

# HOW DO UNDERGRADUATES DO MATHEMATICS?

A guide to studying mathematics at Oxford University

Charles Batty  
St. John's College, Oxford

with the assistance of

Nick Woodhouse  
Wadham College, Oxford

April 1994

# CONTENTS

<b>Preface</b>	ii
PART I	
<b>Chapter 1: University study</b>	1
1.1 Pattern of work	1
1.2 Lectures	3
1.3 Tutorials	5
1.4 Cooperation with fellow-students	8
1.5 Books and libraries	9
1.6 Vacation work	10
<b>Chapter 2: University mathematics</b>	11
2.1 Studying the theory	11
2.2 Problem-solving	15
2.3 Writing mathematics	16
<b>Chapter 3: The perspective of applied mathematics</b> by Nick Woodhouse	19
3.1 Pure and applied mathematics	19
3.2 Solving problems in applied mathematics	20
3.3 Writing out the solution	24
PART II	
<b>Chapter 4: The formulation of mathematical statements</b>	29
4.1 Hypotheses and conclusions	30
4.2 “If”, “only if”, and “if and only if”	33
4.3 “And” and “or”	37
4.4 “For all” and “there exists”	41
4.5 What depends on what?	43
<b>Chapter 5: Proofs</b>	46
5.1 Counterexamples	46
5.2 Constructing proofs	49
5.3 Understanding the problem	50
5.4 Experimentation	53
5.5 Making the proof precise	58
5.6 What can you assume?	64
5.7 Proofs by contradiction	65
5.8 Proofs by induction	70
<b>Appendix: Some symbols</b>	76

## Preface

In one sense, mathematics at university follows on directly from school mathematics. In another sense, university mathematics is self-contained and requires no prior knowledge. In reality, neither of these descriptions is anything like complete. Although it would be impossible to study mathematics at Oxford without having studied it before, there is a marked change of style at university, involving abstraction and rigour. As an undergraduate in any subject, your pattern and method of study will differ from your schooldays, and in mathematics you will also have to master new skills such as interpreting mathematical statements correctly and constructing rigorous proofs.

These notes are an introduction to some of these aspects of studying mathematics at Oxford. Their purpose is to introduce you to ideas which you are unlikely to have met at school, and which are not covered in textbooks. **It is not expected that you will understand everything on first reading**; for instance, some of the examples may be taken from topics which you have not covered at school, and some of the material in Part II needs time to be absorbed. With the exception of a few fleeting references to physical applied mathematics, the notes discuss topics which are common to all the mathematics courses in Oxford, so they should be almost as useful for students in Mathematics & Philosophy and Mathematics & Computation as for those in Mathematics. It is hoped that after a term or two you will have absorbed all the advice in the notes and that they will become redundant. For this reason, there is no attempt to cover any topics which will not arise until later in your course, e.g. examinations and options.

It is important to appreciate that the notes are an introduction to *studying* mathematics, not an introduction to the mathematics; indeed, there is very little genuine mathematics in the notes. In particular, the notes attempt to answer the question: How to study mathematics? but not the questions: Why study mathematics? or: What topics do we study in university mathematics? Occasionally, the notes consider the question: Why do we do mathematics the way we do? but the emphasis is on How? not Why? Consequently, the notes have some of the style of a manual, and they will not make the most exciting of reads. You are therefore recommended to seek another source for the other questions. At present, the best such book known to me is *Alice in Numberland*, by J. Baylis and R. Haggarty (Macmillan). Indeed, I recommend that you read *Alice in Numberland* (at least a substantial part of it) before you read these notes (at least before you read Part II).

Your tutors are likely to offer their own advice which sections of these notes you should read at what time and what else you should read, and to set you some problems. I will therefore refrain from proffering guidance except to say that Part I will be most useful if it is read before your courses start in the first full week after your arrival in Oxford. It should not take long to read this part as it contains very little mathematics. Chapter 1 is the only chapter where some of the content is specific to Oxford (on the other hand, much of its content is not specific to mathematics).

Part I contains many instructions and imperatives: *Do this; Do that; You should*

... They are to be regarded as recommendations or advice, rather than rules which must be religiously observed. (Indeed, it is unlikely that anyone would have time or inclination to carry out all the tasks suggested.) Students differ in temperament and intellect, so what suits one undergraduate will not necessarily suit another. It is up to you to work out your own work-style; it is hoped that these notes will assist you. It is possible that your tutor will disagree with some of the advice. What I have included is what I regard as good advice in general. I believe that all tutors will agree with most of what I have written, and I am confident that students who follow most of the advice will be more successful, in general, than those who ignore it. Nevertheless, any opinions expressed in these notes are my own (and, in the case of Chapter 3, those of Nick Woodhouse); they do not necessarily represent the views of any other individual or institution.

Part II is much more technical than Part I, and will need to be read carefully. Many of the individual sentences may need to be read several times to understand their precise meaning and significance. Since some of the ideas are sophisticated, there may be some points which remain unclear to you even after hard thought, particularly if you are reading them before the course has started. If so, do not be alarmed—all may become clear when the course starts and you have a context in which to consider the point. You should be able to understand later sections without having grasped every point earlier on. The most difficult sections are probably 4.4, 4.5 and 5.7, so you may want to start Chapter 5 before tackling Sections 4.4 and 4.5; in Sections 5.2–5.5, it will be more profitable if you pick out some of the examples which are at the appropriate level for you than if you work through them all at one reading. Apart from the list of contents, you can use the summaries at the end of each section to find where a topic is discussed. In order to facilitate using the notes for reference, there are many cross-references between sections, and a few places where material is repeated.

Before writing the notes, I read two booklets: *Study Skills in Mathematics*, edited by Diana Burkardt and Des Rutherford of the Shell Centre for Mathematical Education at the University of Nottingham, and *Proofs and What is a proof?* appendices of lecture notes of La Trobe University, Australia. Although only faint traces of those sources may appear in these notes, they were very helpful in guiding me where to start and how to express the ideas. I am grateful to several colleagues for their comments on earlier drafts, especially to Nick Woodhouse for contributing a chapter explaining the differences, and many similarities, between pure and applied mathematics.

St. John's College  
Oxford

C.J.K.B.  
April 1994

# PART I

## CHAPTER 1

### University study

In this chapter, we shall discuss the general way in which you are likely to go about your academic work.

#### 1.1 Pattern of work

There are three main aspects of university study:

- (1) Lectures
- (2) Tutorials (or classes)
- (3) Private study (reading books, working through lecture notes, doing problems)

This ordering is arbitrary; it corresponds neither to importance nor to duration of these activities. Tutorials are the most important in terms of the ratio of potential benefit to duration, but they will occupy the least time, private study the most.

The total amount of time spent on study should be consistent with your status as a full-time student. This means that it should at least equal the amount of time that you would work in a full-time job, and the amount of time which you spent at school, including homework. There are 168 hours in the week, so this leaves plenty of time for other activities, even after allowing for eating, sleeping, etc. Provided that you organise your time well, you will be able to socialise and participate in sport, or arts, or whatever else you want to do, as well as study mathematics successfully.

Where university education differs from school, and from most employment, is that most of your-working time is organised by yourself. In your first two terms, lectures will account for 10 hours a week, and tutorials for about 2 hours, at appointed hours. It is up to you to choose at what time to do your private study. You may like to work from 9 to 5 Monday to Friday, or you may prefer to work in the early hours of the morning and at weekends. It is not important when you choose to work, provided that you meet the following constraints:

- (i) you will have to perform certain tasks between lectures and before tutorials (see Sections 1.2 and 1.3),
- (ii) unless you have exceptional stamina, it is probably not a good idea to work for very long periods without at least a short break (even if it is only to make yourself a cup of coffee),
- (iii) you should be alert (i.e., not half-asleep) for lectures and tutorials.

Although it is not important when you choose to work, it is vital that you should get into a regular and full pattern of work soon after your arrival in Oxford (before you commit yourself to too many other activities). You are recommended to write down a weekly timetable for yourself. The main purpose of this is to avoid the unhappy situation, into which too many students fall, where increasing involvement in other activities, and/or general frittering away of time, erodes your work without you really noticing it. It will take some experimentation to work your timetable out, but it should be possible for you to have a tentative timetable within your first week, which you can adjust as your pattern of activities settles down. It is not expected that you will stick slavishly to the same timetable week after week. Of course, there will be occasions when unexpected or irregular events cause you to miss a session when you intended to work, or when you are feeling under the weather so that working would be pointless. If you are conscious that you have slipped behind a planned schedule, you are much more likely to make time to catch up later. You should not be afraid to tell fellow-students who interrupt your work that you cannot join in social activities for the time being. You won't lose any friends by doing this—at any rate, not worthwhile friends!

It is also important that at an early stage you should find a place to work where you will not be greatly distracted. This is most likely to be your own room in college, but if you find that your attention drifts too often towards the non-mathematical pursuits available in the room, or if you are interrupted too often by friends eager to socialise, you may prefer to work in a library or elsewhere.

Your work is likely to get on better if you discuss mathematics with your peers (see Section 1.4). So, it will be beneficial to find one or two other mathematicians in your year with whom you get on sufficiently well that you can discuss any difficulties which you may have with lectures, exercises, etc.

## Summary

1. Your total working-time should be consistent with your status as a full-time student.
2. You should get into a regular pattern of work as soon as possible.
3. You should find a regular place to work, free from distractions.
4. Find other students with whom you can discuss your work.

## 1.2 Lectures

In your first two terms, you will have about 10 lectures a week (usually, two or three lectures a week on each of about four topics). Subsequently, the number of lectures will tend to decrease gradually. Lectures last a nominal hour; in practice this means about 50 minutes, to allow time for arrival and departure.

You should go to all the lectures on all your courses. If by necessity you miss a lecture, get hold of the lecture notes from somebody else and work through them. (Don't just photocopy them and file them; that way, you learn nothing.)

### During lectures

Many of the lectures will have a rather formal style. The lecturer will speak and write on the board (some lectures may have interludes such as computer demonstrations, jokes, etc.). You will have to learn by experience what form of notes to take. Ideally, you should concentrate on trying to follow the mathematical argument, and take notes only of the essential points; if you succeed in understanding the argument, you will have no difficulty in filling in the gaps afterwards (provided that you do so before you have forgotten everything). However, in practice, you may well find that you often fail to keep up with the lecturer's reasoning. In that case, rather than coming out of the lecture with little understood and little written down, you should copy whatever is written on the board, plus anything else which seems important. This seems a crazy, inefficient, way to study—why don't the lecturers provide handouts of what they write on the board? The answer is that experience shows that most students do not work through impersonal handouts thoroughly, whereas the copying method does help to concentrate the mind, ensures that something is absorbed if only by a process of osmosis, and provides students with something which they can regard as their personal version of the mathematics.

The lectures will vary in quality; some will be much more entertaining than others; some will be much clearer than others; some will be much better prepared than others. Often, you will lose the thread of the mathematical argument for a while during a lecture. Whatever happens, you should stick it out, and make as much of the lecture as you can. In the long run, you often learn more from a lecture which at the time appeared obscure than from one which was crystal clear. This may be because the material was deep and difficult, or it may be because you are obliged to work on the lecture afterwards. You can be sure that it will be much easier to understand the material by reading a book afterwards if you have been to a lecture (however bad the lecture may have seemed) than if you skipped the lecture. Reciprocally, it will be much easier to follow a lecture if you read ahead in a book.

Please do not talk or otherwise disturb the rest of the audience during lectures, unless you want to ask the lecturer a question. There may be moments when you are bored and would prefer to chat to your friends, but other people sitting near you may want to listen to the lecturer and you should let them do so.

## Between lectures

As soon as possible after each lecture, and in any case before the next lecture on the same course, you should go through your notes carefully and make sure that you understand them. If the notes are untidy, you may decide to rewrite them completely (but it is time-consuming if you always do this). Otherwise, annotate your notes by highlighting the points which seem important, and by adding marginal notes to clarify anything which is obscure. Compare your notes with corresponding material in whichever book(s) you are using for the course; you may want to make cross-references. If the books and notes formulate some result in different ways, make sure that you understand the relationship between the various versions. (Does each result imply the other? Is one version more general?) We shall return to this subject in more detail in Section 2.1.

Next, try some problems which appear to be related to the material in the lecture, using exercises from textbooks or from problem sheets given to you by the lecturer or your tutor. Each sheet is likely to cover more than one lecture, so some of the questions may depend on material which you have not yet covered. Therefore, you should not spend large amounts of time stuck on one problem—set it aside until after the next lecture, or until you have read more of the book, when it may become much easier; or just set it aside until tomorrow, when inspiration may come for no apparent reason. This method of attempting problems at an early stage may appear to be wasteful of your time by inducing you to attempt some problems before you have learned the relevant techniques. However, that time will not be wasted; the surest way of understanding a piece of theory is to find out how it can be applied to solve a difficult problem which you have previously attempted unsuccessfully! Moreover, by doing some problems at each stage, you will make it less likely that you will be short of time before a tutorial. We shall return to this subject in more detail in Section 2.2.

If you read ahead in a book, you will greatly improve your chances of following the next lecture, and therefore be likely to benefit much more from the lecture than if you encounter the material for the first time in the lecture.

## Resolving obscure points

It may be that, either during or after a lecture, you will come across something which you think is wrong (lecturers do make mistakes, some more often than others), or which you do not understand, even after consulting textbooks. Don't just put it aside saying "It's only a small point; it doesn't matter". Do one or more of the following:

- (i) interrupt the lecture to ask a question (this takes courage, but it may help your fellow-students),
- (ii) ask the lecturer at the end of the lecture (or the next lecture),
- (iii) ask fellow-students (you may find that they do not understand too clearly, either!),
- (iv) ask your tutor in your next tutorial.

## Summary

1. Go to all the lectures and take notes.
2. After the lecture, work through your notes, make sure that you understand them, and add clarifications of difficult points.
3. Attempt some problems related to the lecture.
4. Make sure that all points which you cannot work out for yourself are explained to you.

## 1.3 Tutorials

The arrangements for tutorials vary from college to college, but you are likely to have about two tutorials a week. Typically, there are two students (and one tutor) in each tutorial, but there may be variations. In some colleges, some tutorials may be replaced by classes of 4–10 students. For brevity, we shall only use the word “tutorial” here, but almost everything applies equally to classes.

Although you may be doing several lecture courses simultaneously, any one tutorial is likely to cover at most two of them, with the other courses being deferred until a later tutorial probably with a different tutor. This means that tutorials on any particular course will occur at intervals of one or two weeks. However, if you are seriously stuck on a problem, it is sensible to take any opportunity to get some help, for example by contacting your tutor between tutorials, or by asking another of your tutors at the end of a tutorial on other courses.

### During the tutorial

The principal function of a tutorial is for the tutor to fill you in on any parts of the course which are causing you difficulty. It is important to realise that the tutor is there to help you, and not to examine you. However, in order to help you, the tutor has to find out what you do not know or do not understand. Naturally, exposure of your ignorance may make you uncomfortable, particularly if you do not know the tutor very well, but it is likely to be disastrous if you try to cover up gaps in your knowledge or understanding, particularly if you succeed for a time. In the long run, you will be exposed, if not by the tutor, then by the examiner; meanwhile, much time will have been wasted.

There are various ways in which the tutor may find out what you need to have explained:

- (i) you can tell the tutor (this is the best way—it saves time!),

- (ii) you may have failed to solve correctly some of the set questions,
- (iii) the tutor may ask oral questions.

The purpose of oral questions is to see if you are following the argument, so that the tutor can recapitulate if necessary, or move on if not. These questions are not intended to be a test, or to trick you. You should not be afraid to give a partial answer, or even (after thinking for a while) to say that you do not know the answer or you do not understand the question. The worst possible outcome is 5 minutes of complete silence which the tutor is eventually obliged to break by answering his/her own question.

In practice, the majority of any tutorial is likely to be taken up with going through problems given to you in advance either by the lecturer or by the tutor. However, there should be an opportunity for you to raise other questions. Occasionally, if time permits, you may find the tutor takes the initiative to expand on some topic of his/her own choice. Some tutors will have told you to hand in your work beforehand so that they can mark it in advance.

You may find that tutors discourage you from taking notes during tutorials, as they want you to use the time to think about the mathematics. Most tutors will invite you to take away anything that they have written down on paper during the tutorial (if they don't, you should ask for it). There will only be one copy of this, so you will have to come to an arrangement with your partner about photocopying or about passing the notes on to each other. Some tutors may use a blackboard or whiteboard (and may expect you to participate on the board); feel free to ask their advice about taking notes.

### **Before the tutorial**

However you organise your work, you should start work set for a tutorial well before it need be completed. If you leave it all until the day before, you will very likely not leave yourself enough time, and you will certainly not give yourself any chance of having a second attempt if the first is unsuccessful. If the tutorials on the course are at fortnightly intervals, you should attempt some of the problems within the first week, interleaving with your work on other courses, in order to keep up with the lectures and to prevent a backlog. None of this means that you will have more work in the long run—you will have to do the problems sometime. It is just a matter of arranging your schedule in such a way as to get maximum benefit both from your tutorials and also from working consistently (and deeply) rather than in crises (when inevitably work tends to become superficial). See Section 1.2 for further comments on the value of doing problems regularly between lectures.

It is very good practice to meet your tutorial partner shortly before the tutorial to find out which parts of the set work each of you has done or failed to do. If there is some problem which your partner has solved and you have not, ask him/her to explain it to you. If he/she succeeds in this, you will have saved time in the tutorial. Surprisingly often, you will find that your partner cannot explain it satisfactorily, either because there is something definitely wrong in the solution, or because he/she does not properly

understand what he/she is doing. In that case, you will both discover the error or gap in your partner's work. Apart from avoiding considerable confusion in the tutorial, this process will greatly improve the understanding of both of you. Explaining a piece of mathematics to other people is a very effective way of getting to understand it. We shall return to this subject in Section 1.4.

If you and your partner have arrived at contradictory solutions, examine each other's solutions and see if you can find a mistake. Often, you will both find a mistake!

### **After the tutorial**

As soon as possible after the tutorial (preferably immediately), you should go through the problems which you failed to solve, and write out proper solutions. Any notes which you bring out of the tutorial will be sketchy; you should write out detailed solutions, as this will ensure that you think about the mathematics involved. Finding out how to solve a problem which had previously defeated you is a highly effective way of gaining a deep mathematical understanding (which you will retain for future use), provided that you take the trouble to understand the argument fully. Don't just file away the tutor's notes with your own (failed) solutions—that way, you learn nothing more.

### **Arranging tutorials**

The timings for tutorials are arranged at the beginning of term, but plans often have to be changed later. You will have some opportunity to express preferences over times, but this is likely to be severely restricted by tutors' many other commitments. Please don't say "I can only manage a tutorial at 4 o'clock on Wednesday"; but it is very proper to say "I'd prefer not to have a tutorial on Thursday, because I have already arranged one that day with Dr X"; you are likely to find tutors willing to adjust their arrangements to meet your non-academic commitments, within reason. Some tutors prefer to teach in the morning, some in the afternoon; you may find they are not keen to alter such a pattern.

The initial choice of pairings for tutorials in your first term will be highly arbitrary and experimental, as tutors have little reliable information by which to decide. After a while, you may find pairings changed as tutors find out who fits well with whom. These choices depend on many factors; mathematical ability is one of them, but it is certainly not the only one. Tutors will also take into account personality and compatibility; it helps if the students in a tutorial are on good terms with each other.

If you have a strong wish not to be paired with some individual, you should let your tutors know. Please, do not be excessively restrictive—"I refuse to have tutorials with anybody except Y" may not make you too popular with your tutor, even if it does please Y! However, it is legitimate to express positive preferences as to tutorial partners. Often, your preference will be granted, but occasionally the tutor may have some overriding reason to pair you in some other way.

## Summary

1. Tutorials are intended to fill in gaps in your knowledge, so you should not be afraid to ask questions or say that you do not understand.
2. Start work for a tutorial well in advance.
3. Discuss with your partner which topics need to be raised in a tutorial.
4. After the tutorial, make sure that you understand everything which was explained to you.

## 1.4 Cooperation with fellow-students

We have already described in Sections 1.1 and 1.3 the advantages of constructive cooperation with your fellow-students. There is no doubt that students who discuss their work sensibly with their peers enhance their work by doing so. Usually, these discussions take place in an informal setting, so it helps if you make friends with some fellow mathematicians. Of course, this does not mean that you should merely copy out other people's solutions—such a procedure on its own teaches you almost nothing, and it may be positively damaging to you if it deceives your tutors and causes tutorials to be pitched at a higher level than is appropriate for your own ability.

Cooperation between students can be highly constructive if it is done selectively. You should always try the problems on your own first, but if you are seriously stuck, try asking other students. Preferably, you should first explain to them what you have tried—sometimes, the act of doing this makes the way forward obvious—then you should ask for a hint or sketch of the solution. If you do have to go so far as asking another student to tell you the full answer, you should ask him/her to explain things orally (simultaneously writing down formulae, etc.), rather than asking for the written solution to be handed over. This process of explaining a solution can be very valuable for both parties, as it encourages them both to understand the mathematics involved. Cooperation can only be two-sided, so you should be prepared to help other students in a constructive way—make sure that they have made a decent attempt at the problems.

As mentioned in Section 1.3, it is particularly beneficial if you discuss work with your tutorial partner, especially if you discover beforehand which topics need to be raised in each tutorial.

## Summary

1. You are recommended to discuss your work with fellow-students, in a constructive way.

2. Do not ride on other students' backs by using their answers without making a determined effort yourself.

## 1.5 Books and libraries

Students sometimes seem to think that a complete set of lecture notes removes the need for books, or vice versa. This is a mistake; the content of books and lecture notes often overlap, but they fulfil different roles. Books last; lectures end within an hour. Books have room for more material than can be fitted into lectures, so you will often find extra explanation or information in the books; on the other hand lecturers can use tools such as vocal emphasis and visual aids to bring out points which are difficult to put across in print. While doing problems, you will sometimes find that you want to look up something which is not in, or not easily found in, your lecture notes; or perhaps you will just want to look in a book to see if you can find some useful idea (see Section 2.2 for more discussion of doing problems). Moreover, it can be a very valuable exercise to compare two or more different formulations of a result, one in your lecture notes and others in books, and to work out the relationship between them (see also Section 1.2).

In order to use books efficiently, it is necessary that you should have them readily to hand. Although college libraries will have a copy of all the basic undergraduate textbooks, they are likely to have only one copy of each book, so not all undergraduates can have access to the right library books at the right time. For these reasons, it is highly desirable that you should obtain your own copies of some books. You will be given lists of recommended books, and tutors will advise you which are essential. Both Blackwells and Dillons shops in Oxford are likely to have these books in stock when you arrive in Oxford (but they may disappear quickly); you may be able to buy second-hand copies, for example from Blackwells second-hand department, or from the book sale usually held in the Mathematical Institute a few days after your arrival (but second-hand copies will disappear extremely quickly).

You will probably find that some of the books which you are recommended to obtain during your first term will immediately be of great value for the courses being given at that time, while others appear at first not to be so relevant, and perhaps appear to be rather too advanced. Books in the former category will often only be useful for one particular course, while those in the latter category are likely to be important in the long run as reference texts which you may use intermittently for various courses throughout your undergraduate career.

Apart from college libraries, you can use the Radcliffe Science Library for reference and the Hooke Library for borrowing. The RSL and the Hooke are both parts of the Bodleian Library, although they are in separate buildings. In order to use the RSL or the Hooke, it is necessary to register at the Bodleian. Since you never know when

you will want to use these libraries, it is advisable to register there when you have the opportunity to do so during your first week in Oxford.

## Summary

1. Books play a valuable role in studying mathematics, beyond that of lecture notes.
2. Obtain your own copy of the important books, so that you always have them readily available.

## 1.6 Vacation work

“Vacation” and “holiday” are far from synonymous. You are expected to do some academic work during each vacation, although this will certainly be at a reduced level of intensity and the vacation may well include a period of holiday.

You will be given specific guidance about vacation work by your tutors, but the following are general principles:

- (i) First, complete work on the courses which were being given at the end of term; for example, you may have problem sheets still to do. These tasks should be done during the first week of the vacation.
- (ii) Next, recapitulate the whole of the courses which you did during the term. A good way to do this is to do some old examination questions on the course.
- (iii) Finally, read ahead for next term’s courses.

To encourage you to carry out these tasks, particularly (ii), tutors will sometimes set “Collections” (informal college examinations) at the beginning of term.

## Summary

1. Use the vacations to consolidate what you have done during the preceding term, and to read ahead for the coming term.

## CHAPTER 2

### University mathematics

Mathematics at university level involves (a) studying the theory, and (b) solving problems; you cannot be a good mathematician without mastering both these skills. Although there is, in some respects, a fairly clear distinction between these two aspects of university mathematics, it is wrong to think of them as separate activities. In order to solve any problem, you are likely to have to apply some theory, and it is essential that you understand the theory properly, both in order that you know which piece of theory to apply, and also to ensure that you do not misapply it. Moreover, many problems will themselves be rather theoretical, requiring you to prove things about general classes of mathematical objects, not just specific examples. It is equally true (but less obvious) that, in order to understand the theory in a way which impresses on you its significance and which will therefore lodge itself in your mind (as opposed to merely understanding the statements line-by-line), it is essential to solve some problems which either apply the theory or extend it in some direction.

In this chapter, we shall give some general guidance about these two aspects of mathematics, and about the need for good exposition in written work. In the next chapter, you will find an applied mathematician's version of similar topics. In Part II of the notes, we shall discuss in more detail two particular aspects of the theory of (pure) mathematics (the correct interpretation and formulation of mathematical statements, and proofs and their construction) which frequently cause difficulty for new undergraduates. Many aspects of Part II will also be relevant to problem-solving and exposition.

#### 2.1 Studying the theory

There are two words which are commonly used to sum up the difference between university mathematics and school mathematics: *abstraction* and *rigour*. University mathematics is abstract in the sense that, most of the time, we do not study specific examples (and perform calculations involving numbers), but we study classes of mathematical objects such as functions, matrices, vectors, and groups (and perform calculations involving symbols and variables, such as  $f, x, A, a_{ij}, V, \mathbf{v}, G, g, \dots$ , whose values are never specified). Our mathematics is rigorous in the sense that most of what we say or write has to be proved to be true.

The purpose of abstraction in mathematics (as opposed to arithmetical calculations) is to obtain general results which can be applied over and over again to many different problems. This is far more efficient in the long run than developing methods ad hoc for each different problem, or repeating similar calculations many times. An example which occurs in school mathematics is the formula for the roots of a quadratic equation. Apart from a few carefully-chosen examples, it is much quicker to apply the formula than it is to solve an equation by experimentation.

The purpose of rigour is partly to assure ourselves (and anyone who might want to apply our results, and is sufficiently interested) that our results are correct, partly to ensure that we understand why our results are true sufficiently well that we can adapt them to other situations which may arise, and partly to enable students to learn how to prove things in mathematics so that they can go on to prove new things. It is not essential that everyone who drives a car should understand the detailed workings of the internal combustion engine. However, it is essential that there should be designers and mechanics who do understand the engine, in order to be sure that the engine works properly, that it can be adapted to correct faults, and that its design can be improved. Analogously, it is not essential that every physicist, engineer, economist, or computer programmer who uses the products of mathematics should understand the detailed mathematics, but it is essential that there should be trained mathematicians available to help with these applications in times of need.

Your lectures and books will contain many, many, theorems and proofs, particularly in pure mathematics. Theorems are a convenient way to formulate in writing the mathematical ideas and theories which the course will explain to you; they are also important tools in solving problems (see Section 2.2). The abstraction is built into the formulation of the theorems, and the rigour appears in the proofs of the theorems and also in the precision with which they are stated. The fact that the roots of the quadratic equation are given by the formula is itself a theorem, but you will soon see that it is a rather elementary theorem. (A much more interesting theorem, which appears in the third year of the Oxford course, is that there is, in general, no formula of this type for the roots of a “quintic”, a polynomial of degree 5.)

Some of the theorems will state things which seem to be obvious (for example, that every positive real number has a positive square root), some of them will state things which you have been told some time ago, but have never known why they are true (for example, you may be aware that every polynomial has a complex root, but do you know why?), and some of them will state things which are far from obvious and which you would never have guessed in a month of Sundays, either because the statement is too complicated for you to have an immediate insight into it, or because it is a genuinely surprising fact (is it not surprising that there there is no formula for the roots of a quintic, when there are formulae for the roots of quadratics, cubics, and quartics?).

Some courses in applied mathematics (mechanics, methods, etc.) will not contain many theorems, and will have more of the flavour of school mathematics. They will be

developing “theory”, some of which could be formulated as theorems if you wished, and they will frequently be applying theorems from pure mathematics, without necessarily making this explicit. For more information about applied mathematics, see Chapter 3. The various degree courses contain different selections of topics in applied mathematics, and Part II of these notes is directed primarily towards topics in pure mathematics which are common to all mathematics degree courses.

You will have two main sources for the theory, lectures and books (see Sections 1.2 and 1.5), and you should use them both. In the case of theory presented to you in a lecture, you should sit down afterwards to compare the material with the textbooks, and to work around the theorems in the ways which are described below. When you are learning from books without lectures, you should sit at a desk with pen and paper to hand so that you can try out the ideas listed below.

The best way to get to understand theory is to experiment with it, and then to do some problems. When studying a theorem, you should aim to do as many as possible of the following:

- (i) Make sure that you understand the formal statement (see Chapter 4).
- (ii) Compare different versions of the theorem; does each imply the other? is one version more general?
- (iii) Try to prove the theorem yourself without reading any given proof.
- (iv) Make sure that you understand the given proof line-by-line, i.e. you understand the meaning of each statement, and why it follows from the previous statements.
- (v) Identify where in the proof each assumption is used.
- (vi) Identify the crucial ideas in the proof (mark them in the margin of your lecture notes, or note them on a separate sheet; this will be useful for revision).
- (vii) Try omitting one of the assumptions; does the conclusion of the theorem still hold? can you find an example to show that it does not?

The importance of (i) should be self-evident; Chapter 4 will give some guidance about difficulties which you may encounter. Steps (ii) and (vii) will help you to understand the significance of the hypotheses and the generality of the conclusions; (vii) will also give you practice in the skill of constructing counterexamples (see Section 5.1). The best way to understand a proof is to find it yourself, which explains (iii). The purposes of (iv) are to convince yourself that the theorem is true (if you are not already satisfied) and to teach yourself how to set out a proof accurately, clearly, and unambiguously. When you are reading somebody else’s proof, you should think consciously about the author’s reasons for writing the proof as he/she did, and you should look out for the points of difficulty such as those which will be discussed in Chapter 4. Step (v) will assist your general understanding of the theorem and its proof.

The purpose of (vi) is to develop your insight, and to enable you to reconstruct the

proof, and to construct variants of it, if required to do so. Any proof, however long it may be, depends on only one or two crucial ideas; the rest of the argument consists of details. It is futile (and counter-productive) to try to remember by heart all the details of all the hundreds of proofs which you will see, so the only practical way is to get to grips with the proof sufficiently deeply that you appreciate the essential ideas. These crucial ideas may be hidden in some preliminary lemma or proposition, so you may need to look around. If you can recall them (even if you have forgotten the details), you will be able to reconstruct the proof when required, and you will have improved your insight into the mathematics.

Let us illustrate this last point by considering the following standard proof that  $\sqrt{2}$  is irrational.

Suppose that  $\sqrt{2}$  is rational. Then  $\sqrt{2} = \frac{m}{n}$ , where  $m, n$  are positive integers. By dividing  $m$  and  $n$  by powers of 2, we may assume that they are not both even. Now  $m = \sqrt{2}n$ , so  $m^2 = 2n^2$ . Thus  $m^2$  is even, so  $m$  is even. Therefore  $m = 2k$  for some integer  $k$ , so  $4k^2 = 2n^2$ . Hence,  $n^2 = 2k^2$ , so  $n$  is even. Thus both  $m$  and  $n$  are even, a contradiction.

Hence, our assumption must be false, so  $\sqrt{2}$  is irrational.

The essential idea of this proof is that it is an argument by contradiction; the main subsidiary idea is that we can assume that  $m$  and  $n$  are not both even. The rest of the proof (everything after the first three sentences) is routine manipulation which follows naturally. If you remember the essential and subsidiary idea, you will undoubtedly be able to reconstruct the whole proof.

Let us return to the list (i)–(vii) above. You should be aware of the difference between (v) and (vii). The fact that a proof uses an assumption does not necessarily mean that the assumption is essential for the theorem; the theorem may be true even if the assumption is omitted, because there might be a cleverer proof which does not require the assumption. We shall return to this point in Section 5.1.

Finally, when you think that you have mastered the theorem (or a piece of theory), do some problems on it, before going on to the next topic.

## Summary

1. Mathematical theory at this level is usually abstract and rigorous.
2. When studying theory, carry out the tasks listed in (i)–(vii) above, to improve your insight into the theory and the proof.
3. If you understand the essential ideas of a proof, you are likely to be able to reconstruct the whole proof when required, and to construct variants of it.
4. When reading a proof line-by-line, think consciously about why the proof is written as it is.
5. When you think that you have understood a piece of theory, try some relevant

problems, to test yourself, and to improve further your understanding of the significance of the theory, and its applicability.

## 2.2 Problem-solving

Since one cannot be a mathematician without solving problems, you should do some at every occasion, between lectures, and at frequent intervals while reading. They may be exercises from a book, or problems from a sheet which you have been given. Where you get the problems from is not so important, provided that they span a range of difficulty, so that there are some which you expect to solve, and some which will really stretch you.

In solving a mathematical problem, there are usually three stages:

- (1) understanding the problem;
- (2) experimentation (to find a method which looks likely to work);
- (3) verifying and writing out the solution.

These stages will be examined more closely in Chapters 3 and 5, but for the moment we will merely make outline some general principles.

Suppose then that you are faced with a taxing problem. Here are some guidelines as to how you might proceed if you find the problem difficult. These suggestions are listed in roughly the order in which they should be tried, but not all of them will be suitable for any given problem.

- (i) While searching for a method, do not worry about details; fill them in later, if you can.
- (ii) Try working backwards from the answer (if this is given); this may show you the connection between the assumptions and the conclusion; eventually, you will have to turn the argument around so that it runs in the right direction.
- (iii) Draw a diagram, if appropriate (many students are very reluctant to do this). Although diagrams rarely constitute a solution in themselves, they often show you something which you would not otherwise have noticed.
- (iv) If the problem is a general one, try special cases, or simpler versions of the problem.
- (v) Make sure that you have read as much as possible of the relevant theory. Re-read your lecture notes and books; write down on a single sheet of paper all the information which appears to be relevant.

- (vi) Look in your notes and books for similar worked examples.
- (vii) If you are still stuck, put the problem aside and come back to it later (e.g. the next day). Surprisingly often, the solution will then stare you in the face.
- (viii) If you are still stuck, ask your fellow-students (see Section 1.4).
- (ix) If the worst comes to the worst, ask your tutor.

Let's be cheerful, and suppose that somehow or other you have found a method which you think may work. Inevitably, there will be some details which you have skipped over, or which you are not completely confident about, so there remains the task of writing out a complete and accurate solution. For example, you may have to reverse the direction of your argument, or you may have to set out a formal proof by induction. This stage also serves to verify your solution, in the sense that it shows you whether the method actually works. It may be that when you try to write the solution out, you find there is some unexpected difficulty which you had overlooked, in which case you will have to return to experimentation. We shall describe some of the general principles of writing mathematics in Section 2.3, and some particular aspects will be discussed in detail in Chapters 4 and 5.

Even if you think that you have solved a problem correctly and arrived at the correct answer, pay attention to any solution given in the back of the book or in tutorials. If the other solution looks different from yours, think about whether yours is equally valid; if your solution now looks dubious, make sure that you understand what is wrong.

### Summary

1. While studying mathematics in any form, always do some problems at frequent intervals.
2. When attempting a difficult problem, try the methods listed in this section ((i)–(ix)).

## 2.3 Writing mathematics

The third aspect of doing mathematics is the written exposition of your work. As a student, you should always aim to write out your work in a complete, accurate, and clear fashion. This applies equally to writing out theory and to solutions of problems.

There are several reasons why good exposition is important. Firstly, when you come to write down details of problems, you may find some difficulties which you had overlooked in your rough working. Secondly, any tutor or examiner who is going to read your work will need evidence that you really understood what you were doing. Thirdly, you should aim to acquire the ability to communicate mathematics to other people; this

will obviously be a valuable skill if you follow a career which involves mathematics, and it will also enhance your ability to write unambiguously and logically in any occupation. In mathematics, accuracy is much more important than literary quality!

Although your work may be read only by your tutor, it is a good idea, when you are writing your solutions, to imagine that you are writing for a reader who has a similar level of mathematical knowledge to yourself but who has not encountered the argument which you are writing down. Such a reader should be able to understand what you have written without danger of ambiguity, to see that it is correct, and to see why it is correct.

The English language (in common with other languages) is prone to ambiguity, and mathematicians have to be particularly careful to use language (and symbols, which are part of mathematical language) in an unambiguous way. Good punctuation can be very helpful; for example, commas can be used to separate clauses. Sometimes, mathematicians go beyond normal punctuation by using brackets to remove ambiguity. The use of brackets inside algebraic expressions is commonplace (consider the difference in meaning between  $(xy) + z$  and  $x(y + z)$ ); it is a mathematical convention that the unbracketed expression  $xy + z$  means  $(xy) + z$ , so if we mean  $x(y + z)$ , we have to include the brackets). However, brackets can also be used around entire clauses, to remove ambiguity. For example, the statement:

$$x = 0 \text{ and } y = 1 \text{ or } z = 2 \text{ and } t = 3,$$

is highly ambiguous, but the following are not:

$$[x = 0 \text{ and } y = 1] \text{ or } [z = 2 \text{ and } t = 3],$$

$$x = 0, \text{ and } [y = 1 \text{ or } z = 2], \text{ and } t = 3.$$

When you write out your final solution, make sure that it is fully comprehensible, and that each step follows logically from previous ones. What you write should be clear, intelligible and unambiguous, as well as being logically correct and mathematically accurate. In particular, when all the symbols are converted into English, your script should consist of complete sentences. You should pay particular attention to points such as the following:

- (i) If you are arguing by contradiction or induction (see Sections 5.7 and 5.8), say so clearly.
- (ii) If you make some assumptions, say so clearly, by including phrases such as “Suppose that ...”.
- (iii) Be careful to use phrases such as “if”, “only if”, “if and only if”, correctly (see Section 4.2). If you use symbols such as  $\implies$ ,  $\impliedby$ , and  $\iff$ , make sure that they go in the direction which is both logically correct and relevant to the problem (this may be the opposite of the direction in your rough working).
- (iv) Include phrases such as “and”, “or”, “for all”, “for some” (see Sections 4.3 and 4.4); you cannot expect the reader to know which you mean if you omit

these phrases.

- (v) Make sure that your argument shows correctly which quantities depend on which other quantities (see Section 4.5).
- (vi) If the statement which you are writing down depends on something several lines earlier, make this clear to the reader, for example by labelling the earlier statement with (\*) and writing “It follows from (\*) that ... ”.
- (vii) If your proof involves a division into two or more cases, make it clear what the different cases are (see Section 4.3).

See Chapter 4 for further discussion of several of these points.

### **Summary**

1. When writing out a solution, ensure that your argument is complete and accurate, and that, when the symbols are converted into English, what you have written makes full sense grammatically and is logically correct.
2. When writing out a solution, pay attention to the points listed above ((i)–(vii)).
3. Good punctuation, particularly the use of commas and brackets, may remove ambiguities.

## CHAPTER 3

### The perspective of applied mathematics

by Nick Woodhouse

This chapter describes some of the differences between pure and applied mathematics. Nevertheless, almost all the advice is relevant to all branches of the subject.

#### 3.1 Pure and applied mathematics

Although the other chapters of these notes are written mainly from the viewpoint of pure mathematics, much of what is said is also relevant to applied mathematics. Nonetheless there are some important points of difference, and you should bear these in mind when you start working for tutorials in applied mathematics.

The first is an obvious point about motivation. Pure mathematicians solve problems to answer questions in mathematics, applied mathematicians solve problems to answer questions outside mathematics—in physics, biology, economics, and so on. However, this distinction is often blurred and in many cases amounts to no more than a difference of emphasis. It is certainly not one of mathematical content: questions about differential equations in physics generated the development of new fields of pure mathematics; questions in logic about the axiomatic foundations of mathematics sparked the development of the computer industry. You will not see at university the clear boundaries between ‘pure’ and ‘applied’ that were familiar to you in your schoolwork, and some of the lectures (for example, on geometry) will be hard to classify as belonging exclusively to one camp or the other.

The same problem can be considered from both a ‘pure’ and an ‘applied’ perspective. Pure mathematics will be more concerned to answer questions like: What is going on here? How does it fit in with the rest of mathematics? Does the result generalize? This leads to an emphasis on establishing general propositions, and on the importance of precise formulation and of formal proof. Applied mathematics will concentrate more on questions like: What can I do with this? Is it a useful technique?. But different questions can often lead in the same direction.

The second point is that ‘applied mathematics’ covers a very wide range of disciplines. At one extreme, it is not very different from pure mathematics: you are studying questions that have been thrown up by particular practical problems, but which are of such universal importance and are so intriguing to mathematicians that they have taken on a life of their own and are no longer treated in the context of the original

applications. The theory of differential equations is a good example. At the other extreme, there are subjects like statistics and numerical analysis, where the applications are always at the forefront. Here very considerable skill is often needed to formulate the questions in appropriate mathematical terms.

With this diversity goes a wide range of reasons for including ‘applications’ in the undergraduate syllabus. First, there is the ‘cultural’ one. Mathematics is not a pure and abstract creation: every branch can be traced back, ultimately, to some practical problem. Good mathematicians should be aware of the origins of their subject. Much of modern mathematics (including calculus) has its roots in the classical areas of applied mathematics that you study in the first year (mechanics, heat, gravity, and so on). Second, it is there to develop skill in ‘methods’, such as algebraic manipulation, integration, equation solving, and geometric reasoning. These are skills that are needed to underpin all later development in mathematics, pure or applied. Third, it is there to develop skills in modelling—in translating into mathematical language problems posed in non-mathematical terms.

## Summary

1. The difference between pure and applied mathematics is often one of motivation and emphasis, rather than mathematical content.
2. In pure mathematics, you will concentrate on techniques for developing proofs. In applied mathematics you will concentrate on modelling and methods of problem solving. The two disciplines are not disjoint.

## 3.2 Solving problems in applied mathematics

The general principles described in Section 2.2, apply to all of mathematics. In particular, there are three stages in problem-solving:

- (1) understanding the problem;
- (2) experimentation;
- (3) writing out the solution.

Parallel with these there are three different levels of understanding the solution: first, seeing the solution for yourself; second, seeing the solution clearly enough to explain it to a friend; and, third, seeing it clearly enough to write it down in good mathematical style.

## Understanding the problem

Whereas in pure mathematics, stage (1) is often just a matter of understanding a

formal statement (see Section 5.3), in applied mathematics it can be more complicated, and it often contains the main content of the question. You have to understand the terms and to identify what it is you are being asked to do, but in addition, you may well have to begin by translating the problem into mathematical language. Sometimes this involves a simple ‘decoding’ of a fairly standardized description. For example, if you are told that a rod is “light”, then you set its mass to zero; if you are told that something is chosen “at random”, then it means that all choices are equally probable. At other times, you will not be given in explicit terms all the information that you need, and you will have to decide for yourself on the most appropriate assumptions. For example, you may be asked to find the frequencies at which a violin string can oscillate. You will have to think out for yourself that it is sensible to assume that the endpoints of the string are fixed, that the tension and density are constant along the string, and that the transverse displacement is small compared with the length of the string.

Setting up the problem in mathematical language will involve introducing notation. Clumsy notation can mislead you and can make it hard for others to understand your work. Follow standard conventions wherever possible (“ $m$ ” is a mass, “ $F$ ” is a force, not the other way round). Explain your notation clearly, if possible at the beginning of your solution (“Let  $A$  denote the event that ...”). Respect the symmetries of the problem: if, for example, you are considering the motion of three particles with very similar equations of motion, it will be more helpful to denote their masses by  $m_1$ ,  $m_2$ ,  $m_3$  than by  $m$ ,  $M$  and  $x$ . Otherwise, at a later stage, you may find yourself writing out the same argument three times over, and may miss the opportunity to take a shortcut by saying “by permuting the subscripts, ...”. Remember that notation is your servant, not your master. You do not have to use the same notation as the lecturer if you do not like it, and you should not hesitate to rework a problem in a different notation if your initial choices have not turned out well.

A very important rule is to keep clear in your mind the distinction between setting up a problem and solving it. If, for example, you are set a probability problem about rolling dice, you will find it very difficult to think clearly and to express yourself clearly if you write things like

$$P(\text{Three sixes given the first throw was 2}) = \dots$$

This will look messy and long-winded. Abbreviation will lead to a lack of precision, and to mistakes. You will be shown in the lectures how to set up a problem like this in the language of probability theory. You will be shown how to begin your solution by introducing a probability space and defining the relevant events as subsets of the probability space (e.g. “ $A$  is the event that there are three sixes in the first four throws”). The mathematical argument will then be within the framework of abstract probability theory. You will write things like,

$$P(A|B) = \dots,$$

and the logical basis of your argument will be very much clearer. (The “|” in the expression on the LHS denotes conditional probability: you read it as “the probability of  $A$  given  $B$ ”.)

Similarly, in mechanics, you should keep separate the physical principles that you need to translate the problem (which usually involves setting up a system of differential equations), and the subsequent analysis (solving them). For example, if you start by translating Newton’s second law into a differential equation such as

$$m \frac{d^2x}{dt^2} = -\frac{k}{x^2},$$

you should not, if you subsequently get stuck, inject into the argument at a later point

$$\text{“by conservation of energy, } \frac{1}{2}m\dot{x}^2 - \frac{k}{x} = \text{constant”}$$

because this is actually a consequence of the original second-order equation. Of course, it is perfectly acceptable to prove from the second-order equation that the expression  $\frac{1}{2}m\dot{x}^2 - k/x$  is constant, or to quote a theorem about second-order equations that implies that it is constant. It is also helpful to point out that the conserved quantity is the total energy. If the wording of the problem allowed it, you could also use energy conservation instead of Newton’s second law to make the initial translation into mathematical language. The mistake is to confuse the process of setting up the model with the mathematical analysis, by using a physical argument in the middle of a mathematical argument. The price you pay for such confusion is to lose track of the assumptions from which your conclusions are drawn; you also risk making contradictory assumptions.

At the end of the setting up process, you should have something that looks much more like a mathematical question. For example, a problem about smooth rods and particles has now been translated into a more straightforward instruction: “Solve the following differential equations subject to the following initial conditions”. In other contexts, where you have been set an exercise in method as opposed to application, the problem may be given to you in this form in the first place.

Before launching into the ‘experimental’ stage, you should first ask “is it plausible that this problem has a solution?” One very important guide, which can be used in a wide variety of situations, is to see if the number of equations matches the number of unknowns. If you have introduced three coordinates to describe the position of a particle, and your application of Newton’s second law has given you only two differential equations, then no amount of manipulation will enable you to solve the problem. Go back and see where the extra information is going to come from. If you have more equations than unknowns, then it is possible that the equations are inconsistent. Double check!

## Experimentation

The advice given in Section 2.2 applies to both pure and applied mathematicians. “Try some special cases, examples, or simpler versions of the problem” is especially

good advice for applied mathematicians. If you cannot solve a differential equation as set, try putting some of the parameters to zero or dropping an awkward term. This may suggest a clever substitution that solves the original problem. If you are set a complicated probability question involving the tossing of  $n$  coins, get a feel for the problem by solving it first when  $n$  has some definite low value.

I shall just add a few tips on how to avoid one trap that undergraduate mathematicians often fall into: covering many sheets of paper with long calculations that do not seem to get anywhere. You will meet very few problems in your first year that cannot be solved on a single side of A4.

- (i) If you seem to be getting nowhere, pause and check that what you are trying to do is sensible. Are you trying to solve a system of equations with more unknowns than equations? Is what you are trying to show obviously untrue, because of the existence of a simple counterexample? Are you, for example, trying to find an algebraic expression for a transcendental function? Has the problem been carelessly set? (This happens from time to time—you must learn to spot nonsense problems, so that you do not waste time.) If it is not sensible, go back to the beginning and check that you have not missed something in the first stage.
- (ii) If what you are trying does not seem to work, do not tear it up. Try a few other approaches, and compare them.
- (iii) It is very easy, when you are not fully confident about what you are doing, to rush and calculate carelessly. There is no cure for this—it is human nature—but remember that you waste far more time trying to track down mistakes than in carrying out the calculation slowly and carefully in the first place.
- (iv) Check that you are not trying to do more than you have been asked to do. Two examples will illustrate this. In probability theory, there are some very slick methods for finding expectations of random variables without first finding their distributions (for example, by using indicator random variables). If you try to find a distribution when all you are asked for is an expectation, you may be adding unnecessarily to your task. In the theory of central orbits, you can often find the perihelion (point of closest approach to the centre) by deriving conservation of energy and then setting the radial velocity to zero, instead of following what can be a very much harder route of finding the trajectory and then minimizing the radial distance.
- (v) Put the problem aside and come back to it the next day. (This is one of many reasons for not leaving your work until the last minute.)
- (vi) Ask a friend or go and knock on your tutor's door.

Remember: the Oxford course is designed to stretch very able students. You learn more from solving one hard problem for yourself, even if it takes several days, than in knocking off twenty easy ones.

## Summary

1. In applied mathematics, understanding the problem is often the main task.
2. Separate the stage of setting up the problem from the stage of solving it.
3. Check that you have enough information to solve the problem.
4. Do not pursue long calculations if you are not making progress.

### 3.3 Writing out the solution

Section 2.3 discusses in general terms the task of writing out the solution accurately and completely. When you read Chapter 5, you will find some examples taken mainly from pure mathematics, where the task is to write out a proof at the appropriate level of ‘rigour’. In applied mathematics, very little (but still some) of your work will involve writing out formal proofs of precisely formulated general propositions. You will spend much more time developing techniques to solve particular problems, without trying to push up against the limits of their applicability. Aiming for rigour in your applied work can lead you into inappropriate and distracting arguments.

In your analysis courses you may well work through proofs of very simple and familiar propositions, such as that the derivative of the function  $x \mapsto x^2$  is the function  $x \mapsto 2x$ . It takes a while to see why it is important to do this: it is that you need to put the foundations of calculus on a sound basis in order to build on them later. If you rely on intuitive and vague definitions of concepts like limit and derivative, you will eventually come up against the same barriers that stalled the development of mathematics in the eighteenth century. In pure mathematics, the danger is that, in exploring the foundations, you accidentally allow in intuitive arguments based on your informal understanding of the concepts you are playing with. Hence the emphasis on the discipline of rigour, of taking great care to check that your argument follows logically, step-by-step from the definitions. There is a corresponding danger in your applied work that you will be so unnerved by concern for rigour that you will be unable to write down “if  $f(x) = x^2$ , then  $f'(x) = 2x$ ” without proving it. Taken to its extreme, this attitude would imply that you could not use integrals at all in your first two terms because you do not begin the rigorous theory of integration until your third term.<sup>1</sup> This would be to misunderstand the purpose of rigour, which is precisely to allow you to use the familiar rules of calculus (and the less familiar ones that you will meet in the coming year) with confidence. The purpose is to reinforce that confidence, not to undermine it.

---

<sup>1</sup>I am told that there is a French mathematical text that does take this attitude to its extreme, by apologising at the beginning of the book for using the natural numbers to label the pages before they have been defined.

It is very easy to misinterpret these remarks as saying that “rigour has no place in applied mathematics; any argument, however sloppy, will do”. That is wrong. All that is being said is that to sum up the instructions for writing good *applied* mathematics under the heading of ‘rigour’ would be misleading. The important point is that your written work should be *clear*, *accurate*, and *concise* (I shall expand on this below). Rigour is, in any case, a relative term. One of the disconcerting lessons of twentieth century mathematics is that, if you insist on proving everything by strict rules of inference from a finite set of axioms, then your mathematical horizons will be very limited. A celebrated theorem of Gödel’s implies that you will not even be able to prove rigorously all the propositions in arithmetic that you know to be true by informal argument. In almost all your work, pure and applied, you will construct arguments by building on other results that you take for granted. Read what is said about this in Section 5.6 and look carefully at some of the examples in Chapter 5 when you work through them. In your applied work, you will build on the results you prove in your pure work.

### Clear writing

A solution to a mathematical problem takes the form of an argument. It should be written in good grammatical English with correct punctuation. It should make sense when you read it out loud. This applies to the mathematical expressions as well as to the explanatory parts of the solution. A solution should never take the form of a disconnected sequence of equations. When you are writing up a calculation, you must make clear the logical connection between successive lines (does the first equation imply the second, or the other way around?), and you must make the equations fit in a coherent way into your sentences and paragraphs (see the remarks in Section 5.5). One minor point: many mathematicians are poor spellers. If you are not confident about your spelling, keep a dictionary on your desk, and use it. A good way to improve your style is to browse occasionally in Fowler’s *Modern English usage*.

You will never write a clear argument unless you are clear in your own mind what it is that the argument is proving, and you make this clear to your reader. Sometimes it is simplest to do this by using the ‘proposition-proof’ style of pure mathematics, but more often this is inappropriate in applied mathematics (it looks very artificial to formulate as a ‘proposition’ a result that is special to a particular problem). It is very helpful, both for yourself and for your reader, to say what you are going to do before you do it, for example, by writing something like “We shall now show that every function given by an expression of the form ... satisfies the differential equation”. It is also helpful to break up a long argument into short steps by using sub-headings. For example, if you are asked to show that “ $X$  is true if and only if  $Y$  is true”, then you might head the first part the argument with “Proof that  $X$  implies  $Y$ ” and the second with “Proof that  $Y$  implies  $X$ ” (see Section 4.2).

Of course it is in the nature of much of applied mathematics that you have to do quite a lot of thinking to extract from the original formulation of the problem the precise statements that you have to establish. Even something as apparently simple as

*Solve the equation  $y'' = y$*

contains potential pitfalls. The correct answer is:  $y = Ae^x + Be^{-x}$ , where  $A$  and  $B$  are arbitrary constants. Your argument in this case will establish two things: first, that every function of the form  $y = Ae^x + Be^{-x}$  satisfies the equation, and, second, that every function that satisfies the equation is of the form  $y = Ae^x + Be^{-x}$ . In a simple example like this, you can establish both statements by the same short calculation, and it would be rather fussy to break the argument up into two parts. But you do not have to go much further to find an example in which you can go badly wrong if you simply manipulate without thinking. Consider the following problem.

*Solve the simultaneous differential equations  $y' = z$ ,  $z' = y$ .*

If you ignore what has been said above, you might be tempted to write the following.

*Solution.*

$$y' = z, \quad z' = y$$

$$y'' = z', \quad z'' = y'$$

$$y'' = y, \quad z'' = z$$

Answer  $y = Ae^x + Be^{-x}, \quad z = Ce^x + De^{-x}.$

However, if you take particular values for the ‘arbitrary constants’  $A$ ,  $B$ ,  $C$ ,  $D$ , say,  $A = 1$ ,  $B = 0$ ,  $C = 0$ ,  $D = 0$ , then you rapidly find that your ‘answer’ does not work. The problem is that it is not clear what the ‘argument’ in the solution is establishing. If you look carefully, you will see that the first line implies the second line, that the first two lines together imply the third, and that the third implies the fourth. However, when you try to go backwards, you cannot get from the second line to the first. To solve this problem without confusing yourself and your reader, you should break it up into two parts, along the following lines.

*Solution.* Suppose that  $y$  and  $z$  are functions such that  $y' = z$  and  $z' = y$ . Then

$$y'' = z' = y, \quad z'' = y' = z.$$

But the general solution to the differential equation  $y'' = y$  is  $y = Ae^x + Be^{-x}$ , where  $A$  and  $B$  are constant. Therefore  $y$  and  $z$  must be of the form

$$y = Ae^x + Be^{-x}, \quad z = Ce^x + De^{-x},$$

for some constants  $A$ ,  $B$ ,  $C$  and  $D$ . [This does *not* say that all  $y$  and  $z$  of this form are solutions.]

Conversely, suppose that  $y$  and  $z$  are of this form. Then

$$y' = Ae^x - Be^{-x}, \quad z' = Ce^x - De^{-x}.$$

Therefore  $y' = z$  and  $z' = y$  if  $A = C$  and  $B = -D$ .

We conclude that the general solution is  $y = Ae^x + Be^{-x}$ ,  $z = Ae^x - Be^{-x}$ , where  $A$  and  $B$  are arbitrary constants.

Even in such a very simple problem, you must think carefully about the logical structure of your argument if you are going to avoid mistakes. Two final remarks: (1) you should have spotted in the initial stage of thinking about this problem that the solution would contain two arbitrary constants, and not four; (2) if you think about rigour, then there are a lot of other questions that come up here, such as “What are the domains of the functions? Are they twice differentiable?”, and so on. A skill that you must learn by practice is that of finding the appropriate level of rigour. There are no absolute rules. What is appropriate depends on the context. Too little rigour, and you will end up saying things that are untrue; too much will be distracting.

### Accurate manipulation

For the most part, this is a matter of taking trouble, and giving yourself time to check your work carefully. However, the following tips may help.

- (i) If you have sketched out a solution in rough, do not just make a fair copy of your notes. Having seen how to solve the problem in rough, put your notes aside, and write out the argument from scratch, checking each step carefully as you go. It is very hard to see mistakes when you are simply copying, rather than reworking.
- (ii) If you are prone to making careless slips, treat your work with suspicion, and look for ways of seeing that bits of it must be wrong. In mechanics problems, you can often pick up sign errors by thinking about whether your answer makes good physical sense. If your trajectory for a projectile implies that it falls upwards rather than downwards, then you have made a mistake. If you have found a formula involving parameters for the probability of some event, check that it always gives an answer between 0 and 1. If a particular choice of parameter values gives a negative probability, then you have made a mistake. Test your conclusions against common sense. If your calculation gives 2.31 as the expected number of tosses of a coin needed to get three heads in a row, then you have made a mistake.
- (iii) Wherever possible, check your answer. For example, always substitute your solution back into a differential equation (this will also pick up the sort of problem we met in the example above). If this is too messy, try doing it with special values of the constants. It may be too much of a labour to check that

$$y = \cos^{-1}(\sqrt{k \log(\sec^2 x + x^x) + x^2})$$

satisfies some complicated differential equation, but it may be much easier when the constant  $k$  vanishes.

- (iv) Never bring a rough and sloppy piece of work to a tutorial in the hope that your tutor will not spot the mistakes. The chances are that your tutor will not

spot the mistakes, in which case nobody learns anything; if your tutor does spot them, then your tutor learns something about you, but you learn little about mathematics. If you are uncertain about something you have written, even after spending a lot of time on it, draw it to your tutor's attention, and ask to go through it carefully together.

## Concise writing

Pascal wrote in a letter to a friend "I have made this letter longer than usual, only because I have not had time to make it shorter". Long-winded arguments in mathematics are hard to follow. It is well worth spending time and effort to extract the essential steps from your original rough notes and to present your argument in as concise as form as is consistent with clarity. But do not be surprised if you find it hard work, sometimes much harder than spotting the solution in the first place.

There are two points to keep in mind here: first, tutors very rarely criticise undergraduate work for being over-concise. Second, a good test to apply is the 'text-book test': "if my solution were printed as a worked example in a text-book, would I find it helpful and easy to follow"

## Diagrams

One other topic needs special mention: the use of diagrams. It is a classic howler, and one that is almost too easy for a tutor to spot, to argue "from the diagram it is obvious that . . . ". If you try it, you will be left in no doubt that such arguments cannot be rigorous (incidentally, you should in any case avoid phrases like "it is obvious that": either it is obvious, in which case you do not have to say that it is obvious, or it is not, in which case you should be more honest). Undergraduates who have been chastened by the scornful reaction to a 'diagramatic' proof sometimes draw the false conclusion that diagrams have no place in mathematics. This is quite wrong. First, you should use diagrams freely in the experimentation stage. Second, it is quite legitimate to use diagrams to help your reader to follow an argument, even in the most abstract parts of pure mathematics. The mistake is to make deductions from a diagram. It is similar to the mistake of arguing a general proposition from a single example. In fact, it is bad practice to suppress diagrams which have played an important part in the construction of your solution, particularly if you give away the fact that you have a diagram in mind by using phrases like "above the  $x$ -axis" or "in the upper half-plane". Finally, it is quite legitimate to use a diagram to set up notation. For example, there is nothing wrong with using a diagram in a geometric proof to specify the labelling of points ("where  $A$ ,  $B$ ,  $C$  are as shown . . . "); or in analytical proof to give the definition of a function that takes different constant values in different ranges (a picture here may be much easier to take in at a glance than a long sequence of expressions with inequalities defining the various parts of the domain). The only test you must satisfy is: is the use of the diagram clear and unambiguous?

## Summary

1. In applied mathematics, write your solutions in a way which makes it clear what you are showing, but avoid excessive concern for rigour.
2. Write your answers grammatically and concisely.
3. Check the accuracy of your calculations by verifying that the answer is sensible.
4. Use diagrams freely for experimentation and illustration, but not for deduction.

# PART II

## CHAPTER 4

### The formulation of mathematical statements

In everyday language, many statements do not mean literally what they say, either by design or by oversight on the part of the speaker or writer. For example, “I’ll be with you in a second” never means that the speaker will be available in exactly one second; after all, it takes more than a second to say it. However, the situation in mathematics is very different—mathematical statements have very precise meanings, and there is no room for metaphor, hyperbole, or other figures of speech. Moreover, some of the terms used will have specific mathematical meanings, which should have been defined (for example, “vector”, “group”, “continuous function”), while others will express the same logical concepts as in ordinary English language, but be used in a very precise way (for example “if”, “for all”, “or”). The statements are often quite complicated in structure, with several dependent clauses; a small change in the syntax of a statement often produces a large change in the mathematical meaning, which may change a true statement into a fallacy. At first, undergraduates often have difficulty grasping the precise meaning of mathematical statements made in their courses, but it is extremely important to learn to master the formulation of mathematical statements, both in order to understand the meaning of statements made in books and lectures, and also in order to write down mathematics accurately and unambiguously. In this chapter, we discuss various points concerning the way mathematical statements are commonly formulated, in particular the precise use of everyday words to express logical ideas.

There are two points of detail which we should make about the material in this chapter. Firstly, the phrase “mathematical statement” or “statement” is not intended to encompass everything that is written in a textbook or said in a lecture or tutorial. It refers only to formal mathematical statements, such as the statement of a theorem, proposition, lemma, or corollary, or part of a formal proof, or a formal definition, or certain other statements. Books and lectures will include other material whose role is to motivate the theory, or to explain it informally; this material may include some imprecise statements. Tutorials, and other forms of discussion of mathematics, may well include many very imprecise statements!

Secondly, we shall sometimes refer to two statements, A and B, as having the same meaning, or being synonyms. This indicates that the statements are equivalent for purely formal logical reasons, involving only use of language, notation, and definitions; it has nothing to do with any mathematical content of the statement. In other words, A and B are different ways of expressing the same logical concept. This concept may

or may not be true. However, if A is true then B must be true (because they have the same meaning); if A is false, then B is false. Any statement can be expressed in many different ways in English (i.e., there are many statements which have the same meaning); in our examples, we shall only be able to mention a few.

To assist the reader, we shall label synonymous statements in the form S1a, S1b, S1c, ... whereas statements with different meanings will have different numbers (e.g. S1a and S2).

## 4.1 Hypotheses and conclusions

Most mathematical statements, in particular most theorems, take the form of making certain assumptions (hypotheses) and drawing certain conclusions (consequences). The hypotheses and/or conclusions may be quite complicated; each may come in several parts. For ease of language, the statement may involve several sentences, and the hypotheses and/or conclusions may themselves be spread over more than one sentence. The statement is true provided that whenever all the hypotheses are satisfied then all the conclusions are satisfied.

For example, consider the following statements:

S1a: *Every positive real number has a unique positive square root.*

S1b: *Let  $x > 0$ . There exists a unique  $y > 0$  such that  $y^2 = x$ .*

Provided that the reader is aware that the context of S1b is real numbers, these statements are synonymous, because, for example, a square root of  $x$  is, by definition, a number  $y$  such that  $y^2 = x$ , and “ $x > 0$ ” is conventional notation for “ $x$  is a positive real number”. Because S1a and S1b mean the same, it follows (without thinking about the mathematical truth of the statements) that if one of them is a true mathematical statement, then so is the other. Of course, as you well know, both statements are true, but that is a special fact about real numbers. Indeed the following statement, which is very similar to S1a in structure, is false:

S2: *Every positive rational number has a unique positive rational square root.*

Thus, the fact that S1a and S1b are true is a fact of mathematics, depending on properties of the real numbers, while the fact that they are synonyms is a fact of language and logic (given the definitions and notation).

Statement S1b is rather more formal than S1a, and the distinction between the hypothesis ( $x > 0$ ) and conclusion (there exists a unique  $y > 0$  such that  $y^2 = x$ ) is clearer in S1b; nevertheless, S1a has the same hypothesis as S1b, and the same conclusion as S1b. Moreover, S1a has the merit of mentioning explicitly the context of

real numbers. Which you choose to write is mainly a matter of taste and context.

You should be aware that the conclusion of statement S1b (or S1a) actually says two things:

- (i) there exists *at least one*  $y > 0$  such that  $y^2 = x$  (existence),
- (ii) there exists *at most one*  $y > 0$  such that  $y^2 = x$  (uniqueness).

Statements S1a and S1b are not only true statements, they are theorems, or more precisely two equivalent statements of the same theorem. You may be surprised about this: the existence of square roots of positive real numbers is not a law of nature, but is something which can be deduced from a few axioms about the system of real numbers. (Most of the axioms are simple laws of arithmetic, e.g.  $a(b + c) = ab + ac$ , or laws concerning the ordering, e.g. if  $a \leq b$  and  $0 \leq c$ , then  $ac \leq bc$ ; however, there is one more complicated axiom which distinguishes the real numbers from the rational numbers, say—see *Alice in Numberland*, Chapter 8, for example.)

To show that S1a and S1b are true statements, it is necessary to prove the theorem. Since the statements have the same meaning, it is sufficient to prove one of them. Since the conclusion has two parts, (i) and (ii) above, the proof will also have two parts. The proof of (i) depends on the complicated axiom of the real numbers, and we will not give it here. However, the proof of (ii) is quite short, and uses only the simple axioms, or familiar properties of real numbers easily derived from the axioms, so we will give the proof here, for the interest of readers who may be becoming bored by the absence of mathematics.

Suppose that  $y_1 > 0$ ,  $y_1^2 = x$ ,  $y_2 > 0$ , and  $y_2^2 = x$ . [Our objective is to show that  $y_1 = y_2$ , which will complete the proof of (ii).] Then

$$0 = x - x = y_1^2 - y_2^2 = (y_1 - y_2)(y_1 + y_2).$$

But  $y_1 + y_2 > 0$  (since  $y_1 > 0$  and  $y_2 > 0$ ). Since  $y_1 + y_2 \neq 0$ , we may divide by  $y_1 + y_2$ , and deduce that  $0 = y_1 - y_2$ , so  $y_1 = y_2$ , as required.

To show that S2 is false, it is sufficient to give one example which satisfies the hypothesis of S2, but not the conclusion. In other words, it is sufficient to exhibit one positive rational number  $x$  which does not have a (unique, positive) rational square root. This can be achieved by taking  $x = 2$ . For completeness, we should *prove* that 2 does not have a rational square root (in common parlance, that  $\sqrt{2}$  is irrational), but the reader probably knows how to do this, and the argument was given in Section 2.1.

Another aspect of S1b which merits comment is the use of the word “Let” in the phrase “Let  $x > 0$ ”. This is shorthand notation for “Let  $x$  be a positive real number”, but what does this mean? It tells the reader that we are going to consider *one* positive real number, which we will call  $x$ , but we do not know which positive real number it is. In other words,  $x$  is an *arbitrary* positive real number  $x$ . Similarly, the statement

S3: *Let  $x$  be a real number such that  $0 < x < 1$*

tells the reader that we are going to consider one real number which we will call  $x$ , that  $x$  lies (strictly) between 0 and 1, but that otherwise  $x$  is arbitrary.

Consider the statements:

S4a: *Let  $\mathbf{R}$  be the set of real numbers;*

S5a: *Let  $A$  be a set of real numbers.*

These statements are almost identical, but their meanings are different; S4a means that  $\mathbf{R}$  is the set of *all* real numbers, while S5a means that  $A$  is a set of *some* real numbers. The clue here is the use of "the" in S4a, indicating that we are discussing a unique set, while "a" in S5a indicates that  $A$  is one of many possible sets. Actually instead of S4a and S5a, it would be better to write:

S4b: *Let  $\mathbf{R}$  be the set of real numbers;*

S5b: *Let  $A$  be a subset of  $\mathbf{R}$ .*

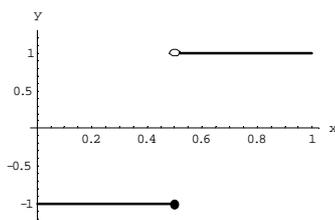
[The notation  $\mathbf{R}$  is standard for the set of all real numbers.]

We have already pointed out that the hypotheses of a statement may be spread over several sentences. One particular trap that may catch you out is that sometimes a standing hypothesis may be made, which is not repeated in every theorem. For example, consider the following statement:

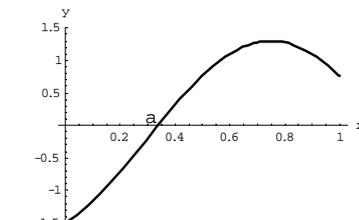
S6: *Suppose that  $f(0) < 0$  and  $f(1) > 0$ . Then there exists a such that  $0 < a < 1$  and  $f(a) = 0$ .*

When you read S6, it is natural to assume that  $f$  is a real-valued function of a real variable between 0 and 1, i.e. that  $f$  specifies a real number  $f(x)$  for each value of  $x$  with  $0 \leq x \leq 1$ . However, if that is all that you assume, apart from the hypotheses  $f(0) < 0$  and  $f(1) > 0$  which are given explicitly, then S6 is false. Indeed, one may take

$$f(x) = \begin{cases} -1 & \text{for } 0 \leq x \leq \frac{1}{2} \\ 1 & \text{for } \frac{1}{2} < x \leq 1. \end{cases}$$



Graph of  $y = f(x)$



Graph of a continuous  $f(x)$

Clearly  $f$  satisfies the hypotheses of S6, but not its conclusion. Thus S6 is false.

However, suppose that the writer of S6 had at some previous stage, maybe many pages earlier, written: *From now on,  $f$  will be a continuous function.* Then S6 would be a true statement. This is a theorem which is sufficiently deep to have a name (the Intermediate Value Theorem), and the proof is beyond us here. It is one of those theorems which appears obvious (if you have to get from below the  $x$ -axis to above it with a continuous curve, then you have to cross the  $x$ -axis somewhere — see the diagram on the right), but one cannot even begin to prove it until one has made a precise definition of *continuous*.

This shows that it is very important to identify all the hypotheses in a statement. You will also need to distinguish correctly between the hypotheses and the conclusions, both when reading mathematics, and when writing. This may sound like a simple task, but there are many pitfalls surrounding the word “if”, which we shall discuss in the next section.

## Summary

1. Mathematical statements have very precise meanings.
2. Typically, a mathematical statement has certain hypotheses and certain conclusions. The statement is true if, whenever the hypotheses are satisfied, then the conclusions are satisfied.
3. A hypothesis such as “Let  $x$  be a whatsit” means that  $x$  is one of many whatsits, but we do not specify which whatsit.
4. The hypotheses of a statement may be spread over several sentences, and may include standing hypotheses made a long time earlier.

## 4.2 “If”, “only if”, and “if and only if”

Suppose that we wish to formulate a statement in which we make some hypotheses and draw certain conclusions from them. Let  $P$  stand for all the hypotheses and  $Q$  for all the conclusions. For example, in S1b,  $P$  stands for:  $x > 0$  and  $Q$  for: *there exists a unique  $y > 0$  such that  $y^2 = x$* . We can then write our desired statement as:

S7a: *If  $P$ , then  $Q$ .*

Thus we can rewrite S1a and S1b in the format of S7a:

S1c: *If  $x > 0$ , then there exists a unique  $y > 0$  such that  $y^2 = x$ .*

In principle, any statement with a set of hypotheses and a set of conclusions can be written in the form S7a, but in practice it may be clumsy to do this, especially if  $P$  and/or  $Q$  is complicated. In this section, we are going to discuss logical aspects of statements containing the word “if”.

As we mentioned in Section 4.1, the statement S7a is true provided that whenever the hypotheses  $P$  are satisfied, then the conclusions  $Q$  are satisfied. In mathematics, to avoid ambiguity we have to stick rigidly to this principle. This differs from everyday conversation, where people often do not distinguish carefully between S7a and its *converse*:

S8a: *If  $Q$ , then  $P$ .*

A synonymous way of expressing the converse is:

S8b: *If  $P$  does not hold, then  $Q$  does not hold.*

(The statement  *$P$  does not hold* or  *$P$  is false* is known as the *negation* of  $P$ ; we shall have more to say about negations in Section 5.7.)

For example, a statement such as:

S9: *If you get three As in A Level, I'll buy you a computer*

is often taken to include the converse (*if you don't get three As, I won't buy you a computer*). In fact, strictly speaking (and mathematicians often have to speak strictly), there is no reason to infer the converse from S9, so if someone did make the statement S9 to you and you did not get three As, you are quite entitled to ask them whether they are going to buy you a computer. On the other hand, if someone made this statement and you did get three As, then you are entitled to *demand* your computer.

Since mathematics is a very precise subject, we do have to be very careful what is meant when reading, and more particularly when writing, mathematics. So, we can write:

S10: *If  $x > 1$ , then  $x^2 > 1$*

(a true statement), but we should not get confused and write the converse:

S11: *If  $x^2 > 1$ , then  $x > 1$*

(a false statement). Equally, we should not confuse the following two statements, which are converses of each other:

S12: *If  $x > 0$ , then there exists  $y > 0$  such that  $y^2 = x$ .*

S13: *If there exists  $y > 0$  such that  $y^2 = x$ , then  $x > 0$ .*

Both S12 and S13 are true, but they have different meanings.

There are many statements which have the same meaning as S7a. For example, the following are all synonyms:

S7a: *If  $P$ , then  $Q$ .*

S7b:  *$P$  implies  $Q$ .*

S7c: *Let  $P$  hold. Then  $Q$  holds.*

S7d:  *$P$  only if  $Q$ .*

S7e:  *$P$  is a sufficient condition for  $Q$  (or: In order that  $Q$  holds, it is sufficient that  $P$  holds).*

S7f:  *$Q$  is a necessary condition for  $P$  (or: In order that  $P$  holds, it is necessary that  $Q$  holds).*

S7g: *If  $Q$  does not hold, then  $P$  does not hold.*

When making a mathematical statement, the choice of format for the statement is up to the writer, but one should try to choose a style which is readable, clear, and unambiguous. Depending on the context, some of the formats listed above are likely to be unsuitable, and there may be convenient formats which are not listed.

To revert to a particular example, the following are all synonymous with S1a:

S1a: *Every positive real number has a unique positive square root.*

S1c: *If  $x > 0$ , then there exists a unique  $y > 0$  such that  $y^2 = x$ .*

S1b: *Let  $x > 0$ . Then there exists a unique  $y > 0$  such that  $y^2 = x$ .*

S1d:  *$x > 0$  only if there exists a unique  $y > 0$  such that  $y^2 = x$ .*

S1e:  *$x > 0 \implies$  there exists a unique  $y > 0$  such that  $y^2 = x$ .*

S1f: *In order that there exists a unique  $y > 0$  such that  $y^2 = x$ , it is sufficient that  $x > 0$ .*

S1g: *A necessary condition that  $x > 0$  is that there exists a unique  $y > 0$  such that  $y^2 = x$ .*

S1h: *If there does not exist a unique  $y > 0$  such that  $y^2 = x$ , then  $x \leq 0$ .*

Take your pick! Both S1d and S1g are very peculiar ways to make this statement, and you are unlikely to choose them.

We have mentioned above that S7a is synonymous with S7d. Interchanging the roles of P and Q, we see that

S8c: *Q only if P*

is synonymous with S8a. Thus S8c is the converse of S7a. In short, “only if” is the converse of “if”. In practice, the phrase “only if” is rarely used on its own, but it is often used in the phrase “if and only if”.

If one wishes to make a statement “If P, then Q” and its converse “If Q, then P” simultaneously, one may write

S14: *P if and only if Q.*

(occasionally, you may see this abbreviated to “P iff Q”). Thus S14 is really two statements. So, to prove S14, you will usually have to give two separate proofs, one of “If P, then Q” and one of “If Q, then P”; only rarely is it possible to prove both parts simultaneously (see also Section 5.7). Note that S14 is synonymous with “Q if and only if P”.

For example, we could combine the statements S12 and S13 into a single statement:

S15: *There exists  $y > 0$  such that  $y^2 = x$  if and only if  $x > 0$ .*

However, S15 could be ambiguous; it might mean (as it is supposed to):

S16a: [*There exists  $y > 0$  such that  $y^2 = x$* ] *if and only if* [ $x > 0$ ],

or it might mean:

S17: *There exists  $y > 0$  such that* [ $y^2 = x$  *if and only if*  $x > 0$ ].

The reader could probably work out that we do not mean S17, either from the context, or by observing that S17 is false (S17 implies that there is some positive number whose square equals all positive numbers, which is clearly nonsense). Nevertheless, we should try to avoid ambiguity if it is possible to do so without becoming unreadable, so instead of S15, it would be better to write S16a or one of the following:

S16b:  $x > 0$  *if and only if there exists  $y > 0$  such that  $y^2 = x$ ,*

S16c: *A necessary and sufficient condition that there exists  $y > 0$  such that  $y^2 = x$  is that  $x > 0$ ,*

S16d: *The following are equivalent:*

(i)  $x > 0$ ,

(ii) *there exists  $y > 0$  such that  $y^2 = x$ .*

The format of S16d is most appropriate for complicated statements, or if there are more than two equivalent statements ((i),(ii),(iii), ... ).

In Section 2.3, we discussed the need to choose language and punctuation in such a way as to remove ambiguity. One useful tip is that any statement starting with “If” should have a matching “then”. This makes it clear when the hypotheses are completed and the conclusions start. For example, what does the following mean?

S18: *If  $n$  is prime,  $n \neq 2$ ,  $n$  is odd.*

It could mean:

S19: *If  $n$  is prime, then  $n \neq 2$  and  $n$  is odd,*

or

S20: *If  $n$  is prime and  $n \neq 2$ , then  $n$  is odd.*

Of course, S19 is false and S20 is true, so you (an educated mathematician) know which interpretation to take. However, there is no point in only making statements whose truth is obvious to the reader (it frequently happens that mathematicians find statements which they wrote themselves are far from obvious after an interval in which they have forgotten the context). An uneducated person might misinterpret S18, by taking it to mean S19 or even

S21: *If  $n$  is prime or  $n \neq 2$ , then  $n$  is odd,*

or

S22: *If  $n$  is prime, then  $n \neq 2$  or  $n$  is odd.*

So the moral is that you should include all those little words like “then”, “and”, “or”, etc.

## Summary

1. Distinguish carefully between a statement “If P, then Q”, its converse “If Q, then P” and the combined statement “P if and only if Q”.
2. The same statement may be written in many different ways.
3. To avoid ambiguity when writing mathematics, construct your statements carefully and include words such as “then”, “and”, “or”, etc.

### 4.3 “And” and “or”

A mathematical statement is likely to have various clauses, linked by logical conjunctions such as “if” and “then” discussed in Section 4.2. Another common pair of logical conjunctions are “and” and “or”, which arose briefly at the end of Section 4.2. For example in a statement of the form “If P, then Q” the hypotheses P and/or the conclusions Q may be composed of smaller clauses linked by these words. For example, consider the statement

S23: *If  $p$  is prime and  $p$  divides  $n^2 - 1$ , then  $p$  divides  $n - 1$  or  $p$  divides  $n + 1$ .*

Here, the hypothesis consists of two parts, “ $p$  is prime” and “ $p$  divides  $n^2 - 1$ ” linked by “and”. This means that we assume that both parts of the hypothesis are true simultaneously. The conclusion also consists of two parts “ $p$  divides  $n - 1$ ” and “ $p$  divides “ $n + 1$ ”, but they are linked by “or”. This means that the conclusion is that  $p$  divides at least one of  $n - 1$  and  $n + 1$ .

In S23, it is implicit that  $p$  and  $n$  are positive integers and that “prime” and “divides” have their usual meanings. Then statement S23 is true (you should be able to see why). However if either clause of the hypothesis were omitted, then the statement would be false. If  $p$  is a non-prime which divides  $n^2 - 1$ , then  $p$  may not divide either  $n - 1$  or  $n + 1$  (e.g.,  $p = 8$ ,  $n = 5$ ). Hence the statement

S24: *If  $p$  divides  $n^2 - 1$ , then  $p$  divides  $n - 1$  or  $p$  divides  $n + 1$*

is false. (However, if it had already been established that  $p$  is prime, then S24 would be true; if other constraints on  $p$  and  $n$  had been established, then S24 might be true.) Similarly, the statement

S25: *If  $p$  is a prime and  $p$  divides  $n^2 - 1$ , then  $p$  divides  $n - 1$*

is false (unless some suitable constraints on  $p$  and  $n$  had previously been established)—

consider  $p = 3$ ,  $n = 2$ , for example.

A statement of the form “P and Q” is true if (and only if) both P and Q are true, simultaneously; “P or Q” is true if (and only if) at least one of P and Q is true. Note that it is a mathematical convention that “or” is used inclusively, so that “P or Q” means “P or Q or both”. (Occasionally, this convention may be breached; in such circumstances, the writer should give explicit warning to the reader.) Thus the following statements are true:

S26: *Everyone in Britain is a European or a foreigner,*

S27: *If  $x$  is a real number, then  $x > 0$  or  $x < 1$ ,*

while the following is false:

S28: *If  $-2 < x < 1$  or  $-1 < x < 2$ , then  $1 \leq |x| \leq 2$ .*

The logical concept of “and” or “or” is often implicit in a mathematical statement, without the words appearing explicitly. Consider the following two problems:

**Example 1.** *Solve the simultaneous equations:*

$$(1) \quad (x - 1)(x - y) = 0,$$

$$(2) \quad (y - 3)(x^2 - y^2 + 1)y = 0.$$

Neither the word “and” nor “or” appears in the problem, but both concepts emerge in the solution below.

*Solution.* Firstly, we are told that the solutions are simultaneous. This means that we should assume (1) *and* (2), and see what we can deduce about  $x$  and  $y$ .

Now, (1) tells us that  $x - 1 = 0$  *or*  $x - y = 0$ , *and* (2) tells us that  $y - 3 = 0$  *or*  $x^2 - y^2 + 1 = 0$  *or*  $y = 0$ . So there are two possibilities for (1) and three (independent) possibilities for (2), making six in all.

*Case 1:*  $x - 1 = 0$  *and*  $y - 3 = 0$ . This gives one solution:  $x = 1$  *and*  $y = 3$ .

*Case 2:*  $x - 1 = 0$  *and*  $x^2 - y^2 + 1 = 0$ . This gives  $x = 1$  *and*  $y^2 = 2$ , so  $x = 1$  *and*  $[y = \sqrt{2}$  *or*  $y = -\sqrt{2}]$ . This provides two solutions:  $x = 1$  *and*  $y = \sqrt{2}$ ;  $x = 1$  *and*  $y = -\sqrt{2}$ .

*Case 3:*  $x - 1 = 0$  *and*  $y = 0$ . This gives one solution:  $x = 1$  *and*  $y = 0$ .

*Case 4:*  $x - y = 0$  *and*  $y - 3 = 0$ . This gives  $x = y$  *and*  $y = 3$ , which provides one solution:  $x = 3$  *and*  $y = 3$ .

*Case 5:*  $x - y = 0$  *and*  $x^2 - y^2 + 1 = 0$ . This gives  $x = y$  *and*  $1 = 0$ , which is impossible! There are no solutions in this case.

*Case 6:*  $x - y = 0$  *and*  $y = 0$ . This gives one solution:  $x = 0$  *and*  $y = 0$ .

The six cases were independent alternatives, so any solution of any of the six cases is a solution of the original problem, and any solution of the original problem is a solution of at least one of the six cases. Counting up, we find that there are six solutions in all:

$$(x, y) = (1, 3) \quad \text{or} \quad (1, \sqrt{2}) \quad \text{or} \quad (1, -\sqrt{2}) \quad \text{or} \quad (1, 0) \quad \text{or} \quad (3, 3) \quad \text{or} \quad (0, 0).$$

Note that although the total number of solutions equals the total number of cases, we did not find one solution for each case; the equality is just a coincidence.

**Example 2.** Which real numbers  $x$  satisfy the following inequality?

$$(*) \quad 4|x + 1| + 5 > 3|x + 2| + 2|x - 1|$$

It is not easy to manipulate the modulus function  $|y|$ , because its definition has different forms for positive and negative values of  $y$ :

$$|y| = \begin{cases} -y & \text{if } y < 0 \\ y & \text{if } y \geq 0. \end{cases}$$

The easiest way to handle it is to consider the cases  $y < 0$  and  $y \geq 0$  separately. In this example, we therefore consider various cases, depending on whether  $x + 1$ ,  $x + 2$ , and  $x - 1$  are positive or negative.

*Solution.* We shall divide the solution into four cases: (1)  $x < -2$ , (2)  $-2 \leq x < -1$ , (3)  $-1 \leq x < 1$ , (4)  $x \geq 1$ .

*Case 1:* Suppose that  $x < -2$ . Then  $|x + 1| = -x - 1$ ,  $|x + 2| = -x - 2$ , and  $|x - 1| = 1 - x$ . Thus (\*) becomes:

$$4(-x - 1) + 5 > 3(-x - 2) + 2(1 - x).$$

Rearrangement gives:  $x > -5$ . We are also assuming that  $x < -2$ , so the solutions of (\*) in this case are the real numbers satisfying  $x < -2$  and  $x > -5$ , in other words those  $x$  such that  $-5 < x < -2$ .

*Case 2:* Suppose that  $-2 \leq x < -1$ . Then (\*) becomes:

$$4(-x - 1) + 5 > 3(x + 2) + 2(1 - x).$$

Rearrangement gives:  $-7 > 5x$ . In this case, the solutions are the real numbers  $x$  satisfying  $-2 \leq x < -1$  and  $x < -7/5$ , in other words those  $x$  such that  $-2 \leq x < -7/5$ .

*Case 3:* Suppose that  $-1 \leq x < 1$ . Then (\*) becomes:

$$4(x + 1) + 5 > 3(x + 2) + 2(1 - x).$$

Rearrangement gives  $3x > -1$ . In this case, the solutions are the real numbers  $x$  satisfying  $-1 \leq x < 1$  and  $x > -1/3$ , in other words those  $x$  such that  $-1/3 < x < 1$ .

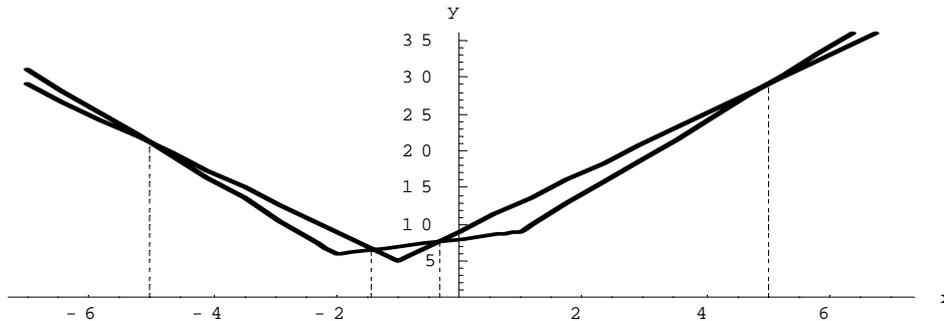
Case 4: Suppose that  $x \geq 1$ . Then (\*) becomes:

$$4(x + 1) + 5 > 3(x + 2) + 2(x - 1).$$

Rearrangement gives:  $5 > x$ . In this case, the solutions are the real numbers  $x$  satisfying  $x \geq 1$  and  $x < 5$ , in other words those  $x$  such that  $1 \leq x < 5$ .

Now for any number  $x$ , exactly one of the four cases applies. Hence  $x$  satisfies (\*) if  $-5 < x < -2$  or  $-2 \leq x < -7/5$  or  $-1/3 < x < 1$  or  $1 \leq x < 5$ ; otherwise  $x$  does not satisfy (\*). Thus,  $x$  satisfies (\*) if and only if either  $-5 < x < -7/5$  or  $-1/3 < x < 5$ .

Our solution is confirmed by the diagram below of the graphs of each side of (\*).



The solutions to Examples 1 and 2 both involved a separation into a number of cases. This is quite common in mathematics. It is not possible to give hard and fast rules as to when this is likely to happen, but it is almost inevitable that it will occur when you are required to prove a statement of the form:

S29: *If  $P$  or  $P'$ , then  $Q$ ,*

when you will probably divide your argument into two cases (Case 1 being when  $P$  holds, and Case 2 when  $P'$  holds). In some problems such as Examples 1 and 2 above, and "or" occurs naturally at some intermediate point of the argument, and then the argument is divided into cases. In some problems, you may choose to divide the argument into cases even though no "or" has explicitly appeared.

The logical concepts of "and" and "or" are in some sense dual to each other. The situation dual to a proof of S29 concerns a statement of the form:

S30: *If  $P$ , then  $Q$  and  $Q'$ .*

To prove S30, it is likely that your argument will consist of two parts, one part proving  $Q$  and the other part proving  $Q'$  (in both cases assuming  $P$ ). It may be easier to prove  $Q$  first, because this will help you to prove  $Q'$ , or it may be easier to prove  $Q'$  first, because this will help to prove  $Q$ , or it may not make any difference which you prove first.

## Summary

1. Look out for situations where mathematical statements are linked (implicitly or explicitly) by “and” or “or”.
2. “P or Q” means “P or Q or both”.
3. Problems involving “or” may need to be solved by division into cases.
4. Problems where the conclusion has more than one part may need separate proofs of the various parts.

### 4.4 “For all” and “there exists”

Two other phrases which occur frequently in mathematics are “for all” (and its synonyms) and “there exists” (and its synonyms). The typical structure of a statement containing them is:

S31a: *For all  $x \in S$ ,  $P(x)$  holds.*

S32: *There exists  $x \in S$  such that  $P(x)$  holds.*

Here,  $S$  is some set,  $\in$  is the standard set-theoretic symbol denoting either “belonging to” or “belongs to”, and  $P(x)$  is a statement which depends on some parameter  $x$ . For example, if  $S$  is the set of all real numbers, and  $P(x)$  is the statement  $x^2 = 2$ , then S31a becomes the false statement:

S33a: *For all real numbers  $x$ ,  $x^2 = 2$ ,*

and S32a becomes the true statement:

S34a: *There exists a real number  $x$  such that  $x^2 = 2$ .*

Often, the distinction between “for all” and “there exists” is very clear, for example in S33a and S34a, and in the following:

S35: *All undergraduates are female;*

S36: *There exist undergraduates who are female.*

However, with complicated mathematical statements, mistakes are sometimes made, so you need to be careful when reading. Moreover, you cannot expect the reader to guess which quantifier you mean, so you should always include them (or some equivalent wording) in your written work.

The phrases “for all” and “there exists” (and their synonyms) are known as *quantifiers*; they are often denoted by the symbols  $\forall$  and  $\exists$ , respectively. In S31a and S32a, the variable  $x$  is said to be *subordinate* to the quantifier; the set  $S$  is known as the

range of the quantifier. **Any statement containing the quantifier  $\forall$  or  $\exists$  should include a subordinate variable and a range.** You are recommended not to use these symbols as general abbreviations for “all” and “exists”.

In both S33a and S34a, the variable is  $x$  and the range is the set  $\mathbf{R}$  of all real numbers. Thus, S33a and S34a may be rewritten as:

$$\text{S33b: } [\forall x \in \mathbf{R}] x^2 = 2,$$

$$\text{S34b: } \exists x \in \mathbf{R} \text{ such that } x^2 = 2.$$

Phrases which may be synonymous with “for all” include

*Given, whenever, for any, every, for each, etc.*

Phrases which may be synonymous with “there exists” include

*for some, for at least one, there is, has, etc.*

It is also possible to rewrite S31a in an “If ... , then ... ” format (see Section 4.2 for further reformulations):

S31b: *If  $x \in S$ , then  $P(x)$  holds.*

Note that in statements S31a and S32,  $x$  is merely a dummy variable, and can be replaced by any other symbol representing a variable, provided that this is done consistently and provided that the new symbol is not used for any other purpose. Thus, S31a is synonymous with:

S31c: *For all  $y \in S$ ,  $P(y)$  holds.*

When writing mathematics, it is important to observe the following general principle:

**The name (i.e. symbol) given to a variable subordinate to a quantifier may be chosen arbitrarily, provided that the name chosen is not already in use in the argument.**

It is particularly important to observe this principle in the case of the quantifier  $\exists$ , when all sorts of logical chaos is likely to break out if the principle is not observed. For example, the following is a true statement:

S37: *For any integer  $n \geq 2$ , there is a prime  $p$  which divides  $n$ .*

However, the following argument becomes gobbledygook:

S38: *Let  $n$  be any integer with  $n \geq 2$ , and let  $p$  be any prime. By S37, there is a prime  $p$  which divides  $n$ .*

At this point, there is a danger that you will deduce that the arbitrary prime  $p$  divides  $n$ , i.e. that  $n$  is divisible by all primes! Instead of S38, one should write:

S39: *Let  $n$  be any integer with  $n \geq 2$ , and let  $p$  be any prime. By S37, there is a prime  $q$  which divides  $n$ .*

S39 is true, and clear, even if it is not very interesting.

In advanced mathematics, one often comes across statements with several quantifiers embedded inside each other. For example, the following is a typical mathematical statement:

S40a: *Given  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $|f(y) - f(x)| < \epsilon$  whenever  $|y - x| < \delta$ .*

S40a includes one explicit “there exists” and two implicit “for all”s, signalled by “given” and “whenever”. We can write S40a in the following formal fashion:

S40b:  $\forall \epsilon > 0 \exists \delta > 0$  such that  $\forall [(x, y) \text{ such that } |y - x| < \delta] \quad |f(y) - f(x)| < \epsilon$ .

In S40b, it is presumed that  $f$  is a function (of one variable) which has already been introduced;  $\epsilon$  and  $\delta$ , and the pair  $x, y$  are variables which are subordinate to quantifiers, so they can be replaced by other symbols if you wish, provided this is done consistently. The range of  $\epsilon$ , and the range of  $\delta$ , are both the set of all positive real numbers; the range of the pair  $x, y$  is the set of all pairs of real numbers satisfying  $|y - x| < \delta$ .

In fact, in saying that S40a and S40b are synonymous, we have cheated slightly. This is correct, provided that neither  $x$  nor  $y$  has already been introduced. In that case, they are both subordinate to the quantifier “whenever”, and rewriting S40a as S40b is correct. However, if  $x$  had been introduced at some stage before S40a is stated (but  $y$  had not been), then  $x$  cannot be subordinate to the quantifier. In that case, S40a should be rewritten as:

S41:  $\forall \epsilon > 0 \exists \delta > 0$  such that  $\forall [y \text{ such that } |y - x| < \delta] \quad |f(y) - f(x)| < \epsilon$ .

This apparently small change affects the meaning of the statement significantly. We shall return to this subtle, but important, point in the next section.

## Summary

1. Many mathematical statements involve the quantifiers “for all” and “there exists”, but they may be disguised.
2. Every quantifier has a subordinate variable and a range, both of which should be given explicitly if the symbol  $\forall$  or  $\exists$  is used.
3. The meaning of a statement is unchanged if the name of a variable which is subordinate to a quantifier is changed, provided that every occurrence of the variable is changed in the same way, and the new name is not used for any other purpose.
4. When writing a statement involving the quantifier  $\exists$ , choose a name (symbol) for the subordinate variable which is not already in use.
5. If a variable has been introduced before a statement, then that variable cannot be subordinate to any of the quantifiers in the statement.
6. All relevant quantifiers should be included in your written work.

## 4.5 What depends on what?

Consider the following statement:

S42:  $\forall x > 0 \exists y > 0$  such that  $y^2 = x$ .

Is it synonymous with the following statement?

S43:  $\exists y > 0$  such that  $\forall x > 0 \quad y^2 = x$ .

Of course not! When you convert the symbols into English, it should be clear that S42 is the true statement that every positive number has a positive square root. However, S43 is the nonsensical statement that there is one positive number whose square equals all positive real numbers simultaneously.

This example shows that in statements involving quantifiers, the order in which the variables are introduced matters crucially. This is not really a point of mathematics at all, merely one of language, but it is of great significance for mathematicians, and it often causes difficulty for undergraduates.

In S42, the clause  $\exists y > 0$  appears *after* the introduction of the variable  $x$ . It therefore follows from the structure of the sentence (“For all  $x$ , there exists  $y$  such that ...”) that the variable may depend on  $x$ . Similarly, in S40a, S40b, and S41,  $\delta$  may depend on  $\epsilon$ . On the other hand in S43 (“There exists  $y$  such that for all  $x$  ...”),  $y$  cannot possibly depend on  $x$ , because  $x$  has not even been introduced to the reader’s mind when  $y$  is supposed to be determined. We can summarise this principle:

**A variable subordinate to a quantifier  $\exists$  (“there exists”, “for some”, etc.) may depend on variables which have been introduced earlier.**

For example, consider the statement:

S44a: *Every undergraduate in my college owns a bicycle.*

This can be formalised as:

S44b:  $\forall x \in \{\text{undergraduates in my college}\} \exists y \in \{\text{bicycles}\}$  such that  $x$  owns  $y$ .

Of course, the bicycle  $y$  depends on the undergraduate  $x$ . Since the undergraduate is introduced in the statement before the bicycle, this is consistent with the principle. On the other hand, in the statement:

S45: *Every college has a library which is available for all students in the college,*

the library depends on the college (introduced earlier) but not on the student (introduced later in the statement).

This principle applies even when the other variables have been introduced in earlier sentences, including those which appeared considerably earlier. For example, consider the following statement (compare S40a and S41):

S46: *Let  $x \in \mathbf{R}$ . A function  $f : \mathbf{R} \rightarrow \mathbf{R}$  is said to be continuous at  $x$  if, for all*

$\epsilon > 0$ , there exists  $\delta > 0$  such that  $|f(y) - f(x)| < \epsilon$  whenever  $|y - x| < \delta$ .

Here,  $\delta$  may depend on  $x$ ,  $f$  and  $\epsilon$  (all previously introduced), but not on  $y$  (introduced later). On the other hand, in the statement:

S47: *Let  $f : \mathbf{R} \rightarrow \mathbf{R}$  be a function. Then  $f$  is said to be uniformly continuous if, for all  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $|f(y) - f(x)| < \epsilon$  whenever  $|y - x| < \delta$ ,*

$\delta$  should be independent of both  $x$  and  $y$  (not previously introduced), but  $\delta$  may depend on  $f$  and  $\epsilon$ .

To avoid ambiguity, and to stress what depends on what, you may find that authors write a statement such as S40a in the form:

S40c: *Given  $\epsilon > 0$ , there exists  $\delta = \delta_\epsilon > 0$  such that  $|f(y) - f(x)| < \epsilon$  whenever  $|y - x| < \delta$ .*

This is a notational device which can be helpful in reminding both the writer and the reader that  $\delta$  may depend on  $\epsilon$ . Sometimes, the author might mention explicitly that, in S47,  $\delta$  is independent of  $x$  and  $y$ . These are good practices to adopt yourself in your notes and written solutions, at least until you are confident that you will not get confused.

Incidentally, understanding why S46 is a sensible definition of continuity is outside the theme of these notes, but you might like to think about it (see *Alice in Numberland*, Chapter 11, for example).

## Summary

1. In statements involving the quantifier “there exists”, the subordinate variable(s) may depend on variables which have already been introduced, but must be independent of any variables introduced later in the statement.
2. If you think that you might get confused, include reminders such as “ $\delta = \delta_\epsilon$ ” or “independent of  $x$  and  $y$ ” in your written work.

## CHAPTER 5

### Proofs

Apart from understanding the statements of abstract theorems (see Chapter 4), you should become used to proving them rigorously. The most important theorems will be proved in books and lectures, but you may be asked to reproduce the proofs in examinations. Moreover, you will frequently be set problems which require you to prove abstract statements which you have not seen before. In all cases, the aim will be to write out accurate and efficient proofs. There are three basic skills concerning proofs:

- (i) identifying the essential points of a standard proof,
- (ii) constructing proofs of facts which were previously unknown to you,
- (iii) setting out proofs accurately, clearly, and efficiently.

We have discussed (i) in Section 2.1, and (iii) in Section 2.3. Many of the principles described in Section 2.2 are applicable to (ii) (see Section 5.4), but this chapter will be devoted to a more detailed discussion of (ii). The examples in Section 5.5 will also serve to illustrate (iii).

#### 5.1 Counterexamples

In real-life problems, you do not know the answer until you have solved the problem. Undergraduate mathematics is rather different, in that in many problems you are told the answer, and required only to find out *why* it is the answer. However, from time to time, you will be asked questions in which you are required to determine whether a given statement is true or false. In these cases, it is always implicit that you should justify your answer (there are no prizes for guessing correctly). This means that if the statement is true, then you have to prove it; if the statement is false, then you have to disprove it. We shall give you some ideas in the subsequent sections of this chapter how to go about proving things, but in this section we tell you how to disprove (false) statements.

Suppose that you are considering a typical mathematical statement which assumes some hypotheses and claims some conclusion (see Section 4.2); suppose moreover that you think that the statement may be false. In order to disprove it, all you have to do is to provide a single counterexample to the statement, i.e. a single example where all

the hypotheses of the statement are satisfied, but the conclusion is not. In principle, you may have to prove that your example satisfies the hypotheses and does not satisfy the conclusion, but this will often be self-evident. For example, in order to show that the statement:

*If  $m$  and  $n$  are positive integers such that  $m$  divides  $n^2 - 1$ , then  $m$  divides  $n - 1$  or  $m$  divides  $n + 1$*

is false, it is enough to observe that taking  $m = 8$  and  $n = 5$  produces a counterexample (see S34 in Section 4.3).

Although it sounds easy to find just one example (much easier than giving a complete proof), most students find it difficult, even if they are told to look for a counterexample. When they are not told whether the statement is true or false, students have even more difficulty deciding, particularly when the statement is in fact false.

The skill of deciding whether a statement is true or false depends largely on mathematical insight, which can only be acquired gradually by doing mathematics over a long and continuous period. Otherwise, the best advice that can be given is that you should first try to prove the statement; if you run into difficulties, look for a counterexample—the difficulties which you encountered in trying to prove the statement should tell you where to start looking. If you run into difficulties looking for a counterexample, you may acquire new ideas for a proof. If they do not work, the new difficulties may give you new ideas for a counterexample . . . etc., etc. Advice on how to set about searching for a proof is given in Sections 2.2 and 5.2–5.5.

For example, you may be asked to determine whether the following is true:

*Let  $S$  be the set of all numbers of the form  $a + b\sqrt{2}$  where  $a$  and  $b$  are rational. Let  $f : S \rightarrow \mathbf{R}$  be a function such that  $f(x + y) = f(x) + f(y)$  for all  $x$  and  $y$  in  $S$ . Then  $f(x) = f(1)x$  for all  $x \in S$ .*

On trying to prove the statement, you may succeed in proving that  $f(a) = f(1)a$  for all rational  $a$  (see Example 6 in Sections 5.2–5.5), and perhaps go on to prove that  $f(a + b\sqrt{2}) = f(1)a + f(\sqrt{2})b$  for all rational  $a$  and  $b$ . At this point, you may notice that it would suffice to prove that  $f(\sqrt{2}) = f(1)\sqrt{2}$ . After a brief period of experimentation, you should come to realise that there is no way of proving this from the given information, and that you can show the statement to be false, by choosing  $f(1)$  and  $f(\sqrt{2})$  in such a way that  $f(\sqrt{2}) \neq f(1)\sqrt{2}$ . Thus, you can write down your solution:

The statement is false. Let  $f$  be the function defined by  $f(a + b\sqrt{2}) = a$  for rationals  $a, b$ . Then  $f(x + y) = f(x) + f(y)$  for all  $x, y$  in  $S$ . However,  $f(\sqrt{2}) = 0 \neq \sqrt{2} = f(1)\sqrt{2}$ .

Actually, to be truly complete, you should show that this example satisfies the identity  $f(x + y) = f(x) + f(y)$  (by putting  $x = a + b\sqrt{2}$ ,  $y = c + d\sqrt{2}$ , and doing some manipulations), but we will omit this detail.

Apart from acquiring good mathematical insight, which is a process which it is not easy to hasten, there is only one good way to learn the skill of constructing counterexamples. That is to practise it yourself; seeing somebody else provide counterexamples is not the same thing at all. This is one reason why we have recommended (Section 2.1) that you should take standard theorems, omit part of the hypotheses, and try to find an example to show that the conclusion of the theorem becomes false. This is not always easy, so if you have difficulty finding such counterexamples, do not be afraid to ask your tutor or fellow-students.

Students often make the mistake of trying to disprove a statement by discussing some argument which might be hoped to provide a proof, and saying this “This does not work because ...”, or words to that effect. This is not adequate to show the statement to be false, because it could be that you overlooked a cleverer argument which would in fact have produced a proof. Thus to disprove a statement, one should always provide a counterexample. (In fact, there are certain situations in advanced mathematics where one can prove a statement to be false by an indirect argument, without being able to provide a specific counterexample. However, you are unlikely to encounter such a situation in undergraduate work, certainly not in the first year.)

To illustrate the difference between failure of an attempted proof, and disproof, we can recall two famous mathematical problems. The first is the Four Colour Theorem (that every map can be coloured in such a way that no two adjacent countries have the same colour, using only four colours). A “proof” of this was published by Kempe in 1879. In 1890, Heawood pointed out that Kempe’s proof contained a gap, but that it could be modified to prove the Five Colour Theorem. The fact that the proposed proof of the Four Colour Theorem failed did not mean that the Four Colour Theorem was false—to establish that, it would have been necessary to give an example of a map which could not be coloured in four colours (and show that it could not be). Indeed, the truth or falsity of the Four Colour Theorem was not resolved until 1976, when it was proved to be true. Remarkably, the basic idea of the proof was the same as Kempe’s, but the argument was much more complicated, requiring many hours of computer time to check various cases.

The second example is Fermat’s Last “Theorem”, which states that, for  $n \geq 3$ , there do not exist positive integers  $x$ ,  $y$ , and  $z$ , such that  $x^n + y^n = z^n$ . Although Fermat (1601?–1665) made a note that he was able to prove this, he did not write down the proof and nobody has subsequently been able to find one. It has been shown to be true for all  $n \leq 100,000$ , and for many larger values of  $n$ . Thousands of “proofs” of Fermat’s Last Theorem have been put forward, but so far all of them have been found to contain errors.<sup>2</sup> None of this means that Fermat’s Last Theorem is false. In fact, most mathematicians would be extremely surprised if it were shown to be false, even though

---

<sup>2</sup>One of the most publicised instances occurred in June 1993 when Andrew Wiles, a former Oxford undergraduate (and tutorial partner of the author of these notes), announced a “proof”. Although the mathematical community at first thought it likely that he was right, he announced five months later that there was a gap in his argument. At the time of writing these notes, the gap has not been filled, although Wiles’s method represents a great advance.

they may be sceptical about Fermat's famous claim to have found a proof. To show Fermat's Last Theorem to be false, somebody would have to provide a counterexample (which would involve gigantic numbers) or provide a rigorous indirect argument that a counterexample must exist. Meanwhile, the search for a proof continues.

## Summary

1. To disprove a mathematical statement, you should provide a specific counterexample.
2. If asked to decide whether a given statement is true or false and if your insight does not immediately provide the answer, try to prove the statement; then use your experience of the difficulties encountered to try to construct a counterexample; then use your experience of those difficulties to try again to prove the statement, etc.
3. To gain experience in constructing counterexamples, you should practise by taking standard theorems, omitting part of the hypothesis, and trying to find counterexamples to the conclusions.

## 5.2 Constructing proofs

In pure mathematics, a typical problem is quite theoretical, requiring you to prove something of greater or lesser generality. We saw in Section 2.2 that problem-solving involves three stages which in this context can best be described as:

- (i) understanding the problem (understanding the statement, and what is to be proved);
- (ii) experimentation (to find a method which looks likely to work);
- (iii) making the proof precise (making sure that the proof does work, and writing it out accurately and completely).

We shall discuss these three stages in turn in the next three sections. Meanwhile, we take the opportunity here to introduce several examples which we will use in the next three sections. These examples are arranged roughly in increasing order of sophistication. **It is not expected that you will be able to solve all these problems, or even to understand all of them, on first reading of these notes.** Do not despair: part of the purpose is to illustrate the need to find out what the statements mean before you start to prove them (see Section 5.3); some of the examples may be left until a second reading. The next two paragraphs give you more guidance how you might use the examples.

First, read the examples. It is likely that you will understand the statements of some of them (you may even be familiar with some of the results). For each of those examples, try to solve the problem before you read Sections 5.3–5.5. If you succeed, write out a careful solution of your own, and then compare it with the solution given in Section 5.5. If you cannot solve the problem after several attempts, read the discussion of the example in Section 5.3 to make sure that you have understood it correctly. Then read the treatment of the example in Sections 5.4 and 5.5.

There will probably be some examples whose statements you do not understand (e.g., the “trace” of a matrix in Example 8 may be unfamiliar). Choose one (or two) of them (preferably the first). Read that example in Section 5.3 to try to understand the problem (if the difficulty is notational, look also in the Appendix). Then try to solve the problem before looking at the solution in Sections 5.4 and 5.5. Ignore any remaining examples until a later reading of the notes.

**Example 1.** *Let  $A$ ,  $B$ , and  $C$  be sets. Prove that  $(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$ .*

**Example 2.** *Let  $a$ ,  $b$ ,  $c$  and  $d$  be positive real numbers. Prove that*

$$\frac{a + b + c + d}{4} \geq (abcd)^{1/4}.$$

**Example 3.** *Let  $n$  be a positive integer, and suppose that  $2^n - 1$  is prime. Prove that  $n$  is prime.*

**Example 4.** *Let  $n \geq 1$ . Prove that  $\left(1 - \frac{x}{n}\right)^n \leq e^{-x}$  whenever  $0 \leq x \leq n$ .*

**Example 5.** *Show that  $|\sin x - \sin y| \leq |x - y|$  whenever  $x, y \in \mathbf{R}$ .*

**Example 6.** *Let  $f : \mathbf{Q} \rightarrow \mathbf{R}$  be a function such that  $f(x + y) = f(x) + f(y)$  for all  $x, y \in \mathbf{Q}$ . Prove that  $f\left(\frac{m}{n}\right) = \frac{m}{n}f(1)$  for all integers  $m$  and all strictly positive integers  $n$ .*

**Example 7.** *Let  $(a_n)_{n \geq 0}$  be a sequence of positive real numbers defined recursively by:*

$$(*) \quad a_0 = 1, \quad a_{n+1} = \sqrt{2 + a_n}.$$

*Prove that  $a_n$  converges to a limit as  $n \rightarrow \infty$ , and find the value of that limit.*

**Example 8.** *Let  $A$  and  $B$  be  $n \times n$  matrices. Prove that the traces of  $AB$  and  $BA$  are equal.*

### 5.3 Understanding the problem

The main things to be done here are:

- (i) Look up the definition of any terms (words or symbols) about which you are uncertain. Likely places to look are your lecture notes, and textbooks; the index of a textbook usually includes a reference to each definition; some textbooks have a separate index of symbols. By the time that you have done this, you should be able to translate the statement into ordinary English.
- (ii) Decide which parts of the statement are hypotheses which you are supposed to assume. Write them down (“We assume that ... ”); include any reformulations of the hypotheses which look as though they might be useful; for example, you may replace a technical mathematical term by its definition.
- (iii) Identify the conclusions which you are supposed to prove; write them down (“We have to prove that ... ”). If there is more than one part to the conclusion, write them down separately. If appropriate, write down any reformulations of the conclusion which appear promising; for example, it may be useful to put the conclusion in a more mathematical form, by replacing technical terms by their definitions.

Now, we consider the examples introduced in Section 5.2.

**Example 1.** *Definitions:* It is likely that the reader is familiar with the notions of union and intersection, and the symbols  $\cup, \cap$  which denote them.

*Hypotheses:* The only hypothesis is that  $A, B$  and  $C$  are sets. (It is implicit that they are all considered to be subsets of some universal set.)

*Conclusions:* We have to prove that  $(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$ . To do this, we should show:

- (a) If  $x \in (A \cup B) \cap (A \cup C)$ , then  $x \in A \cup (B \cap C)$ ;
- (b) If  $x \in A \cup (B \cap C)$ , then  $x \in (A \cup B) \cap (A \cup C)$ .

**Example 2.** *Definitions:* The reader is unlikely to have any difficulty with the terminology or notation.

*Hypotheses:* The hypothesis is that  $a, b, c$  and  $d$  are positive real numbers.

*Conclusions:* We have to prove that  $\frac{a + b + c + d}{4} \geq (abcd)^{1/4}$ .

**Example 3.** *Definitions:* As usual, a positive integer  $k$  is said to be prime if  $k$  is not divisible by any positive integer, except 1 and  $k$ .

*Hypotheses:* The hypotheses are that  $n$  is a positive integer and that  $2^n - 1$  is prime.

*Conclusions:* We have to prove that  $n$  is prime; in other words, we have to show that if  $n = mk$  for some positive integers  $m$  and  $k$ , then  $m = 1$  or  $k = 1$ .

**Example 4.** *Definitions:* It is not specified whether  $n$  is supposed to be an integer, or a real number. It is common practice to reserve letters such as  $j$ ,  $k$ ,  $m$ ,  $n$ , and  $r$  to denote integers (unless specified otherwise), so we might assume that  $n$  is an integer here. However, in this example, it turns out not to matter whether we assume  $n$  is an integer or allow it be a real number.

The notation  $e^{-x}$  represents the exponential of  $-x$ .

*Hypotheses:* The hypotheses are that  $n \geq 1$  and  $0 \leq x \leq n$ .

*Conclusions:* We have to prove that  $\left(1 - \frac{x}{n}\right)^n \leq e^{-x}$ .

**Example 5.** *Definitions:* It is assumed that the reader is familiar with the sine function. The symbol  $\mathbf{R}$  denotes the set of all real numbers, and  $\in$  is the standard set-theoretic symbol meaning “belongs to”. Thus, “for all  $x, y \in \mathbf{R}$ ” means “for all real numbers  $x$  and  $y$ ”.

*Hypotheses:* The only hypothesis here is that  $x$  and  $y$  are real numbers. We have to prove the conclusion for all such  $x$  and  $y$ , so we are not allowed to choose them to have specific values.

*Conclusions:* The conclusion to be proved is that  $|\sin x - \sin y| \leq |x - y|$ .

**Example 6.** *Definitions:* The symbol  $\mathbf{Q}$  denotes the set of all rational numbers;  $\mathbf{R}$  again denotes the set of all real numbers; the notation  $f : \mathbf{Q} \rightarrow \mathbf{R}$  means that, to each rational number  $x$ ,  $f$  assigns a real value  $f(x)$ .

*Hypotheses:* The hypothesis is that  $f(x + y) = f(x) + f(y)$  for all rational numbers  $x$  and  $y$ .

*Conclusions:* We have to prove that  $f\left(\frac{m}{n}\right) = \frac{m}{n}f(1)$  for all integers  $m$  and strictly positive integers  $n$ , in other words, that  $f(x) = xf(1)$  for all rational numbers  $x$ .

**Example 7.** *Definitions:* The word “recursively” indicates that the value of each  $a_k$  can be found by applying the formula  $a_{n+1} = \sqrt{2 + a_n}$  successively for  $n = 0, 1, 2, \dots, k-1$ . The concept of convergence to a limit has a precise mathematical definition, which can be found in textbooks on analysis, but we will not need to know it for this problem.

*Hypotheses:* The only hypothesis is that  $(a_n)$  is the specific sequence defined by (\*).

*Conclusions:* We have to prove that  $a_n$  converges to a limit  $l$ , and we have to find the value of  $l$ .

**Example 8.** *Definitions:* The reader may not know the definition of the trace of a matrix. It can be found in books on linear algebra. The trace of an  $n \times n$  matrix  $A$  is the sum of the entries on the leading diagonal. If the entry in row  $i$ , column  $j$  of  $A$  is

denoted by  $a_{ij}$ , then

$$\operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^n a_{ii}.$$

*Hypotheses:* The only hypotheses are that  $A$  and  $B$  are  $n \times n$  matrices.

*Conclusions:* We have to prove that  $\operatorname{tr} AB = \operatorname{tr} BA$ .

If you failed to solve some of these examples because you did not understand the statements, you are now recommended to try them again before going on to read the next two sections.

## Summary

1. Look up the definitions of any terms and symbols about which you are doubtful.
2. Identify the hypotheses, write them down, and write down any reformulations which appear useful.
3. Identify the conclusions to be proved, write them down, separating them into various parts; write down any reformulations which appear promising.

## 5.4 Experimentation

This part is the real core to the solution, where you do most of the hard thinking. What you do at this stage may not form part of your final solution, at least not in the form in which you do it at this stage, but it should lead you to see why the statement to be proved is true, and also show you how to put together a rigorous argument. Another description for “experimentation” might be “rough working”, or “exploration”, but we have in mind not just calculations and algebraic manipulations, but also the whole process of finding a method which works to solve the problem.

The general principles which should apply while exploring the problem have already been listed in Section 2.2. We repeat some of them here:

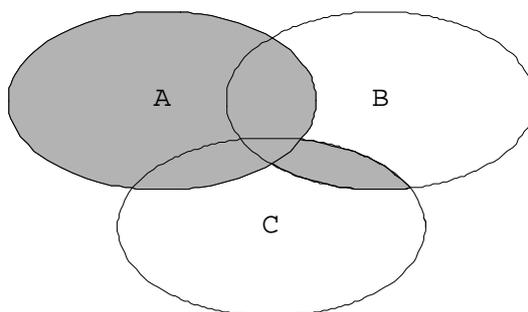
- (i) While searching for a method, do not worry about details.
- (ii) Try working backwards from the conclusion.
- (iii) Draw a diagram, if appropriate.
- (iv) Try some special cases, examples, or simpler versions of the problem.
- (v) If appropriate, break the problem up into several cases.

(vi) Consider the possibility of a proof by contradiction (see Section 5.7).

It is quite likely that your first attempt will not be successful. Keep trying different methods until you find one which works or until you run out of ideas. Look in your lecture notes and books for relevant worked examples. If none of this works, put the problem aside and return to it later.

We now consider the process of experimentation for the examples introduced in Section 5.2, making use of the clarifications described in Section 5.3. Of course, what we do here is unlikely to be typical of your work in that we shall not describe many methods which lead to dead ends!

**Example 1.** Let's draw a Venn diagram of  $(A \cup B) \cap (A \cup C)$ :



This is exactly  $A \cup (B \cap C)$ , as required. So, a formal argument should work without serious difficulty.

**Example 2.** In this example, there are many things which you might try, but which turn out to be not very helpful. For example, you might take the fourth powers of both sides of the inequality, which then becomes:

$$(a + b + c + d)^4 \geq 256abcd.$$

Expanding the left-hand side looks hopeless.

After several unsuccessful attempts, you might try some simpler cases. For example, you could try putting  $b = c = d = 1$ , but even this does not seem very promising.

Eventually, you might wonder at the significance of the number 4 in this example. Is there an analogue with 4 replaced by 2? Perhaps the analogue would be:

$$\frac{a + b}{2} \geq (ab)^{1/2}. \quad (*)$$

Can we prove (\*)? Squaring and multiplying through by 4, (\*) becomes:  $(a + b)^2 \geq 4ab$ . After some manipulation, this becomes:  $a^2 - 2ab + b^2 \geq 0$ , or  $(a - b)^2 \geq 0$ . This is certainly true, so we should be able to prove (\*) by reversing the argument.

Replacing  $a$  and  $b$  by  $c$  and  $d$  in (\*) gives:  $\frac{c+d}{2} \geq (cd)^{1/2}$ . Hence

$$\frac{a + b + c + d}{4} = \frac{1}{2} \left( \left( \frac{a + b}{2} \right) + \left( \frac{c + d}{2} \right) \right) \geq \frac{(ab)^{1/2} + (cd)^{1/2}}{2}.$$

Replacing  $a$  and  $b$  in (\*) by  $(ab)^{1/2}$  and  $(cd)^{1/2}$  shows

$$\frac{(ab)^{1/2} + (cd)^{1/2}}{2} \geq (abcd)^{1/4}.$$

Combining these two inequalities gives the required result.

**Example 3.** Suppose that  $n$  can be factorised as  $n = mk$  where  $m$  and  $k$  are integers greater than 1. Can we factorise  $2^n - 1$ ? Now,

$$2^n - 1 = 2^{mk} - 1 = (2^m)^k - 1.$$

This is an expression of the form  $x^k - 1$ . This polynomial has a root  $x = 1$ , so it must have a factor  $x - 1$ . A little calculation shows that

$$x^k - 1 = (x - 1)(x^{k-1} + x^{k-2} + \dots + 1).$$

Hence

$$2^n - 1 = (2^k - 1)(2^{m(k-1)} + 2^{m(k-2)} + \dots + 1).$$

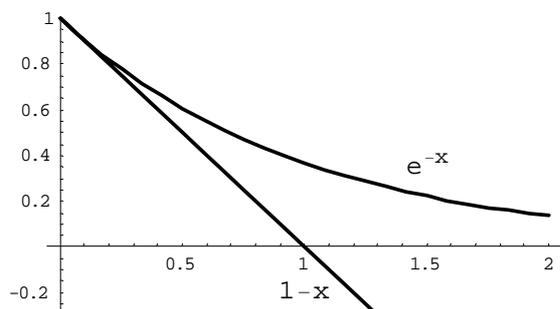
Thus, if  $2^n - 1$  is prime, we cannot have such a factorisation, so there cannot be a factorisation of  $n$ ; in other words,  $n$  is prime.

**Example 4.** You might try the binomial expansion of the left-hand side of the inequality, but this does not prove to be very fruitful.

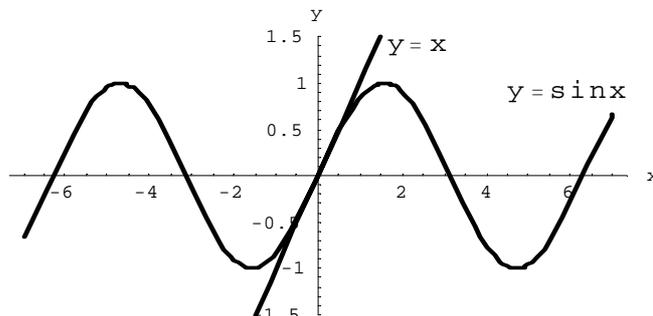
Let's look at the simplest case,  $n = 1$ . Then we have to prove that  $1 - x \leq e^{-x}$  whenever  $0 \leq x \leq 1$ . A sketch of the graph of  $e^{-x}$  shows that this is true; it should be easily proved by calculus.

For general  $n$ , we can take  $n$ th roots, so the inequality to be proved becomes:  $1 - \frac{x}{n} \leq e^{-x/n}$ . This can be deduced from the inequality in the preceding paragraph, if we replace  $x$  by  $x/n$ .

**Example 5.** First let's draw a diagram of the graph of  $\sin x$ .



Graph for Example 4



Graph for Example 5

Doing this may bring two points to mind. Firstly, since the sine function is periodic ( $\sin(x + 2\pi) = \sin x$ ), it should be sufficient to assume that  $0 \leq x, y \leq 2\pi$ . Secondly, the inequality  $|\sin x - \sin y| \leq |x - y|$  says that the (modulus of) the gradient of the chord is less than 1. Since  $\frac{d}{dx}(\sin x) = \cos x$ , the gradient of the tangent lies between  $-1$  and  $+1$ , so the gradient of the chord may well lie in the same range.

Since you probably do not know yet a theorem relating the gradient of the chord to the gradient of tangents, the argument so far is purely heuristic. To make it rigorous, we shall have to find a precise analytic argument, and we now experiment to look for this.

To deal with  $\sin x - \sin y$ , it may be useful to use a trigonometric formula. This can be found from memory, by working it out, or by looking it up in a book:

$$\sin x - \sin y = 2 \sin \left( \frac{x - y}{2} \right) \cos \left( \frac{x + y}{2} \right).$$

Hence

$$|\sin x - \sin y| = 2 \left| \sin \left( \frac{x-y}{2} \right) \right| \left| \cos \left( \frac{x+y}{2} \right) \right|.$$

Now, the case which is going to cause most difficulty is when  $|x - y|$  is small, so  $x$  is close to  $y$ . We have to ensure that the right-hand side is small, and what makes that true is the term  $\left| \sin \left( \frac{x-y}{2} \right) \right|$ . It is therefore not a good idea to replace this term by

1. However, it may not be harmful to replace  $\left| \cos \left( \frac{x+y}{2} \right) \right|$  by 1. Thus,

$$|\sin x - \sin y| \leq 2 \left| \sin \left( \frac{x-y}{2} \right) \right|.$$

Now, is it true that  $2 \left| \sin \left( \frac{x-y}{2} \right) \right| \leq |x-y|$ ? Yes! Because  $|\sin t| \leq t$  for all real numbers  $t$  (you probably know this; if not, look again at the graph of the sine function and its tangent at the origin). Thus

$$|\sin x - \sin y| \leq 2 \left| \frac{x-y}{2} \right| = |x-y|,$$

as required.

**Example 6.** We want to prove that  $nf\left(\frac{m}{n}\right) = mf(1)$  for all  $m$  and  $n$ . Putting  $n = 1$ , we need to show that  $f(m) = mf(1)$ . We will consider this special case first and then think about the rest of the problem.

Using the given identity  $f(x+y) = f(x) + f(y)$  with  $x = 1$  and  $y = 1$ , we find that  $f(2) = 2f(1)$ . Then putting  $x = 2$  and  $y = 1$ , we get  $f(3) = f(2) + f(1) = 3f(1)$ . Iterating this (the formal proof will be by induction), we will get  $f(m) = mf(1)$  for all positive integers  $m$ .

What about negative integers? Putting  $x = y = 0$ , we find that  $f(0) = 2f(0)$ , so  $f(0) = 0$ . Now putting  $x = m$  and  $y = -m$  gives  $0 = f(0) = f(m) + f(-m)$ , so  $f(-m) = -f(m) = -mf(1)$  (whenever  $m$  is a positive integer).

Now, what about the case  $n \geq 2$ ? We want to show that  $nf\left(\frac{m}{n}\right) = f(m)$ . The argument used to show that  $f(m) = mf(1)$  should also show that  $f(nx) = nf(x)$  for any rational  $x$  and any positive integer  $n$ . Putting  $x = \frac{m}{n}$  will then give the result.

**Example 7.** First, assume that  $a_n \rightarrow l$  as  $n \rightarrow \infty$ . Then  $a_{n+1} \rightarrow l$  and  $\sqrt{2+a_n} \rightarrow \sqrt{2+l}$ . Hence  $l = \sqrt{2+l}$ , so  $l^2 = 2+l$ ,  $l^2 - l - 2 = 0$ ,  $(l-2)(l+1) = 0$ , so  $l = 2$  or  $l = -1$ . However, since  $a_n > 0$  for all  $n$ ,  $l$  cannot be  $-1$ . Hence,  $l = 2$ .

This argument does not prove that  $a_n$  converges to 2; what it shows is that, if  $a_n$  converges to anything, then it converges to 2. It still remains to prove that  $a_n$  does

converge. For this, let's calculate a few values:

$$\begin{aligned} a_1 &= \sqrt{2+1} = 1.732050 \\ a_2 &= \sqrt{2+1.732050} = 1.931851 \\ a_3 &= \sqrt{2+1.931851} = 1.982889 \end{aligned}$$

This looks promising. If you carry on for a while, you will find that  $a_{15} = 1.999999999$  (to 9 decimal places), which looks even more promising, but none of this will constitute a *proof* that  $a_n$  converges to 2.

To prove that  $a_n$  does converge to 2, we might consider  $2 - a_{n+1}$ . However, this involves square roots, so it is better to look at  $4 - a_{n+1}^2$ . Now,

$$4 - a_{n+1}^2 = 4 - (2 + a_n) = 2 - a_n,$$

so

$$2 - a_{n+1} = \frac{2 - a_n}{2 + a_{n+1}}.$$

But  $a_{n+1} \geq 0$ , so  $2 + a_{n+1} \geq 2$ , so  $|2 - a_{n+1}| \leq \frac{1}{2}|2 - a_n|$  for all  $n$ . Iterating this gives

$$|2 - a_{n+1}| \leq \left(\frac{1}{2}\right)^{n+1} |2 - a_0| = \left(\frac{1}{2}\right)^{n+1} \rightarrow 0.$$

Hence,  $a_{n+1} \rightarrow 2$ , so  $a_n \rightarrow 2$ .

**Example 8.** First, consider the case of  $2 \times 2$  matrices. Then we may write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad B = \begin{pmatrix} w & x \\ y & z \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} aw + by & ax + bz \\ cw + dy & cx + dz \end{pmatrix} \quad BA = \begin{pmatrix} wa + xc & wb + xd \\ ya + zc & yb + zd \end{pmatrix}.$$

So,

$$\begin{aligned} \text{tr}(AB) &= aw + by + cx + dz, \\ \text{tr}(BA) &= wa + xc + yb + zd. \end{aligned}$$

These agree (since addition and multiplication of real numbers are commutative).

Next,  $3 \times 3$  matrices. To avoid using too many letters, let's write

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & \dots & \dots \\ \dots & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & \dots \\ \dots & \dots & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{pmatrix}$$

$$BA = \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} + b_{13}a_{31} & \dots & \dots \\ \dots & b_{21}a_{12} + b_{22}a_{22} + b_{23}a_{32} & \dots \\ \dots & \dots & b_{31}a_{13} + b_{32}a_{23} + b_{33}a_{33} \end{pmatrix}$$

$$\text{tr}(AB) = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} + a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} + a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33},$$

$$\text{tr}(BA) = b_{11}a_{11} + b_{12}a_{21} + b_{13}a_{31} + b_{21}a_{12} + b_{22}a_{22} + b_{23}a_{32} + b_{31}a_{13} + b_{32}a_{23} + b_{33}a_{33}.$$

These agree. This should be enough to convince you that the result is true, and the proof is a routine algebraic calculation. The main problem in giving a proof for general  $n$  will be an efficient choice of notation.

## Summary

1. The process of experimentation (rough working) varies enormously from problem to problem. Some general principles are listed in (i)–(vi) in this section.
2. Some of your experimentation may lead to dead ends. Try as many different approaches as you can think of until you find one that leads forward.

## 5.5 Making the proof precise

Having found a method which looks as though it will provide the required proof, the final stage is to fill in the details and to write out a complete proof. Typically, these two tasks can be carried out simultaneously. It may happen that while you are doing this, you encounter some difficulty which you had not previously noticed—you may have made a slip in calculation, or you may have overlooked some logical difficulty, or you may have knowingly made an unwarranted assumption (for simplicity) which you now find hard to remove, or there may be a point where you want to reverse your argument but it is not clear how to do so. If this happens, you will have to return to experimentation until you find a way around the difficulty.

As explained in Section 2.3, the purpose of writing out a rigorous solution is both to ensure that you have not made any errors or overlooked any difficulties, and to provide the reader with an argument which is complete, clear, accurate, and unambiguous. Your solution should provide a proof which will satisfy someone with a similar level of mathematical knowledge to yourself (but someone who has not solved the problem).

In writing your solution, it is a good idea to imagine that you are writing it for such a person, even if you expect the solution to be read only by your tutor.

In writing your formal solution, you should aim to observe the following general points:

- (i) Each statement should be mathematically accurate; in particular, the principles of Sections 4.2, 4.3, 4.4, and 4.5 should be observed.
- (ii) When the symbols are converted into words, each statement should make sense as a sentence in English, and be unambiguous.
- (iii) Each statement should follow logically from previous statements in your argument, or from standard results.
- (iv) It should be clear to the reader why each statement is true. Where appropriate, include phrases such as “By So-and-So’s Theorem, . . . ” or “It follows from (\*) above that . . . ”.
- (v) All assumptions being made at any stage should be made clear to the reader.

We can illustrate (v) with two specific cases. If you are arguing by contradiction, write something like:

The proof will be by contradiction. Suppose that the result is false. Then

...

If your argument is divided into several cases, say explicitly what the cases are. For example:

We shall divide the argument into three cases which together cover all possible integers  $n \geq 2$ :

Case 1:  $n$  has two distinct prime factors  $p_1$  and  $p_2$ ,

Case 2:  $n = 2^k$  for some  $k \geq 1$ ,

Case 3:  $n = p^k$  for some odd prime  $p$ , and some  $k \geq 1$ .

Case 1: ...

What we have said so far in this section concerns your *formal* solution; this material must not be omitted from your written solution. However, sometimes it is appropriate to include some extra explanation which will help the reader understand what you are doing, even though it is not strictly necessary and may even be imprecisely expressed (just as books sometimes include informal explanations of their arguments). This may be a diagram, or it may be some extract from your process of experimentation. Indeed, including material of this type in your written solutions is to be recommended; if you make small errors in the solution, your informal comments often satisfy the reader that they were not errors of substance.

We now give rigorous, complete, solutions to the examples introduced in Section

5.2. These solutions are based on the informal working carried out in Section 5.4, but they are set out in a precise and logical way. In these examples, material written in italics does not form part of the solution; this material contains comments included in these notes for the purposes of explaining why we are setting out the proof as we are; in normal circumstances, these comments would be omitted.

**Example 1.** Let  $x \in (A \cup B) \cap (A \cup C)$ . Then  $x \in A \cup B$  and  $x \in A \cup C$ . Hence  $[x \in A \text{ or } x \in B]$  and (simultaneously)  $[x \in A \text{ or } x \in C]$ . Thus, if  $x \notin A$ , then  $x \in B$  and  $x \in C$ , so  $x \in B \cap C$ . Hence,  $x \in A \text{ or } x \in B \cap C$ , so  $x \in A \cup (B \cap C)$ . Thus,

$$(1) \quad (A \cup B) \cap (A \cup C) \subseteq A \cup (B \cap C).$$

Conversely, let  $x \in A \cup (B \cap C)$ . Then  $x \in A$  or  $x \in B \cap C$ . We consider these two cases separately. If  $x \in A$ , then  $x \in A \cup B$  and  $x \in A \cup C$ , so  $x \in (A \cup B) \cap (A \cup C)$ . If  $x \in B \cap C$ , then  $x \in B$  and  $x \in C$ , so  $x \in A \cup B$  and  $x \in A \cup C$ , so  $x \in (A \cup B) \cap (A \cup C)$ . In either case, we have shown that  $x \in (A \cup B) \cap (A \cup C)$ . Thus,

$$(2) \quad A \cup (B \cap C) \subseteq (A \cup B) \cap (A \cup C).$$

It follows from (1) and (2) that  $(A \cup B) \cap (A \cup C) = A \cup (B \cap C)$ , as required.

*Note how the unions and intersections in the problem have produced statements involving “or” and “and”, so we have had to apply the principles of Section 4.3.*

*It would be sensible to include the Venn diagram given in Section 5.4 with the formal solution.*

**Example 2.** *We first establish a version of the inequality (\*) of Section 5.4. Since we will use this inequality three times, we introduce new variables  $x$  and  $y$ .*

We will first prove that

$$(1) \quad \frac{x^2 + y^2}{2} \geq xy$$

for all real numbers  $x$  and  $y$ . This follows from the inequality

$$0 \leq (x - y)^2 = x^2 - 2xy + y^2.$$

Putting  $x = a^{1/2}$  and  $y = b^{1/2}$  in (1) gives:

$$(2) \quad \frac{a + b}{2} \geq (ab)^{1/2}.$$

Similarly,

$$(3) \quad \frac{c + d}{2} \geq (cd)^{1/2}.$$

Now putting  $x = (ab)^{1/4}$  and  $y = (cd)^{1/4}$  in (1), and then using (2) and (3) gives:

$$(abcd)^{1/4} \leq \frac{(ab)^{1/2} + (cd)^{1/2}}{2} \leq \frac{a + b + c + d}{4}.$$

*There are many other ways of solving this problem.*

**Example 3.** Suppose that  $2^n - 1$  is prime and that  $n = mk$  where  $m$  and  $k$  are positive integers. We have to show that either  $m = 1$  or  $k = 1$ .

Note that

$$x^k - 1 = (x - 1)(x^{k-1} + x^{k-2} + \cdots + 1)$$

for all  $x$ . Hence

$$2^n - 1 = (2^m)^k - 1 = (2^m - 1)(2^{m(k-1)} + 2^{m(k-2)} + \cdots + 1).$$

Since  $2^n - 1$  is prime, one of the factors on the right-hand side must equal 1. If  $2^m - 1 = 1$ , then  $m = 1$ ; if  $2^{m(k-1)} + 2^{m(k-2)} + \cdots + 1 = 1$ , then  $k = 1$ . This completes the proof.

*In the experimentation in Section 5.4, it may have appeared that we were arguing by contradiction. However, this was due to imagining that  $m$  and  $k$  were proper factors of  $n$ . In the formal proof, we avoid the need for a contradiction, and thereby simplify the logic.*

**Example 4.** We first prove that  $1 - y \leq e^{-y}$  whenever  $y \geq 0$ . Let  $f(y) = y + e^{-y}$ . Then  $f'(y) = 1 - e^{-y} \geq 0$  for all  $y \geq 0$ , so  $f$  is increasing in this range. Now,  $f(0) = 1$ , so  $1 \leq f(y) = y + e^{-y}$  for all  $y \geq 0$ . This gives the required inequality.

Now suppose that  $n \geq 1$  and  $0 \leq x \leq n$ . Putting  $y = x/n$  in the inequality obtained in the preceding paragraph gives

$$1 - \frac{x}{n} \leq e^{-x/n}.$$

Since  $x \leq n$ , the left-hand side of this inequality is non-negative. We may therefore raise both sides to the power  $n$ , to obtain:

$$\left(1 - \frac{x}{n}\right)^n \leq e^{-x},$$

as required.

*It would be wise to include the sketch given in Section 5.4 with your solution.*

**Example 5.** By a standard trigonometric formula,

$$\sin x - \sin y = 2 \sin \left( \frac{x - y}{2} \right) \cos \left( \frac{x + y}{2} \right).$$

Moreover,  $|\cos t| \leq 1$  for all  $t$  and  $|\sin t| \leq |t|$  for all  $t$ , so

$$|\sin x - \sin y| \leq 2 \left| \frac{x - y}{2} \right| = |x - y|.$$

You may like to include the diagram shown in Section 5.4.

There are other valid solutions. For example, one can observe that  $\sin x - \sin y = \int_y^x \cos t \, dt$  and use the inequalities  $-1 \leq \cos t \leq 1$ .

**Example 6.** Rather than proving first that  $f(m) = mf(1)$ , and then repeating the argument to prove that  $f(nx) = nf(x)$ , it is more efficient to prove the general result first. For a general discussion of proofs by induction, see Section 5.8.

We assume that

$$(1) \quad f(x + y) = f(x) + f(y) \quad \text{for all } x, y \in \mathbf{Q}.$$

We shall first prove the following, by induction on  $n$ :

For any integer  $n \geq 1$  and any rational  $x$ ,

$$(2) \quad f(nx) = nf(x).$$

For  $n = 1$ , (2) is trivial. Suppose that (2) holds for  $n = k$ . Then by (1) (with  $y = kx$ ),

$$f((k + 1)x) = f(x + kx) = f(x) + f(kx) = f(x) + kf(x) = (k + 1)f(x).$$

Thus (2) holds for  $n = k + 1$ . By induction, (2) holds for all  $n \geq 1$ . Putting  $x = 1$  in (2) shows that

$$(3) \quad f(n) = nf(1).$$

Now, put  $x = y = 0$  in (1). This gives:

$$f(0) = f(0 + 0) = f(0) + f(0),$$

so  $f(0) = 0$ . Next, put  $y = -x$  in (1). This gives:

$$f(x) + f(-x) = f(x + (-x)) = f(0) = 0,$$

so  $f(-x) = -f(x)$ . If  $n$  is a positive integer, then it follows from this and (3) that  $f(-n) = -f(n) = -nf(1)$ . It follows from this and (3) that

$$(4) \quad f(m) = mf(1)$$

for all integers  $m$ .

Now let  $m$  be any integer and  $n \geq 1$ . Then

$$\begin{aligned} mf(1) &= f(m) && \text{(by (4))} \\ &= f\left(n \cdot \frac{m}{n}\right) \\ &= nf\left(\frac{m}{n}\right) && \text{(by (2)).} \end{aligned}$$

Thus  $f\left(\frac{m}{n}\right) = \frac{m}{n}f(1)$ . This completes the proof.

**Example 7.** *The first part of the solution is not strictly necessary, but it will help the reader to see how we found the limit to be 2.*

Suppose that  $a_n \rightarrow l$ . Then  $a_{n+1} \rightarrow l$ , and  $a_{n+1} = \sqrt{2 + a_n} \rightarrow \sqrt{2 + l}$ . Hence,  $l = \sqrt{2 + l}$ , so  $l^2 = 2 + l$ . It follows that  $l = 2$  or  $l = -1$ . Since  $a_n \geq 0$ ,  $l \geq 0$ . Thus the only possible value of the limit is 2.

Now

$$2 - a_{n+1} = \frac{4 - a_{n+1}^2}{2 + a_{n+1}} = \frac{4 - (2 + a_n)}{2 + a_{n+1}} = \frac{2 - a_n}{2 + a_{n+1}}.$$

Since  $a_{n+1} \geq 0$ , it follows that

$$|2 - a_{n+1}| \leq \frac{1}{2}|2 - a_n|.$$

By iteration,

$$|2 - a_{n+1}| \leq \frac{1}{2}|2 - a_n| \leq \frac{1}{4}|2 - a_{n-1}| \cdots \leq \frac{1}{2^{n+1}}|2 - a_0| = \frac{1}{2^{n+1}} \rightarrow 0.$$

Thus  $a_{n+1} \rightarrow 2$ , so  $a_n \rightarrow 2$ .

**Example 8.** Let  $a_{ij}$  denote the entry in the  $i$ th row,  $j$ th column of  $A$ , and  $b_{ij}$  be the corresponding entry of  $B$ . By definition of matrix multiplication, the entry in the  $i$ th row,  $i$ th column (the  $i$ th entry on the diagonal) of  $AB$  is  $\sum_{j=1}^n a_{ij}b_{ji}$ . Hence

$$\text{tr}(AB) = \sum_{i=1}^n \sum_{j=1}^n a_{ij}b_{ji}.$$

Similarly,

$$\text{tr}(BA) = \sum_{i=1}^n \sum_{j=1}^n b_{ij}a_{ji}.$$

In these double sums,  $i$  and  $j$  are dummy variables which may be replaced by any other letters. Replacing them by  $r$  and  $s$  respectively in one case, and by  $s$  and  $r$  respectively in the other, gives:

$$\text{tr}(AB) = \sum_{r=1}^n \sum_{s=1}^n a_{rs}b_{sr}, \quad \text{tr}(BA) = \sum_{s=1}^n \sum_{r=1}^n b_{sr}a_{rs}.$$

Since addition of real numbers is commutative, we can reverse the order of summation in the second double sum. Using also the commutativity of multiplication of real numbers, this gives:

$$\operatorname{tr}(BA) = \sum_{r=1}^n \sum_{s=1}^n b_{sr} a_{rs} = \sum_{r=1}^n \sum_{s=1}^n a_{rs} b_{sr} = \operatorname{tr}(AB),$$

as required.

*Students who are inexperienced with the  $\sum$  notation are often puzzled why it is legitimate to replace  $i$  and  $j$  by  $r$  and  $s$  in this argument, and why it is legitimate to reverse the order of the double summation here. As indicated above, the first is justified because the choice of names for these dummy variables is entirely arbitrary (cf.  $\int f(x) dx = \int f(y) dy$ ); the second is justified because all it amounts to is adding up the same terms in a different order. If you are in any doubt, write out the sums longhand. For  $n = 3$ , this is exactly what we did in Section 5.4; for general  $n$ , you would have to use  $\dots$  to denote strings of missing terms and use your imagination to see that the missing terms match up. It would be legitimate to use such notation in your solution instead of the  $\sum$  signs.*

## Summary

1. Your final proof should observe the principles described in (i)–(v) in this section.
2. In certain circumstances, it may be useful to include some of your rough working and/or informal explanation (for example, a diagram) with your formal proof.

## 5.6 What can you assume?

Students often worry a great deal about this question, particularly in the early stages of their undergraduate careers. There is indeed some difficulty in judging what is permissible to assume, but students are inclined to exaggerate the problem. There are two reasons for saying this. One is that tutors will be much less concerned about a solution in which everything is correct except that you have (knowingly) assumed something which you were really intended to prove than about a solution which contains fallacies or non-sequiturs.

The second reason is that, in some of the courses in your first term or two, you will, for a while, be expected to prove things which you have either known for years (such as the existence of  $n$ th roots of positive numbers) or which are in some sense obvious (such as the fact that a continuous function which takes positive and negative values must also take the value zero somewhere). Simultaneously, you will be doing

other (more applied) courses where these and other familiar or obvious facts are freely used without comment. This can be confusing, but the situation will not last very long. After a while, all courses will be proving facts which you did not know and which are not obvious. When you reach that stage, it should be clearer what you can assume and what you should prove.

Nevertheless, it is possible to give some broad guidelines about this question, all of which are valid in some situations:

- (i) What you are allowed to assume depends on the context; for example, if you are answering questions on a course which is concerned with deriving fundamental properties of the real numbers using only basic axioms, then you should *not* assume these properties without proof; in courses on mechanics etc., you may assume such properties.
- (ii) When answering tutorial problems, you may assume any results which have been proved in lectures.
- (iii) If you are in doubt, a reasonable compromise is to give a clear statement, without proof, of what you are assuming.
- (iv) If you think carefully about the relationship between various results in lectures and the question you are required to answer, you may find that the point of the question becomes clear and that your doubts about what to assume are resolved.
- (v) Use your common sense.

## Summary

1. The problem of what can be assumed may be less serious than you imagine.
2. Bear in mind the principles listed in this section ((i)–(v) above).

## 5.7 Proofs by contradiction

We saw in Section 2.1 one example of a proof by contradiction. This was rather a simple example—the statement (that  $\sqrt{2}$  is irrational) had no hypotheses and a simple conclusion—so there was no great difficulty in setting up the argument. In more advanced situations, proof by contradiction can be a powerful method, but it is a dangerous one, as it is easy to introduce fallacies.

Recall from Section 4.2 that typically one has to prove a statement of the form:

Z1: *If P, then Q.*

Here, P represents the hypotheses and Q the conclusion. In principle, there are three

ways in which Z1 can be proved:

- (i) Assume that P is true, and deduce by some logical process that Q is true (a *direct* proof);
- (ii) Assume that Q is false, and deduce that P is false (i.e., prove the *contrapositive* of Z1);
- (iii) Assume that P is true and Q is false, and deduce some obviously false statement (e.g.,  $1 = 0$ ) (i.e., proof by *reductio ad absurdum*).

The term “proof by contradiction” is commonly applied to proofs of type (ii) or (iii) (although some would say that it should be applied only to (iii)).

Sometimes, it is possible to prove a given statement by two of these three methods, or even by all three. For example, we will now prove the following statement by each method in turn:

Z2: *If  $a^3 + 3a^2 + 3a \geq 0$ , then  $a \geq 0$ .*

In this statement, it is assumed that  $a$  is a real number. The hypothesis is that  $a^3 + 3a^2 + 3a \geq 0$ ; the conclusion is that  $a \geq 0$ .

*First proof:* We assume that  $a^3 + 3a^2 + 3a \geq 0$ . Then  $(a + 1)^3 = a^3 + 3a^2 + 3a + 1 \geq 1$ . Taking cube roots, it follows that  $a + 1 \geq 1$ , so  $a \geq 0$ . This completes a direct proof of Z2.

*Second proof:* Suppose that  $a < 0$ . Then  $a + 1 < 1$ . It follows that  $(a + 1)^3 < 1$ , so  $a^3 + 3a^2 + 3a = (a + 1)^3 - 1 < 0$ . This contradicts the hypothesis of Z2. Thus we have proved Z2 by proving its contrapositive.

*Third proof:* Suppose that  $a^3 + 3a^2 + 3a \geq 0$  and  $a < 0$ . Dividing by the negative number  $a$  gives:

$$(*) \quad a^2 + 3a + 3 \leq 0.$$

But

$$a^2 + 3a + 3 = \left(a + \frac{3}{2}\right)^2 + \frac{3}{4} \geq \frac{3}{4}.$$

This contradicts (\*), so we have proved Z2 by reductio ad absurdum.

If you have *Alice in Numberland*, you will find (p.6) three proofs of the statement: *If  $n^2$  is odd, then  $n$  is odd.*

A principle which is generally, but not universally, true is that your first approach should be to look for a proof of type (i) (a direct proof). The reason why this is a good principle is that both (ii) and (iii) require you to assume the *negation* of Q, i.e. to assume that Q is false. Now, Q is likely to be a fairly complicated mathematical statement, and it is not always easy to write down a mathematical formulation of the negation of Q. There is one obvious exception to this—when Q is itself a negative statement. For

example, the conclusion  $\sqrt{2}$  is irrational is a negative statement, because, by definition, it means  $\sqrt{2}$  is not rational, and the negation of this is  $\sqrt{2}$  is rational. Thus, in this example, and in others where the conclusion  $Q$  is a negative statement, it is natural to look for a proof by contradiction.

Even where the conclusion  $Q$  is a positive statement, you will find many instances where it just does not seem possible to give a direct proof of  $Z1$ , and you are obliged to consider the possibility of a proof by contradiction. The first problem then is to formulate the negation of  $Q$  correctly, in mathematical terms. We shall devote the remainder of this section to how to do this. The negation of  $Q$  is the statement  $Q$  is false, or  $Q$  does not hold, but the problem is to put such a statement in a form which is suitable for mathematical argument.

The first step is to formulate  $Q$  in basic mathematical terms. For example, if  $Q$  is the statement  $f$  is continuous, then the negation of  $Q$  is  $f$  is discontinuous. However, this is useless for the purpose of proving things, unless we have a mathematical definition of discontinuous. In fact, continuity is defined mathematically, and then one has to deduce the definition of discontinuity by taking the negation. The mathematical definition of  $f$  is continuous is:

Z3: For all  $x$ , and for all  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $|f(y) - f(x)| < \epsilon$  whenever  $|y - x| < \delta$ .

At this stage, one should forget about the meaning of the statement, and proceed formally. This may seem surprising advice, but it is the most reliable way of obtaining the correct negation of a complicated mathematical statement. It is likely that the statement can be built up out of a few standard constructions such as the following:

Z4:  $\forall x \in S, P(x)$ ;

Z5:  $\exists x \in S$  such that  $P(x)$ ;

Z6:  $P$  and  $P'$ ;

Z7:  $P$  or  $P'$ .

Here,  $S$  denotes a certain set,  $x$  is a variable,  $P(x)$  is a statement which depends on the variable  $x$  (see Section 4.4), and  $P$  and  $P'$  are statements. All the statements  $P(x)$ ,  $P$  and  $P'$  may be complicated, involving several quantifiers  $\forall$  and  $\exists$  as well as “and” and “or”. The next step is to write  $Q$  in a formal fashion using statements of the form Z4–Z7. For example, we can write Z3 using three constructions of the type Z4 and one of Z5:

Z8:  $\forall x \in \mathbf{R} \forall \epsilon > 0, \exists \delta > 0$  such that  $\forall y \in (x - \delta, x + \delta), |f(y) - f(x)| < \epsilon$ .

To find the negation of a statement  $Q$  constructed out of components Z4–Z7, it is enough to know how to negate each of statements Z4–Z7, and then to apply those principles to each of the components in turn. It is easy to find the negations of Z4–Z7. Using the notation  $\sim P$  to denote the negation of  $P$ , the negations of Z4–Z7 are:

$\sim Z4: \exists x \in S$  such that  $\sim P(x)$ ;

$\sim Z5: \forall x \in S, \sim P(x)$ ;

$\sim Z6: \sim P$  or  $\sim P'$ ;

$\sim Z7: \sim P$  and  $\sim P'$ .

It should be clear to you after a little thought that this list is correct. If you are worried, try taking the following specific examples:

Z9: *Every undergraduate in my college is male;*

Z10: *There is an undergraduate in my college who is less than 10 years old;*

Z11:  $x > 0$  and  $x < 1$ ;

Z12:  $x > 1$  or  $x < 0$ .

Now, if Q is composed of several clauses of the type Z4–Z7, we can find the negation of Q by applying the negation clause by clause until eventually we have only to take the negation of a simple statement. For example, the statement Z8 has the form:

Z13:  $\forall x \in \mathbf{R}, P(x)$ ,

where, for the moment, we are not bothered what  $P(x)$  is. So the negation of Z13 is:

$\sim Z13: \exists x \in \mathbf{R}$  such that  $\sim P(x)$ .

Now, we have to work out what  $\sim P(x)$  is. Since  $P(x)$  is of the form:

Z14:  $\forall \epsilon > 0, Q(x, \epsilon)$ ,

its negation is:

$\sim Z14: \exists x \in \mathbf{R}$  such that  $\sim Q(x, \epsilon)$ .

Hence, the negation of Z8 is:

Z15:  $\exists x \in \mathbf{R} \exists \epsilon > 0$  such that  $\sim Q(x, \epsilon)$ .

Continuing in this way, we eventually find that the negation of Z8 is:

$\sim Z8: \exists x \in \mathbf{R} \exists \epsilon > 0$  such that  $\forall \delta > 0 \exists y \in (x - \delta, x + \delta)$  such that  $|f(y) - f(x)| \geq \epsilon$ .

We can convert this back into a more readable form to find what it means to say that  $f$  is discontinuous:

$\sim Z3: There$  exist  $x \in \mathbf{R}$  and  $\epsilon > 0$  such that, for all  $\delta > 0$ , there exists  $y$  such that  $|y - x| < \delta$  and  $|f(y) - f(x)| \geq \epsilon$ .

It is very easy to summarise this procedure of taking negations:

1. Convert the statement into basic mathematical form by replacing technical terms by their definitions.
2. Write the statement in a formal style using the constructions Z4, Z5, Z6, and Z7.

3. Change each  $\forall$  to  $\exists$ , each  $\exists$  to  $\forall$ , each logical “and” to “or”, and each logical “or” to “and”; take the negations of the remaining simple statements (this should be easy, for example reversing an inequality).
4. Convert the negation into more readable form.

We should warn you about two points. Firstly, when we say that each “and” should be replaced by “or”, and vice versa, this applies only when the word “and” or “or” is used as in Z6 or Z7. It does not apply when “and” or “or” is subordinate to a quantifier. For example, the negation of the statement:

Z16: *For all  $x \in S$  and all  $a \in A$ ,  $P(x, a)$*  (formally,  $\forall x \in S \forall a \in A, P(x, a)$ )

is:

$\sim$ Z16: *There exist  $x \in S$  and  $a \in A$  such that  $\sim P(x, a)$*  (formally,  $\exists x \in S \exists a \in A$  such that  $\sim P(x, a)$ ).

Secondly, in Section 4.5, we recommended that in a statement such as Z3 or Z8, you might write  $\delta_\epsilon$  instead of  $\delta$  to remind yourself that  $\delta$  may depend on  $\epsilon$ . However, if you are going to take the negation of the statement, you should not do such a thing; if you do, the negation will come out somewhat garbled.

At the beginning of this section, we described the three strategies for proving a statement of the form Z1: *If  $P$ , then  $Q$*  (direct, contrapositive, reductio ad absurdum). Often you will be required to prove a statement of the form:

Z17:  *$P$  if and only if  $Q$ .*

This means that you have to prove both of the following (see Section 4.2):

Z1: *If  $P$ , then  $Q$ ;*

Z18: *If  $Q$ , then  $P$ .*

There are three strategies for proving Z1. Whichever way you choose to prove Z1, you are free to choose any of the three strategies to prove Z18. Thus, there are (at least) nine different strategies for proving Z176. For example, you might assume  $P$  and deduce  $Q$  (a direct proof of Z1), and then assume  $Q$  and deduce  $P$  (a direct proof of Z18). Alternatively, you could assume  $P$  and deduce  $Q$ , then assume that  $P$  is false and deduce that  $Q$  is false (a proof of the contrapositive of Z18).

This is an opportunity to repeat a warning made in Section 4.2. It is tempting to try to save time by proving Z1 and Z18 simultaneously. You are advised to resist the temptation, as this process very rarely works. Only occasionally is it possible to give a complete proof which works logically accurately in both directions at once. In attempting to construct such a proof, you may well confuse yourself and end up by proving neither direction satisfactorily.

## Summary

1. To prove the statement *If  $P$ , then  $Q$* , you may give a direct proof, a proof of the contrapositive, or a proof by reductio ad absurdum.

2. To find the negation of a complicated statement, it is advisable to formulate the statement in a very formal manner and to construct the negation step by step.
3. The negations of the basic statements Z4–Z7 in this section are given by  $\sim Z4$ – $\sim Z7$ .
4. To prove the statement *P if and only if Q*, you should give separate proofs of *If P, then Q* and of *If Q, then P*. Each of these may be proved directly, by contrapositive, or by reductio ad absurdum.

## 5.8 Proofs by induction

We have seen one example of a proof by induction in Section 5.5 (Example 6). Here, we want to explain the method in more detail. (Induction is also discussed in *Alice In Numberland*, Chapter 5, and in various other books.)

Induction is useful when it is required to prove a statement where the conclusion is, or can be formulated as:

(\*): *For all integers  $n \geq 1$ ,  $P(n)$  holds.*

Examples include the following:

**Example 1.** *For  $n \geq 1$ ,  $\sum_{r=1}^n r = \frac{n(n+1)}{2}$ ;*

**Example 2.** *For  $n \geq 1$ ,  $\int_0^{2\pi} (\cos \theta)^{2n} d\theta = \frac{(2n)! \pi}{2^{2n-1} (n!)^2}$ .*

**Example 3.** *Let  $f : \mathbf{Q} \rightarrow \mathbf{R}$  be a function such that  $f(x+y) = f(x) + f(y)$  for all  $x$  and  $y$  in  $\mathbf{Q}$ . Then  $f(nx) = nf(x)$  for all  $x \in \mathbf{Q}$  and all  $n \geq 1$ .*

**Example 4.** *Let  $n$  be an integer with  $n \geq 2$ . Then  $n$  is a product of (one or more) primes.*

**Example 5.** *Suppose that  $(a_n)_{n \geq 1}$  is a sequence of real numbers such that  $a_1 = 3$ ,  $a_2 = 6$ , and  $a_{n+2} = a_{n+1} + 2a_n - 2n + 1$  for all  $n \geq 1$ . Then  $a_n = 2^n + n$  for all  $n \geq 1$ .*

**Example 6.** *Suppose that  $n$  is an integer which is not a perfect square. Then  $\sqrt{n}$  is irrational.*

The statement in Example 4 is a well-known fact, and the reader may not previously have thought of it as needing proof.

In some cases such as Example 6, it is not immediately apparent that we are asked to prove a conclusion of the form (\*), until the problem is reformulated. Indeed, Example 6 is equivalent to:

Let  $n \geq 1$ . Then  $\sqrt{n}$  is either an integer or is irrational.

### Simple induction

One way to prove (\*) is as follows:

- (i) Show that  $P(1)$  is true;
- (ii) Show that if  $P(n)$  is true, then  $P(n + 1)$  is true.

It is easy to see that this is adequate to prove that  $P(n)$  is true for all  $n$ . Firstly, (i) ensures that  $P(1)$  is true; then, (ii) with  $n = 1$  ensures that  $P(2)$  is true; then, (ii) with  $n = 2$  ensures that  $P(3)$  is true, etc. Thus, if someone were to specify a positive integer (say, 319), you could tell them that  $P(319)$  is true, by (i) and 318 applications of (ii). In other words,  $P(n)$  is true for all  $n = 1, 2, \dots$ . Note, however, that this method does not tell you that  $P(\infty)$  is true, even if the statement  $P(n)$  makes sense, when  $n$  is replaced by  $\infty$ ; the fact that  $P(n)$  is true for all finite (but arbitrary)  $n$  does not imply that  $P(\infty)$  is true.

A proof by this method is known as a proof by (*simple*) *induction*, and  $P(n)$  is the *inductive statement* to be proved. Stage (i) of the proof is known as the *base case*, and stage (ii) as the *inductive step*. At the beginning of the inductive step, one makes the *inductive hypothesis*: *Suppose that  $P(n)$  holds for some  $n$ .*

You should note that in (\*),  $n$  is a dummy variable, which can be replaced by another variable, i.e. proving that  $P(n)$  is true for all  $n$  is the same as proving that  $P(k)$  is true for all  $k$ . When you are presenting a proof by induction, you should make a clear statement of the inductive hypothesis  $P(n)$ , and you should make clear the name of the dummy variable.

Let us see how this method works in some of our examples.

**Example 1.** We shall prove the following by induction on  $n$ :

$$P(n): \quad \sum_{r=1}^n r = \frac{n(n+1)}{2}.$$

Firstly,  $P(1)$  holds, since for  $n = 1$  both the left and right hand sides equal 1.

Now, suppose that  $P(n)$  holds for some  $n \geq 1$ . Then

$$\sum_{r=1}^{n+1} r = \sum_{r=1}^n r + (n+1) = \frac{n(n+1)}{2} + n+1 = \frac{(n+1)(n+2)}{2},$$

so  $P(n+1)$  holds.

It follows, by induction, that  $P(n)$  holds for all  $n \geq 1$ .

**Example 2.** We shall prove the following by induction on  $n$ :

$$P(n): \quad \int_0^{2\pi} (\cos \theta)^{2n} d\theta = \frac{(2n)! \pi}{2^{2n-1} (n!)^2}.$$

For  $n = 0$ , the left-hand side of the equation  $P(0)$  is:  $\int_0^{2\pi} d\theta = 2\pi$ , and the right-hand side is:  $\frac{\pi}{2^{-1}} = 2\pi$ . Thus,  $P(0)$  holds.

Now, suppose that  $P(n)$  holds for some  $n \geq 0$ . Then

$$\begin{aligned} \int_0^{2\pi} (\cos \theta)^{2(n+1)} d\theta &= \int_0^{2\pi} (\cos \theta)^{2n+1} \cos \theta d\theta \\ &= [(\cos \theta)^{2n+1} \sin \theta]_0^{2\pi} - \int_0^{2\pi} (2n+1)(\cos \theta)^{2n} (-\sin \theta) \sin \theta d\theta \\ &= (2n+1) \int_0^{2\pi} (\cos \theta)^{2n} (\sin \theta)^2 d\theta \\ &= (2n+1) \int_0^{2\pi} \left( (\cos \theta)^{2n} - (\cos \theta)^{2(n+1)} \right) d\theta, \end{aligned}$$

where we used integration by parts in the second line and the formula  $(\cos \theta)^2 + (\sin \theta)^2 = 1$  in the fourth line. Hence,

$$\int_0^{2\pi} (\cos \theta)^{2(n+1)} d\theta = \frac{2n+1}{2n+2} \int_0^{2\pi} (\cos \theta)^{2n} d\theta = \frac{2n+1}{2n+2} \frac{(2n)! \pi}{2^{2n-1} (n!)^2} = \frac{(2n+2)! \pi}{2^{2n+1} ((n+1)!)^2},$$

where we used  $P(n)$  in the second equality. Thus  $P(n+1)$  is true.

It follows by induction that  $P(n)$  is true for all  $n \geq 1$ .

For a formal proof of Example 3 by induction, you are referred to Example 6 of Section 5.5.

### Compound induction

Another way to prove that  $P(n)$  is true for all positive integers  $n$  is as follows:

- (i) Show that  $P(1)$  is true;
- (ii) Show that if  $P(n)$  is true for all  $n = 1, 2, \dots, k$ , then  $P(k+1)$  is true.

Again, it should be clear that this is adequate to prove that  $P(n)$  is true for all integers  $n$ . A proof of this type is said to be a proof by (*compound*) *induction*; in this case, the inductive hypothesis is:

*Suppose that  $P(n)$  is true for all  $n = 1, 2, \dots, k$ .*

Indeed, a proof of  $P(n)$  by compound induction is essentially the same as a proof of  $P(k)^*$  by simple induction, where:

$P(k)^*$ :  $P(n)$  is true for all  $n = 1, 2, \dots, k$ .

**Example 4.** We shall prove, by induction on  $n$ , that every integer  $n \geq 2$  is a product of primes.

For  $n = 2$ , this is trivial, since 2 is itself prime.

Now, let  $k \geq 2$ . We assume as inductive hypothesis that every integer  $n$  such that  $2 \leq n \leq k$  is a product of primes. Now, there are two cases:

*Case 1:*  $k + 1$  is prime. Then  $k + 1$  is the product of one prime, itself.

*Case 2:*  $k + 1$  is not prime. Then  $k + 1 = mn$  for some  $m, n \geq 2$ . Then  $m, n \leq k$ . By the inductive hypothesis,  $m$  and  $n$  are both products of primes, say  $m = p_1 p_2 \dots p_r$ ,  $n = q_1 q_2 \dots q_s$ . Then  $k + 1 = p_1 p_2 \dots p_r q_1 q_2 \dots q_s$ , which is a product of primes.

In each case, we have shown that  $k + 1$  is a product of primes. This completes the inductive step. It follows, by induction, that every integer  $n \geq 2$  is a product of primes.

**Example 5.** We shall prove that  $a_n = 2^n + n$  for  $n \geq 1$ , by induction on  $n$ . For  $n = 1$  and  $n = 2$ , this is trivial since we are given that  $a_1 = 3$  and  $a_2 = 6$ .

Now suppose that  $a_n = 2^n + n$  for all  $n = 1, 2, \dots, k$ , for some  $k \geq 2$ . Then, in particular,  $a_{k-1} = 2^{k-1} + (k-1)$  and  $a_k = 2^k + k$ . Now taking  $n = k-1$  in the given recurrence relation,

$$\begin{aligned} a_{k+1} &= a_k + 2a_{k-1} - 2(k-1) + 1 \\ &= 2^k + k + 2(2^{k-1} + k - 1) - 2k + 3 \\ &= 2^k + 2^k + k + 1 \\ &= 2^{k+1} + (k+1). \end{aligned}$$

This completes the inductive step. Hence, by induction,  $a_n = 2^n + n$  for all  $n \geq 1$ .

## Iterative arguments

We have already explained that a proof by induction works because the inductive step (ii) may be applied iteratively as many times as is needed. For this reason, it is not surprising that some proofs which might be given by induction (and perhaps *should* be given by induction, if one is being extremely fussy) can be set up as arguments by iteration. An example of this is the proof of Example 2 above, which we now set out in a slightly less formal way.

**Example 2.** Let

$$I_n = \int_0^{2\pi} (\cos \theta)^n d\theta \quad (n \geq 0).$$

Then

$$\begin{aligned}
 I_n &= \int_0^{2\pi} (\cos \theta)^{2n-1} \cos \theta \, d\theta \\
 &= [(\cos \theta)^{2n-1} \sin \theta]_0^{2\pi} - \int_0^{2\pi} (2n-1)(\cos \theta)^{2n-2}(-\sin \theta) \sin \theta \, d\theta \\
 &= (2n-1) \int_0^{2\pi} (\cos \theta)^{2n-2} (\sin \theta)^2 \, d\theta \\
 &= (2n-1) \int_0^{2\pi} \left( (\cos \theta)^{2(n-1)} - (\cos \theta)^{2n} \right) \, d\theta,
 \end{aligned}$$

where we used integration by parts in the second line and the formula  $(\cos \theta)^2 + (\sin \theta)^2 = 1$  in the fourth line. Hence,

$$2nI_n = (2n-1)I_{n-1},$$

and, by iteration,

$$\begin{aligned}
 I_n &= \frac{2n-1}{2n} I_{n-1} \\
 &= \frac{(2n-1)(2n-3)}{2n(2n-2)} I_{n-2} \\
 &\dots\dots\dots \\
 &= \frac{(2n-1)(2n-3)\dots 1}{2n(2n-2)\dots 2} I_0.
 \end{aligned}$$

But  $I_0 = \int_0^{2\pi} d\theta = 2\pi$ , so

$$I_n = \frac{2n(2n-1)(2n-2)(2n-3)\dots 2.1}{2n(2n)(2n-2)(2n-2)\dots 2.2} 2\pi = \frac{(2n)!\pi}{2^{2n-1}(n!)^2}.$$

In many examples, the iterative argument is so simple that it would be futile to formulate it as a proof by induction (see the iteration near the end of Example 7 in Section 5.5). In some examples such as Example 2 here, there is not much to choose between the two methods. In other examples where a proof by induction can be given, it is not feasible to set this out as an iterative argument, e.g., Examples 1 and 5 in this section. There is no hard-and-fast rule which tells you when an iterative argument is possible and when it is not. However, the process of experimentation (Section 5.4) which led you towards a proof by induction or iterative argument, and your attempts to give a rigorous argument (Section 5.5), should also indicate whether you need to give a formal proof by induction.

## Least criminals

There is one more variant of induction which is worth mentioning here. Another way to prove (\*) is the following:

Assume that there is a smallest integer  $k$  such that  $P(k)$  is false, and deduce a contradiction (e.g., by showing that there is an integer  $n < k$  such that  $P(n)$  is false).

This argument is valid because if there were a positive integer  $n$  such that  $P(n)$  is false, then there would be a smallest positive integer  $k$  such that  $P(k)$  is false. If we call an integer  $n$  *criminal* if  $P(n)$  is false, then  $k$  is a least criminal, so such a proof is sometimes known as a *least criminal* argument. Essentially, a least criminal argument is what you get by combining the concept of proof by induction with the concept of proof by contradiction (Section 5.7). An example of a least criminal argument follows.

**Example 6.** Suppose that there is a positive integer  $n$  such that  $\sqrt{n}$  is neither an integer nor irrational. Then there is a least such integer  $k$ . Now,  $\sqrt{k}$  must be rational, but not an integer, so  $\sqrt{k} = m/n$ , where  $m$  and  $n$  are positive integers and  $n \geq 2$ . We may assume that  $m$  and  $n$  have no common factor. Let  $p$  be a prime dividing  $k$ . Then  $p$  divides  $n^2k$ . But  $n^2k = m^2$ , so  $p$  divides  $m^2$ . Since  $p$  is prime, it follows that  $p$  divides  $m$ . Hence  $p^2$  divides  $n^2k$ . Since  $m$  and  $n$  are coprime,  $p^2$  must divide  $k$ . Thus  $k = p^2a$  and  $m = pb$  for some positive integers  $a$  and  $b$ . Now  $\sqrt{a} = b/n$ , which is neither an integer nor irrational. But  $a < k$ . This contradicts the choice of  $k$  as the *least* such integer. It follows that, for all  $n \geq 1$ ,  $\sqrt{n}$  is either an integer or is irrational.

Note that this least criminal argument is not the only solution of Example 6. In principle, it is always possible to reorganise a least criminal argument as induction (try doing this for Example 6), but there are some cases where least criminals are the best approach. A striking example is the successful (1976) proof of the Four Colour Theorem (see Section 5.1).

## Summary

1. To prove that a statement  $P(n)$  is true for all positive integers  $n$ , it may be appropriate to prove the statement by induction.
2. Proofs by induction may be by simple induction, or compound induction, or may be formulated as inductive arguments or as the least criminal method.
3. When you intend to prove a statement by induction, you should say so clearly, and you should make it very clear what statement is being proved.

## APPENDIX

### Some symbols

This appendix gives some of the standard mathematical symbols which you are likely to meet early in your course; some of them have been used in the text of these notes.

#### Greek letters

There are certain conventions which symbols are used for which purpose in mathematics, for example  $x$  and  $y$  usually denote real variables;  $a$ ,  $b$  and  $c$  usually denote real constants;  $i$ ,  $j$ ,  $k$ ,  $m$  and  $n$  usually denote integers. Consequently, we soon exhaust our usual alphabet and we have to turn to Greek. The following lists those which are used in mathematics (two lower case letters, omicron and upsilon, and several upper case letters, are rarely used because of their similarity to other symbols).

#### Lower case

$\alpha$	alpha	$\iota$	iota	$\rho$	rho
$\beta$	beta	$\kappa$	kappa	$\sigma$	sigma
$\gamma$	gamma	$\lambda$	lambda	$\tau$	tau
$\delta$	delta	$\mu$	mu	$\phi$ or $\varphi$	phi
$\epsilon$ or $\varepsilon$	epsilon	$\nu$	nu	$\chi$	chi
$\zeta$	zeta	$\xi$	xi	$\psi$	psi
$\eta$	eta	$\pi$	pi	$\omega$	omega
$\theta$	theta				

#### Upper case

$\Gamma$	Gamma	$\Xi$	Xi	$\Phi$	Phi
$\Delta$	Delta	$\Pi$	Pi	$\Psi$	Psi
$\Theta$	Theta	$\Sigma$	Sigma	$\Omega$	Omega
$\Lambda$	Lambda				

## Some special sets

The following are standard ways of denoting some special sets of numbers.

<b>C</b> or $\mathbb{C}$	The set of all complex numbers
<b>N</b> or $\mathbb{N}$	The set of all positive integers
<b>Q</b> or $\mathbb{Q}$	The set of all rational numbers
<b>R</b> or $\mathbb{R}$	The set of all real numbers
<b>Z</b> or $\mathbb{Z}$	The set of all integers
$(a, b)$	The set of all real numbers $x$ such that $a < x < b$
$[a, b]$	The set of all real numbers $x$ such that $a \leq x \leq b$
$(a, \infty)$	The set of all real numbers $x$ such that $a < x$

N.B:  $\infty$  is not a real number; it is a symbol used in certain specific contexts with specific meanings.

## Set-theoretic symbols

The following is standard notation for the set of all those elements  $x$  which have a given property P:

$$\{x : x \text{ has property P}\}.$$

The following is a list of some of the most commonly used set-theoretic symbols.

Symbol	Name	Example	Interpretation
$\in$	belongs to	$a \in A$	$a$ is a member of the set $A$
$\cup$	union	$A \cup B$	$\{x : x \in A \text{ or } x \in B\}$
$\cap$	intersection	$A \cap B$	$\{x : x \in A \text{ and } x \in B\}$
$\setminus$	difference of sets	$A \setminus B$	$\{x : x \in A \text{ and } x \notin B\}$
$\times$	product of sets	$A \times B$	The set of all ordered pairs $(a, b)$ , where $a \in A$ and $b \in B$
$\subseteq$	subset	$A \subseteq B$	$A$ is a subset of $B$ , i.e. every member of $A$ is a member of $B$
$\emptyset$	empty set		the set with no members

## Other notation

$f : A \rightarrow B$	$f$ is a function from $A$ to $B$ , i.e. to each member $a$ of $A$ , $f$ assigns a value $f(a)$ in $B$
$\implies$	implies (see Section 4.2)
$\iff$	if and only if (see Section 4.2)

N.B: Do not use  $\implies$  as shorthand for “Therefore” or “It follows that ...”.