PART I
Studying Analysis

This part of the book discusses productive ways to think and study when learning Analysis. Chapter 1 is very short—it demonstrates what a page of Analysis notes looks like and gives initial comments on notation and on the form of mathematics at this level. Chapter 2 discusses axioms, definitions and theorems, demonstrating ways to relate abstract statements to examples and diagrams. Chapter 3 discusses proofs—it explains how mathematical theories are structured and provides research-based guidance on how to read and understand logical arguments. Chapter 4 discusses what it feels like to study Analysis, how to keep up, how to avoid wasting time, and how to make good use of resources such as lecture notes, fellow students, and support from lecturers and tutors.

CHAPTER 1

What is Analysis Like?

This chapter demonstrates what definitions, theorems and proofs in Analysis look like. It introduces some notation and explains how symbols and words in Analysis are used and should be read. It points out differences between this type of mathematics and earlier mathematical procedures, and gives initial comments on learning about mathematical theories in a lecture course.

Analysis is different from earlier mathematics, and students who want to understand it therefore need to develop new knowledge and skills. This chapter demonstrates this by showing, on the next page, a typical section of Analysis lecture notes. I do not expect you to understand these notes—the aim of the book is to teach the skills you'll need in order to do that, and Chapter 5 covers the relevant material on sequence convergence. But I do want it to be clear that Analysis is demanding. So turn the page, read what you can, then continue.

Definition: $(a_n) \to a$ if and only if

$$\forall \varepsilon > 0\ \exists N \in \mathbb{N} \text{ such that } \forall n > N,\ |a_n - a| < \varepsilon.$$

Theorem: Suppose that $(a_n) \to a$ and $(b_n) \to b$. Then $(a_n b_n) \to ab$.

***Proof*:** Let $(a_n) \to a$ and $(b_n) \to b$.

Let $\varepsilon > 0$ be arbitrary.

Then $\exists N_1 \in \mathbb{N}$ such that $\forall n > N_1$, $|a_n - a| < \dfrac{\varepsilon}{2|b| + 1}$.

Also (a_n) is bounded because every convergent sequence is bounded.

So $\exists M > 0$ such that $\forall n \in \mathbb{N}$, $|a_n| \leq M$.

For this M, $\exists N_2 \in \mathbb{N}$ such that $\forall n > N_2$, $|b_n - b| < \dfrac{\varepsilon}{2M}$.

Let $N = \max\{N_1, N_2\}$.

Then $\forall n > N$,

$$\begin{aligned}
|a_n b_n - ab| &= |a_n b_n - a_n b + a_n b - ab| \\
&\leq |a_n(b_n - b)| + |b(a_n - a)| \\
&\qquad\qquad\qquad \text{by the triangle inequality} \\
&= |a_n||b_n - b| + |b||a_n - a| \\
&< \frac{M\varepsilon}{2M} + \frac{|b|\varepsilon}{2|b| + 1} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Hence $(a_n b_n) \to ab$.

Practically every page of your Analysis notes will look like this. On the one hand, that's exciting—you'll be learning some sophisticated mathematics. On the other hand, as you can probably imagine, students who do not know how to interpret such material cannot make sense of Analysis at all. To them, every page looks the same: full of symbols like 'ε', '$\mathbb{N}$', '$\forall$' and '$\exists$', and empty of meaning. By the end of this book, you will be equipped to understand such material: to identify its key components, to recognize how these fit together to form a coherent theory, and to appreciate the intellectual achievements of the mathematicians who created that theory. Right now, I just want to draw your attention to a few important features of the text.

The first feature is that text like this contains a lot of symbols and abbreviations. Here is a list stating what each one means:

(a_n)	a general sequence (usually read as '*a n*')
$\rightarrow$	'tends to' or 'converges to'
$\forall$	'for all' or 'for every'
ε	epsilon (a Greek letter, used here as a variable)
$\exists$	'there exists'
$\in$	'in' or '(which) is an element of'
$\mathbb{N}$	the natural numbers (the numbers $1, 2, 3, \ldots$)
max	'(the) maximum (of)'
$\{N_1, N_2\}$	the set containing the numbers N_1 and N_2

Such a list gives you immediate power in that you might not understand the text, but at least now you can read it aloud. Try it: pick a few lines, refer to the list where necessary, and read out what those lines say. You should be able to do this with fairly natural inflections, even if it takes a few attempts. That's because mathematicians write in sentences, so the page might look like a jumble of symbols and words, but it can be read aloud like other text. It might be a while before you can read such material fluently, but fluent reading should be your goal, because if all your energy is taken up with remembering symbol meanings you have little chance of understanding the content. So take opportunities to practise, even if it feels a bit slow and unnatural at first. Don't let lecturers[1] be the only ones who can 'speak' mathematics—aim to own it yourself.

While on the subject of symbols, I would like to comment on their use in this book. Symbols function as abbreviations: they allow us to express mathematical ideas in a condensed form. For this reason, I like them. However, not all lecturers share this view. Some worry that learning new symbols takes up students' mental resources and thereby interferes with their understanding of new concepts. Such lecturers prefer to avoid the symbols and write everything out in words. They are right, of course: it does take a while to get used to new symbols. However, I think that this would be true whenever the symbols were introduced, that there really aren't that many, and that the power they confer makes it worth mastering them early. So I'm going to use them straightaway. I'd like to tell you that I have evidence that this is the best approach, but in this case I don't—it's just personal preference. You can find a full list of symbols used in this book in the Symbols section on page xiii.

The second thing to notice about the page of notes is that it contains a definition, a theorem and a proof. The definition states what it means for a sequence to converge to a limit. This might be far from obvious at this point, but don't worry about that—I'll discuss it in detail in Chapter 5. The theorem is a general statement about what happens when we combine two convergent sequences by multiplying together their respective terms. You can probably see this, and you might be ready to agree that the theorem seems reasonable. The proof is an argument[2] showing that the theorem is true. This argument uses the definition of convergence—notice that some of the symbol strings used in the definition reappear in the proof. The proof starts by assuming that the two sequences (a_n) and (b_n) satisfy the definition, and ends by concluding that the sequence $(a_n b_n)$ satisfies the definition too. It takes some thought to see exactly how the argument fits together, but this book will teach you to look

[1] As noted in the Introduction, British people use the word 'lecturer' to refer to anyone who teaches undergraduate students.

[2] When mathematicians say 'argument', they don't mean a verbal fight between two people, they mean a single chain of logically valid reasoning. People use the word in this way in everyday life when they say things like 'That's not a very convincing argument.'

for structures on that level, and I will refer you back to this proof from Section 5.10.

What the page of notes does not contain is a procedure to follow. It is extremely important to recognize this. Students whose mathematical experience to date has consisted mostly of following procedures are often slow to do so. They look for procedures everywhere; they are mystified when they don't find many, and they fail to meaningfully interpret what *is* there. Analysis, like much undergraduate pure mathematics, can be understood as a theory: a network of general results linked together by valid logical arguments known as proofs (see Chapter 3—in particular Section 3.2). The fact that a proof is valid for all objects that satisfy the premises of the associated theorem (see Section 2.2) means that it could be applied repeatedly to particular objects. However, Analysis does not focus on repetitious calculations. Rather, its focus is the theory: it is the theorems, proofs and ways of thinking about them that you are supposed to understand.

The final thing to know is that developing this understanding is your responsibility. You will, of course, have an Analysis lecturer, and maybe an academic tutor or a graduate teaching assistant to offer further face-to-face teaching. These people will do their best to support your learning, but at least some of the time you will be part of a large class where opportunities for individual attention are limited, and you will leave numerous lectures with only a partial understanding of the new material. You therefore need to get good at working out for yourself what it all means. This book is designed to help you do that, starting in the next chapter with some information on the components of mathematical theories.

CHAPTER 2

Axioms, Definitions and Theorems

This chapter is about the building blocks of mathematical theories: axioms, definitions and theorems. It explains their typical logical structures and describes strategies for relating them to examples and diagrams. It illustrates these strategies using Rolle's Theorem and the definition of 'bounded above', and it discusses the utility and limitations of diagrams in general. Finally, it discusses counterexamples and the importance of recognizing the difference between a theorem and its converse.

2.1 Components of mathematics

The main components of a mathematical theory like Analysis are axioms, definitions, theorems and proofs. This chapter discusses the first three of these. Proofs are discussed separately in Chapter 3, but I recommend that you start here, even if you have already begun an Analysis course and you think you are struggling primarily with the proofs—at least some difficulties with proofs arise when people have not fully understood the relevant axioms, definitions and theorems or have not fully understood how a proof should relate to these.

Lots of the axioms, definitions and theorems in Analysis can be represented using diagrams, though people vary in the extent to which they do this. I like diagrams because I find them helpful for understanding abstract information. So I will use a lot of diagrams in this book, and in this chapter I will explain how they can be used to represent both specific and generic examples. I will also offer some words of warning about the limitations of diagrams and the importance of thinking beyond the examples that first come to mind. People who have read *How to Study for/as a Mathematics Degree/Major* will recognize some ideas in this chapter; here the discussion is briefer but more specific to Analysis.

2.2 Axioms

An *axiom* is a statement that mathematicians agree to treat as true; axioms form a basis from which we develop a theory. In Analysis axioms are used to capture intuitive notions about numbers, sequences, functions and so on, so your earlier experience will usually lead you to recognize them as true. They include things like these:

$$\forall a, b \in \mathbb{R}, \quad a + b = b + a;$$
$$\exists\, 0 \in \mathbb{R} \text{ s.t. } \forall a \in \mathbb{R}, \quad a + 0 = a = 0 + a.$$

Don't forget to practise reading aloud. Here is a list of the relevant symbols and abbreviations:

$\forall$	'for all' or 'for every'
$\in$	'in' or '(which) is an element of'
$\mathbb{R}$	the real numbers (often read as 'the reals' or simply as 'R')
$\exists$	'there exists'
s.t.	'such that'

Thus, for instance,

$$\forall a, b \in \mathbb{R}, \quad a + b = b + a$$

is read as

'For all *a*, *b* in the reals, *a* plus *b* is equal to *b* plus *a*.'

Axioms sometimes have names, so you might see bracketed information before or after each one, like this:

$$\forall a, b \in \mathbb{R}, \quad a + b = b + a \qquad \text{[commutativity of addition];}$$
$$\exists\, 0 \in \mathbb{R} \text{ s.t. } \forall a \in \mathbb{R}, \quad a + 0 = a = 0 + a \quad \text{[existence of an additive identity].}$$

Can you infer the meanings of 'commutativity' and 'additive identity' by looking at these axioms? Can you explain these concepts in your own words without sacrificing accuracy?

Axioms for the real numbers will be discussed in more detail in Chapter 10, which also explains the philosophically interesting shift we make when thinking about mathematical theories in these terms.

2.3 Definitions

A definition is a precise statement of the meaning of a mathematical word. In Analysis you will encounter definitions of new concepts and definitions of concepts that are already familiar. It is the second kind, believe it or not, that will cause you more bother. This is for two reasons. First, some of these definitions will be complicated compared with your existing understanding. They are only as complicated as they need to be and you will come to appreciate their precision, but they take some effort to master and you might have to work through a stage of wondering why things aren't simpler. Second, some of the defined concepts will not quite match your intuitive understanding, so your intuition and the formal theory will occasionally tell you different things, and you will have to sort out the conflict and override your intuitive responses if necessary.

Because of this, I will postpone discussion about definitions of familiar concepts until Part 2. In this chapter I will introduce some definitions of concepts that are likely to be unfamiliar—at least to readers who have not yet studied much undergraduate mathematics—and use these to illustrate skills for interacting with definitions: relating definitions to multiple examples, thinking in terms of diagrams, and being precise.

We will start with the definition below, which I provide in two forms, the first using symbols and the second with (almost) everything written out in words. This should help with your reading aloud, but I'll stop doing it soon so keep up the practice.

Definition: A function $f : X \to \mathbb{R}$ is *bounded above on* X if and only if $\exists M \in \mathbb{R}$ s.t. $\forall x \in X, f(x) \leq M$.

Definition (in words): A function f from the set X to the reals is *bounded above on* X if and only if there exists M in the reals such that for all x in X, $f(x)$ is less than or equal to M.

Definitions like this appear routinely in Analysis lectures. They have a predictable structure, and there are two things to notice. First, each definition defines a single concept—this one defines what it means for a certain kind of function to be *bounded above*. In printed material the concept being defined is commonly italicized as here or printed in bold; in handwritten notes, you might see and use underlining instead. Second, this term is said to apply *if and only if* something is true. It is probably easier to see why this is appropriate by considering a simpler definition like this one ('integer' is the proper mathematical name for a whole number):

Definition: A number n is *even* if and only if there exists an integer k such that $n = 2k$.

Splitting this up should enable you to see why both the 'if' and the 'only if' are appropriate:

A number n is even *if* there exists an integer k such that $n = 2k$.
A number n is even *only if* there exists an integer k such that $n = 2k$.

That said, you might see definitions written with just the 'if'. I think this is not ideal, but lots of mathematicians do it because they all know what is intended.

Did you understand the definition of *bounded above*? We will take it apart in detail in the following sections.

2.4 Relating a definition to an example

One way to understand new definitions is to relate them to examples. That sounds simple, but it is important to understand that when mathematicians use the word *example*, they do not usually mean a worked example that shows how to carry out a type of calculation. Rather, they mean a specific object (a function, perhaps, or a number or a set or a sequence) that satisfies a certain property or combination of properties. This can cause miscommunication between lecturers and students. Students say 'We want more examples,' meaning that they want more worked examples, and lecturers think, 'What are you talking about? I've given loads of examples,' meaning examples of objects that satisfy the properties under discussion. Because advanced mathematics is less about learning and applying procedures and more about understanding logical relationships between concepts, worked examples are fewer and further between. And examples of objects are more important—knowledge of a few key examples can clarify logical relationships and help you to remember them. Because of this, your lecturers will almost certainly illustrate definitions using examples. But I want you to develop confidence in generating your own so that you don't have to rely on lecturers for this; in this section and the next I'll describe some ways to go about it.

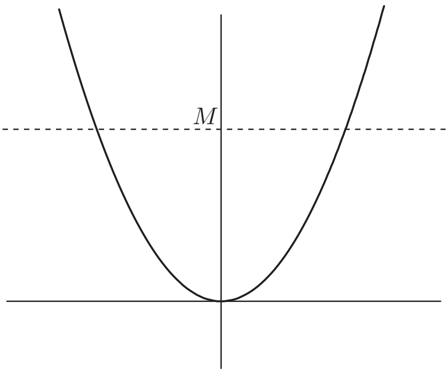
To get us started, here again is the definition of *bounded above* (if you understood this definition immediately, good, but you might like to read the following explanation anyway as it includes advice on thinking beyond your initial understanding).

Definition: A function $f : X \to \mathbb{R}$ is *bounded above on* X if and only if $\exists M \in \mathbb{R}$ s.t. $\forall x \in X, f(x) \leq M$.

This definition defines a property of a function $f : X \to \mathbb{R}$, meaning that f takes elements of the set X as inputs and returns real numbers as outputs. Many people, when asked to think about a function, think about $f(x) = x^2$, so we will start with that. Notice that this function is defined for every real number, so its domain is $X = \mathbb{R}$ and it is a function $f : \mathbb{R} \to \mathbb{R}$. To establish whether or not this function is bounded above, we ask whether or not the definition is satisfied. Substituting in all the appropriate information, $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is *bounded above on* $\mathbb{R}$ if and only if $\exists M \in \mathbb{R}$ s.t. $\forall x \in \mathbb{R}, x^2 \leq M$. Check to make sure you can see this.

So, is the definition satisfied? Does there exist a real number M such that for every real number x, $x^2 \leq M$? Even if you can answer immediately, it is worth noting that it is often easier to start making sense of a definition like this from the end rather than from the beginning. Here the last part says '$f(x) \leq M$', which can be thought of in terms of checking whether values on the vertical axis[1] are less than or equal to M:

[1] You might want to call this the y-axis. That's fine, but I will tend to use the notation $f(x)$ instead of y because it generalizes better when working with multiple functions (as we often do in Analysis) or with functions of more than one variable (as we do in multivariable calculus).



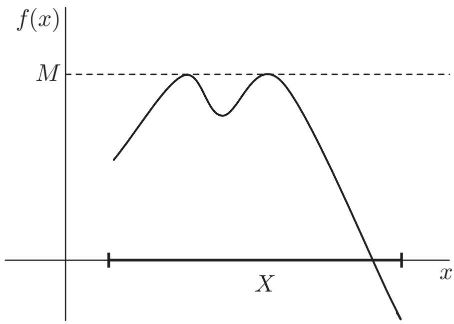
For the M shown, some numbers x in the domain $\mathbb{R}$ have $f(x) \leq M$ and some don't. So for this M it is not true that $\forall x \in \mathbb{R},\ f(x) \leq M$. However, we are interested in whether or not *there exists* M such that $\forall x \in \mathbb{R}, f(x) \leq M$. Does there exist such a number? No. Even for a really big M, there will still be domain values x for which $f(x) > M$. So this function does not satisfy the definition, meaning that it is not bounded above on the set $X = \mathbb{R}$.

2.5 Relating a definition to more examples

To think about a function that *is* bounded above on a set, we can do three things. The first is probably the most obvious: think about different functions. Can you think of a function that is bounded above? Can you think of lots of different ones, in fact? One that might come to mind is $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \sin x$, which is bounded above by $M = 1$ because $\forall x \in \mathbb{R}, \sin x \leq 1$. It is also bounded above by $M = 2$, notice, because it is also true that $\forall x \in \mathbb{R}, \sin x \leq 2$ (there is nothing in the definition to say that M has to be the 'best' bound). We might also consider really simple functions like the constant function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 106$ (meaning that $f(x) = 106\ \forall x \in \mathbb{R}$). This isn't very interesting but it is a perfectly good function, and it is certainly bounded above. Or we might consider $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 3 - x^2$. This is bounded above by 3, for instance. It happens not to be bounded below—could you write down a definition of *bounded below* and confirm this? And can you think of a function that is not bounded above and not bounded below?

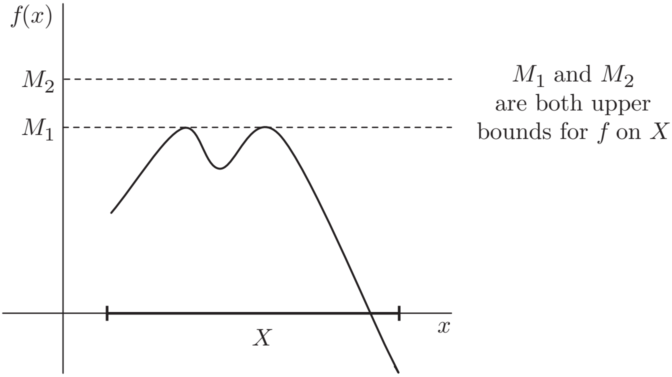
A second thing we might consider is changing the set X. This is less likely to occur to new undergraduate students, because earlier mathematics tends to involve functions from the reals to the reals. But there is nothing to stop us restricting the domain to, say, the set $X = [0, 10]$ (the set containing the numbers 0, 10 and every number in between). The function $f : [0, 10] \to \mathbb{R}$ given by $f(x) = x^2$ is bounded above on $[0, 10]$ by $M = 100$, because $\forall x \in [0, 10], f(x) \leq 100$. What other numbers could play the role of M here?

Finally, we might stop thinking about specific functions, and instead imagine a generic one. To get a general sense of what this definition says, I might draw or imagine something like this:

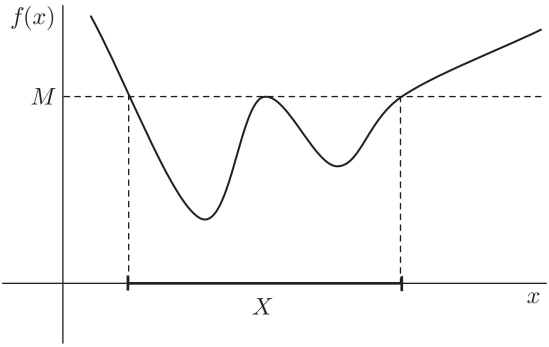


This diagram represents a function on a set X, because for every $x \in X$ (the thickened bar on the x-axis) there is a corresponding $f(x)$. But it isn't supposed to represent a function for which I have a formula in mind, which is appropriate because the definition applies to all functions, not just to those that can be specified by nice formulas. I have also made an effort to capture all aspects of the definition. The diagram shows a restricted set X, for instance, rather than assuming that $X = \mathbb{R}$. In fact, I have drawn a function that is defined only on this set X. Students usually tend to sketch functions that are defined on the whole of $\mathbb{R}$, but that isn't necessary. I have also shown a specific M on the vertical axis, and extended a horizontal line across so that it's clear that all the $f(x)$ values lie below this. Finally, I have made $f(x)$ equal to M in a couple of places to illustrate the fact that this is allowed.

These things all relate to information that appears explicitly in the definition. However, I could add more to the diagram, either to exhibit my own understanding or to explain the definition to someone else. I could, for instance, illustrate the fact that greater values of M are also upper bounds by adding another one, perhaps with some commentary:



Or I could emphasize the fact that the function is bounded above *on the set X* by extending the graph upwards elsewhere; this would illustrate the idea that the definition says nothing about $f(x)$ values for $x \notin X$ (the symbol '$\notin$' means 'not in'):



Or I could think about a more complicated set X. You don't have to do this type of exploration, but I think that it provides a fuller sense of the meaning of the definition, which is important because your lecturer will not have time to give an extensive explanation for every concept. Most likely, he or she will introduce a definition and show how it applies (or doesn't) to just one or two examples, assuming that you will do thinking like this for yourself.

2.6 Precision in using definitions

Later chapters will contain information about specific definitions and guidance on how to recognize where these are used in proofs. Here I want to emphasize the importance of precision when working with definitions. To demonstrate what I mean, here is another definition:

Definition: M is an *upper bound* for the function f on the set X if and only if $\forall x \in X, f(x) \leq M$.

This definition and the previous one are about the same core ideas. But the previous one defines what it means for a function to be bounded above on a set—it is about the *function*. This one defines what it means for a number to be a bound for a function on a set—it is about the *number*. This is a subtle distinction but it is one that mathematicians take care over. Imagine that an exam or test asks what it means for M to be an upper bound for a function f on a set X. This requires the second definition, and there are two ways in which a student might fail to answer well. The first is to give an informal answer, saying something like 'It means the function is below M'. When I read this kind of thing I sigh, because the student has clearly understood something about the concept but failed to grasp the fact that mathematicians work with precise definitions.[2] The second is to give the first definition, defining what it means for a function to be bounded rather than what it means for M to be a bound. This would be better but it would still not merit full marks because it doesn't answer the question as asked.

To illustrate how things can go more badly wrong, consider this third definition, which is also to do with boundedness:

[2] For more on why, see Chapter 3 in *How to Study for/as a Mathematics Degree/Major.*

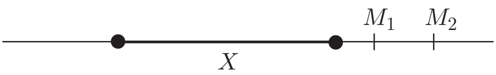
Definition: The set X is *bounded above* if and only if $\exists\, M \in \mathbb{R}$ s.t. $\forall x \in X, x \leq M$.

Students often confuse this with the definition of a function f being bounded on a set X. But look carefully: in this case *there is no function*. This definition is not about a function being bounded above on a set, it is about a *set* being bounded above; it is x-values that are related to M. Here is an example of a set that is bounded above:

$\{x \in \mathbb{R} \mid x^2 < 3\}$ ('the set of all x in R such that x^2 is less than 3').

This set is bounded above by, for example, $\sqrt{3}$, or by 522.

Here is an appropriate generic diagram:



Notice that this definition is just about sets of real numbers so there is no need for a two-dimensional graph—everything of interest can be represented on a single number line. I hope this convinces you that attention to detail is important if we are to distinguish related concepts. And note that the need for precision makes it risky to memorize definitions by rote—much better to understand them properly so you can reconstruct them meaningfully.

2.7 Theorems

A theorem is a statement about a relationship between concepts. Usually this is a relationship that holds *in general*, where I use this phrase in the mathematical sense: when mathematicians says 'in general', they often mean in all cases, not just in the majority of cases.[3] In this section and the next I will explain how to understand theorems by identifying their *premises* and *conclusions* and by systematically seeking examples

[3] As a student you should pay attention to differences between everyday and mathematical English so that you do not get confused or misinterpret what someone is saying. If you do, I predict that you will find these differences strange for a couple of months, then you will stop noticing them, then you will become someone who naturally uses words in a mathematical way.

that demonstrate why each premise is needed. We will work with this theorem, which is about functions (the notation is explained below):

Rolle's Theorem:
Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), and that $f(a) = f(b)$. Then $\exists\, c \in (a, b)$ such that $f'(c) = 0$.

All theorems have one or more *premises*—things that we assume—and a *conclusion*—something that is definitely true if the premises are true. In this case, the premises are flagged by the word 'Suppose'. They are:

- that f is a function defined on an interval $[a, b]$;
- that f is continuous on the interval $[a, b]$;
- that f is differentiable on the interval (a, b);
- that $f(a) = f(b)$.

That's quite a few premises; each will be discussed in detail below.

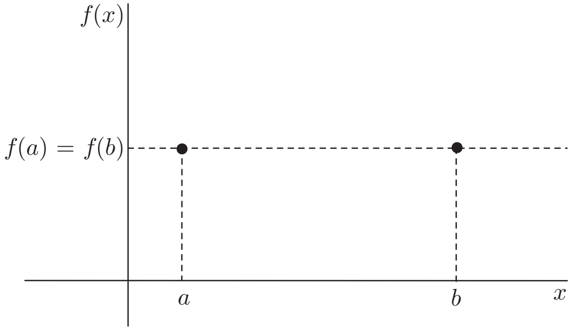
The conclusion is flagged by the word 'Then'; in this case it is that there exists $c \in (a, b)$ such that $f'(c) = 0$. The notation $f'(c) = 0$ means that the derivative of f at c is zero,[4] and the theorem tells us that there exists a point c in (a, b) with this property (the *open interval* (a, b) is the set containing all the numbers between a and b but not including a or b). The theorem does not tell us exactly where c is—*existence theorems* like this are quite common in advanced mathematics.

As with definitions, we can think about how theorems relate to examples. In this case, we can ask how the theorem applies (or doesn't) to specific functions. To satisfy the premises, a function needs to be defined on a *closed interval* $[a, b]$ (the notation $[a, b]$ means the set containing a and b and every number in between). So we need to decide on a function and on values for a and b as well. For instance, if we take $f(x) = x^2$ with $a = -3$ and $b = 3$, then $f(a) = f(b)$ and f is continuous and differentiable everywhere, so all the premises are satisfied. Thus the conclusion holds: there exists c in (a, b) such that $f'(c) = 0$. In this case the derivative happens to be 0 at $c = 0$, which is certainly between -3 and 3.

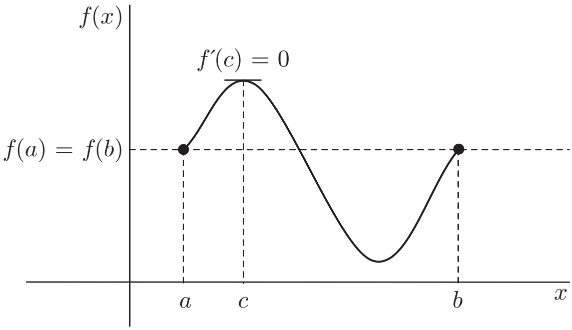
Again, you could think of more examples. But I would suggest that when dealing with theorems like this one, you might as well think

[4] Many students are more accustomed to the notation $\mathrm{d}f/\mathrm{d}x$ for derivatives, but the $f'(x)$ notation is briefer and is more common in Analysis.

straightaway about a generic diagram. In this case that is doubly beneficial because it requires more careful thought about the premises. To draw a generic diagram, the obvious thing might be to start by drawing a function, but in fact it's often easier to start with the simpler premises. Here, for instance, we can start with points a and b such that $f(a) = f(b)$:

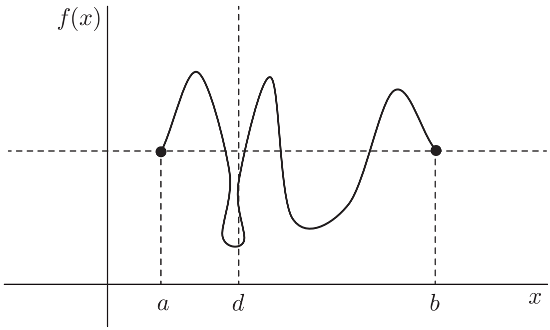


If you just draw what comes naturally in this case you'll end up with a diagram showing a function with the required properties—something like that shown below. Labelling can help to indicate exactly how various parts of the diagram relate to the theorem, so I've labelled an appropriate point c and a little line indicating that the derivative at c is zero. Notice that there are two possible values of c in this diagram, and that it would be straightforward to draw a function that has more.



Does the diagram convince you that the theorem is true? Can you see why, given the premises, there must always be a c where $f'(c) = 0$? If your immediate answer is 'yes', that's good, although you might still have something to learn about the technical meanings of continuity and differentiability. If you hesitated because you're not completely sure what we mean by these concepts, that's even better, and you will appreciate the discussion in the next section.

One pernickety point before we move on, though: don't get carried away with your loops when making sketches like this. The diagram below does *not* show a function because there is not a uniquely defined value for $f(d)$, for instance (the graph fails the vertical line test, if you've heard it put that way). I know that when students draw things like this it is usually just carelessness—they usually intend to draw something appropriate. But, again, precision matters in advanced mathematics, so do pay attention to this sort of thing.

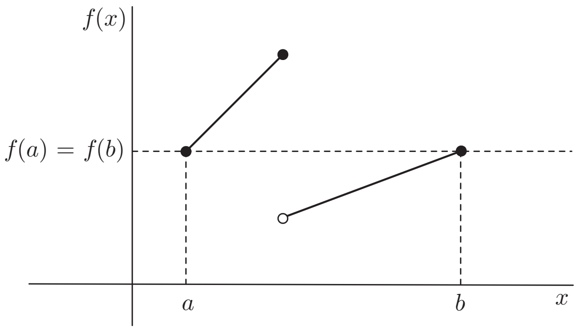


2.8 Examining theorem premises

Rolle's Theorem provides an opportunity to think about the concepts of Analysis in a more serious way, and to learn to think about theorems in depth by asking why all the premises are included. Here is the theorem again:

Rolle's Theorem:
Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), and that $f(a) = f(b)$. Then $\exists\, c \in (a, b)$ such that $f'(c) = 0$.

One premise is that the function is continuous on the interval $[a, b]$. Most people naturally think about continuous functions, because most functions they have worked with before are continuous (if not everywhere, then at least for most values of x). This book will urge you, however, to avoid thinking only about continuous functions, because the assumption of continuity is sometimes unwarranted—Chapters 7, 8 and 9 include functions that are discontinuous in a variety of interesting ways. Also, we can often gain insight into why a premise is included by thinking about what would go wrong if it were not. Considering Rolle's Theorem, it is quite easy to construct a function that has $f(a) = f(b)$ but that is not continuous on $[a, b]$, and for which the conclusion does not hold. In this diagram, for instance, there is no point c where $f'(c) = 0$:



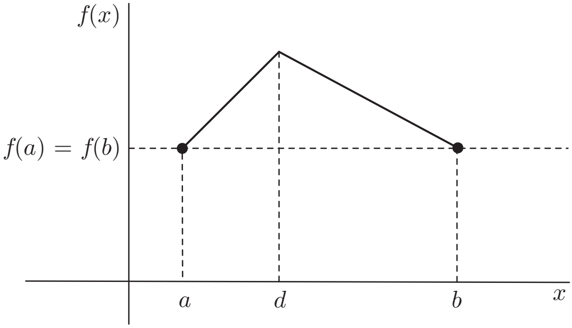
So we need the continuity premise—the theorem would not be true without it. For conceptual insight this diagram is probably enough, but functions like this can be expressed using formulas and it is good practice to give specific examples where possible. Specifying a, b and f like this, for instance, gives a graph like that above:

$$\text{Let } f : [1,4] \to \mathbb{R} \text{ be defined by } f(x) = \begin{cases} x+1 & \text{if } 1 \leq x \leq 2 \\ x/2 & \text{if } 2 < x \leq 4 \end{cases}.$$

This is a *piecewise-defined* function—it is specified differently on different parts of its domain. Notice that it is nevertheless a perfectly good function from $[1,4]$ to $\mathbb{R}$, because for every x in the interval $[1,4]$ there is a single specified value of $f(x)$ (sometimes students think this sort of thing is two functions, which is wrong). What do you think is meant by the filled-in dot and the non-filled-in dot in the diagram? Can you come up with other specific examples for which the continuity premise does not hold and the conclusion, again, fails?

Another premise is that the function is differentiable on the interval (a,b). Again, most people naturally think about differentiable functions, because most functions they have worked with before are differentiable. In fact, many new Analysis students are not even aware that they are thinking about differentiable functions, because they have done a lot of differentiation but have not thought in a theoretical way about what it means for a function to be differentiable. Differentiability will be discussed formally in Chapter 8, but for an approximate informal understanding you might think of it as meaning that the graph of the function has no 'sharp corners'. With that in mind, what might go wrong for Rolle's Theorem if the continuity premise holds but the differentiability premise does not? How might the conclusion fail?

Here is a diagram showing one simple case:



Here $f(a) = f(b)$ and the function is continuous. But it is not differentiable at $x = d$, and nowhere is there a point c with $f'(c) = 0$. I will pause here because, for some readers, it might not be obvious that these things are true. Some students, for instance, are unsure about whether a function like this is continuous at d. They see that 'you can draw it without taking your pen off the page', but they hesitate because they are accustomed to graphs of continuous functions being nice and curvy, not pointy. In fact, this function is continuous, and there will be more about such issues in Chapter 7.

Similarly, some students are unsure about the idea of a derivative at the point d. Again, they are accustomed to thinking about derivatives for functions with nice curvy graphs, and some wonder whether a function like this does have a derivative at the 'corner'. This gets to the essence of differentiability, which is about whether or not we could draw a single sensible tangent line at a point. In this case, we can't (what would be its gradient/slope?[5]), and this issue will be explored in detail in Chapter 8. In the meantime I ask you to take my word for it, and again to note that the differentiability premise is necessary; without it, the conclusion might not hold.

[5] People in the UK use the word 'gradient'; people in the US use the word 'slope' to mean the same thing.

One formula specifying a function like that in the diagram is

$$\text{Let } f : [1,4] \to \mathbb{R} \text{ be defined by } f(x) = \begin{cases} x+1 & \text{if } 1 \leq x \leq 2 \\ 4 - x/2 & \text{if } 2 < x \leq 4 \end{cases}.$$

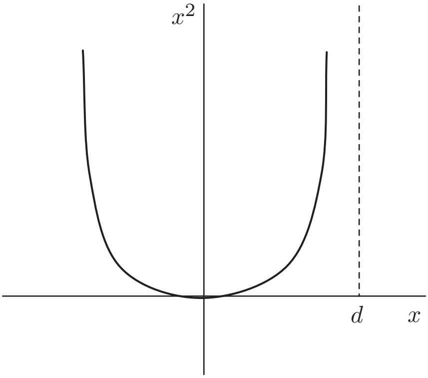
A simpler example with similar properties would be the function $f(x) = |x|$, on, say, the set $[-5, 5]$. In fact, $f(x) = |x|$ is everyone's favourite example of a function that (at the point $x = 0$) is continuous but not differentiable. You will probably see it introduced as an example of such, but you should bear in mind that when mathematicians give a single, simple example like that, they often intend you to see it as a representative of a general class. They show you $f(x) = |x|$ and perhaps a proof of some claim about it, but they intend that you will generalize the thinking to other functions with similar properties.

2.9 Diagrams and generality

Astute readers might have noticed that I've glossed over three subtleties about diagrams. The first is that, although I've been talking about some diagrams as generic, in a technical sense they are not. As soon as we commit a graph to paper, we are looking at a specific function. However, I think most readers will agree that some diagrams can be thought of as generic in the sense that they are not supposed to call to mind a formula; they don't tempt us to get distracted by knowledge of specific functions in the way that a graph of $f(x) = x^2$ or $g(x) = \sin x$ might.

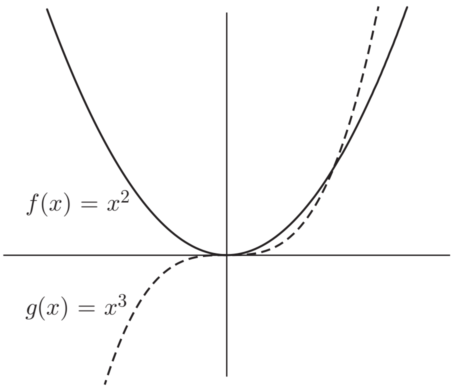
The second subtlety is that a diagram might not represent a 'whole' function. It is often easy enough to sketch a function on an interval, but a finite diagram cannot fully represent a function defined on the whole of the reals. This probably doesn't bother you, and in most cases it shouldn't: you'll be accustomed to imagining graphs that 'carry on forever' in a predictable way. But do be aware that because any particular diagram is finite (and specific), graphs in and of themselves don't prove very much. They might provide insights that are valuable for constructing proofs, but mathematicians look for definition-based arguments as well.

The third subtlety is that people are often a bit careless about some aspects of their drawings, so they can be misled by local properties of graphs that represent what is going on near $(0, 0)$. For instance, some people tend to sketch a graph of $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ so that it appears to have a U-shape, like this:

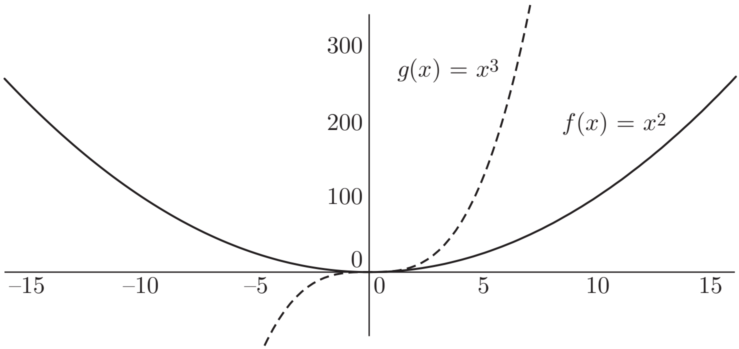


That's misleading because it appears to have vertical asymptotes. What would be the value of x^2 for $x = d$, for instance? It looks like there isn't one, which obviously isn't appropriate. Again, this is a bit pernickety, but you want to draw your graphs so that mathematicians know that you are aware of such issues.

Even when taking care to make the graph look parabolic rather than U-shaped, though, we can be misled by other things. For instance, sketching the graph of f along with that of $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = x^3$ gives something like this:

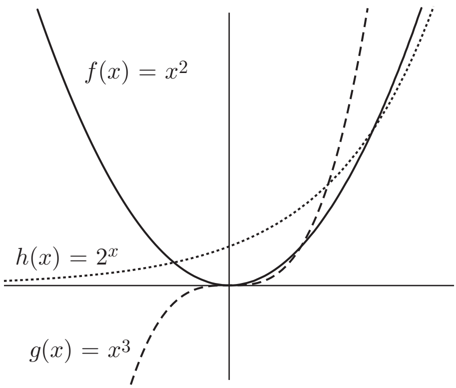


In this diagram, the shapes of the graphs for $x \geq 0$ look more or less alike; it looks like the functions change in a similar way. But, of course, they really don't. We don't have to zoom out much before they start to look very different, as can be seen here:

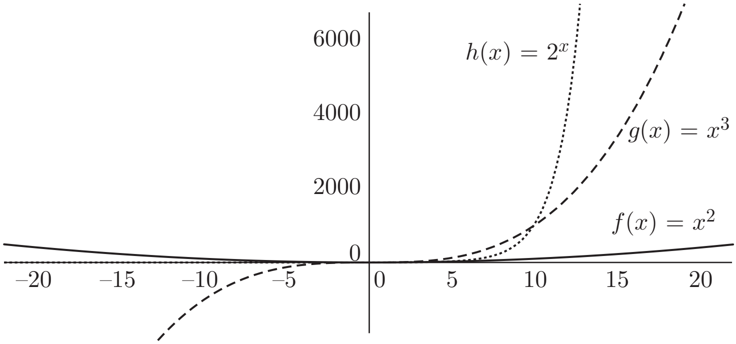


If you tend to be a bit lazy about your sketching, this should make you pause for thought. It should also make everyone think in a more nuanced way about how functions behave for 'big' values.

Similarly, how about comparing the graphs of f and g with that of the exponential function $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = 2^x$? The function h crosses the vertical axis in a different place but again, other than that, we might be inclined to make them look quite similar:



Do you know what happens for bigger values of x, though? The exponential function gets bigger *much* faster. Here is a zoomed-out view that gives a completely different sense of the relative behaviours of these functions:



The issues this raises are central in Analysis, where one thing we study is limiting properties: what happens to a function (or a sequence) as x (or n) tends to infinity. Graphs can be useful for developing intuition about this kind of thing, but you should also get out of the habit of treating computer or calculator output as the ultimate arbiter of truth. In Analysis we ask for more than that, developing understanding of the formal theory so that we can actually prove results about our observations.

2.10 Theorems and converses

This final section extends my earlier comments about the structures of theorems and points out some important differences between related *conditional statements*. A conditional statement is an 'if. . . then. . . ' statement, like this:

If f is a constant function then $f'(x) = 0\ \forall x \in \mathbb{R}$.

This is a true statement. Here is its *converse*:

If $f'(x) = 0\ \forall x \in \mathbb{R}$ then f is a constant function.

This is also a true statement. It is not the same statement, though. Here is a *biconditional statement* that captures both things together:

f is a constant function if and only if $f'(x) = 0\ \forall x \in \mathbb{R}$.

This is also a true statement, but it is a different statement again. You might like to think about the phrase *if and only if* and to consider how this biconditional statement captures both of the conditional statements.

There are two technical things to be aware of at this point. The first is that we could use an alternative notation, writing the three statements like this:

f is a constant function $\Rightarrow f'(x) = 0\ \forall x \in \mathbb{R}$.
$f'(x) = 0\ \forall x \in \mathbb{R} \Rightarrow f$ is a constant function.
f is a constant function $\Leftrightarrow f'(x) = 0\ \forall x \in \mathbb{R}$.

The symbol '$\Rightarrow$' is read aloud as 'implies (that)' and the symbol '$\Leftrightarrow$' is read as 'if and only if' or 'is equivalent to'. These are specific, standard meanings—don't use the arrows unless you intend exactly these meanings.

The second technical thing is that the first conditional statement should really be written like this:

For every function $f : \mathbb{R} \to \mathbb{R}$, if f is a constant function then $f'(x) = 0$ $\forall x \in \mathbb{R}$.

The new bit at the beginning just clarifies that we are talking about all functions of a certain kind. Probably you assumed that in any case, and most mathematicians would too, so extra phrases like this are often omitted. But mathematicians interpret conditional statements as though they were there.

Now, in everyday life we tend to be a bit sloppy in our use of conditional statements. We do not always distinguish a statement from its converse, and we often interpret a conditional statement as though it were a biconditional.[6] In fact, this is so common that there is extensive literature in psychology devoted to people's everyday interpretations of, and reasoning with, conditional statements.

In mathematics, we are not sloppy. When mathematicians write a conditional statement, they mean it exactly as written. This is very important, for two reasons. First, proving a statement is different from proving its converse.

[6] For a detailed explanation see Section 4.6 of *How to Study for/as a Mathematics Degree/Major.*

To prove that

if f is a constant function then $f'(x) = 0\ \forall x \in \mathbb{R}$,

we would assume that f is a constant function and deduce from this that $f'(x) = 0\ \forall x \in \mathbb{R}$. To prove that

if $f'(x) = 0\ \forall x \in \mathbb{R}$ then f is a constant function,

we would assume that $f'(x) = 0\ \forall x \in \mathbb{R}$ and deduce from this that f is a constant function. This does not necessarily amount to doing the same thing: a proof that works in one direction cannot necessarily be reversed.[7] In this case, one statement can be proved directly from the definition of the derivative, but the other requires more serious theoretical machinery—see Section 8.7 for details.

The second reason is more basic: sometimes a conditional statement is true but its converse is not. For instance, here is another conditional statement:

If f is differentiable at c then f is continuous at c.

This is true. Here is its converse:

If f is continuous at c then f is differentiable at c.

This is **not** true. We have already seen that the function $f(x) = |x|$ is continuous at 0 but not differentiable at 0, meaning that it constitutes a *counterexample* to the conditional statement, demonstrating that it is not universally true. This explains, incidentally, why people have favourite examples of functions and other mathematical objects. Some examples are particularly valuable for remembering key theorems and for avoiding mixing up theorems and their converses. This is handy because Analysis is awash with true theorems that have false converses. Here are a few to be going on with—what is the converse in each case? And do you know enough at present to see why the theorem is true but the converse is not?

[7] For a straightforward algebraic example, see Section 8.3 of *How to Study for/as a Mathematics Degree/Major*.

Theorem: If $(a_n) \to \infty$ then $(1/a_n) \to 0$.

Theorem: If $\sum_{n=1}^{\infty} a_n$ converges then $(a_n) \to 0$.

Theorem: If f is continuous on $[a, b]$ then f is bounded on $[a, b]$.

Theorem: If f and g are both differentiable at a then $f + g$ is differentiable at a.

Theorem: If f is bounded and increasing on $[a, b]$ then f is integrable on $[a, b]$.

Theorem: If $x, y \in \mathbb{Q}$ then $xy \in \mathbb{Q}$.

Some of these theorems appear later in the book. Some don't, but you will likely see them in an Analysis course. There might be lots more in your course as well. Whenever you see a conditional statement, I would advise you to think about its converse and think about whether either or both are true; doing so should help you to understand the structure of any accompanying proof. I also have lots more advice about understanding proofs, which you can find in the next chapter.

CHAPTER 3

Proofs

This chapter discusses the meaning of proof in mathematics and the place of proofs in mathematical theories. It discusses ways in which theories and proofs are structured, and ways in which they are taught. It also provides self-explanation training, which has been shown in research studies to improve students' proof comprehension.

3.1 Proofs and mathematical theories

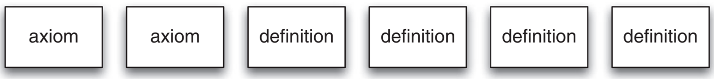
Undergraduate students often think that proofs are mysterious. They're really not. A specific proof might be difficult to understand because of its logical complexity, or because a student doesn't have a good enough grip on the definitions of the relevant concepts. But the idea is not difficult at all: a proof is just a convincing argument that something is true. The apparent mysteries, I think, arise because proofs in a subject like Analysis have to fit within a mathematical theory so, in addition to being convincing, they have to be structured according to the appropriate definitions and theorems. Part 2 of this book is about specific definitions and theorems associated with key concepts in Analysis, about how to identify where they are used in proofs, and about how to use them to structure proofs of your own. In this chapter, I will discuss general strategies for making sense of proofs presented in lectures or textbooks—these strategies can (and should) be applied whenever you encounter a proof in Analysis. First, though, I will briefly explain how proofs fit into mathematical theories.

One thing to get out of the way is that a *theory* is different from a *theorem*. A *theorem*, as discussed in Chapter 2, is a single statement about a relationship between some mathematical concepts. A mathematical *theory* is a network of interconnected axioms, definitions, theorems and proofs. This network might be huge. My current Analysis course contains 16 axioms, 32 definitions and 60 theorems with accompanying proofs. That's not nearly as scary as it sounds, because many of them are very simple. But this is only a small part of what might be considered the 'whole' theory of Analysis. As you might imagine, then, theories can be very complex. But they have some features that make them simpler to understand, and knowing what to look for should help you to appreciate what proofs are for and how they work.

3.2 The structure of a mathematical theory

Mathematical theories are developed over time, and this development is not linear. Mathematicians try to solve problems and to state and prove theorems, and to do so they formulate axioms and definitions to capture the concepts they wish to use. But mathematicians also value theory building—they want everything to fit into a coherent overall structure, which means that axioms, definitions, theorems and proofs are adjusted as groups of mathematicians come to agree on effective ways to capture both individual concepts and important logical relationships.

As a student, you might also solve problems. But unless you take an unusual Analysis course, you will not often be involved in formulating definitions and theorems. Rather, your job will be to learn the established theory of Analysis as it is understood by the contemporary mathematical community. This means that you can think of mathematical definitions and axioms[1] as 'basic' in the sense that they form the bottom layer of the theory's building blocks.



[1] See Section 2.2 for a brief introduction to axioms, and Chapter 10 for a more detailed discussion of axioms for the real numbers.

With the bottom layer in place, new blocks at higher levels take the form of theorems, where each theorem says something about a relationship between concepts from the preceding levels. In the initial stages of an Analysis course, theorems might be about just one concept. They will say, for instance, that a property that applies to objects x and y also applies to an object created by combining x and y; by adding them if they are numbers or functions or sequences, for instance, or by taking their union if they are sets. Here are some theorems like that:[2]

Theorem: If $x, y \in \mathbb{Q}$ then $xy \in \mathbb{Q}$.

Theorem: Suppose that $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are both differentiable at a. Then $f + g$ is differentiable at a with $(f + g)'(a) = f'(a) + g'(a)$.

Theorem: If $X, Y \subseteq \mathbb{R}$ are both bounded above, then $X \cup Y$ is bounded above.

Proving such a theorem would involve just one definition. The third theorem, for instance, says that if two subsets X and Y of $\mathbb{R}$ are both bounded above, then their union (the set of all elements in X or in Y or in both) is also bounded above. To prove it, we would do this:

> Suppose that X and Y are both bounded above.
>
> *Say what this means in terms of the definition of bounded above.*
>
> *Use algebraic manipulations and logical deductions to construct an argument showing that $X \cup Y$ also satisfies the definition of bounded above.*
>
> Conclude that $X \cup Y$ is bounded above.

Because I introduced the definition of *bounded above* in Section 2.6, I will show how the detail is filled in here. Guidance on reading and

[2] There is a notation list in the Symbols section at the start of the book, on page xiii.

understanding proofs is provided in Section 3.5; you might like to see what sense you can make of this proof now, then come back to it after that.

Theorem: If $X, Y \subseteq \mathbb{R}$ are both bounded above, then $X \cup Y$ is bounded above.

***Proof*:** Suppose that X and Y are both bounded above.

Then $\exists M_1 \in \mathbb{R}$ s.t. $\forall x \in X, x \leq M_1$

and $\exists M_2 \in \mathbb{R}$ s.t. $\forall y \in Y, y \leq M_2$.

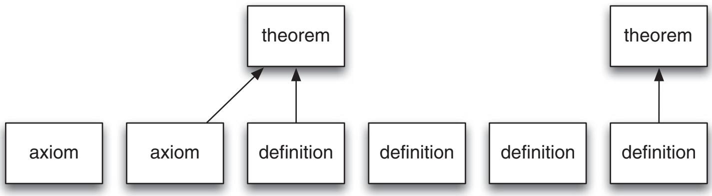
Now consider $M = \max\{M_1, M_2\}$.

Then $\forall x \in X, x \leq M_1 \leq M$ and $\forall y \in Y, y \leq M_2 \leq M$.

So every element of $X \cup Y$ is less than or equal to M.

So $X \cup Y$ is bounded above.

We could think of a theorem like this as adding a new block to the theory that sits above just one definition, or perhaps that uses one definition and an axiom (maybe an axiom about addition or inequalities).



Later theorems will involve multiple concepts. They will say, for instance, that an object with one property must also have another one, or that an object with a combination of properties must also have another one. Here are some theorems like that:

Theorem: Let (a_n) be a convergent sequence. Then (a_n) is bounded.

Theorem: Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), and that $f(a) = f(b)$. Then $\exists\, c \in (a, b)$ such that $f'(c) = 0$.

Theorem: If f is bounded and increasing on $[a, b]$ then f is integrable on $[a, b]$.

Proving such a theorem would use all the relevant definitions. To prove that every convergent sequence is bounded, for instance, we would do this:

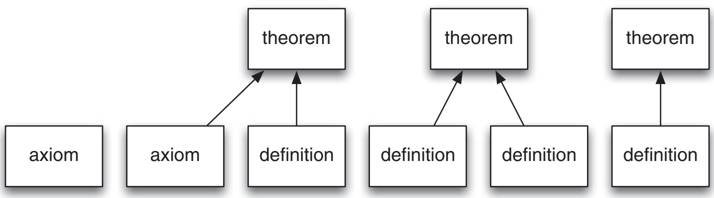
Assume that a sequence (a_n) is convergent.

Say what this means in terms of the definition of convergent.

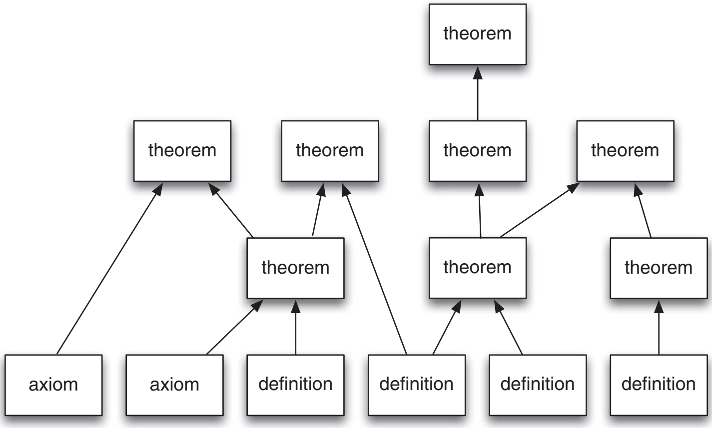
Use algebraic manipulations and logical deductions to construct an argument showing that (a_n) *also satisfies the definition of bounded.*

Conclude that (a_n) is bounded.

This time we don't yet have the machinery to fill in the details, but I will return to this theorem in Section 5.9. However, because of its structure, we could think of such a theorem as adding a new block to the theory like this, where again the arrows indicate what was used in the proof:



Does this mean that all proofs use definitions directly? No, because once we have proved a theorem, it stays proved. This means that we can use the established theorems to prove new ones, so our theory will build up and up like this:



3.3 How Analysis is taught

Although theories are structured as above, it would be a bit weird and disjointed to teach all the definitions and axioms first, then to teach a first 'layer' of theorems, and so on. If we did that, most students would have forgotten the first definition by the time it was used. So lecturers usually introduce a couple of key definitions first and then immediately state and prove some theorems that require only these. They then introduce another definition and build up some more theorems using this together with existing material, and so on. Thinking of theory development in this way should help you to understand the structure of Analysis.

There is, however, one more thing you need to know in order to make sense of an Analysis course. You need to know how this theory fits with earlier mathematics. Most new Analysis students already know a lot of calculus: they know about functions and about how to differentiate and integrate them, and they might also know some things about sequences and series. Many anticipate that Analysis will start from what they know and go upwards; that they will learn fancier and more complicated techniques for integration and differentiation and for working with sequences and series. In fact, that is not what happens at all. Analysis doesn't build upwards from calculus—it sits *underneath* it. In Analysis, we explore the theory that underlies calculus, picking apart our assumptions and understanding why it all works. Other courses do build upwards: mathematics students will probably learn about solving differential equations, and about integration and differentiation for functions of more than one variable or for functions of complex variables. But Analysis, on the whole, builds *down* from calculus, not up.

Analysis doesn't necessarily start with what you know and build down one layer at a time, however. That might make psychological sense, and it might be more reflective of the historical development of the subject, but it's hard to turn it into a logical presentation. Because theories are built on basic axioms and definitions—because proofs use those axioms and definitions—it makes more logical sense to start at the bottom and build back up towards the stuff you already know. So your lecturer might do that and, if so, you're likely to find the experience a bit weird. Analysis will seem a long way from what you've just been studying, and some of the early work will seem insultingly basic. But that's the point: a theory in advanced mathematics should start from basic things and build up a coherent theory.

That said, starting at the very bottom does tend to make students feel disoriented. So, when I teach Analysis, I tend to operate a sort of hybrid in which we start with definitions and study those in detail, but I don't mention axioms at first. Rather, I just get on and use axioms that I think students will take for granted (I'm right—I've never had anyone complain that we haven't specifically stated that $\forall a, b \in \mathbb{R}, a + b = b + a$). Then, once the class has had a crack at proving things on the basis of definitions, we might take another step downward and examine the more basic axiomatic assumptions. Students tend to be more ready for it by then, because they're able to appreciate the importance of systematic reasoning within a network of results. Your lecturer, of course, might do something different.

3.4 Studying proofs

The preceding sections should clarify the roles of definitions, theorems and proofs within a mathematical theory. On a page, however, definitions and theorems are small—typically one line or two—and proofs are bigger—perhaps five lines, or ten or fifteen. As a result, proofs tend to draw the eye and to seem more important, and students often talk about proofs as though they have an independent existence. But a proof is always a proof *of* something, and the something will be stated as a theorem (though it might be called a *proposition* or a *lemma* or a *claim*). Unsurprisingly, then, if you don't understand the theorem, you won't understand its accompanying proof; if you don't know what the author of the proof was trying to establish, how will you know when they've convinced you that they've done it?

So, don't think of a proof as an isolated entity, think of it as belonging to a theorem, and make sure that you understand what the theorem says first. This will often involve thinking at two levels, the first intuitive and the second formal. For instance, if a theorem says that every convergent sequence is bounded, you might find that you have an immediate intuitive sense of what this means. Nevertheless, it is advisable to pause and think properly about what it means in relation to the formal definitions of *convergent* and *bounded*: these are technical concepts, and the theorem is about those technical concepts, not about your overlapping but probably slightly woolly intuitive understanding. See Sections 2.7 and 2.8 for more on how you might do this.

Once you understand what the theorem says, you are in a position to study the proof. But how should you do this? How will you know when something is proved? For many undergraduate students, the obvious answer is that you know something is proved when your lecturer or textbook says it is. Obviously you have no reason to doubt a presented proof: it's entirely reasonable to believe something on the basis that someone in authority tells you that it's valid. It's not very intellectually satisfying, though—much better to understand something in detail than just to believe it. The good news from research in mathematics education is that students generally seem to have enough knowledge and logical reasoning skill to develop pretty good understanding of undergraduate proofs; the bad news is that many of them don't mobilize their knowledge very well. They can, however, do better once they have had some simple *self-explanation training*. Mathematics-specific self-explanation training appears in the next section.

3.5 Self-explanation in mathematics

At my university, we have used self-explanation training in several research studies, with positive results. The training is available at <http://setmath.lboro.ac.uk> and it is reproduced below as used in the studies—I have added two footnotes to link to ideas from elsewhere in the book, but other than that I haven't changed anything except the formatting. Because of this, both the style and the content in this section are a bit different—the style is less conversational and more instructional, and the content is more general—it involves concepts from number theory as well as Analysis.

SELF-EXPLANATION TRAINING

The self-explanation strategy has been found to enhance problem solving and comprehension in learners across a wide variety of academic subjects. It can help you to better understand mathematical proofs: in one recent research study students who had worked through these materials before reading a proof scored 30% higher than a control group on a subsequent proof comprehension test.

HOW TO SELF-EXPLAIN

To improve your understanding of a proof, there is a series of techniques you should apply.

After reading each line:

- Try to identify and elaborate the main ideas in the proof.
- Attempt to explain each line in terms of previous ideas. These may be ideas from the information in the proof, ideas from previous theorems/proofs, or ideas from your own prior knowledge of the topic area.
- Consider any questions that arise if new information contradicts your current understanding.

Before proceeding to the next line of the proof you should ask yourself the following:

- Do I understand the ideas used in that line?
- Do I understand why those ideas have been used?
- How do those ideas link to other ideas in the proof, other theorems, or prior knowledge that I may have?
- Does the self-explanation I have generated help to answer the questions that I am asking?

Below you will find an example showing possible self-explanations generated by students when trying to understand a proof (the labels '(L1)' etc. in the proof indicate line numbers). Please read the example carefully in order to understand how to use this strategy in your own learning.

EXAMPLE SELF-EXPLANATIONS

Theorem: No odd integer can be expressed as the sum of three even integers.

***Proof*:** (L1) Assume, to the contrary, that there is an odd integer x, such that $x = a + b + c$, where a, b, and c are even integers.

(L2) Then $a = 2k, b = 2l$, and $c = 2p$, for some integers k, l, and p.

(L3) Thus $x = a + b + c = 2k + 2l + 2p = 2(k + l + p)$.

(L4) It follows that x is even; a contradiction.

(L5) Thus no odd integer can be expressed as the sum of three even integers.

After reading this proof, one reader made the following self-explanations:

- 'This proof uses the technique of proof by contradiction.'[3]
- 'Since a, b and c are even integers, we have to use the definition of an even integer, which is used in L2.'

[3] Proof by contradiction is discussed along with other types of proof in Chapter 6 of *How to Study for/as a Mathematics Degree/Major*.

- ‘The proof then replaces a, b and c with their respective definitions in the formula for x.’
- ‘The formula for x is then simplified and is shown to satisfy the definition of an even integer also; a contradiction.’
- ‘Therefore, no odd integer can be expressed as the sum of three even integers.’

SELF-EXPLANATIONS COMPARED WITH OTHER COMMENTS

You must also be aware that the self-explanation strategy is not the same as *monitoring* or *paraphrasing*. These two methods will not help your learning to the same extent as self-explanation.

Paraphrasing

‘a, b and c have to be positive or negative, even whole numbers.’

There is no self-explanation in this statement. No additional information is added or linked. The reader merely uses different words to describe what is already represented in the text by the words ‘even integers’. You should avoid using such paraphrasing during your own proof comprehension.[4] Paraphrasing will not improve your understanding of the text as much as self-explanation will.

Monitoring

‘OK, I understand that $2(k + l + p)$ is an even integer.’

This statement simply shows the reader’s thought process. It is not the same as self-explanation, because the student does not relate the sentence to additional information in the text or to prior knowledge. Please concentrate on self-explanation rather than monitoring.

[4] I don’t intend this to contradict the advice in Chapter 1 about reading mathematics aloud. You might need to read what is literally on the page first, but you should then think beyond it to self-explanation.

A possible self-explanation of the same sentence would be:

> 'OK, $2(k+l+p)$ is an even integer because the sum of 3 integers is an integer and 2 times an integer is an even integer.'

In this example the reader identifies and elaborates the main ideas in the text. They use information that has already been presented to understand the logic of the proof.

This is the approach you should take after reading every line of a proof in order to improve your understanding of the material.

PRACTICE PROOF 1

Now read this short theorem and proof and self-explain each line, either in your head or by making notes on a piece of paper, using the advice from the preceding pages.

Theorem: There is no smallest positive real number.

***Proof*:** Assume, to the contrary, that there exists a smallest positive real number.

Therefore, by assumption, there exists a real number r such that for every positive number s, $0 < r < s$.

Consider $m = r/2$.

Clearly, $0 < m < r$.

This is a contradiction because m is a positive real number that is smaller than r.

Thus there is no smallest positive real number.

PRACTICE PROOF 2

Here's another more complicated proof for practice. This time, a definition is provided too. Remember: use the self-explanation training after *every* line you read, either in your head or by writing on paper.

Definition: An *abundant* number is a positive integer n whose divisors add up to more than $2n$.

For example, 12 is abundant because $1+2+3+4+6+12 > 24$.

Theorem: The product of two distinct primes is not abundant.

***Proof*:** Let $n = p_1 p_2$, where p_1 and p_2 are distinct primes.

Assume that $2 \leq p_1$ and $3 \leq p_2$.

The divisors of n are $1, p_1, p_2$ and $p_1 p_2$.

Note that $\dfrac{p_1+1}{p_1-1}$ is a decreasing function of p_1.

So $\max\left\{\dfrac{p_1+1}{p_1-1}\right\} = \dfrac{2+1}{2-1} = 3.$

Hence $\dfrac{p_1+1}{p_1-1} \leq p_2$.

So $p_1 + 1 \leq p_1 p_2 - p_2$.

So $p_1 + 1 + p_2 \leq p_1 p_2$.

So $1 + p_1 + p_2 + p_1 p_2 \leq 2 p_1 p_2$.

REMEMBER ...

Using the self-explanation strategy had been shown to substantially improve students' comprehension of mathematical proofs. Try to use it every time you read a proof in lectures, in your notes or in a book.

That's the end of the self-explanation training.[5] Some readers might like to apply it now to the proof in Section 3.2.

3.6 Proofs and proving

This chapter is about studying proofs as they appear in lecture notes or books. Lots of your activity as an undergraduate will involve understanding proofs from such sources. But this does not mean that mathematics is fixed and finished. On the contrary, it is a constantly evolving subject. It happens that Analysis as the mathematical community now understands it was developed (mostly) in the nineteenth century, so it is a fair while since anyone disagreed about its finer details, and modern textbooks all capture the central ideas in essentially the same ways. You will therefore learn about Analysis as a network of results established using standard proofs. But that does not mean that proofs are unique: there might be numerous possible proofs for a single theorem, each using different but valid reasoning. And it does not mean that there is no room for creativity in mathematics. Today, ideas at the current boundaries of mathematics are constructed, compared and debated in thousands of universities around the world. Certainly any student learning in an established area still has plenty of opportunities to solve problems and to develop knowledge independently.

CHAPTER 4

Learning Analysis

This chapter explains what it feels like to study Analysis. It offers advice on how to keep up, how to avoid wasting time, and how to make good use of study resources.

4.1 The Analysis experience

Here is what happens when I teach Analysis. In week 1, everyone is in a good mood because they're starting something new. In weeks 2 and 3, there is a buildup of increasingly challenging material. In week 4, the mood in the lecture theatre is dreadful. The whole class has realized that this is difficult stuff and that it isn't going to get any easier. Everyone hates Analysis and, by extension, quite a few people hate me. I am not fazed by this, though, because I have taught Analysis about twenty times now and I know what will happen next. In week 5, everyone will feel slightly better, even if no one can quite explain why. In week 7, a small number of people will approach me and tell me shyly that, although Analysis is challenging, they're starting to think they might like it. By the end of the course, these people will be telling anyone who will listen that Analysis is brilliant, and lots of other students will admit that now that they're getting the hang of it, they can see why people think it's a great subject.

The question for a new student, then, is how to handle it when the work gets difficult and you start to feel negative. Some students turn the negativity inwards: they lose confidence, experience self-doubt about their mathematical ability ('Perhaps I'm not good enough for this?'), and sometimes become withdrawn. Others turn it outwards, expressing frustration and anger about their lecturers ('He's a terrible teacher!') and sometimes, a bit nonsensically, about the mathematics itself ('I don't know why they're teaching us this rubbish—this isn't maths!'). These reactions both arise naturally when people feel a loss of control and consequently get defensive. But neither is very productive. So what is the alternative?

Well, most people do experience a bit of difficulty when first learning Analysis. This is just a fact of life. So, in my view, the trick is simply to expect this as a normal part of the learning experience, and ride it out. If you are ready for a bit of a challenge, you'll be better placed to handle the emotions without hiding away or acting out—you can say to yourself 'Well, okay, I was expecting this,' and continue to study in a sensible way, knowing that things will gradually come together. This chapter is about practical approaches to doing that.

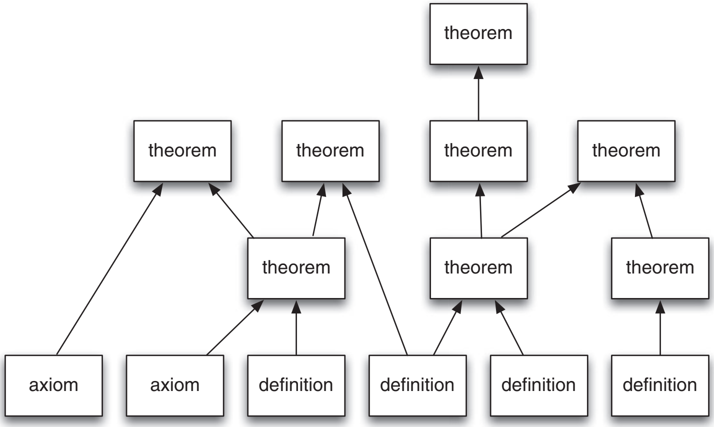
4.2 Keeping up

In Analysis, as in any undergraduate mathematics course, the big challenge is keeping up. If you're taking a decent course, this will be difficult. No one is trying to teach you stuff that you will find easy—what would be the point of that? Also, you will be busy, with other courses and with the rest of your life. So it is very unlikely that you will be on top of everything all the time. You should try not to be distressed by this, because distress doesn't help—negative emotions just impede effective study. The thing to do is to accept that you will not always have perfect knowledge of everything, and work in an intelligent way that allows you to maintain *sufficient* knowledge of the *important* things.

When I say *sufficient* knowledge, I mean enough knowledge to give you a fighting chance of making sense of new material. By the time you are a few weeks into a course, you are unlikely to understand everything in every lecture—I certainly didn't. But you want to have enough under your belt that you can follow the big sweep of the theory development and understand some of the details. When I say the *important* things, I mean the central concepts that come up again and again. At any given time, it is unlikely that you will be able to explain the nuances of every proof, but you want to know the main definitions and theorems so that you can recognize when and how they are used in new work. With that in mind, here is what I would prioritize.

First, you absolutely must know your definitions. In Analysis, it is sometimes tempting to be lax about this, because many of the words used ('increasing', 'convergent', 'limit' etc.) have everyday meanings, and because concepts in Analysis can often be represented using diagrams. Both of these things will tempt you into thinking that intuitive understanding is sufficient. *It isn't*. Definitions are central to any theory in advanced mathematics: as explained in Chapters 2 and 3, they are key to understanding what is really meant by the theorems and what is going on in many of the proofs. If you think you understand the subject without knowing your definitions properly, you are kidding yourself. Because of this, I would start a definitions list on the first day of the course. Keep this on a piece of paper at the front of your folder (even if you keep most of your notes on an electronic device, I would still use paper for this). Every time you encounter a new definition, add it to the list. Study the list regularly, perhaps test yourself on it periodically, and be alert in lectures for defined words—every time lecturers use one, they mean it in exactly the sense captured by the definition.

Second, it is a good idea to be conversant with the main theorems. These capture relationships between concepts, so knowing what they say—even if you don't fully understand the proofs—will give you an overview of the course. Sections 2.7 and 2.8 give advice on thinking thoroughly about theorem meanings—a few minutes spent following this advice is likely to fix a new theorem in your mind. Also, notes are sometimes provided in advance these days, either for a whole course or for some block of it (if your course follows a textbook, you will have the whole thing in advance). So you could get a sense of what theorems are coming by reading ahead. Once the course gets going, I would consider keeping a theorems list too. Indeed, I would go beyond list-making and construct a *concept map* (sometimes called a *mind map* or a *spider diagram*). Because of the way theory is built up, it often makes sense to use a diagram to indicate which theorems (and definitions) are used to prove which other theorems, as discussed in Section 3.2. You could make a concept map that looks something like this, with the words in the boxes replaced by names or abbreviations for the specific definitions and theorems in your course:



Those are the things I would prioritize. If you find yourself getting behind or otherwise short of time, do those before you do anything else. Don't try to go back to where you last understood everything and work forwards from there—doing so will be ineffective, because the course will move on faster than you do, and you will end up in lectures where you do not understand anything. Analysis is hierarchical and thus unforgiving of students who do not keep up with the main building blocks. Prioritizing as suggested here will usually give you sufficient knowledge of the important things; it will allow you to identify key concepts and relationships in new lectures, and will provide you with a framework for more detailed study, as discussed below.

4.3 Avoiding time-wasting

Because keeping up is challenging, you do not want to waste any time. So it is worth thinking about how much study time you've got and how you're going to use it.[1] In the UK, where I work, lecture courses in subjects like Analysis typically involve about three one-hour lectures per

[1] Chapter 11 in *How to Study for/as a Mathematics Degree/Major* discusses large-scale time management for mathematics students; this chapter discusses specific study strategies for an Analysis course.

week. In such a system, I think it reasonable to spend a further three to four hours per week on independent study. If you do that for all your subjects, you'll probably end up with a standard 40-hour working week, which is about right (if you are studying in a different system, you can read the advice below and work out how to adjust it for your situation).

Now, three to four hours isn't very much. You can tell yourself you're going to do more than that if you like, but most people don't, so it's probably more important to make the three to four hours count. During that time, you will have two things to do: study your lecture notes (or your textbook), and work on problems. I put the tasks in that order for a reason. In order to work effectively on the problems, you will need to be familiar with the material in your notes. If you are, you will find that many problems make you think 'Ah, we did something related to this on Wednesday.' If you aren't, you will waste a lot of time having no idea how to start and staring into space. So, notes first.

I suggest spending perhaps sixty to ninety minutes studying your recent notes. This does not mean reading them without really thinking, though. Read everything carefully, following the study suggestions for definitions, theorems and proofs from Chapters 2 and 3, and updating your definitions and theorems lists as you go along. Aim for good self-explanations (see Section 3.5), but don't obsess over anything. Sixty to ninety minutes isn't that long, and you want to be at least somewhat familiar with everything. So, if you have spent a few minutes thinking properly about something but you still don't really get it, get out a piece of paper, write 'Questions about Analysis' at the top, and make a note of where this thing is and what exactly you don't understand. Be precise—sometimes nailing down the problem allows you to sort it out and, if it doesn't, you will have a specific note to come back to so you don't lose the thinking you have already done.

Once you've studied your notes, begin work on the problems. Depending on how you divide up your time, you'll have between two and three hours for this. That will not give you very much time for any given problem, so again you don't want to waste any. Because of this, I suggest a first pass in which you spend perhaps ten minutes on each problem. Some problems you will be able to finish in this time, especially if they involve routine warm-up exercises or direct applications of an idea that you've just studied. (In such cases, see how much you can do without looking at your notes—this might take slightly longer but, if you can construct or reconstruct something for yourself, you will remember it better in the long run.) Other problems you will not be able to finish in ten minutes. If you are making good progress, you might want to carry on for a bit longer. If, on the other hand, you're stuck, and if you've tried a few sensible things to get unstuck, make a note on your 'Questions about Analysis' sheet and move on—those other problems are still waiting.

Now, I said 'at a first pass' because I think problem solving in Analysis should be a multiple-pass task. You want to have a go, then have a break for a day or two, then have another go. Magical things will sometimes happen in the break—your brain will make new connections and you'll see new ways forward. So you probably want to break up your study into at least a couple of blocks. Indeed, you should do that anyway, because thoughtful study is intellectually effortful—if you decide to spend four hours at a stretch studying Analysis, I guarantee that you will waste the last two simply because you will run out of energy.

4.4 Getting your questions answered

Next, what to do with your 'Questions about Analysis' list? For a start, keep an eye on it. Sometimes, working on problems will make you think about an idea in a different way, and you'll be able to cross off something that you added when studying your notes. Sometimes, when you've had a break for a couple of days, a quick re-read of your notes will make something click, and you'll be able to finish a problem and cross that off too. After that, here's what I'd do.

First, get together with a friend or two and work systematically through your respective lists. Everyone thinks a bit differently, so you will probably be able to fill some gaps for one another. Doing this will also force you to speak about Analysis, helping you to become fluent in talking about the concepts and explaining your arguments. Fluency is important, so don't worry if you trip over your words at first. Just have another go—you will only get more confident with practice. Sharing ideas will also help you to become a good mathematical listener. Pay close attention to what your friends are saying and, if you are not sure you understand, say so, and try to specify what is confusing you. Doing this will help your friends to articulate their thoughts more clearly. Again, this is a valuable skill that will help all of you to speak more confidently to lecturers and other tutors. Of course, as in individual work, don't get obsessed—if you can't sort something out between you in a reasonable amount of time, perhaps your effort would be better spent elsewhere.

Once you've shared your knowledge with friends, take your remaining questions to an expert (you can always go to the expert first, of course, but consider the above issues about developing communication skills). Which expert you want will depend on your institution's teaching systems: perhaps your tutor, perhaps your course lecturer, perhaps a mathematics support service. Whoever you see, take your list and your problem sheets and all your relevant notes, and make sure that your list has page or section or question numbers on it—you want to be able to find everything with minimal fuss. If seeing someone involves arranging a specific meeting, consider asking whether you and your friends can go together—that should make the process more efficient. And do not be shy about asking questions, even if you have a long list. Trust me, a student asking specific questions from a well-organized list is always impressive.

Taking this approach should mean that most of your questions are answered most of the time. However, do be realistic. Following this advice will still leave you with gaps. Sometimes there will not be time to sort everything out. Sometimes there will be time to sort everything out, but two weeks later you will realize that you've now forgotten why something works and you need to think it through again. It should be possible to minimize that problem by making decent notes—when you've overcome confusion about something, recording how you changed your thinking will facilitate quick review. Overall, if you get yourself organized at approximately the level suggested in this chapter, you will keep up with the main ideas, you will understand at least some of each new lecture, and you will develop a solid block of knowledge that you can build on when you start preparing for exams.

4.5 Adjusting your strategy

In this chapter I have suggested a specific way to organize your studies. I should say that I don't really expect anyone to behave in precisely this way. You will be subject to constraints about when you can study, to personal preferences about your work habits, and to the shifting requirements of other aspects of your academic work and your social life. So you should reflect occasionally on how things are going, and be ready to adjust. If you need longer to study your notes, adjust your timings; if you need time to study for a test in another subject, cut back to the essentials in Analysis for a week; if one of your friends is great for social outings but a bit rubbish at concentrating on Analysis, quietly make alternative or extra arrangements for discussions with others. And of course, if you're really into a problem, stare into space and think about it for hours, if you like. The advice here should be thought of as a useful place to start, and as a way to develop a routine that will keep you going through the challenging weeks.