surprise you to suddenly see functions popping up? But remember, functions are relations with special properties!

The composition of relations is a pretty useful construction, but it becomes really special when you apply it to functions. We'll see some of the uses later. First we must ask ourselves some obvious questions.

What happens when we form the composition of two functions? Will the result be a function?

Activity 7-7: The composition of functions

Suppose we have functions f: A → B and g: B → C. Is the composition $g \circ f$ a function from A to C?

Yes, indeed. Let's consider a $\in$ A. Since f is a function, there is exactly one b $\in$ B such that (a, b) $\in$ f. But now b belongs to the domain of g, and g is also a function. This means that there is exactly one c $\in$ C such that (b, c) $\in$ g. Clearly the a with which we started has exactly one c linked to it through the intermediary b.

That is, for each a $\in$ A there exists just one c $\in$ C such that (a, c) $\in$ g $\circ$ f.

Therefore we may write g $\circ$ f: A $\rightarrow$ C

We have proved the following theorem:

Theorem 7.1

The composition of two functions is also a function.

Now that we know that the composition of two functions is a function, we provide the following definition:

Definition: Composition of functions

Given the functions f: A $\rightarrow$ B and g: B $\rightarrow$ C the composite function

g $\circ$ f: A $\rightarrow$ C is defined by g $\circ$ f(x) = g(f(x)).

The composite function g $\circ$ f is defined for all x for which f(x) and g(f(x)) exist.

The image of an element x $\in$ A under the function g $\circ$f is denoted by (g $\circ$f)(x). If we think a little about how the composition is defined, it becomes clear that (g $\circ$f)(x) = g(f(x)) since, to find (g $\circ$f)(x), we first feed x to f, which gives f(x), and then feed this result to g, i.e. we feed the element called f(x) to g, and g sends it to the image g(f(x)).

Let's look at some examples.

Examples

Suppose f: $\mathbb{Z} \rightarrow \mathbb{Z}$ is defined by f(n) = 3n + 1 and
g: $\mathbb{Z} \rightarrow \mathbb{Z}$ is defined by g(n) = n^3. We determine g $\circ$ f:

g $\circ$ f: $\mathbb{Z} \rightarrow \mathbb{Z}$ is defined by

(g $\circ$ f)(n) = g(f(n))
= g(3n + 1) (g(**f(n)**) = g(**3n + 1**), i.e f(n) is replaced by 3n+1)
= $(3n + 1)^3$ (g(**n**) = **n**^{3}, thus g(**3n + 1**) = (**3n + 1**)^3)
= $(3n + 1)(3n + 1)^2$
= $(3n + 1)(9n^2 + 6n + 1)$
= $27n^3 + 18n^2 + 3n + 9n^2 + 6n + 1$
= $27n^3 + 27n^2 + 9n + 1$

Note: g $\circ$ f is a *function* on $\mathbb{Z}$ and (g $\circ$ f)(n) is called the *image of n under g $\circ$f.*

Another example: Suppose f: $\mathbb{R} \to \mathbb{R}$ is defined by f(x) = 2x and g: $\mathbb{R} \to \mathbb{R}$ is defined by $g(x) = 3x^2 + 5$. We determine g ∘ f:

g ∘ f: $\mathbb{R} \to \mathbb{R}$ is defined by

$(g \circ f)(x) = g(f(x))$
$= g(2x)$ (g(**f(x)**) = g(**2x**), note that f(x) is replaced by 2x)
$= 3(2x)^2 + 5$ (g(**x**) = 3**x**^{2} + 5, thus g(**2x**) = 3(**2x**)^2 + 5)
$= 3(4x^2) + 5$
$= 12x^2 + 5$

Note: g ∘ f is a function on $\mathbb{R}$ and (g ∘ f)(x) is called the image of x under g ∘ f.

Let's investigate further: Suppose we start off with two surjective functions f: A → B and g: B → C. Is the composition g ∘f: A → C by any chance surjective?

Yes! We mentioned above that the composition of functions is special. To check whether g ∘f: A → C is surjective, one must check that every c ∈ C appears as second co-ordinate in some pair in g ∘f, i.e. that for each c ∈ C there is some a ∈ A such that c = g(f(a)).

Is this reasonable? Well, pick any c ∈ C. Since g: B → C is surjective, we can find some b ∈ B such that c = g(b). Now let's focus on that specific b. Since f: A → B is surjective, we can find an a ∈ A such that b = f(a). But this means that there exists an a ∈ A such that c = g(f(a)), i.e. such that (a, c) ∈ g ∘f. Hence g ∘f: A → C is surjective.

So we have proved the following theorem:

Theorem 7.2

The composition of two surjective functions is surjective.

Activity 7-8: The composition of injective functions

Does the same hold for injective functions?
Suppose f: A → B and g: B → C are both injective. Do we have grounds to claim that g ∘ f: A → C is injective?

Well, suppose that inside C we find elements g(f(a_1)) and g(f(a_2)). Using only the information we have about f and g, we need to show that
if g(f(a_1)) = g(f(a_2)), then $a_1 = a_2$.

But since we know that g is injective, it follows from g(f(a_1)) = g(f(a_2)) that f(a_1) = f(a_2). (Remember that the injective function g spits out equal things only if fed equal things).

And because we know that f is injective, it follows from f(a_1) = f(a_2) that $a_1 = a_2$.

Hence we have proved the following theorem:

Theorem 7.3

The composition of two injective functions is injective.

Activity 7-9: Self-assessment exercises ***Application skills***

Determine $f \circ f$, $g \circ g$, $g \circ f$, and $f \circ g$ in each of the following cases:

(a) $f: \mathbb{Z} \to \mathbb{Z}$ is defined by $f(x) = x + 1$ and
$g: \mathbb{Z} \to \mathbb{Z}$ is defined by $g(x) = x - 1$.

(b) $f: \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = 3x - 2$ and $g: \mathbb{R} \to \mathbb{R}$ is defined by $g(x) = x^2 + x$.

(c) $f: \mathbb{Z}^{\geq} \to \mathbb{Z}^{\geq}$ is defined by $f(x) = 113$ and $g: \mathbb{Z}^{\geq} \to \mathbb{Z}^{\geq}$ is defined by $g(x) = x + 1$.

7.4 Bijective functions and inverses

As we have discussed in study unit 5, if R is a relation from A to B, then we can form a new relation from R by simply *reversing* every ordered pair in R.

For example, if A = {1, 2, 3}, B = {5, 6, 7}, and R = { (1, 7), (3, 5) } then the new relation R^{-1} (called the *inverse relation* of R) is the relation from B to A:
R^{-1} = { (7, 1), (5, 3) }.

Remark: *For a real number x, the notation x^{-1} means 1/x. But if we apply the notation to a relation R, then we must be careful not to read into the notation things that aren't there. We have not defined anything of the form 1/R.*

Definition: Inverse relation

For any relation R, the inverse relation, denoted by R^{-1}, is the set
$\{(y, x) \mid (x, y) \in R\}$.

Now let's return to functions. Given a function $f: A \to B$, then since f is a relation, we can form its inverse relation. But the crucial question is the following: *Under what conditions would the inverse relation of a function $f: A \to B$ itself be a function from B to A?*

We will only be able to answer this question by using the following two definitions and the theorem that follows from them:

Definition: Bijective functions

A function $f: A \to B$ is *bijective* iff f is both surjective and injective.

What does this definition mean? We can say that
if a function is bijective, then it is both surjective and injective, and
if a function is both surjective and injective, then it is bijective.

In this unit we have proved that the function $g: \mathbb{Z} \to \mathbb{Z}$ defined by $g(x) = x + 5$ is injective and surjective, hence we can say that g is a bijective function.

Definition: Invertible functions

A function $f: A \to B$ is *invertible* iff the inverse relation f^{-1} of f is a function from B to A.

This means that if a function $f: A \to B$ is invertible, then the inverse relation f^{-1} of f is a function from B to A, and if the inverse relation f^{-1} of f is a function from B to A, then f is invertible.

Theorem 7.4

Let f: A $\rightarrow$ B be a function. Then f is invertible iff f is bijective.

Proof

Since we have an *if and only if* statement to prove, our argument will have to run in two directions.

First suppose that f: A $\rightarrow$ B is invertible. We want to use this information to show that f is bijective.

But if f: A $\rightarrow$ B is invertible, it means that $f^{-1} = \{(y, x) \mid (x, y) \in f\}$ is a function from B to A.

So the *domain* of f^{-1} is B. But the domain of f^{-1} is also the set of all y's such that $(x, y) \in f$ for some x, i.e. the domain of f^{-1} is just the *range* of f. So the range of f is B, which means that f: A $\rightarrow$ B is surjective.

Furthermore, f^{-1} is a function, so an element $y \in B$ appears just once as first co-ordinate in an ordered pair of f^{-1}. That is, if (y, x_1) and (y, x_2) are both in f^{-1}, then $x_1 = x_2$. In other words, if (x_1, y) and (x_2, y) are both in f, then $x_1 = x_2$.

Thus if $f(x_1) = f(x_2)$, then $x_1 = x_2$, which in turn tells us that f is injective.

Now let us argue in the *reverse* direction. Suppose f: A $\rightarrow$ B is bijective. If we flip around the ordered pairs in f to get f^{-1}, is the result a function from B to A?

The surjectivity of f: A $\rightarrow$ B means that every element of B appears as second co-ordinate in an ordered pair of f, which means that every $b \in B$ appears as first co-ordinate in an ordered pair of f^{-1}, so the domain of f^{-1} is B.

Furthermore, the injectivity of f means that every b in the range of f appears *just once* as second co-ordinate in an ordered pair of f, i.e. every $b \in B$ appears *just once* as first co-ordinate in an ordered pair of f^{-1}, so f^{-1} is a function.

Since the second co-ordinates of f^{-1} are clearly members of A, we may say that f^{-1}: B $\rightarrow$ A, i.e. f^{-1} is the *inverse function* from B to A.

Example

Let us again consider g: $\mathbb{Z} \rightarrow \mathbb{Z}$ defined by g(x) = x + 5. In this unit we have proved that g is injective and surjective, i.e. g is a bijective function. Because g is bijective, it is invertible (from Theorem 7.4), so by definition g^{-1} is a function from $\mathbb{Z}$ to $\mathbb{Z}$, i.e. g^{-1} is the *inverse function* from $\mathbb{Z}$ to $\mathbb{Z}$. We determine the *inverse function* g^{-1}:

$$\begin{aligned} (y, x) \in g^{-1} \quad &\text{iff } (x, y) \in g \\ &\text{iff } y = x + 5 \\ &\text{iff } x = y - 5 \end{aligned}$$

Hence g^{-1}: $\mathbb{Z} \rightarrow \mathbb{Z}$ is defined by $g^{-1}(y) = y - 5$.

Before we close this discussion of functions and their properties, there is one important function we need to introduce namely the identity function.

Definition: Identity function

For any set A, define the function i_A: A → A by requiring that $i_A(x) = x$ for all $x \in A$. This function is called *the identity function* on A, since what it spits out is identical to what it eats.

Activity 7-10: Identity function

Consider the identity function i_B: B → B with B = {1, 2, 3, 4}. Which members live in this set?

Since $i_B(x) = x$ for all $x \in B$, we have $i_B(1) = 1$, $i_B(2) = 2$, $i_B(3) = 3$ and $i_B(4) = 4$, so $i_B = \{(1, 1), (2, 2), (3, 3), (4, 4)\}$.

Can you see that "i_B" is just another name for the equality relation usually called "="?

Activity 7-11: Self-assessment exercises *Application skills*

1. In each of the following cases, write down the bijective (one-to-one correspondence) functions from X to Y (if possible):

(a) X = { ∅, {113} } and Y = { {1} }.
(b) X = { ∅, {113} } and Y = { {1}, {2} }.
(c) X = { ∅, {113} } and Y = { {1}, {2}, {7} }

2. Check the following functions for injectivity (one-to-one), surjectivity (onto) and bijectivity (i.e. functions that are both injective and surjective), and give the inverse function of each:

(a) f: $\mathbb{Z} \to \mathbb{Z}$ is defined by f(x) = x + 1.
(b) f: $\mathbb{Z} \to \mathbb{Z}$ is defined by $f(x) = x^2$.
(c) f: $\mathbb{Z} \to \mathbb{Z}$ is defined by f(x) = 3 – x.
(d) f: $\mathbb{Z} \to \mathbb{Z}$ is defined by f(x) = 4x + 5.

3. Consider an identity function i_C: C → C.

(a) Prove that i_C: C → C is bijective.
(b) Prove that i_C is an equivalence relation on C.

7.5 In summary of the study unit

In this study unit you ensured that you can answer the following questions regarding relations and functions:

- When is a function surjective?
- When is a function injective?
- When is a function invertible?
- How do we map the composition of functions?
- How do we find the inverse of a function?
- What does the term bijective mean?
- How is the identity function defined?

In the following study unit we will learn more about some special operations and their properties.

Study unit 8 Operations

Key questions for this study unit

- *What is a "binary operation"?*
- *How do we know whether an operation is commutative?*
- *How do we add two vectors? How do we add two matrices?*
- *How do we multiply a vector by a scalar?*
- *How do we multiply a matrix by a scalar?*
- *How do we determine the sum and dot product of two vectors?*
- *How do we determine the sum and product of two matrices?*

Activity 8-1: Overview *Study skill*

Draw a mind map of the different sections/headings that you will deal with in this study session. Then page through the study unit with the purpose of completing the map.

Your map should include the concepts of finite and infinite sets, binary operations, vectors, scalars, the vector sum and dot product of two vectors, matrices, matrix addition and multiplication.

Activity 8-2: Concepts *Conceptual skill*

Test your own knowledge and then correct your understanding afterwards. How does your understanding deepen as you jot down the terms used in your home language?

English term	Description	Term in your home language
Finite set		
Infinite set		
A commutative binary operation		
An associative binary operation		
An identity element of a binary operation		
Vector		
Vector sum		
Vector / scalar product		
Dot product		
Matrix		
Matrix addition		
Matrix multiplication		

8.1 Binary operations

Suppose we have a universal set U consisting of objects, variables and numbers. We can take two subsets of U, say A and B, and combine them to form a new subset, namely $A \cup B$.

This is like having a function which takes the ordered pair (A, B) as input and delivers $A \cup B$ as output. What is the domain of this function? Well, if it eats ordered pairs of subsets of U, then the domain must be the Cartesian product $\mathcal{P}(U) \times \mathcal{P}(U)$. And since the function spits out subsets of U, a reasonable choice for the codomain would be $\mathcal{P}(U)$. So if we give the function the name f, we have the function

f: $\mathcal{P}(U) \times \mathcal{P}(U) \rightarrow \mathcal{P}(U)$ defined by $f(A, B) = A \cup B$ for all $A, B \subseteq U$.

Example

Consider some universal set U = {a}, then $\mathcal{P}(U) = \{ \emptyset, \{a\} \}$ and

$\mathcal{P}(U) \times \mathcal{P}(U) = \{ (\emptyset, \emptyset), (\emptyset, \{a\}), (\{a\}, \emptyset), (\{a\}, \{a\}) \}$.

Suppose we want to determine the function g: $\mathcal{P}(U) \times \mathcal{P}(U) \rightarrow \mathcal{P}(U)$ defined by $g(A, B) = A \cup B$ for all subsets A and B of U.

If $(\emptyset, \emptyset)$ is the input, then $\emptyset \cup \emptyset = \emptyset$ is the output, if $(\emptyset, \{a\})$ is the input, then $\emptyset \cup \{a\} = \{a\}$ is the output, the input $(\{a\}, \emptyset)$ delivers {a}, and the input $(\{a\}, \{a\})$ delivers {a}.

We can write g in list notation:
$\{ ((\emptyset, \emptyset), \emptyset), ((\emptyset, \{a\}), \{a\}), ((\{a\}, \emptyset), \{a\}), ((\{a\}, \{a\}), \{a\}) \}$

In mathematics, we frequently find functions such as these, which "eat" pairs and "spit out" single objects. So they deserve a special name. We call them binary operations. *The word "binary" (binary meaning two) reminds us that the inputs are ordered pairs.*

Definition: Binary operation

If f: $X \times X \rightarrow X$ then f is called a *binary operation* on X.

Suppose that f is the name of some binary operation. The image of a pair (x, y) can be given either in *prefix notation*, as f(x, y), or in *infix notation*, as x f y. While infix notation probably looks odd to you, the interesting fact is that we normally use it for binary operations such as the addition and multiplication of real numbers.

Let's think about addition. It is a way to combine two real numbers in order to get a third. So addition is a function (that is traditionally called "+") with $\mathbb{R} \times \mathbb{R}$ as its domain and $\mathbb{R}$ as codomain,

i.e. $+: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$.

But instead of writing +(3, 5) = 8, we are quite accustomed to write 3 + 5 = 8.

Similarly, the multiplication "·" of real numbers is a binary operation:

$\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$

and we usually write $\cdot (x, y)$ as $x \cdot y$.

Of course, the example with which we started, i.e. the binary operation involving the formation of unions of sets, is usually written using infix notation, so we write $A \cup B$ rather than $\cup (A, B)$.

The binary operations we have discussed so far have all been of the form $g: A \times A \to A$ for some set A. That is, the codomain and the relevant sets that were used to form the Cartesian product have been equal. This is no coincidence. Binary operations are usually supposed to take two animals of the same species and use them to produce another animal of the same species.

When we have a binary operation of the form $g: A \times A \to A$ we may say A is closed under g, or g is a binary operation on A, to indicate that g spits out elements of A and not things totally different from the animals living in A.

Quite often one is interested in binary operations on a finite set. We first look at the concepts of finite and infinite sets.

Informal definition: Finite and infinite sets

A *finite* set can informally be defined as a set whose cardinality is a non-negative integer.
If a set is not finite (i.e. its cardinality is not a non-negative integer), then it is an *infinite* set.

Examples

We look at some examples to illustrate the concepts of *finite* and *infinite* sets:

Let the following sets be subsets of a universal set $\mathbb{Z}^+$:

$A = \{1, 2, 3, 4\} = \{x \mid 0 < x \leq 4\}$
$B = \{2, 4, 6, 8, \ldots, 16, 18\} = \{y \mid x \in \mathbb{Z}^+, y = 2x \text{ and } y \leq 18\}$
$C = \{1, 4, 9, 16, \ldots\} = \{y^2 \mid y \in \mathbb{Z}^+\}$

Sets such as A and B are *finite* sets. The cardinality of these sets are respectively $|A| = 4$ and $|B| = 9$. (We can count the number of distinct elements in these two sets.)

The cardinality of C cannot be determined hence it is an *infinite* set.

A very popular way to describe binary operations on finite sets is to use a table. The operators "+", "•" and "∧" in the following examples do not refer to addition, multiplication and conjunction as we know it. The entries in the tables are not determined by any logic calculation.

Examples

1. Suppose A = {a, b, c, d}. Then we could define a binary operation (which we shall call "+") on A by providing the following table:

+	**a**	**b**	**c**	**d**
a	a	b	c	d
b	b	c	d	**a**
c	c	d	a	b
d	d	a	b	c

How does one interpret such a table? The idea is that if we want to know what "+" does to a pair such as (b, d) for example, then we look at the row labelled b and the column labelled d, and the entry at the *intersection* tells us what +(b, d) is. In this case +(b, d) = a.

Note: Here the symbol + does not stand for ordinary addition. We use the symbol "+" to denote the function defined by the table. It need not even be the case that a, b, c and d are numbers. When you work with a binary operation, you have to make a deliberate effort to forget that ordinary addition on $\mathbb{R}$ is also called +.

2. Let us define a different binary operation on the set A = {a, b, c, d}. Let's use "•" to represent the new operation, and let us call it the *dot* operation:

•	**a**	**b**	**c**	**d**
a	a	b	c	d
b	b	a	d	c
c	c	d	a	b
d	d	c	b	a

Notation: Just as + is a popular name to give to binary operations, so is •.

Note: Here the symbol • is not being used to denote the ordinary multiplication of real numbers.

3. Consider the binary operation ∇: {T,F} × {T,F} → {T,F} defined by
∇ = { ((T, T), T), ((T, F), T), ((F, T), F), ((F, F), T) }.
(The domain elements are members of the Cartesian product {T, F} × {T, F}.)

In tabular form, this operation is presented as follows:

∇	**T**	**F**
T	T	T
F	F	T

8.2 The properties of binary operations

In this section we investigate properties of binary operations by using the following binary operation:

Define the function *: $\{1, 2\} \times \{1, 2\} \rightarrow \{1, 2\}$ as the binary operation
$\{ ((1, 1), 1), ((1, 2), 2), ((2, 1), 2), ((2, 2), 1) \}$.
(The domain elements are all the members of the Cartesian product $\{1, 2\} \times \{1, 2\}$.)

We can write this operation in tabular format:

*	1	2
1	1	2
2	2	1

Definition: A commutative binary operation

A binary operation $\Diamond$: $X \times X \rightarrow X$ is commutative iff $x \Diamond y = y \Diamond x$ for all $x, y \in X$.

Refer to the table provided in the above example, then it is clear that * is commutative:

1 * 1 = 1 * 1 = 1,
1 * 2 = 2 * 1 = 2, and
2 * 2 = 2 * 2 = 1.

*We can also observe from the table that * is commutative, because there is symmetry about the diagonal from the top left to the bottom right corners of the table.*

Definition: An associative binary operation

The binary operation $\Diamond$: $X \times X \rightarrow X$ is associative iff $(x \Diamond y) \Diamond z = x \Diamond (y \Diamond z)$ for all $x, y, z \in X$.

In the context of the given example we can determine whether * is associative:

(1 * 1) * 1 = 1 * 1 = 1 and 1 * (1 * 1) = 1 * 1 = 1,
(1 * 1) * 2 = 1 * 2 = 2 and 1 * (1 * 2) = 1 * 2 = 2,
(1 * 2) * 1 = 2 * 1 = 2 and 1 * (2 * 1) = 1 * 2 = 2
(1 * 2) * 2 = 2 * 2 = 1 and 1 * (2 * 2) = 1 * 1 = 1,
(2 * 2) * 1 = 1 * 1 = 1 and 2 * (2 * 1) = 2 * 2 = 1,
(2 * 2) * 2 = 1 * 2 = 2 and 2 * (2 * 2) = 2 * 1 = 2,
(2 * 1) * 1 = 2 * 1 = 2 and 2 * (1 * 1) = 2 * 1 = 2, and
(2 * 1) * 2 = 2 * 2 = 1 and 2 * (1 * 2) = 2 * 2 = 1.

This proves that * is associative.

Definition: An identity element of a binary operation

An element *e* of X is *an identity element* in respect of the binary operation $\Diamond$: $X \times X \rightarrow X$ iff $e \Diamond x = x \Diamond e = x$ for all $x \in X$.

In the given example of the binary operation *, **1** is the identity element:

1 * 1 = 1 * **1** = 1 and
1 * 2 = 2 * **1** = 2.

Another example

Suppose we want to construct a binary operation on the set A = {a, b, c}. Construct a table with columns and rows labelled with the elements of A and put the symbol you want to use for the operation in the upper left-hand corner:

□	**a**	**b**	**c**
a			
b			
c			

Now simply fill in the spaces where the rows and columns intersect with elements of A according to taste, for example:

□	**a**	**b**	**c**
a	b	c	a
b	c	a	b
c	a	b	c

The resulting table is a shorthand description of the function □: A × A → A that is such that
a □ a = b (because the entry where the row labelled "a" and the column labelled "a" intersect, is "b"),
a □ b = c (because the entry where the row labelled "a" and the column labelled "b" intersect, is "c"),
a □ c = a (because the entry where the row labelled "a" and the column labelled "c" intersect, is "a"), and so on…

We are going to investigate the properties of □: Is □ commutative, is □ associative and does □ have an identity element?

Commutativity: Is x □ y = y □ x for all x, y ∈A?

As you can see, we need to use a brute force approach in order to determine whether □ is commutative, so we consider all possible cases:

Case	**x □ y**	**y □ x**
x = y = a	a □ a = b	a □ a = b
x = a, y = b	a □ b = c	b □ a = c
x = a, y = c	a □ c = a	c □ a = a
x = b, y = a	b □ a = c	a □ b = c
x = y = b	b □ b = a	b □ b = a
x = b, y = c	b □ c = b	c □ b = b
x = c, y = a	c □ a = a	a □ c = a
x = c, y = b	c □ b = b	b □ c = b
x = y = c	c □ c = c	c □ c = c

In every case, x □ y = y □ x so □: A × A → A is commutative.

By looking at the table which defines the operation, one can also see that □ is commutative: Draw a diagonal line from the top leftmost corner to the bottom rightmost corner of the table. Then it is clear that you have mirror images of the two triangles.

Associativity: Is $(x \square y) \square z = x \square (y \square z)$ for all $x, y, z \in A$?

To verify that $\square: A \times A \rightarrow A$ is associative is a tedious, but easy task. There are 27 cases to be considered. We look at a few:

Case	**$(x \square y) \square z$**	**$x \square (y \square z)$**
x = a, y = a, z = a	$(a \square a) \square a = b \square a = c$	$a \square (a \square a) = a \square b = c$
x = a, y = a, z = b	$(a \square a) \square b = b \square b = a$	$a \square (a \square b) = a \square c = a$
x = a, y = a, z = c	$(a \square a) \square c = b \square c = b$	$a \square (a \square c) = a \square a = b$
x = a, y = b, z = a	$(a \square b) \square a = c \square a = a$	$a \square (b \square a) = a \square c = a$
x = a, y = b, z = b	$(a \square b) \square b = c \square b = b$	$a \square (b \square b) = a \square a = b$, …

And so one can go on to inspect all the combinations of x, y and z to see that $(x \square y) \square z = x \square (y \square z)$ in all the different cases. This means that □ is associative.

Identity element: Is it possible to identify an element *e* in A such that $e \square x = x \square e = x$ for all $x \in A$?

To determine whether or not $\square: A \times A \rightarrow A$ has an identity element, one has to find an element in A that can connect with any x in A via the binary operation □ without changing x.

We first look at the number set $\mathbb{Z}$. In the set $\mathbb{Z}$, 0 is an additive identity (i.e. $m + 0 = m = 0 + m$ for all $m \in \mathbb{Z}$) and 1 is a multiplicative identity (i.e. $m \cdot 1 = m = 1 \cdot m$ for all $m \in \mathbb{Z}$). So 0 is the identity element with respect to $+: \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ and 1 is the identity element with respect to $\cdot : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$.

Back to the function $\square: A \times A \rightarrow A$: Does the set A = {a, b, c} have an element that can successfully play the role of an identity element with regard to □?

Well, if we inspect the table describing □, we can see similar orderings of the variables (a, b and c) in the *top row* and the *row labelled* ***c***:

□	**a**	**b**	**c**
a	b	c	a
b	c	a	b
c	**a**	**b**	**c**

We can also see similar orderings of the variables in the *left-most column* and the *column labelled* ***c***:

□	a	b	**c**
a	b	c	**a**
b	c	a	**b**
c	a	b	**c**

These similar orderings are an indication that **c** is the identity element.

We can now confirm that c is indeed the identity element:

$c \square a = a \square c = a,$
$c \square b = b \square c = b,$ and
$c \square c = c \square c = c.$

So c acts as an identity element with regard to □.

Activity 8-3: Self-assessment exercises ***Application skills***

1. Let X be {2, 7}.

   (a) Provide 3 binary operations on X, both in list notation and in tabular form.
   (b) Check the 3 operations for commutativity and associativity.

2. Provide 2 binary operations on X = {a, b, c} and check them for commutativity and associativity.

3. Consider the dot operation, "•", defined in Section 8.1. Let us compare the dot operation on A = {a, b, c, d} with ordinary multiplication.

   (a) We know that ordinary multiplication on $\mathbb{R}$ is commutative. Now examine x•y and y•x for each x, y ∈ A. Is • commutative?
   (b) We know that $\mathbb{R}$ has an identity element for multiplication, namely 1.
   This means that $1 \cdot x = x = x \cdot 1$ for all $x \in \mathbb{R}$. Does A have an element that behaves similarly for •?

Activity 8-4: Reflection on binary operations

Think carefully about the following statement: Does it bother you to see old familiar symbols such as + and · being used as names for unfamiliar things?

We have discussed only binary operations so far, i.e. functions of the form

$f: A \times A \to A.$

But the word "operation" is also applicable to functions of the form

$f: A \to A$ *(a unary operation)*
$g: A \times A \times A \to A$ *(a 3-ary, or ternary, operation)*
$h: A \times A \times A \times A \to A$ *(a 4-ary, or quaternary, operation)*

and so on.

8.3 Operations on vectors

The word *vector*, for our purposes, will be understood to mean an ordered n-tuple of numbers.

Definition: Vector

A *vector* is represented by an n-tuple in the following way:

$u = (u_1, u_2, ..., u_n)$ for some $n \geq 2$.

You have already encountered such specimens in study unit 5. In that unit, our discussion of *n-ary* relations should have left you with the feeling that n-tuples (or vectors as we call them in this unit) can be used to *represent information*. What is new in this present discussion is the idea that there are useful ways to combine old vectors in order to get new vectors.

Example

At the Benoni Institute of Technology, Professor Thaddeus Twiddle teaches a course in Creative Television Repair. He has 43 students in his class.

During the year each student earns a year mark (out of a possible total of 100). If the students are ordered alphabetically, the year marks can be represented by a vector

$a = (a_1, a_2, ..., a_{43})$,

where co-ordinate a_i is the year mark of the *i*-th student.

In the final exam, each student earns an exam mark (again out of a possible maximum of 100). The exam marks can be represented by a vector

$b = (b_1, b_2, ..., b_{43})$.

To determine whether a student passes or fails, Professor Twiddle must combine the year mark and exam mark and express the result as a percentage. If the percentage is higher than 79%, the student passes. (This is an imaginary university, so we can afford to have an ideal pass mark.)

Activity 8-5: Vector arithmetic

How should Professor Twiddle work out the percentage that each student obtains? He can do it for each student separately, or he can use the vectors a and b.

If there were some kind of addition on vectors, he could first use it to combine a and b. Then, if there were some sort of multiplication available, he could multiply a + b by some factor to get a vector whose co-ordinates are the percentages.

Suppose we combine a and b by *adding corresponding co-ordinates*. Then we get the following:

$a + b = (a_1 + b_1, a_2 + b_2, ..., a_{43} + b_{43})$.

Now each co-ordinate is a mark out of a possible total of 200. So to change these marks to percentages, all we need to do is to multiply each of them by ½.

So, what we can do, is to define the product of the number ½ and the vector (a + b) to be the new vector

$½ (a + b) = (c_1, c_2, ..., c_{43})$

where each c_i is $½ (a_i + b_i)$.

The example suggests the following definitions:

Definition: Vector sum

If u and v are vectors with the *same number of co-ordinates,* then their *sum, denoted by u + v,* is the vector obtained by adding the corresponding co-ordinates of u and v, i.e.

$u + v = (u_1, u_2, ..., u_n) + (v_1, v_2, ..., v_n) = (u_1 + v_1, u_2 + v_2, ..., u_n + v_n)$.

Definition: Scalar-vector product

If u is a vector and r is some number (scalar), then *the product* of the number r and the vector u is the vector r·u obtained by multiplying each co-ordinate of u by r, i.e.

$r \cdot u = r(u_1, u_2, ..., u_n) = (ru_1, ru_2, ..., ru_n)$.

Important terminology: The number r is often called a scalar*, and the operation defined above is then referred to as the multiplication of a vector by a scalar.*

Activity 8-6: Self-assessment exercise *Application skills*

If u = (3, 1), v = (−4, −4), and w = (0, −1), determine

(a) 2u + v
(b) u − 3v
(c) −3(v + w)

So far we have defined a kind of addition for vectors, and a weird kind of multiplication which combines vectors and things that aren't vectors (i.e. scalars) to produce a new vector.

There is another kind of multiplication which is useful. It is called the dot product, and it combines two vectors *to obtain a* scalar *as the answer.*

Definition: Dot product

The *dot product* of vectors $u = (u_1, u_2, ..., u_n)$ and $v = (v_1, v_2, ..., v_n)$ is denoted by u·v and defined by $u \cdot v = u_1v_1 + u_2v_2 + ... + u_nv_n$.

Note : This operation is sometimes called the *inner product.*

The above definition means that we multiply corresponding co-ordinates together. This is no problem for us, because the co-ordinates are real numbers and we can use ordinary multiplication. Then we add the results together (i.e. $u_1v_1 + u_2v_2 + ... + u_nv_n$), and this constitutes no problem because the results are real numbers and we can use ordinary addition.

Of what use is the dot product? Well, suppose you want to buy groceries: a_1 tins of beans, a_2 tins of peas, and so on. The quantities you need to buy can be

represented by the vector $a = (a_1, a_2, ..., a_n)$. Now suppose that each tin of beans costs b_1 cents, each tin of peas costs b_2 cents, and so on.

The costs of the various items can be represented by the vector

$b = (b_1, b_2, ..., b_n)$.

The total amount you spend can now be determined by taking the dot product of a and b, since

$a \cdot b = a_1b_1 + a_2b_2 + ... + a_nb_n$.

Activity 8-7: Self-assessment exercise ***Application skills***

If $u = (1, 2, 5)$ and $v = (2, 3, 5)$, determine

(a) $u \cdot v$
(b) $v(2u)$

8.4 Operations on matrices

In the previous section we saw that vectors can be used to represent information. Sometimes the information you're working with can best be represented in *tabular form*, i.e. using tables.

Example

Suppose you are going shopping. You want to pop in at the bakery to buy 3 loaves of whole wheat bread and 1 fruit cake. Then you want to go to the local grocer to get 2 of their whole wheat loaves and 5 of their cream cakes. One can display this information in a table as follows:

	Bakery	**Grocer**
loaves	3	2
cakes	1	5

If one rewrites the table in a streamlined form, omitting the labels, one gets:

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

We call this a *matrix*, and the numbers inside the square brackets are called the *entries*. Of course, one can replace the original table with the streamlined version only if the context makes it clear what the various rows and columns represent. In the example above, we need to agree that the first row will deal with loaves, and the second row with cakes, while the first column will deal with the bakery, and the second column with the grocer.

Remember that the term "matrix" carries the connotation that the positions of entries are important. The plural of the word "matrix" is "matrices", pronounced "maytrisseez" and not "matresses".

A matrix, then, is *an array of numbers organised into rows and columns and enclosed within brackets*. If there are m number of rows and n number of columns, we speak of an m × n matrix (an "em-by-en" matrix, when you read it out loud).

For example: $\begin{bmatrix} 3 & 2 \\ 1 & 5 \end{bmatrix}$ is a 2 × 2 matrix,

whereas $\begin{bmatrix} -1 & 3 & 0 & 5 \\ 0 & 2 & 0 & 6 \\ 1 & -1 & 0 & 13 \end{bmatrix}$ is a 3 × 4 matrix.

In general, matrices have the form

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

where a_{ij} is the entry in the *i*-th row and *j*-th column of the above m × n matrix. We can define a kind of addition for matrices which is pretty useful. Let's look at an example.

Suppose our old friend, the matrix $\begin{bmatrix} 3 & 2 \\ 1 & 5 \end{bmatrix}$

represents our shopping plans for Monday, and that the following matrix represents our shopping list for a later date:

$$\begin{bmatrix} 4 & 1 \\ 4 & 2 \end{bmatrix}$$

That is, the first matrix (let's call it A) represents the table

	Bakery	**Grocer**
loaves	3	2
cakes	1	5

while the second matrix (let's call it B) is a streamlined version of the table

	Bakery	**Grocer**
loaves	4	1
cakes	4	2

Then the matrix A + B = $\begin{bmatrix} 7 & 3 \\ 5 & 7 \end{bmatrix}$, which is obtained by adding *corresponding entries,* represents the total purchase for the two occasions.

The previous example suggests the following definition:

Definition: Matrix addition

Let A = $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ . & . & & . \\ . & . & & . \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ and B = $\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ . & . & & . \\ . & . & & . \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$

then the matrix A + B is provided by

$$C = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \quad . & \quad . & & \quad . \\ \quad . & \quad . & & \quad . \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

One can only add matrices if they are of the same "size".

Activity 8-8: Self-assessment exercises ***Application skills***

For each pair A and B given below, determine A + B (if possible):

(a) A = $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ B = $\begin{bmatrix} 5 & 5 \\ 4 & -1 \end{bmatrix}$

(b) A = $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ B = $\begin{bmatrix} 3 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

(c) A = $\begin{bmatrix} 2 & 0 & 3 \\ 0 & 7 & 1 \end{bmatrix}$ B = $\begin{bmatrix} 1 & 1 & -2 \\ 2 & 0 & 6 \end{bmatrix}$

(d) A = $\begin{bmatrix} 3 & 1 \\ -2 & -5 \end{bmatrix}$ B = $\begin{bmatrix} 1 & 2 & 0 \\ 2 & 5 & 1 \end{bmatrix}$

Example

One can also *multiply a matrix by a number to get a new matrix*. For instance, suppose our old friend

A = $\begin{bmatrix} 3 & 2 \\ 1 & 5 \end{bmatrix}$

represents our shopping list for a certain day and we discover that twice as many guests than expected will be visiting. Since the demand for groceries has doubled, we must use a revised shopping list in which each entry has been multiplied by two:

2A = $\begin{bmatrix} 6 & 4 \\ 2 & 10 \end{bmatrix}$

This example leads us to the next definition.

Definition: Matrix multiplication

Given a real number r and a matrix

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

the matrix rA is provided by

$$rA = \begin{bmatrix} ra_{11} & ra_{12} & \cdots & ra_{1n} \\ ra_{21} & ra_{22} & \cdots & ra_{2n} \\ . & . & & . \\ . & . & & . \\ ra_{m1} & ra_{m2} & \cdots & ra_{mn} \end{bmatrix}$$

Activity 8-9: Self-assessment exercise ***Application skills***

Perform the indicated operation:

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

At this stage you might well ask "Is it possible to multiply matrices"? There is actually a special kind of multiplication of matrices that resembles the dot product of vectors, except that in the end we get a matrix, not a real number.

To illustrate matrix multiplication, we'll start with a simple example that shows how similar the dot product of vectors and matrix multiplication really is.

Example

Let A be the matrix [3 ½].

Since A has only a single row, it is common to refer to A as a *row matrix*.

Let B be the column matrix $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$.

Now AB = [3 ½]$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$

is calculated by

(i) *multiplying appropriate entries* of A and B, and
(ii) *adding the results together*.

AB = [3·1 + (½)·4] = [3 + 2] = [5].

Note that the answer is a matrix and therefore has square brackets around it. Of course, a 1 × 1 matrix such as [5] is nothing other than a number, but we will shortly see examples in which the answer is a bigger matrix.

First, let us remind ourselves why this sort of multiplication is useful. (At the same time, we're really reminding ourselves why the dot product of vectors is useful.)

Let's get back to the bakery problem:

The local bakery produces three items, namely brown bread, white bread, and raisin bread. The prices of these items are R7, R9, and R12 per loaf, respectively. In a certain week the bakery sells 3000 loaves of brown bread, 4800 loaves of white bread and 937 loaves of raisin bread.

Now the *total revenue* (i.e. money the bakery gets) is
(7)(3000) + (9)(4800) + (12)(937) Rands.

This can be represented as the product AB of the matrix

$A = [7 \;\; 9 \;\; 12]$

which displays the prices of the items, and the matrix

$$B = \begin{bmatrix} 3000 \\ 4800 \\ 937 \end{bmatrix}$$

which represents the sales of the bakery.

$$AB = [7 \;\; 9 \;\; 12]\begin{bmatrix} 3000 \\ 4800 \\ 937 \end{bmatrix}$$

$= [(7)(3\,000) + (9)(4\,800) + (12)(937)]$
$= [21\,000 + 43\,200 + 11\,244]$
$= [75\,444]$

Note that the multiplication must match brown bread with brown bread, white bread with white bread, and raisin bread with raisin bread.

Well, how do we multiply more general matrices? We'll illustrate the procedure by working out a typical product, namely

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

To obtain the entries of the product, we multiply the rows of the left matrix by the columns of the right matrix, taking care to arrange the products in a specific way to yield a matrix. Start with the first (top) row on the left, $[2 \;\; 1]$, and the first column on the right, $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$.

Their product is [2·1 + 1·4] = [6], so we enter 6 as the entry a_{11} of the product:

$$\begin{bmatrix} \mathbf{2} & \mathbf{1} \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{1} & 1 \\ \mathbf{4} & 2 \end{bmatrix} = \begin{bmatrix} 6 & \\ & \\ & \end{bmatrix}$$

Next, we obtain the product of the first row on the left and the second column on the right, which is [2·1 + 1·2] = [4], so in the resulting product matrix we have $a_{12} = 4$:

$$\begin{bmatrix} \mathbf{2} & \mathbf{1} \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{1} \\ 4 & \mathbf{2} \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ & \\ & \end{bmatrix}$$

Now there are no more columns that can be multiplied by the first row, so we move down to the second row and start the same process again using the first column of the right-hand matrix:

$$\begin{bmatrix} 2 & 1 \\ \mathbf{0} & \mathbf{1} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{1} & 1 \\ \mathbf{4} & 2 \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 4 & \\ & \end{bmatrix}$$

and for the second column:

$$\begin{bmatrix} 2 & 1 \\ \mathbf{0} & \mathbf{1} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{1} \\ 4 & \mathbf{2} \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 4 & 2 \\ & \end{bmatrix}$$

We have now exhausted the second row on the left, so we shift our attention to the third row. Of course, we also move down a row in the resulting product:

Third row, first column:

$$\begin{bmatrix} 2 & 1 \\ 0 & 1 \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{1} & 1 \\ \mathbf{4} & 2 \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 4 & 2 \\ 1 & \end{bmatrix}$$

Third row, second column:

$$\begin{bmatrix} 2 & 1 \\ 0 & 1 \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{1} \\ 4 & \mathbf{2} \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 4 & 2 \\ 1 & 1 \end{bmatrix}$$

Now we have multiplied every row of the left matrix by every column of the right matrix, so we can stop.

Our product is $\begin{bmatrix} 6 & 4 \\ 4 & 2 \\ 1 & 1 \end{bmatrix}$.

Let's do another example: We have to calculate

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

Determine the entries of the resulting matrix:

a_{11}: $\begin{bmatrix} 1 & 5 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 6 & \\ & \end{bmatrix}$

a_{12}: $\begin{bmatrix} 1 & 5 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ & \end{bmatrix}$

a_{21}: $\begin{bmatrix} 1 & 5 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ -1 & \end{bmatrix}$

a_{22}: $\begin{bmatrix} 1 & 5 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ -1 & -6 \end{bmatrix}$

which is our answer.

Note that this method cannot always be used to compute AB. For the procedure to work, a row in A must have as many entries as a column in B. So if we want to multiply A by B, the sizes of A and of B must match up in a special way. We can form AB if A is m × n and B is n × k. The product is an m × k matrix.

Schematically:

A	·	*B*	=	*C*
m × *n*		*n* × *k*		*m* × *k*

(equal)

What would a general definition of matrix multiplication look like? Well, if

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ . & . & & . \\ . & . & & . \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \text{ and}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ . & . & & . \\ . & . & & . \\ b_{n1} & b_{n2} & \cdots & b_{nk} \end{bmatrix}$$

then AB is the matrix C in which C_{ij} is the dot product of the *i*-th row in A with the *j*-th column in B, i.e.

$$c_{ij} = (a_{i1} \cdot b_{1j}) + (a_{i2} \cdot b_{2j}) + \ldots + (a_{in} \cdot b_{nj})$$

Example

As a last example before you tackle some exercises, let's look at the 2 x 2 zero matrix. This is a matrix of which all the entries are zero:

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

What is the product of this zero matrix and the matrix $\begin{bmatrix} 2 & -3 \\ 1 & 5 \end{bmatrix}$?

Work it out step by step and test this for yourself. For instance, the product of the first row of the zero matrix with the first column of the other matrix is (0)(2) + (0)(1) = 0.

Can you see that the product of a zero matrix with any other appropriately sized matrix will always be equal to a zero matrix?

Definition: Identity matrix

If A is a matrix, then an *identity matrix* I with respect to A is a matrix such that IA = AI = A.

This means that if we have an n × n matrix, and we calculate IA, the result is A. Similarly, if we calculate AI, we get A. (I is also a n × n matrix.)

Example

Let A be the 2 × 2 matrix

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

We need to find an identity matrix I such that IA = AI = A.

Let's try $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Calculating AI and IA gives A, so I is an identity matrix.

Activity 8-10: Self-assessment exercises

Perform the indicated matrix operations (if possible):

1.

$$\begin{bmatrix} 31 & -3 & 2 \\ 2 & 5 & 1 \\ 3 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 5 \end{bmatrix}$$

2.

$$\begin{bmatrix} 9 & 3 \\ 1 & 5 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 4 \\ 5 & 1 \end{bmatrix}$$

3.

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

4. Provide examples of matrices X and Y such that XY is a 3×3 matrix, but YX is a 2×2 matrix.

5. Provide examples of matrices X and Y such that both X and Y have at least some nonzero entries, but XY is the 2×2 zero matrix,

   i. e. $XY = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

6. Prove that addition is a commutative operation on the set of 2×2 matrices and that there is a 2×2 matrix that acts as an identity element in respect of *addition*.

7. Prove that multiplication is *not* a commutative operation on the set of 2×2 matrices, and that there is a 2×2 matrix that acts as an identity element in respect of multiplication.

8.5 In summary of the study unit

In this study unit you ensured that you can answer the following questions regarding operations in general, and operations on vectors and matrices in particular.

- What is a binary operation?
- What is a vector?
- What is a scalar?
- How do we add two vectors?
- How do we perform vector multiplication?
- What does the term matrix mean?
- How do we perform matrix addition and matrix multiplication?

In the following study unit we will learn more about logic and how truth tables are used in this field.

NOTES

Study unit 9 Logic: Truth tables

Key questions for this study unit

- *What does the following terms mean: "connectives", "a simple declarative statement (or a proposition)", "truth tables", "compound statements", "disjunction", "conjunction", "negation" and "biconditional"?*
- *What is meant by the term "logical equivalence"?*
- *How does one construct a truth table?*
- *How does one determine whether or not a given statement is a tautology, a contradiction, or neither of the two?*

Activity 9-1: Overview *Study skill*

Draw a mind map of the different sections/headings you will deal with in this study session. Then page through the unit with the purpose of completing the map.

Your map should include the concepts of a simple declarative statement, a compound statement, truth tables, connectives, the conditional, the biconditional, conjunction, disjunction, negation, tautology, contradiction and logical equivalence.

Activity 9-2: Concepts *Conceptual skill*

Test your own knowledge and then correct your understanding afterwards. How does your understanding deepen as you jot down the terms used in your home language?

English term	Description	Term in your home language
A simple declarative statement (or a proposition)		
A compound statement		
Connective		
Conjunction		
Disjunction		
Conditional		
Biconditional		
Negation		
Truth tables		
Tautology		
Contradiction		
Logical equivalence		

9.1 Statements and connectives

Why do mathematicians and logicians spend so much time talking about proofs?

Well, like all scientists they're interested in discovering, as best they can, the truth about things. But it is not good enough for them to establish the facts to their own satisfaction; they also have to establish the facts to the satisfaction of everybody else. So whenever a logician (or computer scientist!) makes a claim, she/he has to be willing to justify it by presenting a convincing argument in support of her/his claim. That is to say, *he/she needs to provide a proof.*

The purpose of a proof is to convey information. One conveys information by making what is called a declarative statement (also called a proposition in some literature). (*Declarative* is a word that comes from the verb *to declare*.)

Some examples of simple declarative statements (or propositions):

"The capital of France is Paris."
"3 is an even integer."
"5 + 3 > b"
"This sentence is false."

Some examples of statements which are not declarative:

"Is 3 an even integer?" (A question, asked in order to acquire information, not to convey it.)
"Add 3 to 5!" (A command, given in order to induce certain behaviour, not to convey information.)
"Inconceivable!" (An exclamation, uttered in order to give vent to some emotion, not to convey information.)

Clearly, in writing proofs we have to restrict ourselves to declarative statements. But not all declarative statements are usable. For instance, there is something peculiar about the declarative statement

"This sentence is false."

When we use a declarative statement in a proof, our purpose is to convey information. The information either gives an accurate picture of the facts (i.e. the statement is true) or it does not (i.e. the statement is false).

Examples

"3 is an even integer" is false, because dividing 3 by 2 leaves a remainder of 1,

but the statement

"The capital of France is Paris" is true, because Paris really is the capital of France.

However the isolated statement "This sentence is false" has no truth value, i.e. is neither true nor false.

To see this, suppose the statement was true. If it is true, its description of the facts is accurate. So the sentence is false. Similarly, if we begin by assuming the statement to be false, we can reason quite logically to reach the conclusion that the sentence is true. But since it makes no sense to say of a statement that it is both true and false, we conclude that this particular statement is neither true nor false.

In proofs, we only use declarative statements that have a truth value of either *true* or *false*.

Such declarative statements are of two kinds:

- The examples we've seen so far are *simple (or atomic) statements*, conveying just one (true or false) fact.
- Other statements are *compound*; they are built up by connecting simple statements.

Example

" $\sqrt{2} > 1$ and $\sqrt{2} < 2$ " is a compound declarative statement which is clearly either true or false (and, in fact, is true). It is a compound statement, because it is built up by *connecting* the two simple statements " $\sqrt{2} > 1$ " and "$\sqrt{2} < 2$ " with the word "and".

An interesting thing is that the compound statement

"$\sqrt{2} > 1$ and $\sqrt{2} < 2$"

is true precisely because *both* of the simple statements from which it is built up are true.

Activity 9-3: The truth value of compound statements

How would we check whether or not a given compound statement is true?

Let's take the statement "$\sqrt{2} > 1$ and $\sqrt{2} < 2$ " as an example:

We would first investigate whether or not $\sqrt{2} > 1$ by some argument such as "$\sqrt{2}^2 = 2$, but $1^2 = 1$, so $\sqrt{2}$ must be greater than 1". Then we would check that $\sqrt{2} < 2$, by a similar process of reasoning.

So the truth value of a compound statement is determined by *the truth values of its component statements*.

How is this done? We want to examine the truth values of compound statements. To begin with, we can list all the *logical connectives* that we are allowed to use in order to build compound statements, together with the symbols by which they are abbreviated, and their official names:

and	$\wedge$	conjunction
or	$\vee$	disjunction
if ..., then ...	$\rightarrow$	the conditional (also referred to as *implication*)
if and only if	$\leftrightarrow$	the biconditional
it is not the case that	$\neg$	negation

Definition: Conjunction

If p and q represent statements, then $p \wedge q$ represents the statement "p and q", and is called the *conjunction* of p and q.

Example

Let p represent "All rational numbers are repeating decimals" and let q represent "All irrational numbers are non-repeating decimals". Then $p \wedge q$ represents "All rational numbers are repeating decimals *and* all irrational numbers are non-repeating decimals".

Conjunction: *Let us analyse the way in which the truth values of the component statements p and q determine the truth value of the compound statement $p \wedge q$. The following example will assist us.*

(a) Grass is green and 1 + 1 = 2.
(b) Grass is green and 1 + 1 = 3.
(c) Grass is blue and 1 + 1 = 2.
(d) Grass is blue and 1 + 1 = 3.

The first compound statement is true (T) since both component statements are true. Each of the compound statements (b) to (d) is false (F), because each has at least one component statement that is false.

This leads us to summarise in tabular form the rule that $p \wedge q$ is true if p and q are both true, but false otherwise.

The *conjunction* of two statements is reflected in the following truth table:

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

It is possible that p and/or q could represent compound statements.

Let p represent a compound declarative statement $r \wedge s$. This means that p is only true when both r and s are true. As a result $p \wedge q$ is only true when r, s and q are true.

Later in this study unit we will see how a truth table for more than two statements can be compiled.

In English there are different ways in which to say the same thing. For instance, the statement "Rationals are repeating decimals and irrationals are non-repeating decimals" conveys the same information as "Rationals are repeating decimals, but irrationals are non-repeating decimals". Perhaps "but" also gives a feeling of opposing things being compared, but (!) for the purposes of logic we ignore vague statements (or statements that are ambiguous) and formalise statements of the form "p but q" as "p and q", that is, "$p \wedge q$".

Definition: Disjunction

If p and q represent statements, then $p \vee q$ represents the statement "p or q", and is called the *disjunction* of p and q.

Example

Let p represent "113 divides 17 304 without a remainder" and let q represent "113 leaves a remainder of 2 when divided into 17 304". Then $p \vee q$ represents "113 divides 17 304 without a remainder *or* 113 leaves a remainder of 2 when divided into 17 304".

Disjunction: *The basic idea behind disjunctions is that we want $p \vee q$ to stand for "either p, or q, or both". For reasons of brevity we write "either ... or ... or both" simply as "or". Now consider the example below.*

(a) Grass is green or 1 + 1 = 2.
(b) Grass is green or 1 + 1 = 3.
(c) Grass is blue or 1 + 1 = 2.
(d) Grass is blue or 1 + 1 = 3.

The first three compound statements are true, because in each case either the first or the second or both component statements are true. The last compound statement is false, because both component statements are false.

We summarise in tabular form the rule that $p \vee q$ is true if at least one of p and q is true – otherwise it is false.

The *disjunction* of two statements is presented in the following truth table:

p	**q**	**$p \vee q$**
T	T	T
T	F	T
F	T	T
F	F	F

The word "or" is used in two different ways in an ordinary English conversation. Sometimes it means "either ... or ... but not both", which is called the exclusive sense of "or".

Sometimes it means "either ... or ... or both", which is called the inclusive sense of "or". We use only the ***inclusive "or" when we write "$\vee$"****. The reason is that we shall be able to express statements of the form "p or q, but not both" in terms of conjunctions, disjunctions, and negations.*

Definition: Conditional

If p and q represent statements, then $p \rightarrow q$ represents the statement "If p, then q", and we may describe $p \rightarrow q$ as a *conditional statement* with *hypothesis* p and *conclusion* q.

Example

Let p represent "$\sqrt{2}$ has a non-repeating decimal expansion", and let q represent "$\sqrt{2}$ is irrational". Then $p \rightarrow q$ represents "If $\sqrt{2}$ has a non-repeating decimal expansion, then $\sqrt{2}$ is irrational".

A conditional: *The essential idea of a conditional statement is that it represents a sort of promise. Suppose a father says to his daughter: "If we go to town today, then I will buy you an ice-cream". This promise will be broken only if they do go to town that day, but the father fails to buy an ice-cream for his daughter. If something has cropped up to prevent them from going to town, we cannot say that the promise was false. So, thinking of the promise as p → q, we get the following truth table:*

A *conditional* statement is presented in the following truth table:

p	**q**	**p → q**
T	T	T
T	F	F
F	T	T
F	F	T

We see that the first row of the truth table corresponds to the case in which they do go to town (since p is true) and she does get her ice-cream (q is true). Then the promise is fulfilled, so p → q ought to be true.

The second row corresponds to the case in which they do go to town (p is true), but she does not get her ice-cream (q is false). The promise has been broken, so p → q should be false.

The third and fourth rows both correspond to situations in which they do not go to town. Whatever else may happen, i.e. whether or not the daughter gets an ice-cream, the father did not break his promise and so p → q ought not to be false, i.e. it must be regarded as being true.

In p → q, ***p*** *is often referred to as the* ***antecedent*** *(which means "an event that happens before another") and* ***q*** *as the* ***consequent****.*

Conditional statements are also called implications.

There are many ways to express p → q in English. Some are listed below. Remember, each statement below really means "If p, then q", i.e. p → q.

- *p implies q*
- *p only if q*
- *q is a necessary condition for p*
- *p is a sufficient condition for q*
- *q if p*
- *q provided that p*
- *q whenever p*

You might require a slightly more mathematical motivation for the truth table of $p \rightarrow q$. Think of it this way: If q is true, then q is true regardless of whether anything else is true or false. That is, "If p then 1 = 1" is true for any statement p, whether p is true or false, because 1 = 1. So we give the conditional $\rightarrow$ the value T in rows one and three in its truth table.

As for row four, which involves the falsity of both antecedent and consequent, consider a statement such as "If 1 = 2 then 3 = 4". This statement deserves to be regarded as true, because if we assume that 1 = 2, then we can prove that 3 = 4:

Assume 1 = 2.
Add 2 to each side of the equation then 3 = 4.

Definition: Biconditional

If p and q represent statements, then $p \leftrightarrow q$ represents the statement "p if and only if q" which, as we saw in earlier study units, can be written as "p iff q". This is referred to as *the biconditional.*

Example

Let p represent "$\sqrt{13}$ + 1 has a non-repeating decimal expansion" and q represent "$\sqrt{13}$+ 1 is irrational". Then $p \leftrightarrow q$ represents "$\sqrt{13}$ + 1 has a non-repeating decimal expansion iff $\sqrt{13}$ + 1 is irrational".

Biconditional: *What is the basic idea behind biconditional statements? Well, as a first approximation think of $p \leftrightarrow q$ as p and q saying the same thing in different words. Then clearly the only circumstance under which $p \leftrightarrow q$ really must be false is when p and q have opposite truth values. The truth table below summarises this idea:*

A *biconditional* statement is presented in the following truth table:

p	**q**	**$p \leftrightarrow q$**
T	T	T
T	F	F
F	T	F
F	F	T

There are many ways to express $p \leftrightarrow q$ in English. Remember that each of the following really means "p iff q".

- *p is a necessary and sufficient condition for q*
- *p implies q and conversely*
- *if p then q, and conversely*
- *p implies q and q implies p*
- *if p then q, and if q then p*
- *p is equivalent to q*

Definition: Negation

If p represents some statement, then ¬p represents the statement "It is not the case that p", or more briefly "not p". This is called the *negation* of a given statement.

Example

Let p represent "$\sqrt{8}$ is irrational". Then ¬ p represents "*It is not the case* that $\sqrt{8}$ is irrational" which could be rewritten as "$\sqrt{8}$ is not irrational".

Clearly the statements p and ¬ p must always have opposite truth values. This leads to the following truth table:

p	**¬ p**
T	F
F	T

Negation is different from the other connectives (such as conjunction) because it doesn't actually connect two statements. However, it is an important way to build compound statements from simple statements.

English can be confusing where negations are involved. For instance, the negation of "I like Brazilian jazz" is "It is not the case that I like Brazilian jazz". Usually we say this more briefly as "I do not like Brazilian jazz". The danger is that we might think that this is the same as "I dislike Brazilian jazz". Of course it isn't. Someone who has no particular preference in this regard may feel completely neutral about Brazilian jazz music and may say "I do not like Brazilian jazz" without meaning "I dislike Brazilian jazz". It is to avoid this sort of confusion that we make a point of putting the "not" in front of the statement, at least until we feel confident. After all, it is not quite so easy to slide from "It is not the case that I like Brazilian jazz" to "I dislike Brazilian jazz".

What have we done so far? We have constructed the fundamental truth tables for conjunctions, disjunctions, conditionals, biconditionals, and negations.

This enables us to take any compound statement (which is, after all, built up from simple statements by means of the connectives) and construct a truth table for it. Such a truth table then displays the way in which the truth values of the simple statements determine the truth value of the whole statement.

Convention: *In the following discussion, lower-case letters of the alphabet such as p, q, and r will be used to denote simple statements. The order in which a compound statement is built up from simple statements will usually be indicated by brackets, except that we reduce the number of brackets required by agreeing that it applies to the shortest statement following it. So we may omit the brackets in* $p \wedge (\neg q)$ *and write it as* $p \wedge \neg q$ *. However, in the case of* $p \wedge \neg (q \vee r)$ *we may not drop the brackets, since* ¬ *applies to* $q \vee r$*, not just q.*

Before we discuss the truth table procedure in general, consider the following two illustrative examples.

Example

We construct a truth table for $p \wedge (\neg q)$ in a number of steps.

Step 1: List the simple statements p, q across the top:

p	**q**	

Step 2: Each simple statement has one of two possible truth values – either T or F. So there are four possible *combinations* of truth values to be entered in successive rows:

p	**q**	
T	T	
T	F	
F	T	
F	F	

Step 3: In building up the compound statement p ∧ (¬ q) from the simple statements p and q, one first forms ¬ q. (After all, expressions in brackets come first.) So we list ¬ q at the top of the third column and enter the truth values by looking at the q column:

p	**q**	**¬ q**	
T	T	F	
T	F	T	
F	T	F	
F	F	T	

(Here we have used our knowledge of the truth table for negation, which tells us that when q is T then ¬ q is F and vice versa.)

Step 4: Finally, to get p ∧ (¬ q) one combines the component statements p and ¬ q by conjunction. List p ∧ (¬ q) at the top of the fourth column and enter the truth values by using our knowledge of conjunction:

p	**q**	**¬ q**	**p ∧ (¬ q)**
T	T	F	**F**
T	F	T	T
F	T	F	F
F	F	T	F

How exactly do we use our knowledge of conjunction here? Well, take row 1. We see that columns 1 and 3 give the truth values T and F for p and ¬ q respectively. By looking again at the following truth table for conjunction, we see that the combination T and F appears in row 2 and we see that the truth value for **p ∧ q** is **F**:

p	**q**	**p ∧ q**
T	T	T
T	F	**F**
F	T	F
F	F	F

By the above reasoning it tells us that **p ∧ (¬ q)** has the truth value **F** because the truth value for **p** is **T** and the truth value for **¬ q** is **F**. So we enter F in row 1 in the fourth column that represents p ∧ (¬ q) in the given table.

Similar reasoning applies to the other rows of the given table.

Activity 9-4: Constructing a truth table

Construct a truth table for $[\neg p \rightarrow (q \wedge r)] \vee r$

There are eight possible combinations of truth values for the simple declarative statements p, q and r. Hence our truth table has eight rows. The fourth column contains the truth values of the negation of p. The fifth column contains the truth values of the conjunction of q and r. The sixth column contains the truth values of the conditional statement $\neg p \rightarrow (q \wedge r)$ Finally the seventh column contains the truth values of the disjunction of $[\neg p \rightarrow (q \wedge r)]$ with r.

Therefore, the truth table for this activity is as follows:

p	q	r	¬ p	(q ∧ r)	[¬ p → (q ∧ r)]	[¬ p → (q ∧ r)] ∨ r
T	T	T	F	T	T	T
T	T	F	F	F	T	T
T	F	T	F	F	T	T
T	F	F	F	F	T	T
F	T	T	T	T	T	T
F	T	F	T	F	F	F
F	F	T	T	F	F	T
F	F	F	T	F	F	F

These examples illustrate the following general approach:

Suppose we want to construct a truth table for a compound statement built up from n simple declarative statements. There will be 2^n rows in such a table. In the above activity, n = 3, thus we have $2^3 = 8$ rows.

To determine the columns, we begin by listing the simple statements. Then we fill up these n columns as follows:

In the first column, enter T in the first half of the rows and F in the second half.

In the second column, for the rows which have T in the first column, enter T in the upper half and F in the lower half. Then do the same for the rows which have F in the first column.

Continue until, in the n-th column, T and F alternate. In the above activity, T and F alternate in the third column.

We then form columns for each statement we have built up, and which forms part of the compound statement that we're interested in. For instance, when we constructed the truth table for $p \wedge \neg q$ we had a column for $\neg q$ since it is a statement built up along the way to building $p \wedge \neg q$. Fill in such columns by consulting the truth table of the relevant connective given previously.

The last column represents the compound statement we are interested in.

Activity 9-5: Self-assessment exercises

1. Suppose that p represents the statement "It is sunny" and q the statement "It is humid". Write each of the following in abbreviated form:

(a) It is sunny and it is not humid.

(b) It is humid or it is sunny.
(c) It is false that it is humid.
(d) It is false that it is sunny and humid.
(e) It is neither sunny nor humid.
(f) It is not the case that if it is sunny then it is humid.
(g) It is humid if it is sunny.
(h) It is humid only if it is sunny.
(i) It is sunny if and only if it is humid.
(j) If it is false that it is either sunny or humid, then it is not sunny.

2. Construct the truth tables for the following compound statements:

(a) $[(\neg q) \rightarrow (\neg p)] \rightarrow (p \rightarrow q)$
(b) $[\neg p \rightarrow (q \wedge (\neg q))] \rightarrow p$
(c) $p \vee (\neg p)$
(d) $[p \wedge (p \rightarrow q)] \rightarrow q$
(e) $(p \vee q) \wedge (\neg p \vee \neg q)$
(f) $(\neg p \rightarrow [\, q \wedge r \,]) \vee r$
(g) $(p \rightarrow [\, q \wedge r \,]) \leftrightarrow ([\, p \rightarrow q \,] \vee [\, p \rightarrow r \,])$

9.2 Relationships between statements

Sometimes a compound statement is always true. For example, the truth table for $p \vee \neg p$ shows that this statement is always true:

p	¬ p	p ∨ ¬ p
T	F	T
F	T	T

We call such a statement a tautology.

Definition: Tautology

Some compound statements are always true. Such a statement is called a *tautology*.

On the other hand, some compound statements are always false. For example, the truth table for $p \wedge \neg p$ shows that this statement is always false:

p	¬ p	p ∧ ¬ p
T	F	F
F	T	F

We call such a statement a contradiction.

Definition: Contradiction

Some compound statements are always false. Such a statement is called a *contradiction*.

Activity 9-6: Self-assessment exercises

1. Express the following sentence symbolically and then determine whether or not it is a tautology:

   If demand has remained constant and prices have been increased, then turnover must have decreased.

   Use p for "demand has remained constant", q for "prices have been increased" and r for "turnover must have decreased".

2. Refer to Activity 9-5, Question 2. From the truth tables you have constructed for (a) to (g), determine whether each of the statements is a tautology, a contradiction or neither of the two.

Suppose that a and b are statements, and not necessarily simple ones. Then we can use the concept of tautology to spell out the idea that a and b have the same meaning or that a and b say the same thing in different words.

Definition: Logical equivalence

The two statements a and b are *logically equivalent*, denoted by $a \equiv b$, if and only if the statement $a \leftrightarrow b$ is a tautology.

Activity 9-7: Truth table for the biconditional

Recall that $a \leftrightarrow b$ has the value T if and only if a and b have the same truth value. So, to check that $a \leftrightarrow b$ is always T, it is enough to check that the final columns in the truth tables of a and b are identical.

Example

Let's look at the truth table for $(p \rightarrow q) \leftrightarrow (\neg q \rightarrow \neg p)$:

p	q	¬ q	¬ p	p → q	¬ q → ¬ p	(p → q) ↔ (¬ q → ¬ p)
T	T	F	F	T	T	**T**
T	F	T	F	F	F	**T**
F	T	F	T	T	T	**T**
F	F	T	T	T	T	**T**

Because there are only T's in the final column, it follows that $(p \rightarrow q) \leftrightarrow (\neg q \rightarrow \neg p)$ is a tautology.

This tells us that $p \rightarrow q$ and $\neg q \rightarrow \neg p$ are logically equivalent,
i.e. $p \rightarrow q \equiv \neg q \rightarrow \neg p$
i.e. $p \rightarrow q$ has exactly the same meaning as $\neg q \rightarrow \neg p$.

Note that $\equiv$ is *not* just another way to write $\leftrightarrow$.

We may write $p \rightarrow q \equiv \neg q \rightarrow \neg p$ only because we have shown that $(p \rightarrow q) \leftrightarrow (\neg q \rightarrow \neg p)$ is a tautology.

Activity 9-8: Important logical equivalences

Take note of the following important logical equivalencies:

(a) $p \vee q \equiv q \vee p$
$p \wedge q \equiv q \wedge p$ (commutative laws)

(b) $p \vee (q \vee r) \equiv (p \vee q) \vee r$
$p \wedge (q \wedge r) \equiv (p \wedge q) \wedge r$ (associative laws)

(c) $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$
$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$ (distributive laws)

(d) $p \vee p \equiv p$
$p \wedge p \equiv p$ (idempotent laws)

(e) $\neg (\neg p) \equiv p$ (law of double negation)

(f) $\neg (p \vee q) \equiv \neg p \wedge \neg q$
$\neg (p \wedge q) \equiv \neg p \vee \neg q$ (De Morgan's laws)

(g) $p \vee \neg p \equiv T_0$, where T_0 is a tautology
$p \wedge \neg p \equiv F_0$, where F_0 is a contradiction (negation)

(h) $\neg F_0 \equiv T_0$
$\neg T_0 \equiv F_0$ (negations of T_0 and F_0)

(i) $p \vee F_0 \equiv p$
$p \wedge T_0 \equiv p$ (identity)

(j) $p \vee T_0 \equiv T_0$
$p \wedge F_0 \equiv F_0$ (universal bound)

We often refer to these as *identities.*

(You can use truth tables to verify that these are indeed logical equivalences.)

Now that we have the notion of logical equivalence, we can derive a rather surprising result: We only require negation plus the connectives $\wedge$ and $\vee$!

To see this, we firstly show that

$p \leftrightarrow q$ is logically equivalent to $(p \rightarrow q) \wedge (q \rightarrow p)$,

which means that we can always work with the latter rather than use $\leftrightarrow$. Then we show that

$p \rightarrow q$ is logically equivalent to $\neg p \vee q$,

so that we can write conditional statements without actually using $\rightarrow$.

Examples

A truth table can show that $(p \rightarrow q) \wedge (q \rightarrow p)$ is logically equivalent to $p \leftrightarrow q$:

p	**q**	**$p \rightarrow q$**	**$q \rightarrow p$**	**$(p \rightarrow q) \wedge (q \rightarrow p)$**	**$p \leftrightarrow q$**
T	T	T	T	T	T
T	F	F	T	F	F
F	T	T	F	F	F
F	F	T	T	T	T

Here we have used a single truth table to show that the truth tables of $(p \rightarrow q) \wedge (q \rightarrow p)$ and of $p \leftrightarrow q$ are identical. Writing out two truth tables would have involved some tedious repetition.

Note: It is important to remember that this means that $(p \rightarrow q) \wedge (q \rightarrow p) \equiv p \leftrightarrow q$. As a result we could (if we wanted to) eliminate the use of $\leftrightarrow$. This is useful for some special applications of logic. However, for our purposes in this module, it is usually convenient to use $\leftrightarrow$.

We also use a truth table to show that $\neg p \vee q$ is logically equivalent to $p \rightarrow q$:

p	**q**	**$\neg p$**	**$\neg p \vee q$**	**$p \rightarrow q$**
T	T	F	T	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

Again we use a single truth table in which the last two columns are identical, rather than write out two separate truth tables.

Note: It is important to remember that this means that $\neg p \vee q \equiv p \rightarrow q$. As a result we could, if we needed to, eliminate the use of the connective "$\rightarrow$". As we said previously, it is sometimes necessary to know that any compound statement can be built up using only the connectives $\neg$, $\wedge$ and $\vee$.

Activity 9-9: Self-assessment exercises

1. Rewrite $p \leftrightarrow q$ as a statement built up using only $\neg$, $\wedge$ and $\vee$.

2. Show that $\equiv$ is an equivalence relation on statements.

3. Suppose we want to define a new connective, *the exclusive disjunction* also referred to as *the "exclusive or"*. By p + q we denote "p or q, but not both". Construct a truth table for this connective.

4. Find a statement that is logically equivalent to $\neg (p \vee \neg q)$.

5. Use the law of double negation and De Morgan's laws to rewrite the following statements so that the not symbol ($\neg$) does not appear outside parentheses.

   (a) $\neg [(p \vee q \vee \neg q) \wedge (q \wedge \neg p)]$
   (b) $\neg [(p \vee (p \rightarrow q)) \vee (p \wedge q)]$

6. Determine whether or not the following statements are equivalent: $\neg p \wedge (\neg p \wedge \neg q)$ and $\neg(p \vee (p \rightarrow q))$.

9.3 In summary of the study unit

In this study unit you ensured that you can answer the following questions regarding logical connectives and truth tables:

- What do we mean by "a declarative statement"?
- What do we mean by "a compound statement"?
- What is a logical connective?
- Which logical connectives did we look at in this study unit?
- How do we construct a truth table?
- How are the truth tables for the different connectives constructed?
- What is a tautology?
- What is a contradiction?
- How do we prove logical equivalence?

In the next unit, we look at quantifiers, some basics for predicate logic are introduced, and we look at different proof strategies.

NOTES

NOTES

Study unit 10 Logic: Quantifiers, predicates and proof strategies

Key questions for this study unit

- *What do the following terms mean: "universal quantifier", "existential quantifier", "a counterexample", "a predicate", "implication"?*
- *What is meant by a proof strategy?*
- *What methods of proof are studied in this study unit?*
- *What is meant by the terms "converse", "contrapositive", "*reductio ad absurdum*" and "vacuous proof"?*
- *How does one rewrite the negation of a quantified statement in a useful form?*

Activity 10-1: Overview *Study skill*

Draw a mind map of the different sections/headings you will deal with in this study session. Then page through the study unit with the purpose of completing the map.

Your mind map should include the concepts of quantification, implication, counterexample, converse and contrapositive, as well as proof strategies.

Activity10-2: Concepts *Conceptual skill*

Test your own knowledge and then correct your understanding afterwards. How does your understanding deepen as you jot down the terms used in your home language?

English term	Description	Term in your home language
Universal quantifier / quantified variable		
Existential quantification		
Predicate		
Direct poof		
Reductio ad absurdum (proof by contradiction)		
Contrapositive		
Converse		
Vacuous proof		

10.1 Quantifiers and predicates

In mathematics, many statements have variables in them which are represented by letters such as x, y, and z, which do not name anything specific.

The great logician Alfred Tarski gives a striking example of how useful variables can be in shortening what we want to say. Consider the arithmetic fact that:

For all real numbers x and y, $x^3 - y^3 = (x - y)(x^2 + xy + y^2)$.

Without the use of variables, the same information would be conveyed more clumsily as follows: The difference of the third powers of any two real numbers is equal to the product of the difference of these numbers and a sum of three terms, the first of which is the square of the first number, the second the product of the two numbers, and the third the square of the second number.

Statements containing variables pose a special problem. Consider, for instance, the statement "x is an even integer".

Is this statement true or false? How can we decide if we don't know what x is?

There are two ways to change "x is an even integer" into a statement that has a truth value.

The first way is to *replace* the variable x with the name of some specific thing. For example, we may substitute $\sqrt{2}$ for x to get the statement

" $\sqrt{2}$ is an even integer", which is false.

This is the kind of process we're involved in when we solve equations: to solve $2x = 6$ means to find all objects which, when their names are substituted for x, give a true statement.

The second way is to *quantify*. One can use either *universal* or *existential quantifiers*.

Description: Universal quantifier

Universal quantifiers are phrases such as:

"For all $x \in \mathbb{R}$ …", or

"For every $x \in \mathbb{Z}$…", or

"For each $x \in \{1, 2, 3\}$…".

We abbreviate these phrases by writing

$\forall x \in \mathbb{R}$ …, or

$\forall x \in \mathbb{Z}$ …, or

$\forall x \in \{1, 2, 3\}$ … respectively.

A variable such as "x" is called "a *quantified variable*".

Think of ∀ as an upside-down A, standing for *"for all".*

E.g. "∀ x ∈ {1, 2, 3}" is read as "*for all* elements x in the set {1, 2, 3}".

Applying universal quantification to "x is an even integer" would give a statement such as

"∀ x ∈ ℝ, x is an even integer",

which is clearly false.

A statement such as "∀ x ∈ ℝ, x is an even integer" is called a "*quantified statement*".

Description: Existential quantifier

Existential quantifiers are phrases such as:

"There exists an x ∈ ℝ such that …", or
"For some x ∈ ℤ …", or
"We can find an x ∈ {1, 2, 3} such that…".

We abbreviate these phrases by writing
∃ x ∈ ℝ …, or
∃ x ∈ ℤ …, or
∃ x ∈ {1, 2, 3} … respectively.

Think of ∃ as a backwards E, standing for "*there exists*".

E.g. "∃ x ∈ {1, 2, 3}" is read as "*there exists* an element x in the set {1, 2, 3}".

Applying existential quantification to "x is an even integer" would give a statement such as

"∃ x ∈ {1, 2, 3} such that x is an even integer"

which is clearly true.

***Notation*:** We may omit "∈ A" and simply write "∀ x ∈ A" as "∀ x" if it is absolutely clear from the context what set is meant. Similarly, "∃ x ∈ B" may be written as "∃ x" if it is absolutely clear from the context what set is meant.

Fundamental rule for quantified statements: *A quantified variable is a "dummy" variable and can be replaced (in all its occurrences) by any other variable.*

For example, the statement $\forall x \in \mathbb{R}, (x > 2) \rightarrow (x^2 > 4)$ *is logically equivalent to*
$\forall y \in \mathbb{R}, (y > 2) \rightarrow (y^2 > 4)$, *and to*
$\forall t \in \mathbb{R}, (t > 2) \rightarrow (t^2 > 4)$. *(Other variables can also be used.)*

Examples

We write down the English equivalences of some statements:

The statement "$\exists\, x \in \mathbb{Z}^+, x > 3$" tells us that "there exists some positive integer that is greater than 3". (This statement is true.)

The statement "$\forall\, x \in \mathbb{Z}^{\geq}, x \geq 0$" tells us that "all non-negative integers are greater than or equal to 0". (Since $\mathbb{Z}^{\geq} = \{0, 1, 2, \ldots\}$, this statement is true.)

Activity 10-3: Self-assessment exercises ***Application skills***

Write down the English equivalent of each of the following statements. Give an opinion on whether or not the statement is true.

(a) $\exists\, y \in \mathbb{Q}, y = \sqrt{2}$

(b) $\forall\, x \in \mathbb{R}, 2x < x^2$

(c) $\forall\, x \in \mathbb{Z}, x > 0$

(d) $\exists\, x \in \mathbb{Z}^+, x = 0$

A quantified statement such as "$\exists\, x \in \mathbb{Z}, x > 113$" is a declarative statement possessing a truth value, i.e. it is the kind of statement we worked with in the previous study unit. And so, of course, we can form conjunctions, disjunctions, conditional statements and biconditional statements in which one or more of the component statements are quantified statements.

When we think carefully about it, we see that universal quantification can be regarded as a *generalisation of conjunction*.

For instance, let $A = \{1, 2, 3\}$, then $\forall\, x \in A, x > \sqrt{2}$

means the same as $(1 > \sqrt{2}) \wedge (2 > \sqrt{2}) \wedge (3 > \sqrt{2})$.

Each one of the *component statements* $(1 > \sqrt{2})$, $(2 > \sqrt{2})$ and $(3 > \sqrt{2})$is either true or false. The *compound statement* $(1 > \sqrt{2}) \wedge (2 > \sqrt{2}) \wedge (3 > \sqrt{2})$, by the way, is false. This will be investigated later in this study unit.

Similarly, existential quantification acts like a *generalised form of disjunction*,

with $\exists\, x \in A, x > \sqrt{2}$ saying the same as

$$(1 > \sqrt{2}) \vee (2 > \sqrt{2}) \vee (3 > \sqrt{2}).$$

The usefulness of quantification arises from the fact that the set A over which we quantify need not be finite. That is, we can say things such as

$$\forall\, x \in \mathbb{Z}, x^2 \geq 0$$

which could not be said with the aid of conjunctions alone, because we would never finish saying
$(0^2 \geq 0) \wedge (1^2 \geq 0) \wedge ((-1)^2 \geq 0) \wedge (2^2 \geq 0) \wedge ((-2)^2 \geq 0) \wedge \ldots$

Similarly

$$\exists\, x \in \mathbb{Z},\ x > \pi$$

could not be said with the help of disjunctions alone, since we would be unable to complete the infinitely long statement

$$(0 > \pi) \vee (1 > \pi) \vee (-1 > \pi) \vee (2 > \pi) \vee (-2 > \pi) \vee \ldots$$

We can also form negations ($\neg$) of quantified statements. Here we have to be aware of an interesting fact: to work effectively with the negation of a quantified statement, one must get the negation *as far inside as possible*. You will see what we mean by this when we do some examples.

But first we must answer the following question: How does one get "$\neg$" inside a quantified statement? After all, "$\neg$" is something we usually put in front of a statement. We first look at the influence "$\neg$" has on compound declarative statements.

First, consider a conjunction, say $p \wedge q$. Negating gives $\neg (p \wedge q)$.

Activity 10-4: Self-assessment exercise ***Application skills***

Prove by means of truth tables that
$\neg (p \wedge q) \equiv (\neg p) \vee (\neg q)$.

Next, consider a disjunction, say $p \vee q$. Negating gives $\neg (p \vee q)$.

Activity 10-5: Self-assessment exercise ***Application skills***

Prove by means of truth tables that
$\neg (p \vee q) \equiv (\neg p) \wedge (\neg q)$.

Now we are ready to tackle the *negation* of *quantification*!

Examples

Consider

$$\forall\, x \in \{1, 2, 3\}, x > \sqrt{2}.$$

As we have seen, this means the same as

$$(1 > \sqrt{2}) \wedge (2 > \sqrt{2}) \wedge (3 > \sqrt{2}).$$

Negating gives $\neg(\forall\, x \in \{1, 2, 3\}, x > \sqrt{2}\,)$ which means the same as

$$\neg\, [(1 > \sqrt{2}) \wedge (2 > \sqrt{2}) \wedge (3 > \sqrt{2})],$$

which, as you know from the activities you have just completed, is logically equivalent to

$$\neg\, (1 > \sqrt{2}) \vee \neg\, (2 > \sqrt{2}) \vee \neg\, (3 > \sqrt{2})$$

i.e. $(1 \leq \sqrt{2}) \vee (2 \leq \sqrt{2}) \vee (3 \leq \sqrt{2})$.

Note that we have simply used the fact that if x is not greater than y, then x is less than or equal to y, for real numbers x and y.

But now $(1 \leq \sqrt{2}) \vee (2 \leq \sqrt{2}) \vee (3 \leq \sqrt{2})$

means the same as $\exists\, x \in \{1, 2, 3\},\ x \leq \sqrt{2}$.

So we conclude that

$$\neg\, [\forall\, x \in \{1, 2, 3\}, x > \sqrt{2}]$$

can be written as

$$\exists\, x \in \{1, 2, 3\}, \neg\, (x > \sqrt{2}\,)$$

i.e. $\exists\, x \in \{1, 2, 3\}, x \leq \sqrt{2}$.

Reasoning in a similar way, we find that

$$\neg\, [\exists\, x \in \{1, 2, 3\}, (x > \pi\,)]$$

means the same as

$$\forall\, x \in \{1, 2, 3\}, \neg\, (x > \pi\,)$$

i.e. $\forall\, x \in \{1, 2, 3\},\ x \leq \pi$.

As we know by now, it can be determined whether a quantified statement is true or false if a finite number of statements are given. Later in this study unit we will investigate some methods of proof which can be applied to determine whether statements of this kind are true or not when an infinite number of statements are given.

Let's determine whether "$\forall x \in \{1, 2, 3\}, x > \sqrt{2}$ " is true or false:

In order to determine whether or not $\forall\, x \in \{1, 2, 3\}, x > \sqrt{2}$ is true, one can determine whether

$(1 > \sqrt{2}) \wedge (2 > \sqrt{2}) \wedge (3 > \sqrt{2})$ is true or false.

We investigate each component of this statement:

Since $\sqrt{2} = 1.4142$, we know $1 > \sqrt{2}$ is false, $2 > \sqrt{2}$ is true, and $3 > \sqrt{2}$ is true.

It follows that $(1 > \sqrt{2},) \wedge (2 > \sqrt{2}) \wedge (3 > \sqrt{2})$ is false because $1 > \sqrt{2}$ is false.

Alternatively, one could say that $\forall\, x \in \{1, 2, 3\}, x > \sqrt{2}$ is false since a *counterexample* can be found, namely x = 1.

If x = 1, then $1 > \sqrt{2}$ is false and hence the compound statement is false. In other words, "$\forall\, x \in \{1, 2, 3\}, x > \sqrt{2}$ " is false because we have shown that x is *not* greater than $\sqrt{2}$ for *all* $x \in \{1, 2, 3\}$. We supplied a *counterexample*.

We have seen that it can be determined whether a statement is true or false for some quantified variables. For instance, in the previous example we determined whether the statement "$x > \sqrt{2}$" is true or false for each $x \in \{1, 2, 3\}$. We name such statements "predicates":

Definition: Predicate

A statement P(x) is called a *predicate* if it expresses some property of a variable $x \in A$, and returns either true or false depending on the value of x. P(x) is true for any variable $x \in A$ that has the property, and P(x) is false if x does not have the property.

For example, the statement "n is even" (with $n \in \mathbb{Z}$), is a predicate that can be written as P(n), such that P(n) is true for all even integers and P(n) is false for all non-even (odd) integers. In this case, statements such as P(−2) and P(6) are true, whereas P(−3) and P(7) are false.

We also re-visit examples discussed previously in this study unit:

Example

If P(x) is the predicate "$x > \sqrt{2}$", then
"$\forall x \in \{1, 2, 3\}, x > \sqrt{2}$" can be written as
"$\forall x \in \{1, 2, 3\}$, P(x)", and

if Q(x) is the predicate "$x > \pi$",
then "$\exists x \in \{1, 2, 3\}, x > \pi$" can be written as
"$\exists x \in \{1, 2, 3\}$, Q(x)".

We summarise the rules for writing the negation of a quantified statement in a useful form (i.e. with the "not" taken as far inside the statement as possible):

Rules: Negation of quantified statements

If P(x) is a predicate containing some variable x, then we can write

(a) $\neg (\forall x \in A, P(x))$ as $\exists x \in A, \neg P(x)$, and

(b) $\neg (\exists x \in A, P(x))$ as $\forall x \in A, \neg P(x)$.

These rules should be applied when x ranges over an infinite set A.

Examples

Determine the negation of the quantified statement "$\forall x \in A, P(x) \vee Q(x)$".

$\neg (\forall x \in A, P(x) \vee Q(x))$
$\equiv \exists x \in A, \neg (P(x) \vee Q(x))$ (from rule (a) above)
$\equiv \exists x \in A, \neg P(x) \wedge \neg Q(x)$ (refer to Activity 10-5)

QED

Another example:

Determine the negation of the quantified statement
"$\exists\, y \in \mathbb{Z}^{\geq}, (y + 1 > 0) \wedge (y^3 \leq 1)$".

$\neg\, (\exists\, y \in \mathbb{Z}^{\geq}, (y + 1 > 0) \wedge (y^3 \leq 1))$
$\equiv \forall\, y \in \mathbb{Z}^{\geq},\ \neg\, ((y + 1 > 0) \wedge (y^3 \leq 1))$
$\equiv \forall\, y \in \mathbb{Z}^{\geq},\ \neg\, (y + 1 > 0) \vee \neg\, (y^3 \leq 1)$
$\equiv \forall\, y \in \mathbb{Z}^{\geq},\ (y + 1 \leq 0) \vee (y^3 > 1)$

Which one is true, the original or the negated statement?

Well, if y = 1, then
$(y + 1 > 0)$ means that $(1 + 1 > 0)$ i.e. $(2 > 0)$, and
$(y^3 \leq 1)$ means that $(1^3 \leq 1)$ i.e. $(1 \leq 1)$.

Both $(2 > 0)$ and $(1 \leq 1)$ are true statements, so there exists an element $y \in \mathbb{Z}^{\geq}$, namely y = 1 such that the original statement is true.

In the previous study unit we proved that $p \rightarrow q \equiv \neg\, p \vee q$. It is convenient to apply this equivalency when we want to get rid of "$\rightarrow$" in some statement.
For example, when we want to determine the negation of $\forall x \in A, P(x) \rightarrow Q(x)$,
*i.e. **$\neg\, (\forall x \in A, P(x) \rightarrow Q(x))$**, we can write this statement as **$\neg\, (\forall x \in A, \neg\, P(x) \vee Q(x))$**.*

We apply this equivalency in the following example:

Example

Determine the negation of the statement
"$\forall\, x \in \mathbb{Z}^{+}, (x \leq 2) \rightarrow (x^2 + 2x > 1)$".

$\neg\, (\forall\, x \in \mathbb{Z}^{+}, (x \leq 2) \rightarrow (x^2 + 2x > 1))$
$\equiv \neg\, (\forall\, x \in \mathbb{Z}^{+}, \neg\, (x \leq 2) \vee (x^2 + 2x > 1))$
$\equiv \exists\, x \in \mathbb{Z}^{+}, \neg\, (\neg\, (x \leq 2) \vee (x^2 + 2x > 1))$
$\equiv \exists\, x \in \mathbb{Z}^{+}, \neg\, \neg\, (x \leq 2) \wedge \neg\, (x^2 + 2x > 1)$
$\equiv \exists\, x \in \mathbb{Z}^{+}, (x \leq 2) \wedge (x^2 + 2x \leq 1)$

Activity 10-6: Self-assessment exercises *Application skills*

Determine the negations of the following quantified statements:
(Show all the steps.)

(a) $\forall\, x \in \mathbb{Z}^{+}, x > 3$

(b) $\exists\, x \in \mathbb{R}, 2x = x^2$

(c) $\forall\, x \in \mathbb{Z}, (x > 0) \vee (x^2 > 0)$

(d) $\exists\, y \in \mathbb{Z}^{+}, (y \leq 10) \wedge (y \neq 0)$

(e) $\exists\, x \in A, P(x) \wedge Q(x)$

(f) $\forall\, x \in \mathbb{Z}^{+}, (x \leq 3) \rightarrow (x^3 \geq 1)$

Activity 10-7: Self-assessment exercises ***Application skills***

For each of (a) to (d) in activity 10-6, try to decide whether the original statement is true, or whether its negation is true, or whether neither of the two is true.

10.2 Proof strategies

In this section we look at a number of standard proof strategies or "methods of proof" frequently used in discrete mathematics. The idea is that we want to prove that a certain statement is always true, i.e. it is true for all variables in the relevant domain. This domain might be infinite. Suppose we want to prove the statement

"For all integers, if n is an even integer, then $n^2 + n + 1$ is odd."

There are a number of different proof strategies we can follow. In this study unit we look at a *direct* proof, a proof by *contradiction* (*reductio ad absurdum*), a proof by *contrapositive* and a vacuous proof.

If we cannot prove that a statement is true, we have to provide only one *counterexample* that illustrates that the statement is false.

Let's look at the individual strategies.

10.2.1 Direct proof

When using this strategy, we start the proof by assuming the truth of p (the "if" part of the statement), and then reason step by step until we can show that q (the "then" part of the statement) is true.

How about an example? Well, there are plenty, but here is one that you haven't encountered before.

Example

Prove that the following statement is true for all $x \in \mathbb{R}$:

"If $x^2 - 4x + 3 < 0$, then $x > 0$."

Remark: We in fact want to prove that the above holds for all real numbers $x \in \mathbb{R}$.

We start the proof by assuming that $x^2 - 4x + 3 < 0$ (the "if" part of the statement) is true.

Assume $x^2 - 4x + 3 < 0$
i.e. $(x - 3)(x - 1) < 0$ (by factorisation)
then

(i) $(x - 3) > 0$ and $(x - 1) < 0$ (since a plus times a minus gives a minus)

i.e. $x > 3$ and $x < 1$, but this cannot be the case since there are no real numbers that are simultaneously greater than 3 and less than 1.

OR

(ii) $(x - 3) < 0$ and $(x - 1) > 0$ (since a minus times a plus gives a minus)
i.e. $x < 3$ and $x > 1$,
i.e. $1 < x < 3$
so $x > 0$
which is what we had to prove.

10.2.2 Proof by contradiction (reductio ad absurdum)

We have already looked at a couple of examples of such a proof, namely in study unit 2 (where we showed that $\sqrt{2}$ is not a rational number) and in the discussion of equivalence relations in study unit 6.

Suppose we have to prove by contradiction that "if p then q". The basic principle we use in this strategy is to assume that p is true. At this point, we have two possibilities – either q is false (the "bad" possibility) or q is true (the "good" possibility). What we do now is to assume that q is false. By using step-by-step reasoning we get a contradiction. This shows that q must be true.

Let's illustrate this with the same example that we used when we applied the direct proof.

Remark: We are going to start the same way as with the direct proof. Then, at the strategic moment, we'll throw in the questionable assumption (the "bad" possibility). We'll show that it leads to a contradiction and will therefore conclude that the questionable assumption was false.

Example

We start the proof by assuming that the antecedent is true:

Assume $x^2 - 4x + 3 < 0$. (initial assumption)

We consider two possibilities:
either the consequent($x > 0$) is true (the "good" possibility), or
the consequent is false (the "bad" possibility), i.e. $x \leq 0$.

Now we'll prove that the "bad" possibility leads to a *contradiction* and then deduce that the "good" possibility must be true.

Assume $x \leq 0$ (the "bad" possibility which is the questionable assumption)
then $-4x \geq 0$ (since a minus times a minus gives a plus)
so $x^2 - 4x + 3 > 0$ (since $x^2 + 3 > 0$, and also $-4x \geq 0$)

However, this *contradicts* the initial assumption. Hence it cannot be the case that $x \leq 0$, and thus we conclude that our questionable assumption was incorrect, and consequently it is true that $x > 0$.

10.2.3 Proof by contrapositive

Another technique, proof by contrapositive is based on the following fact:

Definition: Contrapositive

The *contrapositive* of $p \rightarrow q$ is $\neg q \rightarrow \neg p$. In other words:
$p \rightarrow q$ is logically equivalent to $\neg q \rightarrow \neg p$.

$(p \rightarrow q) \leftrightarrow (\neg q \rightarrow \neg p)$ is a tautology. This means that, in order to prove that $p \rightarrow q$, we may, if we wish, prove that $\neg q \rightarrow \neg p$.

Let's use the same example as before:

Example

We want to prove that "if $x^2 - 4x + 3 < 0$, then $x > 0$".

We can do this (using the contrapositive) by proving that

if $\neg (x > 0)$ then $\neg (x^2 - 4x + 3 < 0)$:

First we assume that $\neg (x > 0)$ is true, i.e. we assume that $x \leq 0$.

Now we factorise the expression $x^2 - 4x + 3$:

$x^2 - 4x + 3 = (x - 3)(x - 1)$

We already know that $x \leq 0$, which means that

$(x - 3) \leq 0$ and $(x - 1) \leq 0$.

So $(x - 3)(x - 1) \geq 0$ (since a minus times a minus gives a plus)
i.e. $x^2 - 4x + 3 \geq 0$
i.e. $\neg (x^2 - 4x + 3 < 0)$.

Let's analyse the proof:

Let p represent "$x^2 - 4x + 3 < 0$" and let q represent "$x > 0$".

We need to prove that $p \rightarrow q$. If we want to prove this using contrapositive, we need to prove that $\neg q \rightarrow \neg p$, i.e. if it is not the case that "$x > 0$", then it is also not the case that "$x^2 - 4x + 3 < 0$".

N.B.: Do not confuse the contrapositive of a statement with the converse, which is defined as follows:

Definition: Converse

The *converse* of $p \rightarrow q$ is written as $q \rightarrow p$.

Proving the converse $q \rightarrow p$ does not establish $p \rightarrow q$. Many people seem to think it does, but this can produce ridiculous arguments such as: "If an exam is too difficult then students fail. So if students fail, the exam was too difficult." It could be that they did not study enough! Let's look at this in more detail.

From the statement

"If $x \in \mathbb{Z}^+$ then $x^2 > 0$"

we cannot conclude

"If $x^2 > 0$ then $x \in \mathbb{Z}^+$", since x might have a value such as –113.

Similarly, from

"If one is hanged by the neck, then one is dead"

we can't conclude that

"If one is dead, then one was hanged by the neck",

for it may well have been a logician enraged by the confusion of converse and contrapositive who battered the poor victim to a pulp with a truth table.

We have encountered the following example of a statement and its contrapositive in the discussion on injective functions:

Example

"If f(x) = f(y), then x = y"

"If x ≠ y, then f(x) ≠ f(y)"

Because of our knowledge of logic, the equivalence of these two statements should now be clear.

Activity 10-8: Implication

Show by means of truth tables that $p \rightarrow q$ and $q \rightarrow p$ are not logically equivalent.

Having covered a number of different ways in which we can prove statements such as $p \rightarrow q$, what other kinds of proof do we still have to consider?

10.2.4 Proofs involving quantifiers

Truth tables cannot be used to prove any quantified statement where the domain is infinite. What should we do when we have to prove a statement such as "$\forall x \in A, P(x)$" where A is finite?

Well, in order to prove "$\forall x \in A, P(x)$", just think of the statement as being equivalent to

"$x \in A \rightarrow P(x)$".

We have discussed the ways in which we can prove statements such as $p \rightarrow q$, so we can apply our knowledge of implications directly to quantified statements.

Let's look at an example:

Example

To prove $\forall x \in \mathbb{R}, x^2 + 1 > 0$, we reason as follows:

If $x \in \mathbb{R}$ then
$x^2 \geq 0$
i.e. $x^2 + 1 \geq 1$
i.e. $x^2 + 1 > 0$

In words this says: "If x is any real number, then $x^2 \geq 0$, i.e. $x^2 + 1 \geq 1$, which means that $x^2 + 1 > 0$." The word "any" can be included, because x is a variable.

Our conclusion: *Proving a statement of the form $\forall x \in A, P(x)$ offers no new problems. We apply the knowledge we already have.*

On the other hand, we sometimes want to *disprove* a statement.

For instance, given the statement

$$\forall x \in \mathbb{R}, x^2 - 4x + 3 \geq 0,$$

we may decide after some thought that it is false. How do we prove that it is false? Well, by proving that *its negation* is true.

The negation of

$$\forall x \in A, P(x)$$

is the statement

$$\neg (\forall x \in A, P(x)),$$

which, in a more useful form, is

$$\exists x \in A, \neg P(x) .$$

And in order to prove an existential statement, it is enough to find *a single element* of A which acts in the right way.

Activity 10-9: Counterexample

Show by means of a *counterexample* that the statement
$\forall x \in \mathbb{R}, x^2 - 4x + 3 \geq 0$ is not true.

To disprove

$$\forall x \in \mathbb{R}, x^2 - 4x + 3 \geq 0,$$

we only have to prove that

$$\exists x \in \mathbb{R}, x^2 - 4x + 3 < 0.$$

Choose x = 3/2.

Then $(3/2)^2 - 4(3/2) + 3$ $= 9/4 - 12/2 + 3$
$= 9/4 - 24/4 + 12/4$
$= -3/4$ which is less than 0.

So we have shown that there exists some x (in this case x = 3/2) such that
$x^2 - 4x + 3 < 0$.

One counterexample is enough to disprove the given statement, but there might exist some other counterexamples that will also show this.

10.2.5 Vacuous proof

We first look at an example to illustrate this proof method:

Example

Suppose we want to prove that
$\emptyset \subseteq X$.

Then, by the subset definition provided in a previous study unit, we have to show that
if $x \in \emptyset$ then $x \in X$.

Proof:
$\emptyset$ is an empty set,
so "$x \in \emptyset$" is false,
thus "if $x \in \emptyset$ then $x \in X$" is *vacuously* true.
QED

What does the last line in the proof say? We can refer to the truth table of the conditional statement "if p then q" in the previous study unit. Whenever p is false, we know that $p \rightarrow q$ is true, no matter whether q is true or false. In our example "$x \in \emptyset$" is false, so no matter whether "$x \in X$" is true or false, we may say that the statement "if $x \in \emptyset$ then $x \in X$" is *vacuously* true.

Let's look at another example:

Example

Let S be a relation on {a, b, c, d}:
S = {(a, b), (a, d)}.

Is S transitive?

Proof:
By the definition of transitivity provided in a previous study unit,
whenever $(x, y) \in S$ and $(y, z) \in S$ then
(x, z) must also live in S.

It is not possible to find two pairs of the form (x, y) and (y, z) in S,
so it is *vacuously* true that S is transitive.

Activity 10-10: Self-assessment exercises ***Application skills***

1. Prove each of the following statements by direct proof, contrapositive and contradiction (*reductio ad absurdum*) respectively. Which strategy works best?

   (a) If $x^2 - 3x + 2 < 0$, then $x > 0$.

   (b) If $x^2 - x - 6 > 0$, then $x \neq 1$.

   (c) For all $a, b \in \mathbb{Z}$, if $a + b$ is odd, then exactly one of a or b is odd.

   (d) For all $x \in \mathbb{Z}$, if x is even, then $x^2 + 4x + 2$ is even.

   (e) If n is a multiple of 3, then $n^3 + n^2$ is a multiple of 3.

2. Provide a counterexample to show that the statement "If $x > 0$, then $x^2 - 3x + 1 < 0$" is not true for all integers $x > 0$.

10.3 In summary of the study unit

In this study unit you ensured that you can answer the following questions:

- What is meant by the term "universal quantifier"?
- What does "existential quantifier" mean?
- What does "implication" mean?
- What is meant by the concept "predicate"?
- How do we approach each method of proof discussed in section 10.2?
- How do we express the converse of an implication?
- How do we express the contrapositive of an implication?
- How do we use a counterexample to show that a statement is not true?

NOTES

INDEX

Bibliography

Bundy, S. 2010. Discrete mathematics. Available from http://pagerankstudio.com/Blog/2010/11/discrete-mathematics/. Accessed on 4 May 2010.

Ensley, DE & Crawley, JW. 2006. *Discrete mathematics: mathematical reasoning.* Available as a free ebook from http://www.ebooktoyou.net/ebook/discrete-mathematics-ensley-pdf.php. Accessed on 4 May 2010.

MindtTools. SQ3R Increasing your retention of written information. Available from http://www.mindtools.com/pages/article/newISS_02.htm. Accessed on 4 May 2010.

Post, T, Behr, M & Lesh, R. 1982. Interpretations of rational number concepts. In Silvey, L & Smart, J (eds). *Mathematics for Grades 5-9*, *1982 NCTM Yearbook* (pp 59-72). Reston, Virginia: NCTM. Available from http://www.cehd.umn.edu/rationalnumberproject/82_1.html. Accessed on 10 May 2010.

Roberts, FS. 2001. *International encyclopedia of the social & behavioral sciences, Discreet mathematics*. Available from: http://dimacs.rutgers.edu/People/Staff/froberts/encyclopediafinal.pdf. Accessed on 14 May 2011. Elsevier Science.