New Revised Edition

Advanced Problems in Mathematics

Preparing for University

Stephen Siklos

ADVANCED PROBLEMS IN MATHEMATICS

Advanced Problems in Mathematics: Preparing for University

Stephen Siklos

OpenBook Publishers
Knowledge is for sharing

http://www.openbookpublishers.com

Any digital material and resources associated with this volume can be found at https://www.openbookpublishers.com/product/1050#resources

STEP questions reproduced by kind permission of Cambridge Assessment Group Archives. Please send any comments or corrections to step@maths.org.

This is the second edition of the first volume in the OBP Series in Mathematics:

ISSN 2397-1126 (Print)
ISSN 2397-1134 (Online)

ISBN Paperback: 9781783747764
ISBN Digital (PDF): 9781783747771
DOI: 10.11647/OBP.0181

Cover image: Photograph © Logan Troxell. Creative Commons Attribution International 4.0, CC BY. https://unsplash.com/photos/DsQmBlbywJ8.

Cover design: Anna Gatti.

Contents

About this book

This book has two aims.

- The general aim is to help bridge the gap between school and university mathematics.

  You might wonder why such a gap exists. The reason is that mathematics is taught at school for various purposes: to improve numeracy; to hone problem-solving skills; as a service for students going on to study subjects that require some mathematical skills (economics, biology, engineering, chemistry — the list is long); and, finally, to provide a foundation for the small number of students who will continue to a specialist mathematics degree. It is a very rare school that can achieve all this, and almost inevitably the course is least successful for its smallest constituency, future mathematics undergraduates.

- The more specific aim is to help you to prepare for STEP or other examinations required for university entrance in mathematics. To find out more about STEP, read the next section.

It used to be said that mathematics and cricket were not spectator sports; this is still true of mathematics. To progress as a mathematician, you have to strengthen your mathematical muscles. It is not enough just to read books or attend lectures. You have to work on problems yourself.

One way of achieving the first of the aims set out above is to work on the second, and that is how this book is structured. It consists almost entirely of problems for you to work on.

The problems are all based on STEP questions. I chose the questions either because they are 'nice'— in the sense that you should get a lot of pleasure from tackling them (I did), or because I felt I had something interesting to say about them.

The first two problems (the 'worked problems') are in a stream of consciousness format. They are intended to give you an idea of how a trained mathematician would think when tackling them. This approach is much too long-winded to sustain for the remainder of the book, but it should help you to see what sort of questions you ought to be asking yourself as you work on the later problems.

Each subsequent problem occupies two pages. On the first page is the STEP question, followed by a comment. The comments may contain hints, they may direct your attention to key points, and they may include more general discussions. On the next page is a solution; you have to turn over, so that your eye cannot accidentally fall on a key line of working. The solutions give enough working for you to be able to read them through and pick up at least the gist of the method; they may not give all the details of the calculations. For each problem, the given solution is of course just one way of producing the required result: there may be many other equally good or better ways. Finally, if there is space on the page after the solution (which is sometimes not the case, especially if diagrams have to be fitted in), there is a postmortem. The postmortems may indicate what aspects of the solution you should be reviewing and they may tell you about the ideas behind the problems.

I hope that you will use the comments and solutions as springboards rather than feather beds. You will only really benefit from this book if you have a good go at each problem before looking at the comment and certainly before looking at the solution. The problems are chosen so that there is something for you to learn from each one, and this will be lost to you for ever if you simply read the solution without thinking about the problem on your own.

I have given each problem a difficulty rating ranging from ✓ to ✓✓✓. Difficulty in mathematics is in the eye of the beholder: you might find a question difficult simply because you overlooked some key step, which on another day you would not have hesitated over. You should not therefore be discouraged if you are stuck on a ✓-question, though you should probably be encouraged if you get through one of the rare ✓✓✓-questions without mishap.

This book is about depth not breadth. I have not tried to teach you any new topics. Instead, I want to lead you towards a deeper understanding of the material you already know. I therefore restricted myself to problems requiring knowledge of the specific and rather limited syllabus that is laid out at the end of this book. This syllabus is for STEP papers that were set before 2019, for 2019 onwards there are new STEP specifications which can be found at www.admissionstesting.org/for-test-takers/step/about-step.

Calculators are not required for any of the problems in this book and calculators are not permitted in STEP examinations. In the early days of STEP, calculators were permitted but they were not required for any question. It was found that candidates who tried to use calculators sometimes ended up missing the point of the question or getting a silly answer. My advice is to remove the battery so that you are not tempted.

I started this section by listing the aims of the book. You may have noticed that teaching you mathematics is not one of them. I can't remember where I heard the following rather nice analogy. In 1464, a huge block of Carrara marble was carefully chosen from a quarry in Tuscany and transported to Florence, where it lay almost untouched for many years. In 1501 it was given to the sculptor Michelangelo. He worked hard on it, chipping away and chipping away for three years, until at last, inside the block, he found a beautiful statue of David. You can see a picture at:
http://www.accademia.org/explore-museum/artworks/michelangelos-david/

And the analogy? I can't teach you mathematics with this book, but I believe that much hard work on your part, chipping away at the problems, will eventually reveal the mathematician that is within you.

I hope you enjoy using this book as much as I have enjoyed putting it together.

STEP

What is STEP?

STEP (Sixth Term Examination Paper) is an examination used by Cambridge University as part of its procedure for admitting students to study mathematics. Applicants are interviewed in December, and may then be offered a place conditional on the results of their public examinations (A-level, International Baccalaureate, etc) and STEP. The examinations are sat in June and offers are confirmed in August when all the examination results are available.

STEP is used for conditional offers not just by Cambridge, but also by Warwick University as part of a range of offers, and to a lesser extent by some other English universities. Many other university mathematics departments recommend that their applicants practise the past papers even if they do not take the examination. In 2019, more than 4500 scripts were marked, compared to about 500 students holding STEP offers from Cambridge.

The first STEPs were taken in 1987, and there were specimen papers before that from which some of the questions in this book are drawn. At that time, there were STEPs in many subjects but by 2001 only the mathematics papers remained. The examination has remained more or less stable over more than 30 years: it has not been blown about by the various fads in the public examination systems that came and went during that time.

I do not want you to think that this book is about examinations; and it is definitely not about how to pass examinations. But since all the questions are taken from STEP papers, I will say a little more about STEP.

There are three STEP papers, 1, 2 and 3, in increasing order of difficulty. More information about STEP, including past papers and up-to-date specifications can be found at www.admissionstesting.org/for-test-takers/step/about-step. The specifications for STEP contain Pure Mathematics, Mechanics and Probability/Statistics.

It has to be said, though, that the statistics questions are very likely to require knowledge of probability rather than statistics (for example, there are very few questions on statistical tests of given data). This is because the underlying theory of statistics is quite difficult, and therefore unsuitable for examination at this level, whereas the application of statistical tests is rather routine and therefore unsuitable for different reasons.

What is the purpose of STEP?

From the point of view of admissions to a university mathematics course, STEP has three purposes.

- It is used as a hurdle for entrance to university mathematics courses, and sometimes for other mathematics-based courses. There is strong evidence that success in STEP correlates very well with university examination results.[1]

[1] Recent studies comparing rank in STEP with rank in first-year Cambridge mathematics examinations reveal a Spearman correlation coefficient of 0.63, which is very high in comparison with other predictors of university examination results.

- It acts as preparation for the university course, because the style of mathematics found in STEP questions is similar to that of undergraduate mathematics.
- It tests motivation. It is important to prepare for STEP (by working through old papers, for example), which can require considerable dedication. Those who are willing to make the effort are more likely to thrive on a difficult university mathematics course.

STEP *vs* A-level

A-level[2] tests mathematical knowledge and technique by asking you to tackle fairly stereotyped problems. STEP asks you to apply the same knowledge and technique to problems that are, ideally, unfamiliar.

Here is an A-level question, in which you follow the instructions in the question:

> By using the substitution $u = 2x - 1$, or otherwise, find
> $$\int \frac{2x}{(2x-1)^2}\,\mathrm{d}x.$$

And here, for comparison, is a STEP question, which requires both competence in basic mathematical techniques and mathematical intuition. Note that help is given for the first integral, so that everyone starts at the same level. Then, for the second integral, candidates have to show that they understand why the substitution used in the first part worked, and how it can be adapted.

> Use the substitution $x = 2 - \cos\theta$ to evaluate the integral
> $$\int_{3/2}^{2} \left(\frac{x-1}{3-x}\right)^{\frac{1}{2}} \mathrm{d}x.$$
> Show that, for $a < b$,
> $$\int_{p}^{q} \left(\frac{x-a}{b-x}\right)^{\frac{1}{2}} \mathrm{d}x = \frac{(b-a)(\pi + 3\sqrt{3} - 6)}{12},$$
> where $p = (3a+b)/4$ and $q = (a+b)/2$.

The differences between STEP and A-level are:

1. STEP questions are much longer.
2. STEP questions are much less routine.
3. STEP questions may require considerable dexterity in performing mathematical manipulations.
4. Individual STEP questions may require knowledge of several different areas of mathematics (especially the mechanics and statistics questions, which will often require advanced pure mathematical techniques).
5. The marks available for each part of the question are not disclosed on the paper.
6. There is a lot of choice on STEP papers (Candidates are expected to attempt 6 questions out of 11 or 12).

---

[2] I use the term ‘A-level’ here as a shorthand for a typical school mathematics examination. The particular examinations you take may well be very different in style and format but, even if that is the case, I am sure some of what follows will strike a chord with you.

7. Calculators are not permitted in STEP examinations.

These differences matter, because in mathematics more than in any other subject it is very important to match the difficulty of the question with the ability of the candidates. For example, you could reasonably set the question 'Was Henry VIII a good king?' on a lower-school history paper, an A-level paper, or as a PhD topic. The answers would (or should) differ according to the level. On mathematics examination papers, the question has to be tailored to the expected level in order to discriminate between the candidates: if it is too easy, nearly all candidates will score very high marks; if it is too hard, nearly all candidates will make little progress on any of the questions.

STEP *vs* Olympiad

STEP questions are not like A-level questions, as we have seen, but they are not like Olympiad questions either. Typically, an Olympiad question comes with no 'scaffolding' (as it is nowadays called) at all. There are no little hints or toe-holds to allow you to find a way into the problem. A typical STEP question (see the next section) does come with a bit of scaffolding; it is more of a learning process. You could say that it is more about the journey than about the arrival.

STEP Questions

STEP questions do not fall into any one category. Typically, there will be a range of types on each of the papers. Here are some thoughts, in no particular order.

- My favourite sort of question is in two (or maybe more) parts: in the first part, you are asked to perform some unfamiliar task and are told how to do it (integration using a given substitution, or expressing a quartic as the algebraic sum of two squares, for example); for the later parts, you are expected to demonstrate that you have understood and learned from the first part by applying the method to a new and perhaps more complicated task.
- Another favourite of mine is the question which has different answers according to the value of a certain number (or *parameter*). A common example involves sketching a graph whose shape depends on whether a parameter is positive or negative. Ideally, the different values of the parameter are not given in the question, and you have to identify them for yourself.
- Another good type of question requires you to do some preliminary special-case work and then prove a general result.
- In another type, you have to show that you can understand and use new notation or a new theorem.
- Questions with several unrelated parts (for example, three integrals using different techniques) are generally avoided; but if they occur, there tends to be a 'sting in the tail' involving putting all the parts together in some way.
- Some questions do not rely on any part of the syllabus: instead they might require 'common sense', involving counting or seeing patterns, or they might involve some aspect of more elementary mathematics with an unusual slant. Such questions try to test capacity for clear and logical thought without using much mathematical knowledge (like the calendar question mentioned in the section on preparation below, or questions concerning islands populated by toads 'who always tell the truth' and frogs 'who always fib').

- Some questions are devised to check that you do not simply apply routine methods blindly. For example, a function might have a maximum value at the end of the interval upon which it is defined, even though its derivative might be non-zero there. Finding the maximum in such a case is not simply a question of routine differentiation.

- There are always questions specifically on integration or differentiation, and many others (including mechanics and probability) that use calculus as a means to an end.

- Graph-sketching is regarded by mathematicians as a fundamental skill and there are nearly always questions that require a sketch.

- Basic ideas from analysis, such as

$$0 \leqslant \mathrm{f}(x) \leqslant k \Rightarrow 0 \leqslant \int_a^b \mathrm{f}(x)\mathrm{d}x \leqslant (b-a)k$$

  or

$$\mathrm{f}'(x) > 0 \Rightarrow \mathrm{f}(b) > \mathrm{f}(a) \text{ for } b > a$$

  or the relationship between an integral and a sum often come up, though knowledge of such results is never assumed — candidates may be told that they can use the result without proof, or a sketch 'proof' may be requested.

- As mentioned above, questions on statistical tests are rare, because questions that require real understanding (rather than 'cookbook' methods) tend to be too difficult. More often, the questions in the Probability and Statistics section are about probability.

- The mechanics questions normally require a firm understanding of the basic principles (when to apply conservation of momentum and energy, for example) and may well involve a differential equation. Projectile questions are often set, but are never routine.

Advice to candidates

First appearances

I am often asked whether STEP is 'difficult'. Of course, it depends on what is meant by 'difficult'; it is not difficult compared with the mathematics I do every day. But to be on the safe side, I always answer 'yes' before explaining further.

Your first impression on looking at a STEP paper is likely to be that it does indeed look very difficult. Don't be discouraged! Its difficult appearance is largely due to it being very different in style from what you are used to.

At the time of writing, a typical A-level examination lasts between 90 mins and two hours and contains between 7 and 13 questions. That is about 10 minutes per question. If you are considering studying mathematics at a top university, it is likely that you will manage to do them all and get them nearly all right in the time available. A STEP examination lasts 3 hours, and you are only supposed to do six questions. This means that each question is designed to take about 30 minutes. If you compare a 10-minute question with a 30-minute question, the 30-minute question will look substantially harder.

When looking through past papers you may be put off by the number of different topics covered on the paper. Before the 2017-18 A-level reforms there was a considerable amount of variation in the different specifications of each examination board, and STEP was designed to provide sufficient questions for all of those specifications. From the summer of 2019 onwards there is a lot less variation (for the pure topics at least). Also, you get to chose which questions to do so you do not have to be overly concerned about the inclusion of some topics you are less fond of.

Once you get used to the idea that STEP is very different from A-level, it becomes much less daunting.

Preparation

The best preparation for STEP (apart, of course, from working through this excellent book) is to work slowly through past papers.[3] Hints and answers are available for the more recent years, but you should use these with discretion: doing a question with hints and answers in front of you is nothing like doing it yourself, and you may well miss the whole point of the question (which is to make you think about mathematics). In general, thinking about the problem is much more important than getting the answer.

It is worth emphasising that there is no 'hidden agenda': a candidate who does two complete probability questions and two complete mechanics questions will obtain the same mark and grade as one who does four complete pure questions. Similarly a candidate who scores 15 on each of 4 questions will get the same grade as one who scores 10 on each of 6 questions.

Just as the examiners have no hidden agenda concerning question choice, so they have no hidden agenda concerning your method of answering the question. If you can get to the end of a question correctly you will get full marks whatever method you use.[4] Some years ago one of the questions asked candidates to find the day of the week of a given date (say, the 5th of June 1905). A candidate who simply counted backwards day by day from the date of the exam would have received full marks for that question (but would not have had time to do any other questions).

You may be worried that the examiners expect some mysterious thing called rigour. Do not worry: STEP is an exam for schools, not universities, and the examiners understand the difference. Nevertheless, it is extremely important that you present ideas clearly, and show working at all stages.

Presentation

You should set out your answer legibly and logically (don't scribble down the first thought that comes into your head)—this not only helps you to avoid silly mistakes but also signals to the examiner that you know what you are doing (which can be effective even if you haven't the foggiest idea what you are doing). It helps if you can read your own writing; be sure that you can distinguish between '2' and 'z' and '5' and 's'.

Examiners are not as concerned with neatness as you might fear. However, if you receive complaints from your teachers that your answers are difficult to follow then you should listen.[5] Remember that more space usually means greater legibility. Try writing on alternate lines (this leaves a blank line for corrections).

Try to read your answers with a hostile eye. Have you made it clear when you have come to the end of

---

[3] These and the other publications mentioned below are obtainable from the Cambridge Assessment Admissions Testing website www.admissionstesting.org/for-test-takers/step/about-step. You may also find the STEP Support Programme (maths.org/step) useful. This has been developed by the University of Cambridge to help students prepare for STEP.

[4] Though you must obey instructions in the question: for example, if it says 'Hence prove ...', then you must use the previous result in your proof. One example is 2018 STEP 3 question 5 which asked candidates to deduce the Arithmetic Mean-Geometric Mean inequality. Those candidates who did not 'deduce' this from previous work did not do as well.

[5] Begin rant: I am very surprised at the scrappy and illegible work that I receive from a few of my students. It seems so disrespectful to expect me to spend ages trying to decipher their work when they could have spent a little more time making it presentable, for example by copying it out neatly or writing more slowly. Why is my time less important than theirs? End rant.

a particular argument? Try underlining your conclusions. Have you explained what you are trying to do? For example, if a question asks 'Is A true?' try beginning your answer by writing 'A is true'—if you think that it is true—so that the examiner knows which way your argument leads. If you used an idea (for example, integration by substitution), did you tell the examiner that this is what you were doing?

You may find that you stop a question and then restart it later—in which case a note to the examiner 'continued later in booklet' will make your solution easier to follow. Also, if you decide to cross out some working do it neatly (for example by putting one diagonal line through it) as then you have the option of writing 'please mark this' if you change your mind.

What to do if you cannot get started on a problem

Try the following, in order:

- Reread the question to check that you understand what is wanted.
- Reread the question to look for clues – the way it is phrased, or the way a formula is written, or other relevant aspects of the question. (You may think that the setters are trying to set difficult questions or to catch you out. Usually, nothing could be further from the truth: they are probably doing all in their power to make it easy for you by trying to tell you what to do).
- Try to work out exactly what it is that you don't understand.
- Simplify the notation – for example, by writing out sums explicitly.
- Look at special cases (choose special values which simplify the problem) in order to try to understand why a result is true.
- Write down your thoughts – in particular, try to express the exact reason why you are stuck.
- Go on to another question and go back later.
- If you are preparing for the examination (but not in the actual examination!) take a short break.[6]
- Discuss it with a friend or teacher (again, better not do this in the actual examination) or consult the hints and answers, but make sure you still think it through yourself.

  BUT REMEMBER: following someone else's solution is not remotely the same thing as doing the problem yourself. Once you have seen someone else's solution to a problem, then you are deprived, for ever, of much of the benefit that could have come from working it out yourself.

  Even if, ultimately, you get stuck on a particular problem, you derive vastly more benefit from seeing a solution to something with which you have already struggled, than by simply following a solution to something to which you've given very little thought.

What if a problem isn't coming out?

If you have got started but the answer doesn't seem to be coming out, then try the following:

---

[6] J. E. Littlewood (1885–1977), distinguished Cambridge mathematician and author of the highly entertaining *A Mathematician's Miscellany* (Cambridge: Cambridge University Press, 1986) used to work seven days a week until an experiment revealed that when he took Sundays off, the good ideas had a way of coming on Mondays.

- Check your algebra. In particular, make sure that what you have written works in special cases. For example: if you have written the series for $\log(1+x)$ as

  $$1 - x + \tfrac{1}{2}x^2 - \tfrac{1}{3}x^3 + \cdots$$

  then a quick check will reveal that it doesn't work for $x = 0$; clearly, the 1 should not be there.[7]

  A note on the subject of algebra. In many of the problems in this book, the algebra is quite stiff: you have to go through many lines of calculation before you get to an expression recognisably close to your target. Really, the only way to manage this efficiently is to check each line carefully before going on the next line. Otherwise, you can waste hours.[8]

- Make sure that what you have written makes sense. For example, in a problem which is dimensionally consistent, you cannot add x (with dimension length, say) to x^2 or to $\exp x$ (which itself does not make sense — the argument of $\exp$ has to be dimensionless). Even if there are no dimensions in the problem, it is often possible to mentally assign dimensions and hence enable a quick check.

  Be wary of applying familiar processes to unfamiliar objects (very easy to do when you are feeling at sea): for example, it is all too easy, if you are not sure where your solution is going, to solve the vector equation $\mathbf{a}.\mathbf{x} = 1$ by dividing both sides by vector $\mathbf{a}$; a bad idea.

- Analyse exactly what you are being asked to do. Try to understand the hints, explicit and implicit. Remember to distinguish between terms such as explain/prove/define/etc. (There is essentially no difference between 'prove' and 'show': the former tends to be used in more formal situations, but if you are asked to 'show' something, a proper proof is required.)

- Remember that different parts of a question are often linked. There may be guidance in the notation and choice of names of variables in the question.

- If you get irretrievably stuck in the exam, state in words what you are trying to do and move on (at A-level, you don't get credit for merely stating intentions, but STEP examiners are generally grateful for any sign of intelligent life).

What to do after completing a question

It is a natural instinct to consider that you have finished with a question once you have got to an answer. However this instinct should be resisted both from a general mathematical point of view and from the much narrower view of preparing for an examination. Instead, when you have completed a question you should stop for a few minutes and think about it. Here is a check list for you to run through:

- Look back over what you have done, checking that the arguments are correct and making sure that they work for any special cases you can think of. It is surprising how often a chain of completely spurious arguments and gross algebraic blunders leads to the given answer.

- Check that your answer is reasonable. For example if the answer is a probability p then you should check that $0 \leqslant p \leqslant 1$. If your answer depends on an integer n, does it behave as it should when $n \to \infty$? Is it dimensionally correct?

---

[7] Another check will reveal that for very small positive x, $\log$ is positive (since its argument is bigger than 1) whereas the series is negative (once the 1 has been removed), so there is clearly something else wrong.

[8] Which is what I normally do; but I am still hoping to heed my own advice in the not-too-distant future.

If, in the exam, you find that your answer is not reasonable, but you don't have time to do anything about it, then write a brief phrase showing that you understand that your answer is unreasonable (for example 'This is wrong because mass must be positive'). It may be that your time will be more profitably spent doing another question than hunting for a sign error.

- Check that you have used all the information given. In many ways the most artificial aspect of examination questions is that you are given exactly the amount of information required to answer the question. If your answer does not use everything you are given then either it is wrong or you are faced with a very unusual examination question.
- Check that you have understood the point of the question. It is, of course, the case that not all exam questions have a point, but many do. What idea did the examiners want you to have? Which techniques did they want you to demonstrate? Is the result of the question interesting in some way? Does it generalise? If you can see the point of the question would your working show the point to someone who did not know it in advance?
- Make sure that you are not unthinkingly applying mathematical tools which you do not fully understand.[9]
- As preparation for the examination, make sure that you actually understand not only what you have done, but also why you have done it that way rather than some other way. This is particularly important if you have had to use a hint or solution.
- If the question has associated hints, answers or examiner's reports then it is a good idea to read these to see what the common approaches and pitfalls were.
- In the examination, check that you have given the detail required. There often comes a point in a question where, if we could show that A implies B, then the result follows. If, after a lot of thought, you suddenly see that A does indeed imply B the natural thing to do is to write triumphantly 'But A implies B so the result follows'[10]. Unfortunately, unscrupulous individuals (not you, of course) who have no idea why A should imply B (apart from the fact that it would complete the question) could, and do, write the exactly the same thing. Go back through the major points of the question making sure that you have written enough for the examiner to be able to follow your reasoning.

---

[9] Mathematicians should feel as insulted as engineers by the following joke.

A mathematician, a physicist and an engineer enter a mathematics contest, the first task of which is to prove that all odd numbers are prime. The mathematician has an elegant argument: '1's a prime, 3's a prime, 5's a prime, 7's a prime. *Therefore*, by mathematical induction, all odd numbers are prime. It's the physicist's turn: '1's a prime, 3's a prime, 5's a prime, 7's a prime, 11's a prime, 13's a prime, so, to within experimental error, all odd numbers are prime.' The most straightforward proof is provided by the engineer: '1's a prime, 3's a prime, 5's a prime, 7's a prime, 9's a prime, 11's a prime . . . '.

[10] There is an old anecdote about the distinguished Professor X. In the middle of a lecture she writes 'It is obvious that A', suddenly falls silent and after a few minutes she rubs it out and walks out of the room. The awed students hear her pacing up and down outside. Then after twenty minutes she returns and writes 'It is obvious that A' and continues the lecture.

Worked Problems

Worked problem 1 (✓)

Let

$$f(x) = ax - \frac{x^3}{1+x^2},$$

where a is a constant.

Show that, if $a \geqslant 9/8$, then

$$f'(x) \geqslant 0$$

for all x.

2000 Paper I

First thoughts

When I play tennis and I see a ball that I think I can hit, I rush up to it and smack it into the net. This is a tendency I try to overcome when I am doing mathematics. In this question, for example, even though I'm pretty sure that I can find $f'(x)$, I'm going to pause for a moment before I do so. I am going to use the pause to think about two things.

- I'm going to think about what to do when I have found $f'(x)$.
- I'm going to try to decide what the question is really about.

Of course, I may not be able to decide what to do with $f'(x)$ until I actually see what it looks like, and I may not be able to see what the question is really about until I have finished it—maybe not even then—some questions are not really about anything in particular. I am also going to use the pause to think a bit about the best way of performing the differentiation. Should I simplify first? Should I make some sort of substitution? Clearly, knowing how to tackle the rest of the question might guide me in deciding the best way to do the differentiation. Or it may turn out that the differentiation is fairly straightforward, so that it doesn't matter how I do it.

Two more points occur to me as I re-read the question:

- I notice that the inequalities are not strict (they are $\geqslant$ rather than $>$). Am I going to have to worry about the difference?
- I also notice that there is an 'If ... then', and I wonder if this is going to cause me trouble. I will need to be careful to get the implication the right way round. I mustn't try to prove that if $f'(x) \geqslant 0$ then $a \geqslant 9/8$.

  It will be interesting to see why the implication is only one way—why it is not an 'if and only if' question. It may be just 'if' because 'only if' isn't true; or it may be just 'if' because the examiners thought that the question was long enough without the 'only if'.

Don't turn over until you have spent a little time thinking along these lines.

Doing the question

Looking ahead, it is clear that the real hurdle is going to be showing that $f'(x) \geqslant 0$. How am I going to do that? Two ways suggest themselves. First, if $f'(x)$ turns out to be a quadratic function, or an obviously positive multiple of a quadratic function, I should be able to use some standard method: looking at the discriminant ('$b^2 - 4ac$'), etc; or, better, completing the square. But if $f'(x)$ is not of this form, I will have to think of something else: maybe I will have to sketch a graph.

I've just noticed that f is an odd function, i.e. $f(x) = -f(-x)$. (Did you notice that?) That is helpful, because it means that $f'(x)$ is an even function.[11] If it had been odd, there could have been a difficulty, because all odd functions (at least, those with no vertical asymptotes) cross the horizontal axis $y = 0$ at least once and cannot therefore be positive for all x.

Now, how should I do the differentiation? I could divide out the fraction, giving

$$\frac{x^3}{1+x^2} = x - \frac{x}{1+x^2}$$

and

$$f(x) = (a-1)x + \frac{x}{1+x^2} = bx + \frac{x}{1+x^2}\,, \tag{†}$$

where I have set $b = a - 1$ to save writing. (Is this right? I'll just check that it works when $x = 2$. Yes, it does: $f(2) = 2a - \frac{8}{5} = 2(a-1) + \frac{2}{5}$.) This might save a bit of writing, but I don't at the moment see it helping me towards a positive function. On balance, I think I'll stick with the original form.

Another thought: should I differentiate the fraction using the quotient formula or should I write it as $x^3(1+x^2)^{-1}$ and use the product rule? I doubt if there is much in it. I never normally use the quotient rule—it's just extra baggage to carry around. But on this occasion, since the final form I am looking for is a single fraction and the denominator using the quotient rule is a square and therefore non-negative, I will.

One more thought: since I am trying to obtain an inequality, I must be careful throughout not to cancel any quantity which might be negative; or at least if I do cancel a negative quantity, I must remember to reverse the inequality.

Here goes (at last):

$$\begin{aligned} f'(x) &= a - \frac{3x^2(1+x^2) - 2x(x^3)}{(1+x^2)^2} \\ &= \frac{a + (2a-3)x^2 + (a-1)x^4}{(1+x^2)^2}\,. \end{aligned}$$

I had to do a bit of algebra to obtain the second equation.

This is working out as I had hoped: the denominator is certainly positive and the numerator is a quadratic function of x^2. I can finish this off most elegantly by completing the square. There are two ways of doing this. Just to be clear in my mind, I'm going to write the numerator as

$$A + Bx^2 + Cx^4\,,$$

where $A = a$, $B = 2a - 3$ and $C = a - 1$. Then I can complete the square in two ways:

$$A\big(1 + (B/2A)x^2\big)^2 + \big(C - B^2/4A\big)x^4$$

or

$$C\big(x^2 + B/2C\big)^2 + \big(A - B^2/4C\big)\,. \tag{‡}$$

[11] You can see this easily by sketching a 'typical' odd function, or by differentiating the Maclaurin expansion of an odd function.

It doesn't seem to matter which I use. The second expression (‡) looks a bit simpler.

If $a \geqslant \frac{9}{8}$, then certainly $C > 0$. That means that the first of the two terms of (‡) is positive: $C(x^2 + B/2C)^2 > 0$.

The second term of (‡) can be written in terms of a as follows:

$$\begin{aligned} A - B^2/4C &= a - \frac{(2a-3)^2}{4(a-1)} \\ &= \frac{4a(a-1) - (2a-3)^2}{4(a-1)} \\ &= \frac{8a-9}{4(a-1)}\,. \end{aligned} \qquad (*)$$

It is very pleasing to see the numerator $8a - 9$; it is exactly what I want, because it is non-negative if $a \geqslant \frac{9}{8}$. That is what I needed to complete the question.

Post-mortem

Now I can look back and analyse what I have done.

On the technical side, it seems I was right not to use the simplified form of $x^3/(1+x^2)$ given in (†). This would have lead to the quadratic $(b+1) + (2b-1)x^2 + bx^4$, which isn't any easier to handle than the quadratic involving a and just gives an extra opportunity to make an algebraic error.

Although introducing new variables A, B and C seemed at first to complicate the problem, it was useful to have them when it came to completing the square, which would have been a bit of a mess had I worked directly from the quartic expression with a in it.

Actually, I see on re-reading my solution that I have made a bit of a meal out of the ending. I needn't have completed the square at all; I could have used the inequality $a \geqslant 9/8$ immediately after finding $\mathrm{f}'(x)$, since a appears in $\mathrm{f}'(x)$ with a plus sign always:

$$\mathrm{f}'(x) = \frac{a + (2a-3)x^2 + (a-1)x^4}{(1+x^2)^2} \geqslant \frac{\frac{9}{8} + (\frac{9}{4} - 3)x^2 + (\frac{9}{8} - 1)x^4}{(1+x^2)^2} = \frac{(x^2-3)^2}{8(1+x^2)^2} \geqslant 0$$

as required.

However, my unnecessarily elaborate proof, involving completing the square, makes the role of a a bit clearer than it is in the shorter alternative. In fact, I can see how to answer my question about 'if and only if', namely, is it the case that $\mathrm{f}'(x) \geqslant 0$ for all x if *and only if* $a \geqslant \frac{9}{8}$? The answer is: Yes. Looking at (‡), I see that if $\mathrm{f}'(x) \geqslant 0$ for all x, then certainly $C \geqslant 0$, otherwise for large enough x the first (squared) term would be dominant and negative. But if $0 \leqslant C < \frac{1}{8}$, then $B < 0$ and there is a value of x^2 for which the squared term in (‡) vanishes, leaving only the second term which is negative, as can be seen from $(*)$. That means $\mathrm{f}'(x) < 0$ for this value of x, contradicting our assumption. Therefore, $\mathrm{f}'(x) \geqslant 0$ for all x implies that $C \geqslant \frac{1}{8}$ and $a \geqslant \frac{9}{8}$.

I wonder why the examiner wanted me to investigate the sign of $\mathrm{f}'(x)$. The obvious reason is to see what the graph looks like. We can now see what this question is about. It is clear that the examiners really wanted to set the question: 'Sketch the graphs of the function $ax - x^3/(1+x^2)$ in the different cases that arise according to the value of a' but it was thought too long or difficult. It is worth looking back over my working to see what can be said about the shape of the graph of $\mathrm{f}(x)$ when $a < 9/8$.

(I leave that to you to think about!)

Worked problem 2 (✓)

The n positive numbers $x_1, x_2, \ldots, x_n$, where $n \geqslant 3$, satisfy

$$x_1 = 1 + \frac{1}{x_2}\,, \quad x_2 = 1 + \frac{1}{x_3}\,, \quad \ldots\,, \quad x_{n-1} = 1 + \frac{1}{x_n}\,,$$

and also

$$x_n = 1 + \frac{1}{x_1}\,.$$

Show that

(i) $x_1, x_2, \ldots, x_n > 1$,

(ii) $x_1 - x_2 = -\dfrac{x_2 - x_3}{x_2 x_3}\,,$

(iii) $x_1 = x_2 = \cdots = x_n\,.$

Hence find the value of x_1.

1999 Paper I

First thoughts

My first thought is that this question has an unknown number of variables: $x_1, \ldots, x_n$. That makes it seem rather complicated. I might, if necessary, try to understand the result by choosing an easy value for n (maybe $n = 3$). If I manage to prove some of the results in this special case, I will certainly go back to the general case: doing the special case might help me tackle the general case, but I don't expect to get many marks in an exam if I just prove the result in one special case.

Next, I see that the question has three sub-parts, then a final one. The final one begins 'Hence ...'. This means that I must use at least one of the previous parts in my working for the final part. It is not clear from the structure of the question whether the three sub-parts are independent; the proof of (ii) and (iii) may require the previous result(s), or it may not.

Actually, I think I can see how to do the very last part. If I assume that (iii) holds, so that $x_1 = x_2 = \cdots = x_n = x$ (say), then each of the equations given in the question is identical and each gives a simple equation for x.

It is surprising, isn't it, that the final result doesn't depend on the value of n? That makes the idea of choosing $n = 3$, just to see what is going on, quite attractive, but I'm not going to resort to that idea unless I get very stuck.

The notation in part (i) is a bit odd. I'm not sure that I have seen anything like it before.[12] But it can only mean that each of the variables $x_1, x_2, \ldots\ x_n$ is greater than 1.

In fact, now I think about it, I am puzzled about part (i). How can an *inequality* help to derive the *equalities* in the later parts? I can think of a couple of ways in which the result $x_i > 1$ could be used. One is that I may need to cancel, say, x_1 from both sides of an equation in which case I would need to know that $x_1 \neq 0$. But looking back at the question, I see I already know that $x_1 > 0$ (it is given right at the

[12] It is just the sort of thing that is used in university texts; but I'm not sure that it would be used in STEP papers nowadays.

start of the question), so this cannot be the right answer. Maybe I need to cancel some other factor, such as $(x_1 - 1)$. Another possibility is that I get two or more solutions by putting $x_1 = x_2 = \cdots = x_n = x$ and I need the one with $x > 1$. This may be the answer: looking back at the question again, I see that it asks for *the* value of x_1 — so I am looking for a single value. I'm still puzzled, but I will remember to keep a sharp look out for ways of using part (i).

One more thing strikes me about the question. The equations satisfied by the x_i are given on two lines ('and also'). This could be for typographic reasons (the equations would not all fit on one line), but more likely it is to make sure that I have noticed that the last equation is a bit different: all the other equations relate x_i to x_{i+1}, whereas the last equation relates x_n to x_1. It goes back to the beginning, completing the cycle. I'm pleased that I thought of this, because this circularity must be important.

Doing the question

I think I will do the very last part first, and see what happens.

Suppose, assuming the result of part (iii), that $x_1 = x_2 = \cdots = x_n = x$. Then substituting into any of the equations given in the question gives

$$x = 1 + \frac{1}{x}$$

i.e. $x^2 - x - 1 = 0$. Using the quadratic formula gives

$$x = \frac{1 \pm \sqrt{5}}{2},$$

which does indeed give two answers (despite the fact that the question asks for just one). However, I see that one is negative and can therefore be eliminated by the condition $x_i > 0$ which was given in the question (not, I note, the condition $x_1 > 1$ from part (i); I still have to find a use for this).

I needed only part (iii) to find x_1, so I expect that either I need both (i) and (ii) directly to prove (iii), or I need (i) to prove (ii), and (ii) to prove (iii).

Now that I have remembered that $x_i > 0$ for each i, I see that part (i) is obvious. Since $x_2 > 0$ then $1/x_2 > 0$ and the first equation given in the question, $x_1 = 1 + 1/x_2$, shows immediately that $x_1 > 1$ and the same applies to x_2, x_3, etc.

Now what about part (ii)? The given equation involves x_1, x_2 and x_3, so clearly I must use the first two equations given in the question:

$$x_1 = 1 + \frac{1}{x_2}, \quad x_2 = 1 + \frac{1}{x_3}.$$

Since I want $x_1 - x_2$, I will see what happens if I subtract the two equations:

$$x_1 - x_2 = (1 + \frac{1}{x_2}) - (1 + \frac{1}{x_3}) = \frac{1}{x_2} - \frac{1}{x_3}i = \frac{x_3 - x_2}{x_2 x_3} = -\frac{x_2 - x_3}{x_2 x_3}. \qquad (*)$$

That seems to work!

One idea that I haven't used so far is what I earlier called the circularity of the equations: the way that x_n links back to x_1. I'll see what happens if I extend the above result. Since there is nothing special about x_1 and x_2, the same result must hold if I add 1 to each of the suffices:

$$x_2 - x_3 = \frac{x_3 - x_4}{x_3 x_4}.$$

I see that I can combine this with the previous result:

$$x_1 - x_2 = -\frac{x_2 - x_3}{x_2 x_3} = \frac{x_3 - x_4}{x_2 x_3^2 x_4}.$$

I now see where this is going. The above step can be repeated to give

$$x_1 - x_2 = \frac{x_3 - x_4}{x_2 x_3^2 x_4} = -\frac{x_4 - x_5}{x_2 x_3^2 x_4^2 x_5} = \cdots$$

and eventually I will get back to $x_1 - x_2$:

$$x_1 - x_2 = -\frac{x_4 - x_5}{x_2 x_3^2 x_4^2 x_5} = \cdots = \pm \frac{x_1 - x_2}{x_2 x_3^2 x_4^2 x_5^2 \ldots x_n^2 x_1^2 x_2}$$

i.e.

$$(x_1 - x_2)\left(1 \mp \frac{1}{x_1^2 x_2^2 x_3^2 x_4^2 x_5^2 \ldots x_n^2}\right) = 0\ .$$

I have put in a $\pm$ because each step introduces a minus sign and I'm not sure yet whether the final sign should be $(-1)^n$ or $(-1)^{n-1}$. I can check this later (for example, by working out one simple case such as $n = 3$); but I may not need to.

I deduce from this last equation that either

$$x_1^2 x_2^2 x_3^2 x_4^2 x_5^2 \ldots x_n^2 = \pm 1$$

or

$$x_1 = x_2$$

(which is what I want). At last I see where to use part (i): I know that

$$x_1^2 x_2^2 x_3^2 x_4^2 x_5^2 \ldots x_n^2 \neq \pm 1$$

because $x_1 > 1$, $x_2 > 1$, etc. Thus the only possibility is $x_1 = x_2$. Since there was nothing special about x_1 and x_2, I deduce further that $x_2 = x_3$, and so on, as required.

Post-mortem

There were a number of useful points in this question:

1. The first point concerns using the information given in the question. The process of teasing information from what is given is fundamental to the whole of mathematics. It is very important to study what is given (especially seemingly unimportant conditions, such as $x_i > 0$) to see why it has been given. If you find you reach the end of a question without apparently using some given information, then you should look back over your work: it is very unlikely that a condition has been given that is not used in some way. It may not be a necessary condition — and we will see that the condition $x_i > 0$ is not, in a sense, necessary in this question — but it should be sufficient. The other piece of information in the question which you might easily have overlooked is the use of the singular rather than the plural in referring to the solution (' ... find *the value* of ...'), implying that there is just one value, despite the fact that the final equation is quadratic.

2. The second point concerns the structure of the question. Here, the position of the word 'Hence' suggested strongly that none of the separate parts were stand-alone results; each had to be used for a later proof. Understanding this point made the question much easier, because I was always on the look out for an opportunity to use the earlier parts. Of course, in some problems (without that 'hence') some parts may be stand-alone; though this is rare in STEP questions.

   You may think that this is like playing a game according to hidden (STEP) rules, but that is not the case. Precision writing and precision reading is vital in mathematics and in many professions (law, for example). Mathematicians have to be good at it, which is the reason why so many employers want to recruit people with mathematical training.

3. The third point was the rather inconclusive speculation about the way inequalities might help to derive an equality. It turned out that what was actually required was $x_1x_2 \dots x_n \neq \pm 1$. I was a bit puzzled by this possibility in my first thoughts, because it seemed that the result ought to hold under conditions different from those given; for example, $x_i < 0$ for all i (does this condition work??). Come to think of it, why are conditions given on all x_i when they are all related by the given equations? This makes me think that there ought to be a better way of proving the result which would reveal exactly the conditions under which it holds.

4. Then there was the idea (which I didn't actually use) that I might try to prove the result for, say, $n = 3$ to help me understand what was going on. This would not have counted as a proof of the result (or anything like it), but it might have given me ideas for tackling the question.

5. A key observation was that the equations given in the question are 'circular'. It was clear that the circularity was essential to the question and it turned out to be the key to the most difficult part. Having identified it early on, I was ready to use it when the opportunity arose.

6. Finally, I was pleased that I read the whole question carefully before plunging in. This allowed me to see that I could easily do the last part before the preceding parts, which I found very helpful in getting into the question.

Final thoughts

It occurs to me only now, after my post-mortem, that there is another way of obtaining the final result.

Suppose I start with the idea of circularity (as indeed I might have, had I not been otherwise directed by the question) and use the given equations to find x_1 in terms of first x_2, then x_3, then x_4 and eventually in terms of x_1 itself. That should give me an equation I can solve, and I should be able to find out what conditions are needed on the x_i. Try it. You may need to guess a formula for x_1 in terms of x_i from a few special cases, then prove it by induction.[13] You will find it useful to define a sequence of numbers F_i such that $F_0 = F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$. (These numbers are called *Fibonacci numbers*.[14])

You should find that if $x_n = x_1$ for any n (greater than 1), then $x_n = \frac{1 \pm \sqrt{5}}{2}$.

---

[13] I went as far as x_7 to be sure of the pattern: I found that $x_7 = \frac{8x_1+5}{5x_1+3}$.

[14] Fibonacci (short for *filius* Bonacci — son of Bonacci) was called the greatest European mathematician of the Middle Ages. He was born in Pisa (Italy) in about 1175 AD. He introduced the series of numbers named after him in his book of 1202 called *Liber Abbaci (Book of the Abacus)*. It was the solution to the problem of the number of pairs of rabbits produced by an initial pair: *A pair of rabbits are put in a field and, if rabbits take a month to become mature and then produce a new pair every month after that, how many pairs will there be in twelve months time?*

Problems

Problem 1: An integer equation (✓)

(i) Find all sets of positive integers a, b and c that satisfy the equation

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} = 1\,.$$

(ii) Determine the sets of positive integers a, b and c that satisfy the inequality

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} \geqslant 1\,.$$

1993 Paper I

Comments

Age has not diminished the value of this old chestnut. It requires almost no mathematics in the sense of examination syllabuses, but instead tests a vital asset for a mathematician, namely the capacity for systematic thought. For this reason, it tends to crop up in mathematics contests, where competitors come from different backgrounds. I most recently saw it in the form 'Find all the positive integer solutions of $bc + ca + ab = abc$'. (You see the connection?)

You will want to make use of the symmetry between a, b and c: if, for example, $a = 2$, $b = 3$ and $c = 4$, then the same solution can be expressed in five other ways, such as $a = 3$, $b = 4$ and $c = 2$. You do not want to derive all these separately, so you need to order a, b and c in some way.

The reason the question uses the word 'determine', rather than 'find' for part (ii) is that in part (i) you can write down all the possibilities explicitly, whereas for part (ii) there are, in some cases, an infinite number of possibilities which obviously cannot be explicitly listed, though they can be described and hence determined.

Solution to problem 1

Let us take, without any loss of generality, $a \leqslant b \leqslant c$.

(i) How small can a be?

First set $a = 1$. This gives no solutions, because it leaves nothing for $1/b + 1/c$.

Next set $a = 2$ and try values of b (with $b \geqslant 2$, since we have assumed that $a \leqslant b \leqslant c$) in order:

if $b = 2$, then $1/c = 0$, which is no good;
if $b = 3$, then $c = 6$, which works;
if $b = 4$, then $c = 4$, which works;
if $b \geqslant 5$, then $c \leqslant b$, so we need not consider this.

Then set $a = 3$, and try values of b (with $b \geqslant 3$) in order:

if $b = 3$, then $c = 3$, which works;
if $b \geqslant 4$, then $c \leqslant b$, so we need not consider this.

Finally, if $a \geqslant 4$, then at least one of b and c must be $\leqslant a$, so we need look no further.

The only possibilities are therefore $(2, 3, 6)$, $(2, 4, 4)$ and $(3, 3, 3)$.

(ii) Clearly, we must include all the solutions found in part (i). We proceed systematically, as in part (i).

First set $a = 1$. This time, any values of b and c will do (though we decided to choose $b \leqslant c$).

Next set $a = 2$ and try values of b (with $b \geqslant 2$, since we have assumed that $a \leqslant b \leqslant c$) in order:

if $b = 2$, then any value of c will do (with $b \leqslant c$).
if $b = 3$, then 3, 4, 5 and 6 will do for c, but 7 is too big;
if $b = 4$, then $c = 4$ will do, but 5 is too big for c;
if $b \geqslant 5$, then $c \leqslant b$, so we need not consider this.

Then set $a = 3$ and try values of b (with $b \geqslant 3$) in order:

if $b = 3$, then $c = 3$, which works, but $c = 4$ is too big;
if $b \geqslant 4$, then $c \leqslant b$, so we need not consider this.

As before $a = 4$ and $b \geqslant 4$, $c \geqslant 4$ gives no possibilities.

Therefore the extra sets for part (ii) are of the form $(1, b, c)$, $(2, 2, c)$, $(2, 3, 3)$, $(2, 3, 4)$ and $(2, 3, 5)$.

Post-mortem

A slightly different approach for part (i), which would also generalise for part (ii), is to start with the case $a = b = c$, for which the only solution is $a = b = c = \frac{1}{3}$. If a, b and c are not all equal then one of them, which we may take to be a, must be greater than $\frac{1}{3}$, i.e. $\frac{1}{2}$. The two remaining possibilities follow easily.

As long as you are methodical, it doesn't matter how you approach the question.

Having done the question, you might well want to investigate whether it generalises: could we replace the 1 on the right hand side with 2? Could we have four reciprocals instead of three? After a few scribbles, I decided that these other cases are not very interesting; but you might find something I missed.

Problem 2: Partitions of 10 and 20 (✓)

(i) Show that you can make up 10 pence in eleven ways using 10p, 5p, 2p and 1p coins.

(ii) In how many ways can you make up 20 pence using 20p, 10p, 5p, 2p and 1p coins?

1997 Paper I

Comments

I don't really approve of this sort of question, at least as far as STEP is concerned, but I thought I'd better include one in this collection. The one given above seems to me to be a particularly bad example, because there are a number of neat and elegant mathematical ways of approaching it, none of which turn out to be any use.

The quickest instrument is the bluntest: just write out all the possibilities. Two things are important. First you must be systematic or you will get hopelessly confused; second, you must lay out your solution, with careful explanations, in a way which allows other people (examiners, for example) to understand exactly what you are doing. You can probably lay it out using pen and paper much better than I have overleaf, using a word-processing package. If you get an answer and, looking back, find that your work lacks clarity, then do it again.

Since different parts of STEP questions are nearly always related, you might be led to believe that the result of the second part follows from the first: you divide the required twenty pence into two tens and then use the result of the first part to give the number of ways of making up each 10. This would give an answer of 66 (why?) plus one for a single 20p piece. This would be neat, but the true answer is less than 67 because some arrangements are counted twice by this method — and it is not easy to work out which ones.

Solution to problem 2

Probably the best approach is to start counting with the arrangements which use as many high denomination coins as possible, then work down.

(i) We can make up 10p as follows:

10 (one way using only one 10p coin);

5+5 (one way using two 5p coins and no 10p coins);

5+2+2+1, 5+2+1+1+1, 5+1+1+1+1+1 (three ways using one 5p coin and no 10p coins);

2+2+2+2+2, 2+2+2+2+1+1, etc (six ways using no 5p or 10p coins).

That makes 11 ways altogether.

(ii) We can make up 20p as follows:

20 (one way using only a 20p coin);

10 + any of the 11 arrangements in the first part of the question (11 ways using one or two 10p coins);

5+5+5+5 (one way using four 5p coins and no 10p or 20p coins);

5+5+5+2+2+1, etc (3 ways using three 5p coins and no 10p or 20p coins);

5+5+2+2+2+2+2, 5+5+2+2+2+2+1+1, etc (6 ways using two 5p coins and no 10p or 20p coins);

5+2+2+2+2+2+2+2+1, etc (8 ways using one 5p coin and no 10p or 20p coins);

2+2+2+2+2+2+2+2+2+2, etc (11 ways using no 5p, 10p or 20p coins).

That makes 41 ways altogether.

Post-mortem

As in the previous question, the most important lesson to be learnt here is the value of a systematic approach and clear explanations. You should not be happy just to obtain the answer: there is no virtue in that. You should only be satisfied if you displayed your working at least as systematically as I have, above.

On reflection, this question is not quite as bad as I thought. I did in fact use the first part to help with the second. And in the second part I certainly used the method of setting out the different ways systematically that I developed in the less complicated first part.

Here is an interesting thought. For the part (i), consider the expansion of

$$\frac{1}{(1-x^{10})(1-x^5)(1-x^2)(1-x)} \qquad (*)$$

in powers of x. This we can obtain from the binomial expansion of each of the four terms in the denominator, giving $(1+x^{10}+x^{20}+\cdots)(1+x^5+x^{10}+\cdots)(1+x^2+x^4+\cdots)(1+x+x^2+\cdots)$. Now if we multiply out these brackets, we find that the term in x^{10} (say) is the sum of terms of the form $x^{10a}\times x^{5b}\times x^{2c}\times x^d$, one from each bracket, where a, b, c and d are non-negative integers such that $10a+5b+2c+d=10$. It is not hard to see that there is exactly one possibility for these integers for each of the arrangements of the coins in part (i), so the number of arrangements is exactly equal to the coefficient of x^{10} in $(*)$.

Although this is neat, it doesn't help, because there is no easy way of obtaining the required coefficient. Formally, though, we could obtain the coefficient using the Taylor series expansion of $(*)$, which involves differentiating $(*)$ 10 times and setting $x=0$. This is interesting, because we have converted a problem in discrete mathematics, involving only integers, to a problem in calculus involving only smooth functions.

Problem 3: Mathematical deduction (✓)

(i) Write down the average of the integers n_1, (n_1+1), ... , (n_2-1), n_2 . Show that

$$n_1 + (n_1+1) + \cdots + (n_2-1) + n_2 = \tfrac{1}{2}(n_2 - n_1 + 1)(n_1 + n_2)\,.$$

(ii) Write down and prove a general law of which the following are special cases:

$$\begin{aligned} 1 &= 0+1 \\ 2+3+4 &= 1+8 \\ 5+6+7+8+9 &= 8+27 \\ 10+11+\cdots+16 &= 27+64\,. \end{aligned}$$

Hence prove that

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \tfrac{1}{4}n^2(n+1)^2.$$

1994 Paper I

Comments

You could use induction to prove your general law, but it is not necessary. The obvious alternative involves use of part (i). You might think of induction for the second bit of part (ii), but the question specifically tells you to use the previous result.

There are various ways of summing kth powers of integers, besides the trick given here, which does not readily generalise to powers other than the third. A simple way is to assume that the result is a polynomial of degree $k+1$. You can then find the coefficients by substituting in $k+1$ values of n to obtain $k+1$ simultaneous equations for the coefficients of the polynomial. We can guess that the leading coefficient will in general be $(k+1)^{-1}$, because the sum approximates the area under the graph $y = x^k$ from $x = 1$ to $x = n$, which is given by the integral of x^k.

Solution to problem 3

(i) The average is $\frac{1}{2}(n_1 + n_2)$. (This is the mid-point of a ruler with n_1 at one end and n_2 at the other.) The sum of any numbers is the average of the numbers times the number of numbers, as given.

(ii) A general rule is

$$\sum_{k=m^2+1}^{(m+1)^2} k = m^3 + (m+1)^3.$$

We can prove this result by applying the formula derived in the first part of the question:

$$\begin{aligned}\sum_{k=m^2+1}^{(m+1)^2} k &= \tfrac{1}{2}\big[(m+1)^2 - m^2\big]\big[(m+1)^2 + (m^2+1)\big] \\ &= \tfrac{1}{2}\big[2m+1\big]\big[2m^2+2m+2\big] \\ &= 2m^3 + 3m^2 + 3m + 1 \\ &= m^3 + (m+1)^3.\end{aligned}$$

For the last part, we can obtain sums of the cubes by adding together the general law for consecutive values of m:

$$\begin{aligned}1 + (2+3+4) + (5+6+7+8+9) + \cdots + \big((N-1)^2+1) + \cdots + N^2\big) \\ = (0+1) + (1+8) + (8+27) + \cdots + \big((N-1)^3 + N^3\big).\end{aligned}$$

The left-hand side is just the sum of integers from 1 to N^3, so using the first part gives

$$\tfrac{1}{2}N^2(N^2+1) = 2(1^3 + 2^3 + \cdots + N^3) - N^3$$

i.e.

$$\tfrac{1}{2}N^2(N^2+2N+1) = 2(1^3 + 2^3 + \cdots + N^3)$$

as required.

Post-mortem

To find the general rule requires an understanding of the algebra you have just done: why does it work in the way it does?

Of course, there are many other 'general rules' besides the one given in the solution. The situation is similar to the 'find the next number in the sequence' questions which come up in IQ tests. As no less a figure than the philosopher Wittgenstein has pointed out, there is no correct answer. Given any finite sequence of numbers, a formula can always be found which will fit all the given numbers and which makes the next number (e.g.) 42.

Nevertheless, when the above question was set, almost everyone produced the same general result (or no result at all). Mathematicians seem to know what sort of thing is required. Electronic computers might not have the faintest idea what to do, though they would no doubt complete the algebra rapidly.

Note: This question has been changed slightly from the original version in the 1994 paper to include part (i), and to change the wording slightly in the second part. Having now included a part (i) the hint given in the original question has now been removed.

Problem 4: Divisibility (✓)

> **(i)** How many integers greater than or equal to zero and less than 1000 are not divisible by 2 or 5? What is the average value of these integers?
>
> **(ii)** How many integers greater than or equal to zero and less than 9261 are not divisible by 3 or 7? What is the average value of these integers?

1999 Paper I

Comments

There are a number of different ways of tackling this problem, but it should be clear that whatever way you choose for part (i) will also work for part (ii) (especially when you realise the significance of the number 9261). A key idea for part (i) is to think in terms of blocks of 10 numbers, realising that all blocks of 10 are the same for the purposes of the problem.

Solution to problem 4

(i) Only integers ending in 1, 3, 7, or 9 are not divisible by 2 or 5. This is $4/10$ of the possible integers, so the total number of such integers is $4/10$ of 1,000, i.e. 400.

The integers can be added in pairs:

$$\text{sum} = (1 + 999) + (3 + 997) + \cdots + (499 + 501)\ .$$

There are 200 such pairs, so the sum is $1,000 \times 200$ and the average is 500.

A simpler argument to obtain the average would be to say that this is obvious by symmetry: there is nothing in the problem that favours an answer greater (or smaller) than 500.

Alternatively, we can find the number of integers divisible by both 2 and 5 by adding the number divisible by 2 (i.e. 500) to the number divisible by 5 (i.e. 200), and subtracting the number divisible by both 2 and 5 (i.e. 100) since these have been counted twice. To find the sum, we can sum those divisible by 2 (using the formula for an arithmetic progression), add the sum of those divisible by 5 and subtract the sum of those divisible by 10. These results can then be used to find the number of integers which are not divisible by 2 and 5, and the average of these (by first summing all the integers between 1 and 1000).

(ii) Either of the above methods will work. In the first method you consider integers in blocks of 21 (essentially arithmetic to base 21): there are 12 integers in each such block that are not divisible by 3 or 7 (namely 1, 2, 4, 5, 8, 10, 11, 13, 16, 17, 19, 20) so the total number is $9261 \times 12/21 = 5292$. The average is $9261/2$ as can be seen using the pairing argument $(1+9260)+(2+9259)+\cdots$ or the symmetry argument.

Post-mortem

I was in year 7 at school when I had my first encounter with lateral thinking in mathematics. The problem was to work out how many houses a postman delivers to in a street of houses numbered from 1 to 1000, given that he or she refuses to deliver to houses with the digit 9 in the number. It seemed impossible to do it systematically, until the idea of counting in base 9 occurred, then it seemed brilliantly simple. I didn't know I was counting in base 9; in those days, school mathematics was very traditional and base 9 would have been thought very advanced. All that was required was the idea of working out how many numbers can be made from nine digits (i.e. 9^3) instead of from ten digits (i.e. 10^3); and of course it doesn't matter which nine.

Did you spot the significance of the number 9261? It is 21^3, i.e. the number in base 10 that is written as 1000 in base 21. Note how carefully the question is written and laid out to suggest a connection between the two paragraphs (and also to highlight the difference). For the examination, the number 1,000,000 was used in part (i) instead of 1000 in order to discourage candidates from spending the first hour of the examination writing down the numbers from 1 to 1000.

Having realised that part (i) involves $(2 \times 5)^3$ and part (ii) involves $(3 \times 7)^3$ you are probably now wondering what the general result is, i.e.:

How many integers greater than or equal to zero and less than $(pq)^3$ are not divisible by p or q? What is the average value of these integers? Let's make p and q co-prime, so that they have no common factors.

I leave this to you, with the hint that you need to think about only pq, rather than $(pq)^3$, to start with.

Problem 5: The modulus function (✓)

Find all the solutions of the equation

$$|x+1| - |x| + 3|x-1| - 2|x-2| = x+2\ .$$

1997 Paper I

Comments

This looks more difficult than it is.

There is no clever way to deal with the modulus function: you have to look at the different cases individually. For example, for $|x|$ you have to look at $x \leqslant 0$ and $x \geqslant 0$. The most straightforward approach would be to solve the equation in each of the different regions determined by the modulus signs: $x \leqslant -1$, $-1 \leqslant x \leqslant 0$, $0 \leqslant x \leqslant 1$, etc. You might find a graphical approach helps you to picture what is going on (I didn't).

Solution to problem 5

Let

$$\mathrm{f}\,(x) = |x+1| - |x| + 3|x-1| - 2|x-2| - (x+2).$$

We have to solve $\mathrm{f}\,(x) = 0$ in the five regions of the x-axis determined by the modulus functions, namely

$$-\infty < x \leqslant -1\,;\quad -1 \leqslant x \leqslant 0\,;\quad 0 \leqslant x \leqslant 1\,;\quad 1 \leqslant x \leqslant 2\,;\quad 2 \leqslant x < \infty\,.$$

In the separate regions, we have

$$\mathrm{f}\,(x) = \begin{cases} -(x+1) + x - 3(x-1) + 2(x-2) - (x+2) & = & -2x-4 & \text{for} & -\infty < x \leqslant -1 \\ (x+1) + x - 3(x-1) + 2(x-2) - (x+2) & = & -2 & \text{for} & -1 \leqslant x \leqslant 0 \\ (x+1) - x - 3(x-1) + 2(x-2) - (x+2) & = & -2x-2 & \text{for} & 0 \leqslant x \leqslant 1 \\ (x+1) - x + 3(x-1) + 2(x-2) - (x+2) & = & 4x-8 & \text{for} & 1 \leqslant x \leqslant 2 \\ (x+1) - x + 3(x-1) - 2(x-2) - (x+2) & = & 0 & \text{for} & 2 \leqslant x < \infty \end{cases}$$

Solving in each region gives:

1. $-\infty < x \leqslant -1$: here, $\mathrm{f}\,(x) = 0$ only if $x = -2$. This is a solution since the point $x = -2$ lies in the region $x \leqslant -1$.

2. $-1 \leqslant x \leqslant 0$: here, $\mathrm{f}\,(x) = -2 \quad (\neq 0)$ so there is no solution.

3. $0 \leqslant x \leqslant 1$: here, $\mathrm{f}\,(x) = 0$ only if $x = -1$. This is not a solution since the point $x = -1$ does not lie in the region $0 \leqslant x \leqslant 1$.

4. $1 \leqslant x \leqslant 2$: here, $\mathrm{f}\,(x) = 0$ only if $x = 2$, which is a solution since $x = 2$ lies in the range $1 \leqslant x \leqslant 2$.

5. $2 \leqslant x < \infty$: here, $\mathrm{f}\,(x) = 0$ identically (i.e. for all values of x), so the equation is satisfied for all x in the region.

Collecting these results together shows that the equation $\mathrm{f}\,(x) = 0$ is satisfied only by $x = -2$ and by any x in the region $x \geqslant 2$.

Post-mortem

As mentioned before, this should not be found too difficult once you identify the different regions and consider each on its own. Some care is needed, though, and it would be sensible to go back and check that the solutions do indeed satisfy the original equation.

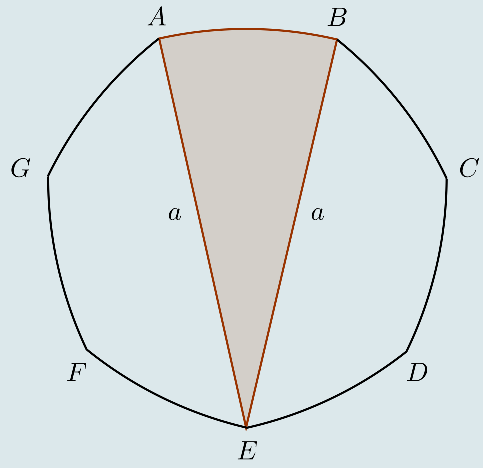
Two small points of technique:

- Since it was necessary to refer to the original equation quite a few times, I found it useful to define a function f so that the equation became $\mathrm{f}\,(x) = 0$.

  Alternatively, you could label the equation with a $(*)$, say, when you first write it down and then later say 'Equation $(*)$ becomes ...'.

- It helped me to set out the different cases very clearly. I numbered them later but it might well have saved writing to number them at the beginning.

Problem 6: The regular Reuleaux heptagon (✓)



The diagram shows a British 50 pence coin. The seven arcs AB, BC, $\ldots$, FG, GA are of equal length and each arc is formed from the circle of radius a having its centre at the vertex diametrically opposite the mid-point of the arc. Show that the area of the face of the coin is

$$\frac{a^2}{2}\left(\pi - 7\tan\frac{\pi}{14}\right).$$

1987 Specimen Paper I

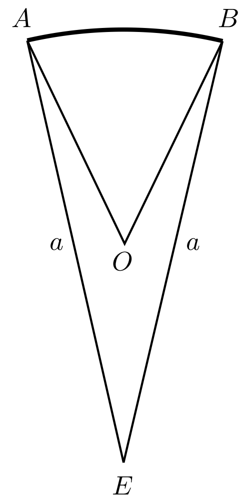
Comments

The first difficulty with this elegant problem is drawing the diagram. However, you can simplify both the drawing and the solution by restricting your attention to just one sector of a circle of radius a.

The figure sketched above has *constant diameter*; it can roll between two parallel lines without losing contact with either. (This looks plausible and you can verify it by sellotaping some 50 pence coins together, but a solid proof is not very easy.) The distance between these lines is the diameter of the figure. Like a circle, the 50 pence piece has circumference equal to π times the diameter, which is in fact always true for a figure with constant diameter.

Reuleaux polygons are general polygons of constant diameter, and a heptagon has seven sides, like, for example, British 50 and 20 pence pieces, Botswanan 50 thebe coins and Jordanian half dinars. The shield on Lancia cars is a Reuleaux triangle.

Solution to problem 6



In the figure, the point O is equidistant from each of three vertices A, B and E. The plan is to find the area of the sector AOB by calculating the area of AEB and subtracting the areas of the two congruent isosceles triangles OBE and OAE. The required area is 7 times this.

First we need angle $\angle AEB$. We know that $\angle AOB = \frac{2}{7}\pi$ and hence $\angle BOE = \frac{1}{2}(2\pi - \frac{2}{7}\pi) = \frac{6}{7}\pi$ (using the sum of angles round the point O). Finally,

$$\angle AEB = 2\angle OEB = 2 \times \tfrac{1}{2}(\pi - \tfrac{6}{7}\pi) = \tfrac{1}{7}\pi\,,$$

using the sum of angles of an isosceles triangle.[15]

Now ABE is a sector of a circle of radius a, so its area is

$$\pi a^2 \times \frac{\frac{1}{7}\pi}{2\pi} = \frac{\pi a^2}{14}.$$

The area of triangle OBE is $\frac{1}{2}BE\times$ height, i.e.

$$\tfrac{1}{2}a \times \frac{a}{2}\tan\angle OBE = \frac{a^2}{4}\tan\frac{\pi}{14}.$$

The area of the coin is therefore

$$7 \times \left(\frac{\pi a^2}{14} - 2 \times \frac{a^2}{4}\tan\frac{\pi}{14}\right),$$

which reduces to the given answer.

Post-mortem

It is simple now to calculate the area of a regular n-sided Reuleaux polygon. You should of course find that the area tends to that of a circle of the same diameter as $n \to \infty$.

---

[15] After the first edition was published, a correspondent pointed out that the following argument gives angle AEB more quickly: clearly A, B and E all lie on a circle with centre O; the angle at O subtended by the chord AB is $\frac{2}{7}\pi$, so the angle at the circumference is $\frac{1}{7}\pi$. Obvious, really—can't think why I didn't see it.

Problem 7: Chain of equations (✓)

Suppose that

$$3 = \frac{2}{x_1} = x_1 + \frac{2}{x_2} = x_2 + \frac{2}{x_3} = x_3 + \frac{2}{x_4} = \cdots .$$

Guess an expression, in terms of n, for x_n . Then, by induction or otherwise, prove the correctness of your guess.

1997 Paper II

Comments

Wording this sort of question is a real headache for the examiners. Suppose you guess wrong; how can you then prove your guess by induction (unless you get that wrong too)? How else can the question be phrased? In the end, we decided to assume that you are all so clever that your guesses will all be correct.

To guess the formula, you need to work out x_1, x_2, x_3 , etc and look for a pattern. You should not need to go beyond x_4 .

Proof by induction is not in the A-level Mathematics specifications. We decided to include it in the specification for STEP 1 because the idea behind it is not difficult and it is very important both as a method of proof and also as an introduction to more sophisticated mathematical thought.

Solution to problem 7

First, let's put the equations into a more manageable form. Each equality can be written in the form

$$3 = x_n + \frac{2}{x_{n+1}}, \quad \text{i.e.} \quad x_{n+1} = \frac{2}{3 - x_n} \,.$$

We find $x_1 = \frac{2}{3}$, $x_2 = \frac{6}{7}$, $x_3 = \frac{14}{15}$ and $x_4 = \frac{30}{31}$. The denominators give the game away. We guess

$$x_n = \frac{2^{n+1} - 2}{2^{n+1} - 1} \,.$$

For the induction, we need a starting point: our guess certainly holds for $n = 1$ (and 2, 3, and 4!).

For the inductive step, we suppose our guess also holds for $n = k$, where k is any integer, so that

$$x_k = \frac{2^{k+1} - 2}{2^{k+1} - 1} \,.$$

If we can show that it then also holds for $n = k + 1$, we are done.

We have, from the equation given in the question,

$$x_{k+1} = \frac{2}{3 - x_k} = \frac{2}{3 - \frac{2^{k+1}-2}{2^{k+1}-1}} = \frac{2(2^{k+1} - 1)}{3(2^{k+1} - 1) - (2^{k+1} - 2)} = \frac{2^{k+2} - 2}{2^{k+2} - 1} \,,$$

as required.

Post-mortem

There's not much to say about this. By STEP standards, it is fairly easy and short. Nevertheless, you are left to your own devices from the beginning, so you should be pleased if you got it out.

Perhaps the key step came in the very first line of the solution, when we had to decide how to separate out the equations. We could have tried instead

$$x_2 + \frac{2}{x_3} = x_3 + \frac{2}{x_4}$$

leading to

$$(x_2 - x_3) = \frac{2(x_3 - x_4)}{x_3 x_4} \,,$$

and so on, but it would not be at all clear what to do next.

Problem 8: Trig. equations (✓✓)

(i) Show that, if $\tan^2\theta = 2\tan\theta + 1$, then $\tan 2\theta = -1$.

(ii) Find all solutions of the equation

$$\tan\theta = 2 + \tan 3\theta$$

which satisfy $0 < \theta < 2\pi$, expressing your answers as rational multiples of π.

(iii) Find all solutions of the equation

$$\cot\phi = 2 + \cot 3\phi$$

which satisfy $-\dfrac{3\pi}{2} < \phi < \dfrac{\pi}{2}$, expressing your answers as rational multiples of π.

1997 Paper II

Comments

There are three distinct parts. It is pretty certain that they are related, but it is not obvious what the relationship is. Part (i) must surely help with part (ii) in some way that will only become apparent once part (ii) is under way.

In the absence of any other good ideas, it looks right to start part (ii) by expressing the double and triple angle tans in terms of single angle tans. You should remember the formula

$$\tan(A+B) = \frac{\tan A + \tan B}{1 - \tan A \tan B}\,.$$

You can use this for $\tan 3\theta$ and hence (later on) for $\cot 3\theta$. If you ever forget the $\tan(A+B)$ formula, you can quickly work it out from the corresponding $\sin$ and $\cos$ formulae:

$$\tan(A+B) = \frac{\sin(A+B)}{\cos(A+B)} = \frac{\sin A\cos B + \cos A \sin B}{\cos A \cos B - \sin A \sin B}\,.$$

You are are not expected to remember the more complicated triple angle formulae (I certainly don't).

You may well find yourself trying to solve cubic equations at some stage in this question; no need to panic — there is sure to be one easily spottable root, in which case you can reduce the cubic to a quadratic.

Interesting, isn't it, that the range of ϕ for part (iii) is not the obvious $0 < \theta < 2\pi$? Maybe that is significant.

Solution to problem 8

We will write t for $\tan\theta$ (or $\tan\phi$) throughout.

(i) $\tan 2\theta = \dfrac{2t}{1-t^2} = -1$ (since $2t = t^2 - 1$ by minor rearrangement of the given equation).

(ii) We first work out $\tan 3\theta$. We have

$$\tan 3\theta = \tan(\theta + 2\theta) = \frac{\tan\theta + \tan 2\theta}{1 - \tan\theta\tan 2\theta} = \frac{t + \frac{2t}{1-t^2}}{1 - t\frac{2t}{1-t^2}} = \frac{3t - t^3}{1 - 3t^2}\,,$$

so the equation becomes

$$\frac{3t - t^3}{1 - 3t^2} = t - 2, \quad \text{i.e.} \quad t^3 - 3t^2 + t + 1 = 0.$$

One solution (by inspection) is $t = 1$. Thus one set of roots is given by $\theta = n\pi + \frac{1}{4}\pi$.

There are no other obvious integer roots, but we can reduce the cubic equation to a quadratic equation by dividing out the known factor $(t-1)$. I would start by writing $t^3 - 3t^2 + t + 1 \equiv (t-1)(t^2 + at - 1)$ since the coefficients of t^2 and of t^0 in the quadratic bracket are obvious. Then I would multiply out the brackets to find that $a = -2$. Now we see the connection with part (i): $t^2 + at - 1 = t^2 - 2t - 1 = 0 \Rightarrow \tan 2\theta = -1$, and hence $2\theta = n\pi - \frac{1}{4}\pi$.

The roots are therefore $\theta = n\pi + \frac{1}{4}\pi$ and $\theta = \frac{1}{2}n\pi - \frac{1}{8}\pi$. The multiples of π in the given range are $\{\frac{1}{4}, \frac{3}{8}, \frac{7}{8}, \frac{5}{4}, \frac{11}{8}, \frac{15}{8}\}$.

(iii) For the last part, we could set $\cot\phi = \dfrac{1}{\tan\phi}$ and $\cot 3\phi = \dfrac{1}{\tan 3\phi}$ in the given equation, thereby obtaining

$$\frac{1}{t} = 2 + \frac{1 - 3t^2}{3t - t^3}\,.$$

This simplifies to the cubic equation $t^3 + t^2 - 3t + 1 = 0$. There is an integer root $t = 1$, and the remaining quadratic is $t^2 + 2t - 1 = 0$. Learning from the first part, we write this as $\dfrac{2t}{1-t^2} = 1$, which means that $\tan 2\phi = n\pi + \frac{1}{4}\pi$. Proceeding as before gives (noting the different range) the following multiples of π: $\{\frac{1}{4}, \frac{1}{8}, -\frac{3}{8}, -\frac{3}{4}, -\frac{7}{8}, -\frac{11}{8}\}$.

Post-mortem

There was a small but worthwhile notational point in this question: it is often possible to use the abbreviation t for $\tan$ (or s for $\sin$, etc), which can save a great deal of writing.

There are two other points worth recalling. First is the way that part (i) fed into part (ii), but had to be mildly adapted for part (iii). This is a typical device used in STEP questions, aimed to see how well you learn new ideas. Second is what to do when faced with a cubic equation. There is a formula for the roots of a cubic, but no one knows it nowadays. Instead, you have to find at least one root by inspection. Having found one root, you have a quick look to see if there are any other obvious roots and, if not, then divide out the known factor to obtain a quadratic equation.

The detectives amongst you will have worked out the reason for the peculiar choice $-\frac{3}{2}\pi < \phi < \frac{1}{2}\pi$ for part (iii). The reciprocal relation between $\tan$ and $\cot$ we used at the start of part (iii) is not the only way to relate these two trigonometric functions. We could have instead used $\cot A = \tan(\frac{1}{2}\pi - A)$.

The equation of part (ii) transforms exactly into the equation of part (iii) if we set $\phi = \frac{1}{2}\pi - \theta$. Furthermore, the given range of ϕ corresponds exactly to the range of θ given in part (ii). We can therefore write down the solutions for part (iii) directly from the solutions for part (ii).

Problem 9: Integration by substitution (✓)

Show, by means of a change of variable or otherwise, that

$$\int_0^\infty \mathrm{f}\big((x^2+1)^{\frac{1}{2}}+x\big)\,\mathrm{d}x = \frac{1}{2}\int_1^\infty (1+t^{-2})\mathrm{f}(t)\,\mathrm{d}t\ ,$$

for any given function f .

Hence, or otherwise, show that

$$\int_0^\infty \big((x^2+1)^{\frac{1}{2}}+x\big)^{-3}\,\mathrm{d}x = \frac{3}{8}\ .$$

1998 Paper I

Comments

Note that 'by change of variable' means the same as 'by substitution'.

There are two things to worry about when you are trying to find a change of variable to convert one integral to another: you need to make the integrands match up and you need to make the limits match up. Sometimes, the limits give the clue to the change of variable. (For example, if the limits on the original integral were 0 and 1 and the limits on the transformed integral were 0 and $\frac{1}{4}\pi$, then an obvious possibility would be to make the substitution $t = \tan x$). Here, the change of variable is determined by the integrand, since it must work for all choices of f.

Perhaps you are worried about the infinite upper limit of the integrals. If you are trying to prove some rigorous result about infinite integrals, you might use the definition

$$\int_0^\infty \mathrm{f}\,(x)\,\mathrm{d}x = \lim_{a\to\infty}\int_0^a \mathrm{f}\,(x)\,\mathrm{d}x\ ,$$

but for present purposes you just do the integral and put in the limits. The infinite limit will not normally present problems. For example,

$$\int_1^\infty (x^{-2}+\mathrm{e}^{-x})\mathrm{d}x = \big(-x^{-1}-\mathrm{e}^{-x}\big)\Big|_1^\infty = -\frac{1}{\infty}-\mathrm{e}^{-\infty}+\frac{1}{1}+\mathrm{e}^{-1} = 1+\mathrm{e}^{-1}\ .$$

Don't be afraid of writing $1/\infty = 0$. It is perfectly OK to use this as shorthand for $\lim\limits_{x\to\infty} 1/x = 0$; but ∞/∞, $\infty-\infty$ and $0/0$ are definitely not OK, because of their ambiguity.

Solution to problem 9

Clearly, to get the argument of f right, we must set

$$t = (x^2+1)^{\frac{1}{2}} + x\,.$$

We must check that the new limits are correct (if not, we are completely stuck). When $x = 0$, $t = 1$ as required. Also, $(x^2+1)^{\frac{1}{2}} \to \infty$ as $x \to \infty$, so $t \to \infty$ as $x \to \infty$. Thus the upper limit is still ∞, again as required.

Using the standard method of changing the variable in an integral (basically the chain rule), the integral becomes

$$\int_1^\infty \mathrm{f}\,(t)\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}t$$

so the next task is to find $\dfrac{\mathrm{d}x}{\mathrm{d}t}$. This we can do in two ways: we find x in terms of t and differentiate it; or we could find $\dfrac{\mathrm{d}t}{\mathrm{d}x}$ and turn it upside down. The snag with the second method is that the answer will be in terms of x, so we will have to express x in terms of t anyway—in which case, we may as well use the first method.

We start by finding x in terms of t:

$$t = (x^2+1)^{\frac{1}{2}} + x \Rightarrow (t-x)^2 = (x^2+1) \Rightarrow x = \frac{t^2-1}{2t} = \frac{t}{2} - \frac{1}{2t}\,.$$

Note that for each value of t there is exactly one value of x; we need this for the substitution to work.

Then we differentiate:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{1}{2} + \frac{1}{2t^2} = \frac{1}{2}(1+t^{-2})$$

which is exactly the factor that we require for the transformed integrand given in the question.

For the last part, we take $\mathrm{f}\,(t) = t^{-3}$. Thus

$$\begin{aligned}\int_0^\infty \left((x^2+1)^{\frac{1}{2}} + x\right)^{-3}\mathrm{d}x &= \frac{1}{2}\int_1^\infty (t^{-3}+t^{-5})\,\mathrm{d}t \\ &= \frac{1}{2}\left(\frac{t^{-2}}{-2} + \frac{t^{-4}}{-4}\right)\Bigg|_1^\infty \\ &= \frac{1}{2}\left(\frac{1}{2}+\frac{1}{4}\right) = \frac{3}{8}\,,\ \text{as required.}\end{aligned}$$

Post-mortem

This is not a difficult question conceptually once you realise the significance of the fact that the change of variable must work whatever function f is in the integrand.

There was a useful point connected with calculating $\dfrac{\mathrm{d}x}{\mathrm{d}t}$. It was a good idea not to plunge into the algebra without first thinking about alternative methods; in particular, we might have used (but didn't)

$$\frac{\mathrm{d}x}{\mathrm{d}t} = 1\Big/\frac{\mathrm{d}t}{\mathrm{d}x}\,.$$

Finally, there was the infinite limit of the integrand, which I hope you saw was not something to worry about (even though infinite limits are excluded from the A-level Mathematics specifications). If you were setting up a formal definition of what an integral is, you would have to use finite limits, but if you are merely calculating the value of an integral, you just go ahead and do it, with whatever limits you are given.

Problem 10: True or false (✓✓)

Which of the following statements are true and which are false? Justify your answers.

(i) $a^{\ln b} = b^{\ln a}$ for all positive numbers a and b.

(ii) $\cos(\sin\theta) = \sin(\cos\theta)$ for all real θ.

(iii) There exists a polynomial P such that $|\mathrm{P}(x) - \cos x| \leqslant 10^{-6}$ for all (real) x.

(iv) $x^4 + 3 + x^{-4} \geqslant 5$ for all $x > 0$.

1998 Paper I

Comments

The four parts are (somewhat annoyingly) related only by the fact that you have to decide whether each statement is true or false.

If true, the justification has to be a proof. If false, you could prove that it is false in some general way, but it is nearly always better to find a simple counterexample — as simple as possible.

Part (iii) might look a bit odd. I suppose it relates to the standard approximation $\cos x \approx 1 - \frac{1}{2}x^2$ which holds when x is small. This result can be improved by using a polynomial of higher degree: the next term is $+\frac{1}{4!}x^4$. It can be proved that $\cos x$ can be approximated as accurately as you like for small x by a polynomial of the form

$$\sum_{n=0}^{N}(-1)^n \frac{x^{2n}}{(2n)!}$$

(the truncated Maclaurin expansion). You have to use more terms of the approximation (i.e. a larger value of N) if you want either greater accuracy or larger x. In part (iii) of this question, you are being asked if there is a polynomial such that the approximation is good for all values of x.

I've awarded the question ✓✓ for difficulty, just because quite a few good ideas are required.

Solution to problem 10

(i) True. The easiest way to see this is to log both sides. For the left hand side, we have

$$\ln(a^{\ln b}) = (\ln b)(\ln a)$$

and for the right hand side we have

$$\ln(b^{\ln a}) = (\ln a)(\ln b)\ ,$$

which agree.

Note that we have to be a bit careful with this sort of argument. The argument used is that $A = B$ because $\ln A = \ln B$. This requires the property of the ln function that $\ln A = \ln B \Rightarrow A = B$. You can easily see that this property holds because $\ln$ is a strictly increasing function; if $A > B$, then $\ln A > \ln B$. The same would not hold for (say) $\sin$ (i.e. $\sin A = \sin B \not\Rightarrow A = B$).

(ii) False. $\theta = \frac{1}{2}\pi$ is an easy counterexample. Even though it is 'obvious' we still need to show that $\cos 1 \neq 0$, which we could do by noting that $0 < 1 < \frac{1}{3}\pi$ and sketching the graph of $\cos x$.

(iii) False. Roughly speaking, any polynomial can be made as large as you like by taking x to be very large (provided it is of degree greater than zero), whereas $|\cos x| \leqslant 1$; and there is obviously no polynomial of degree zero (i.e. no constant number) for which the statement holds.

But how can we write this out formally? First let us knock off the case when the polynomial is of degree zero, i.e. a constant, call it P. Then either $P \geqslant \frac{1}{2}$ or $P < \frac{1}{2}$. In either case, P cannot be close to both $\cos 0$ and $\cos \frac{1}{2}\pi$.

Now suppose

$$\mathrm{P}(x) = a_N x^N + \sum_{n=0}^{N-1} a_n x^n \qquad (*)$$

where $N \geqslant 1$ and assume $a_N > 0$. It is enough to show that $\mathrm{P}(x) > 2$ for some value of x. We can find a number x so large that $a_N x^N > N|a_n|x^n + 2$ for each integer n with $0 \leqslant n \leqslant N-1$. The smallest possible value of $\mathrm{P}(x)$ for any given positive x would be achieved if all the coefficients in the sum were negative. Thus

$$\mathrm{P}(x) \geqslant a_N x^N - \sum_{n=0}^{N-1} |a_n|x^n > a_N x^N + (2 - a_N x^N) = 2$$

and we are done.

(iv) True: $x^4 + 3 + x^{-4} = (x^2 - x^{-2})^2 + 5 \geqslant 5$.

Post-mortem

The important point here is that if you want to show a statement is true, you have to give a formal proof, whereas if you want to show that it is false, you only need give one counterexample. It does not have to be an elaborate counterexample — in fact, the simpler the better.

My proof for part (iii) is more elaborate than could have been expected of candidates in the examination. A sketch of a polynomial of degree N would have been enough, provided the special case of $N = 0$ was dealt with separately.

Problem 11: Egyptian fractions (✓)

A number of the form $\frac{1}{N}$, where N is an integer greater than 1, is called a *unit fraction*. Noting that

$$\frac{1}{2} = \frac{1}{3} + \frac{1}{6} \text{ and } \frac{1}{3} = \frac{1}{4} + \frac{1}{12},$$

guess a general result of the form

$$\frac{1}{N} = \frac{1}{a} + \frac{1}{b} \qquad (*)$$

and hence prove that any unit fraction can be expressed as the sum of two distinct unit fractions.

By writing $(*)$ in the form

$$(a - N)(b - N) = N^2$$

and by considering the factors of N^2, show that if N is prime, then there is only one way of expressing $\frac{1}{N}$ as the sum of two distinct unit fractions.

Prove similarly that any fraction of the form $\frac{2}{N}$, where N is prime number greater than 2, can be expressed uniquely as the sum of two distinct unit fractions.

2000 Paper II

Comments

Fractions written as the sum of unit fractions are called *Egyptian fractions*: they were used by Egyptians. The earliest record of such use is 1900BC. The Rhind papyrus in the British Museum gives a table of representations of fractions of the form $\frac{2}{n}$ as sums of unit fractions for all odd integers n between 5 and 101 — a remarkable achievement when you consider that algebra was 3,500 years in the future.

It is not clear why Egyptians represented fractions this way; maybe it just seemed a good idea at the time. Certainly the notation they used, in which for example $\frac{1}{n}$ was denoted by n with an oval on top, does not lend itself to generalisation to fractions that are not unit. One of the rules for expressing non-unit fractions in terms of unit fractions was that all the unit fractions in should be distinct, so $\frac{2}{7}$ had to be expressed as $\frac{1}{4} + \frac{1}{28}$ instead of as $\frac{1}{7} + \frac{1}{7}$, which seems pretty daft.

It is not obvious that every fraction can be express in Egyptian form; this was proved by Fibonacci in 1202. There are however still many unsolved problems relating to Egyptian fractions.

Egyptian fractions have been called a wrong turn in the history of mathematics; if so, it was a wrong turn that favoured style over utility; no bad thing, in my opinion.

Solution to problem 11

A good guess would be that the first term of the decomposition of $\dfrac{1}{N}$ is $\dfrac{1}{N+1}$, i.e. $a = N + 1$. In that case, the other term is $\dfrac{1}{N} - \dfrac{1}{N+1}$ i.e. $b = N(N+1)$. That proves the result that every unit fraction can be expressed as the sum of two unit fractions.

The only factors of N^2 (since N is prime) are 1, N and N^2. The possible factorisations of N^2 are therefore $N^2 = 1 \times N^2$ or $N^2 = N \times N$. We discount $N \times N$, since this would lead to $a = b$, and this is ruled out in the question. That leaves only $a - N = 1$ and $b - N = N^2$ (or the other way round). Thus $N + 1$ and $N^2 + N$ are the only possible values for a and b and the decomposition is unique.

For the second half, set

$$\frac{2}{N} = \frac{1}{a} + \frac{1}{b}$$

where $a \neq b$. We need to aim for an equation with N^2 on one side, so that we can use the method of the first part. We have $ab - \frac{1}{2}(a+b)N = 0$ which we write as

$$\left(a - \frac{N}{2}\right)\left(b - \frac{N}{2}\right) = \frac{N^2}{4} \quad \text{i.e.} \quad (2a - N)(2b - N) = N^2 \,.$$

Thus $2a - N = N^2$ and $2b - N = 1$ (or the other way round). The decomposition is therefore unique and given by

$$\frac{2}{N} = \frac{1}{\frac{1}{2}(N^2 + N)} + \frac{1}{\frac{1}{2}(N+1)} \,.$$

The only possible fly in the ointment is the $\frac{1}{2}$ in the denominators: a and b are supposed to be integers. However, all prime numbers greater than 2 are odd, so $N + 1$ and $N^2 + N$ are both even and the denominators are indeed integers.

Post-mortem

Another way of getting the last part (less systematically) would have been to notice that

$$\frac{1}{N} = \frac{1}{(N+1)} + \frac{1}{N(N+1)} \implies \frac{2}{N} = \frac{1}{\frac{1}{2}(N+1)} + \frac{1}{\frac{1}{2}N(N+1)}$$

which gives the result immediately. How might you have noticed this? Well, I noticed it by trying to work out some examples, starting with what is given at the very beginning of the question. If $\dfrac{1}{3} = \dfrac{1}{4} + \dfrac{1}{12}$ then $\dfrac{2}{3} = \dfrac{2}{4} + \dfrac{2}{12}$ which works.

Of course, the advantage of the systematic approach is that it allows for generalisation: what happens if N is odd but not prime; what happens if the numerator is 3 not 2?

Problem 12: Maximising with constraints (✓)

> Prove that the rectangle of greatest perimeter which can be inscribed in a given circle is a square. The result changes if, instead of maximising the sum of lengths of sides of the rectangle, we seek to maximise the sum of nth powers of the lengths of those sides for $n \geqslant 2$. What happens if $n = 2$? What happens if $n = 3$? Justify your answers.

1998 Paper I

Comments

Obviously, the perimeter of a *general* rectangle has no maximum (it can be as long as you like), so the key to this question is to use the constraint that the rectangle lies in the circle.

The vertices of a rectangle of greatest perimeter must lie *on* the circle; this seems obvious, but you should devote the first sentence of your solution to justifying it. And draw a diagram (I didn't because there wasn't enough room on the page).

The word 'greatest' in the question immediately suggests differentiating something. The perimeter of a rectangle is expressed in terms of two variables, length x and breadth y, so you must find a way of using the constraint to eliminate one variable. This could be done in two ways: express y in terms of x, or express both x and y in terms of another variable (an angle, say).

Finally, there is the matter of deciding whether your solution is the greatest or least value (or neither). This can be done either by considering second derivatives, or by trying to understand the different situations. Often the latter method, or a combination, is preferable.

The theory of stationary values of functions subject to constraints is very important: it has widespread applications (for example, in theoretical physics, financial mathematics and in fact in almost any area to which mathematics is applied). It forms a whole branch of mathematics called optimisation. Normally, the methods described above (eliminating one variable) are not used: a clever idea, the method of Lagrange multipliers, is used instead.

Solution to problem 12

The vertices of the rectangle must lie on the circle. Suppose not. Then two adjacent vertices (at least) lie in the interior of the circle. We could then increase the perimeter by extending two sides beyond these vertices. [You should include a diagram here!]

Let the circle have diameter d and let the length of one side of the rectangle be x and the length of the adjacent side be y.

Then, by Pythagoras's theorem,

$$y = \sqrt{d^2 - x^2} \tag{†}$$

and the perimeter P is given by

$$P = 2x + 2\sqrt{d^2 - x^2}\ .$$

We can find the largest possible value of P as x varies by calculus. We have

$$\frac{\mathrm{d}P}{\mathrm{d}x} = 2 - 2\frac{x}{\sqrt{d^2 - x^2}}\ ,$$

so for a stationary point we require (cancelling the factor of 2 and squaring)

$$1 = \frac{x^2}{d^2 - x^2} \qquad \text{i.e.} \qquad 2x^2 = d^2\ .$$

Thus $x = d/\sqrt{2}$ (ignoring $x = -d/\sqrt{2}$ for obvious reasons). Substituting this into (†) gives $y = d/\sqrt{2}$, so the rectangle is indeed a square, with perimeter $2\sqrt{2}d$.

But is this the maximum perimeter? The easiest way to investigate is to calculate the second derivative, which is easily seen to be negative for all values of x, and in particular when $x = d/\sqrt{2}$. The stationary point is certainly a maximum. Alternatively, we could argue that the rectangle for which $x = 0$ has perimeter $2d$ which is less than $2\sqrt{2}d$ so $x = d/\sqrt{2}$ cannot correspond to the minimum perimeter.

For the second part, we consider

$$\mathrm{f}(x) = x^n + (d^2 - x^2)^{\frac{n}{2}}\ .$$

The first thing to notice is that f is constant if $n = 2$, so in this case, the largest (and smallest) value is d^2. You can use Pythagoras's theorem to see why this result holds.

For $n = 3$, we have

$$\mathrm{f}'(x) = 3x^2 - 3x(d^2 - x^2)^{\frac{1}{2}},$$

so $\mathrm{f}(x)$ is stationary when $x^4 = x^2(d^2 - x^2)$, i.e. when $2x^2 = d^2$ as before or when $x = 0$. The corresponding stationary values of f are $\sqrt{2}d^3$ and $2d^3$, so this time the largest value occurs when $x = 0$.

Post-mortem

You will almost certainly want to investigate the situation for other values of n yourself, just to see what happens.

Were you happy with the proof given above that the square has the *largest* perimeter? And isn't it a bit odd that we did not discover a stationary point corresponding to the smallest value (which can easily be seen to occur at $x = 0$ or $x = d$)? To convince yourself that the maximum point gives the largest value, it is a good idea to sketch a graph: you can check that $\mathrm{d}P/\mathrm{d}x$ is positive for $0 < x < d/\sqrt{2}$ and negative for $d/\sqrt{2} < x < d$. The smallest values of the perimeter occur at the endpoints of the interval $0 \leqslant x \leqslant d$ and therefore do not have to be turning points.

A much better way of tackling the problem is to set $x = d\cos\theta$ and $y = d\sin\theta$ and find the perimeter as a function of θ. Because θ can take any value (there are no end points such as $x = 0$ to consider), the largest and smallest perimeters correspond to stationary points. Try it.

Problem 13: Binomial expansion (✓)

(i) Use the first four terms of the binomial expansion of $\left(1 - \frac{1}{50}\right)^{\frac{1}{2}}$ to derive the approximation $\sqrt{2} \approx 1.414214$.

(ii) Calculate similarly an approximation to the cube root of 2 to six decimal places by considering $\left(1 + \frac{N}{125}\right)^{\frac{1}{3}}$, where N is a suitable number.

[You need not justify the accuracy of your approximations.]

1998 Paper II

Comments

Although you do not have to justify your approximations, you do need to think carefully about the number of terms required in the expansions. You first have to decide how many you will need to obtain the given number of decimal places; then you have to think about whether the next term is likely to affect the value of the last decimal.

It is not at all obvious where the $\sqrt{2}$ in part (i) comes from until you write $\left(1 - \frac{1}{50}\right)^{\frac{1}{2}} = \sqrt{\frac{49}{50}}$.

For part (ii), you have to choose N in such a way that $125 + N$ has something to do with a power of 2.

You can make the arithmetic of the binomial expansions a bit easier by arranging the denominator to be a power of 10 so that, for example, the expansion in part (i) becomes $\left(1 - \frac{2}{100}\right)^{\frac{1}{2}}$. No need to use a calculator.

Solution to problem 13

(i) First we expand binomially:

$$\begin{aligned}
&\left(1-\frac{2}{100}\right)^{\frac{1}{2}}\\
=\ & 1+\left(\frac{1}{2}\right)\left(-\frac{2}{100}\right)+\left(\frac{1}{2!}\right)\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{2}{100}\right)^2-\left(\frac{1}{3!}\right)\left(\frac{1}{2}\right)\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{2}{100}\right)^3+\cdots\\
\approx\ & 1-\frac{1}{100}-\frac{5}{10^5}-\frac{5}{10^7}=0.9899495\,.
\end{aligned}$$

It is clear that the next term in the expansion would introduce the eighth decimal place, which it seems we do not need. Of course, after further manipulations we might find that the above calculation does not supply the 6 decimal places we need for $\sqrt{2}$, in which case we will have to work out the next term in the expansion.

But $\left(\dfrac{98}{100}\right)^{\frac{1}{2}}=\dfrac{7\sqrt{2}}{10}$, so $\sqrt{2}\approx 9.899495/7\approx 1.414214\,.$

(ii) We have

$$\begin{aligned}
&\left(1+\frac{3}{125}\right)^{\frac{1}{3}}\\
=\ & 1+\left(\frac{1}{3}\right)\left(\frac{3}{125}\right)+\left(\frac{1}{2!}\right)\left(\frac{1}{3}\right)\left(-\frac{2}{3}\right)\left(\frac{3}{125}\right)^2+\left(\frac{1}{3!}\right)\left(\frac{1}{3}\right)\left(-\frac{2}{3}\right)\left(-\frac{5}{3}\right)\left(\frac{3}{125}\right)^3+\cdots\\
\approx\ & 1+\frac{8}{1000}-\frac{64}{10^6}+\frac{5}{3}\frac{8^3}{10^9}\\
=\ & 1.007936+\frac{256}{3}\frac{1}{10^8}=1.007937\,.
\end{aligned}$$

Successive terms in the expansion decrease by a factor of about 1000, so this should give the right number of decimal places.

But $\left(\dfrac{128}{125}\right)^{\frac{1}{3}}=\dfrac{4\sqrt[3]{2}}{5}=\dfrac{8\sqrt[3]{2}}{10}$ so $\sqrt[3]{2}\approx\dfrac{10.0793}{8}\approx 1.259921\,.$

Post-mortem

This question required a bit of intuition, and some accurate arithmetic. You don't have to be brilliant at arithmetic to be a good mathematician, but most mathematicians aren't bad at it. There have in the past been children who were able to perform extraordinary feats of arithmetic. For example, Zerah Colburn, a 19th-century American, toured Europe at the age of 8. He was able to multiply instantly any two 4-digit numbers given to him by the audiences. George Parker Bidder (the Calculating Boy) could perform similar feats, though unlike Colburn he became a distinguished mathematician and scientist. One of his brothers knew the bible by heart.

The error in the approximation is the weak point of this question. Although it is clear that the next term in the expansion is too small to affect the accuracy, it is not obvious that the sum of all the next hundred (say) terms of the expansion is negligible (though in fact it is). What is needed is an estimate of the truncation error in the binomial expansion. Such an estimate is not hard to obtain (first-year university work) and is typically of the same order of magnitude as the first neglected term in the expansion. Without this estimate, the approximation is not justified.

Problem 14: Sketching subsets of the plane (✓✓)

Sketch the following subsets of the x-y plane:

(i) $|x| + |y| \leqslant 1$;

(ii) $|x - 1| + |y - 1| \leqslant 1$;

(iii) $|x - 1| - |y + 1| \leqslant 1$;

(iv) $|x|\,|y - 2| \leqslant 1$.

1999 Paper I

Comments

Often with modulus signs, it is easiest to consider the cases separately, so for example in part (i), you would first work out the case $x > 0$ and $y > 0$, then $x > 0$ and $y < 0$, and so on. Here there is a simple geometric understanding of the different cases: once you have worked out the first case, the other three can be deduced by symmetry.

Another geometric idea should be in your mind when tackling this question, namely the idea of translations in the plane.

Post-mortem

There was no room for a post-mortem over the page, because the diagrams take up so much space.

Don't read this until you have tried the question!

There are two key learning points, if you will excuse this horrible expression.

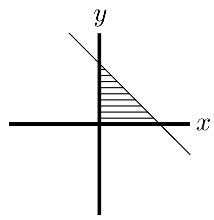
The first is that if you want to sketch a region, it is often best to draw the curves that define the boundary of the region, then just work out whether you want the interior or exterior of the boundary by choosing one interior point and seeing whether it satisfies the inequalities.

The second is that quite complicated inequalities can sometimes be much simplified by translating or rotating the axes.

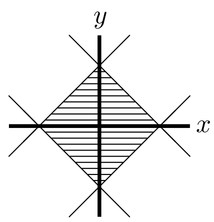
Solution to problem 14

The way to deal with the modulus signs in this question is to consider first the case when the things inside the modulus signs are positive, and then get the full picture by symmetry, shifting the origin as appropriate.

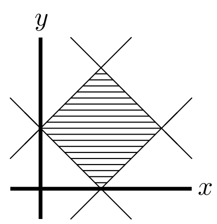
For part (i), consider the first quadrant $x \geqslant 0$ and $y \geqslant 0$. In this quadrant, the inequality is $x+y \leqslant 1$. Draw the line $x+y=1$ and then decide which side of the line is described by the inequality. It is obviously (since x has to be smaller than something) the region to the left of the line; or (since y also has to be smaller than something) the region below the line, which is the same region.



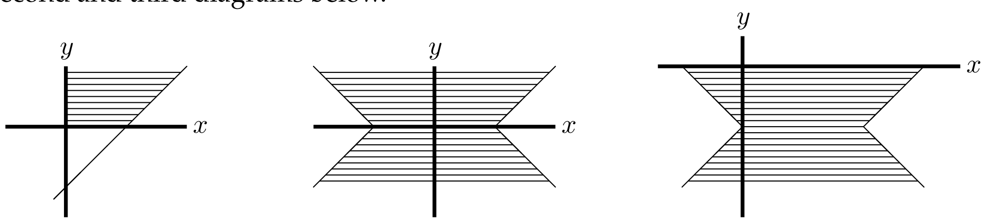
Similar arguments could be used in the other quadrants, but it is easier to note that the inequality $|x|+|y| \leqslant 1$ is unchanged when x is replaced by $-x$, or y is replaced by $-y$, so the sketch should have reflection symmetry in both axes, as shown in the diagram on the right.



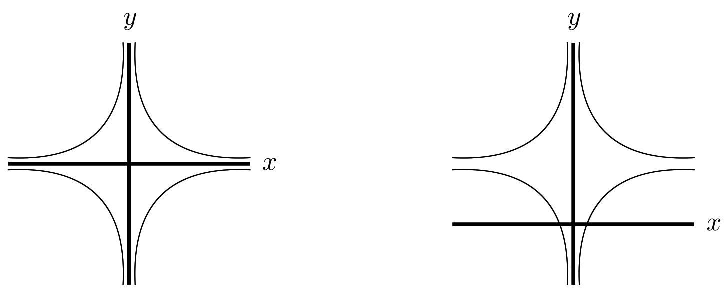
Part (ii) is the same as part (i), except for a translation. The origin of (i) is now at $(1,1)$.



For part (iii), consider first $x-y \leqslant 1$, instead of $|x-1|-|y+1| \leqslant 1$. For $x>0$ and $y>0$, this gives a region that is infinite in extent in the positive y direction, as shown in the first diagram below. Then reflect this in both axes and translate one unit down the y axis and 1 unit along the positive x axis as shown in the second and third diagrams below.



For part (iv), consider first the rectangular hyperbola $xy=1$ reflected in both axes to give four hyperbolas, as in the first diagram below. Then translate the four hyperbolas translated 2 units up the y axis as in the final diagram. The region required is enclosed by the four hyperbolas.



Problem 15: More sketching subsets of the plane (✓✓)

(i) Show that

$$x^2 - y^2 + x + 3y - 2 = (x - y + 2)(x + y - 1)$$

and hence, or otherwise, indicate by means of a sketch the region of the x-y plane for which

$$x^2 - y^2 + x + 3y > 2.$$

(ii) Sketch also the region of the x-y plane for which

$$x^2 - 4y^2 + 3x - 2y < -2.$$

(iii) Give the coordinates of a point for which both inequalities are satisfied or explain why no such point exists.

1995 Paper I

Comments

This question gets a ✓✓ difficulty rating because inequalities always need to be handled with care.

For the very first part, you could either factorise the left hand side to obtain the right hand side, or multiply out the right hand side to get the left hand side. Obviously, it is much easier to do the multiplication than the factorisation. But is that 'cheating' or taking a short cut that might lose marks? No, it is not cheating: it doesn't matter if you start from the given answer and work backwards — it is still a mathematical proof and any proof will get the marks. (But note that if there is a 'hence' in the question, you will lose marks if you do not use the result or results that you have just proved.)

In part (ii), you have to do the factorisation yourself, so you should look carefully at where the terms in the (very similar) first part came from. It will help to spot the similarities between $x^2 - y^2 + x + 3y - 2$ and $x^2 - 4y^2 + 3x - 2y + 2$.

Since no indication is given as to what detail should appear on the sketch, you have to use your judgement: it is clearly important to know where the regions lie relative to the coordinate axes.

Post-mortem

As in the previous sketching question, the solution is so long that there is no room for a post-mortem after the answer.

Don't read this before having tried the question!

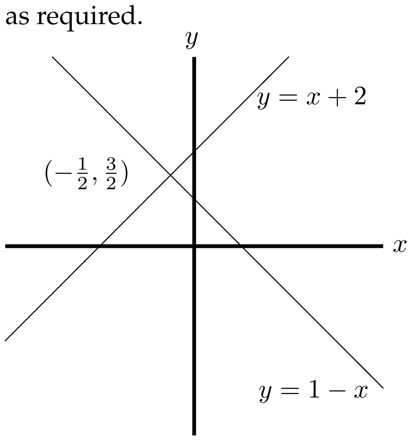
The answer given overleaf is inadequate in two ways, each of which would have probably lost me marks. First, as in the previous question, the answer should definitely have referred to the boundaries. Here, the inequalities are *strict*, which means that the boundary lines are *excluded* from the required regions. Second, justification should have been provided for the claim that the point $(1, 2)$ satisfies both inequalities. This could be provided either by simply indicating the position of the point on both sketches or algebraically by substitution into the inequalities.

Solution to problem 15

To do the multiplication, it pays to be systematic and to set out the algebra nicely:

$$\begin{aligned}(x-y+2)(x+y-1) &= x(x+y-1)-y(x+y-1)+2(x+y-1)\\ &= x^2+xy-x-yx-y^2+y+2x+2y-2\\ &= x^2+x-y^2+3y-2\end{aligned}$$

as required.



For the first inequality, we need both $x-y+2$ and $x+y-1$ either to be positive or to be negative. To sort out the inequalities (or inequations as they are sometimes horribly called), the first thing to do is to draw the lines defined by the corresponding equalities. These lines divide the plane into four regions and we then have to decide which regions are relevant.

The diagonal lines $x-y=-2$ and $x+y=1$ intersect at $(-\frac{1}{2},\frac{3}{2})$; this is the important point to mark on the sketch. The required regions are the left and right quadrants formed by these diagonal lines (since the inequalities mean that the regions are to the right of both lines or to the left of both lines).

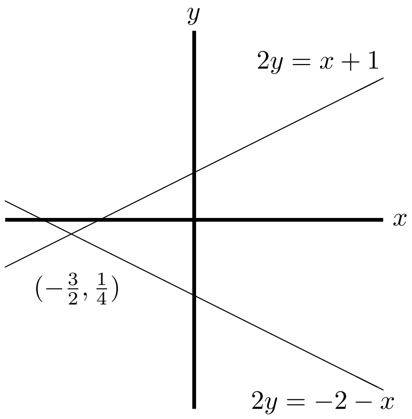
For part (ii), the first thing to do is to factorise $x^2-4y^2+3x-2y+2$. The key similarity with the first part is the absence of 'cross' terms of the form xy. This allows a difference-of-two-squares factorisation of the first two terms: $x^2-4y^2=(x-2y)(x+2y)$. Following the pattern of the first part, we can then try a factorisation of the form

$$x^2-4y^2+3x-2y+2=(x-2y+a)(x+2y+b)$$

where $ab=2$. Considering the terms linear in x and y gives $a+b=3$ and $2a-2b=-2$ which quickly leads to $a=1$ and $b=2$. Note that altogether there were three equations for a and b, so we had no right to expect a consistent solution (except for the fact that this is a STEP question for which we had every right to believe that the first part would guide us through the second part).

Alternatively, we could have completed the square in x and in y and then used difference of two squares:

$$\begin{aligned}x^2-4y^2+3x-2y+2 &= (x+\tfrac{3}{2})^2-\tfrac{9}{4}-(2y+\tfrac{1}{2})^2+\tfrac{1}{4}+2\\ &= (x+\tfrac{3}{2})^2-(2y+\tfrac{1}{2})^2=(x+\tfrac{3}{2}-2y-\tfrac{1}{2})(x+\tfrac{3}{2}+2y+\tfrac{1}{2})\,.\end{aligned}$$



As in the previous case, the required region is formed by two intersecting lines; this time, they intersect at $(-\frac{3}{2},-\frac{1}{4})$ and the upper and lower regions are required, since the inequality is the other way round.

It is easy to see from the sketches that there are points that satisfy both inequalities: for example $(1,2)$.

Problem 16: Non-linear simultaneous equations (✓)

Consider the system of equations

$$\begin{aligned} 2yz + zx - 5xy &= 2 \\ yz - zx + 2xy &= 1 \\ yz - 2zx + 6xy &= 3\,. \end{aligned}$$

Show that

$$xyz = \pm 6$$

and find the possible values of x, y and z.

1996 Paper II

Comments

At first sight, this looks forbidding. A closer look reveals that the variables x, y and z occur only in pairs yz, zx and xy. The problem therefore boils down to solving three simultaneous equations in the variables yz, zx and xy, then using the solution to find x, y and z individually.

What do you make of the equation $xyz = \pm 6$? How could the $\pm$ arise?

There are two ways of tackling such simultaneous equations. You could use the first equation to find an expression for one variable (yz say) in terms of the other two variables, then substitute this into the other equations to eliminate yz from the system. Then use the second equation (in its new form) to find an expression for one of the two remaining variables (zx say), and substitute this into the third equation (in its new form) to obtain an equation for the third variable (xy). Having solved this equation, you can substitute back to find the other variables. This method is called *Gauss elimination*.

Alternatively, you could eliminate one variable (yz say) from the first two of equations by multiplying the first equation by something suitable and the second equation by something suitable and subtracting. You then eliminate yz from the second and third equations similarly. That leaves you with two simultaneous equations in two variables which you can solve by your favourite method.

There is another way of solving the simultaneous equations, which is better in theory than in practice. You write the equations in matrix form $\mathbf{Mx} = \mathbf{c}$, where in this case

$$\mathbf{M} = \begin{pmatrix} 2 & 1 & -5 \\ 1 & -1 & 2 \\ 1 & -2 & 6 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} yz \\ zx \\ xy \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}.$$

The solution is then $\mathbf{x} = \mathbf{M}^{-1}\mathbf{c}$. All you have to do is invert a 3×3 matrix, which is doable but not very pleasant. You might like to try it if you know the formula for the inverse of a matrix.

Solution to problem 16

Start by labelling the equations:

$$2yz + zx - 5xy = 2 , \tag{1}$$
$$yz - zx + 2xy = 1 , \tag{2}$$
$$yz - 2zx + 6xy = 3 . \tag{3}$$

We use Gaussian elimination. Rearranging equation (1) gives

$$yz = -\tfrac{1}{2}zx + \tfrac{5}{2}xy + 1 , \tag{4}$$

which we substitute back into equations (2) and (3) :

$$-\tfrac{3}{2}zx + \tfrac{9}{2}xy = 0 , \tag{5}$$
$$-\tfrac{5}{2}zx + \tfrac{17}{2}xy = 2 . \tag{6}$$

Thus $zx = 3xy$ (using equation (5)). Substituting into equation (6) gives $xy = 2$ and $zx = 6$. Finally, substituting back into equation (1) shows that $yz = 3$.

The question is now plain sailing. Multiplying the three values together gives $(xyz)^2 = 36$ and taking the square root gives $xyz = \pm 6$ as required.

Now it remains to solve for x, y and z individually. We know that $yz = 3$, so if $xyz = +6$ then $x = +2$, and if $xyz = -6$ then $x = -2$. The solutions are therefore either $x = +2$, $y = 1$, and $z = 3$ or $x = -2$, $y = -1$, and $z = -3$.

Post-mortem

There were two key observations which allowed us to do this question quite easily. Both came from looking carefully at the question. The first was that the given equations, although non-linear in x, y and z (they are quadratic, since they involve products of these variables) could be thought of as three linear equations in yz, zx and xy. That allowed us to make a start on the question. The second observation was that the equation $xyz = \pm 6$ is almost certain to come from $(xyz)^2 = 36$ and that gave us the next step after solving the simultaneous equations. (Recall the next step was to multiply all the variables together.)

There was a point of technique in the solution: it is often very helpful in this sort of problem (and many others) to number your equations. This allows you to refer back clearly and quickly, for your benefit as well as for the benefit of your readers.

The three simultaneous quadratic equations (1) – (3) have a geometric interpretation but it is not at all obvious. The equations are quadratic in the variables x, y and z, which means that each equation represents either an ellipsoid or a hyperboloid (or some special cases).[16] The solutions of all three equations represent the points of intersection of the three surfaces. Not very easy to picture.

[16] An *ellipsoid* is roughly the shape of the surface of a rugby ball, or of the giant galaxy ESO 325-G004. A *hyperboloid* can either be the shape of an infinite radar dish (in fact, a pair of such dishes) or it can be the shape of a power station cooling tower. Our equations in fact represent hyperboloids of the cooling-tower type.

Problem 17: Inequalities (✓✓)

Solve the inequalities

(i) $1 + 2x - x^2 > \dfrac{2}{x} \qquad (x \neq 0)$,

(ii) $\sqrt{3x+10} > 2 + \sqrt{x+4} \qquad (x \geqslant -10/3)$.

2001 Paper I

Comments

The two parts are unrelated (unusually for STEP questions), except that they deal with inequalities. Both parts need care, since they contain traps for the unwary.

In part (i) you have to watch out when you multiply an inequality: if the thing you multiply by is negative then the inequality reverses. You might find sketching a graph helpful.

In part (ii) you have to consider the possibility that your algebraic manipulations have created extra spurious solutions. It is worth (after you have finished the question) sketching the graphs of $\sqrt{3x+10}$ and $2 + \sqrt{x+4}$ just to see what is going on. You can get the latter graph by translations of $\sqrt{x}$ (and you can get $y = \sqrt{x}$ by reflecting $y = x^2$ in the line $y = x$). I would have drawn them for you in the post-mortem overleaf had there been room.

In fact, a solution relying on sketches for part (ii) is probably preferable to my solution. With the sketches to hand, you only have to solve the equation $\sqrt{3x+10} = 2 + \sqrt{x+4}$ and then look at your sketches to see what range of values of x you need, thereby saving much anguish.[17]

[17] Now I look at it again, it seems to me that the last sentence of my solution ('Therefore the inequality holds ...') is a bit suspect without sketches to show that the inequality actually *does* hold.

Solution to problem 17

(i) We would like to multiply both sides of the inequality by x in order to obtain a nice cubic expression. However, we have to allow for the possibility that x is negative (which would reverse the inequality). One way of dealing with this is to consider the cases $x > 0$ and $x < 0$ separately.

For $x > 0$, we multiply the whole equation by x without changing the direction of the inequality:

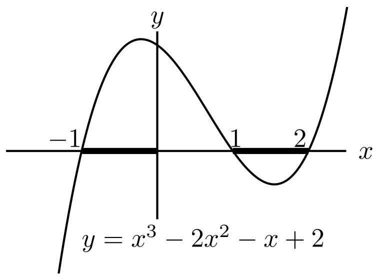
$$\begin{aligned} 1 + 2x - x^2 > \frac{2}{x} &\Rightarrow x + 2x^2 - x^3 > 2 \\ &\Rightarrow x^3 - 2x^2 - x + 2 < 0 \\ &\Rightarrow (x-1)(x+1)(x-2) < 0 \\ &\Rightarrow 1 < x < 2, \end{aligned} \qquad (*)$$

discarding the possibility $x < -1$, since we have assumed that $x > 0$.

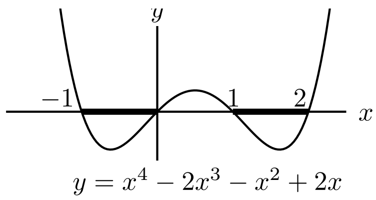
The easiest way of obtaining the result $(*)$ from the previous line is to sketch the graph of $(x-1)(x+1)(x-2) < 0$.

For $x < 0$, we must reverse the inequality when we multiply by x so in this case, $(x-1)(x+1)(x-2) > 0$, which gives $-1 < x < 0$.

In the sketch on the right, the values of x for which the inequality holds are shown as bold lines.



The smart way to do $x > 0$ and $x < 0$ in one step is to multiply by x^2 (which is never negative and hence never changes the direction of the inequality) and analyse $x(x-1)(x+1)(x-2) < 0$. Again, a sketch is useful.



(ii) First square both sides of the inequality $\sqrt{3x+10} > 2 + \sqrt{x+4}$:

$$3x + 10 > 4 + 4\sqrt{x+4} + (x+4) \quad \text{i.e.} \quad x + 1 > 2\sqrt{x+4}\,.$$

Note that both sides of the original inequality are positive or zero (i.e. non-negative), so the direction of the inequality is not changed by squaring. Now consider the new inequality $x + 1 > 2\sqrt{x+4}$. If both sides are non-negative, that is if $x > -1$, we can square both sides again without changing the direction of the inequality. But, if $x < -1$, the inequality cannot be satisfied since the right-hand side is always non-negative. Squaring gives

$$x^2 + 2x + 1 > 4(x+4) \quad \text{i.e.} \quad (x-5)(x+3) > 0\,.$$

Thus, $x > 5$ or $x < -3$. However, we must reject $x < -3$ because of the condition $x > -1$. Therefore the inequality holds for $x > 5$.

Post-mortem

The spurious result $x < -3$ at the end of part (ii) arises through loss of information in the process of squaring: if you square an expression, you lose its sign. After squaring twice, the resulting inequality is the same as would have resulted from $\sqrt{(3x+10)} < 2 - \sqrt{(x+4)}$ and it is to this inequality that $-\frac{10}{3} \leqslant x < -3$ is a solution.

Problem 18: Inequalities from cubics (✓✓)

(i) Sketch, without calculating the stationary points, the graph of the function $\mathrm{f}(x)$ given by

$$\mathrm{f}(x) = (x-p)(x-q)(x-r) ,$$

where $p < q < r$. By considering the quadratic equation $\mathrm{f}'(x) = 0$, or otherwise, show that

$$(p+q+r)^2 > 3(qr+rp+pq) .$$

(ii) By considering $(x^2+gx+h)(x-k)$, or otherwise, show that $g^2 > 4h$ is a sufficient condition but not a necessary condition for the inequality

$$(g-k)^2 > 3(h-gk)$$

to hold.

2001 Paper I

Comments

The idea behind the first part is to obtain an inequality by considering a certain graph. In the second part, we again use graphs to obtain an inequality, which this time holds subject to a different inequality.

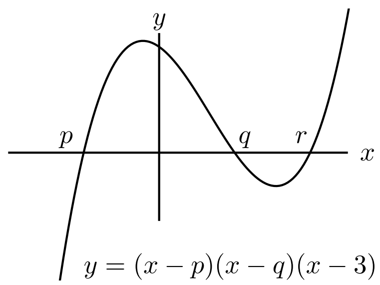
For the sketch in the first part it is not necessary to do more than think about the behaviour for x large and positive, and for x large and negative, and the points at which the graph crosses the x-axis.

For the second part, you should also have a sketch or sketches in mind. The argument is very similar to that of the first part, except that it also tests understanding of the meaning of the terms necessary and sufficient. For this, it is probably best to use $\Rightarrow$ notation.

To show that $g^2 > 4h$ is not a *necessary* condition for the inequality $(g-k)^2 > 3(h-gk)$ to hold, you just have to give an example (the simpler the better) for which $(g-k)^2 > 3(h-gk)$ but $g^2 \leqslant 4h$.

Solution to problem 18

From the sketch, we see that $\mathrm{f}(x)$ has two turning points, so the equation $\mathrm{f}'(x) = 0$ has two real roots.
Now



$$\begin{aligned}\mathrm{f}(x) &= (x-p)(x-q)(x-r)\\ &= x^3 - (p+q+r)x^2 + (qr+rp+pq)x - pqr\end{aligned}$$

so at a turning point

$$\mathrm{f}'(x) = 3x^2 - 2(p+q+r)x + (qr+rp+pq) = 0\ .$$

Using the condition '$b^2 > 4ac$' for this quadratic to have two real roots gives the required result:

$$4(p+q+r)^2 > 12(qr+rp+pq)\ .$$

For the second part, we can use a similar argument. First note that if the quadratic equation $x^2+gx+h = 0$ has two distinct real roots then the cubic equation $(x^2+gx+h)(x-k) = 0$ has three roots, at least two of which are distinct. Thus

$$\begin{aligned}g^2 > 4h &\Rightarrow (x^2+gx+h)(x-k) = 0 \text{ has at least two distinct roots}\\ &\Rightarrow (x^2+gx+h)(x-k) \text{ has at least two distinct turning points (draw graphs!)}\\ &\Rightarrow x^3 + (g-k)x^2 + (h-gk)x - hk \text{ has at least two distinct turning points}\\ &\Rightarrow 3x^2 + 2(g-k)x + (h-gk) = 0 \text{ has two distinct roots}\\ &\Rightarrow 4(g-k)^2 > 12(h-gk) \text{ as required.}\end{aligned}$$

Thus $g^2 > 4h$ is a sufficient condition for this inequality to hold.

Why is it not necessary? That is to say, why is it too strong a condition? The reason is that both turning points could be above the x–axis or both could be below (not one on either side which was the origin of the inequality). Saying this would get all the marks. Or you could give a counterexample: for example, $g = 2$, $h = 1$ and $k = 1000$.

Post-mortem

When I set this question, I got considerable satisfaction from the way that seemingly obscure inequalities are derived from understanding simple graphs and the quadratic formula. It was only when preparing this book that I realised that they are, disappointingly, not at all obscure.

The inequality of the first part is equivalent to $(q-r)^2+(r-p)^2+(p-q)^2 > 0$, the inequality coming from the fact that squares of real numbers are non-negative and in this case the given condition $p < q < r$ means that the expression is strictly positive (not equal to zero).

If we rewrite the second inequality as $(k + \frac{1}{2}g)^2 + 3(\frac{1}{4}g^2 - h) > 0$ we see that it certainly holds if $(\frac{1}{4}g^2 - h) > 0$, as expected. This is not a necessary condition: it also holds if $(\frac{1}{4}g^2 - h) < 0$ provided k is large enough.

Problem 19: Logarithms (✓✓✓)

To nine decimal places, $\log_{10} 2 = 0.301029996$ and $\log_{10} 3 = 0.477121255$.

(i) Calculate $\log_{10} 5$ and $\log_{10} 6$ to three decimal places. By taking logs, or otherwise, show that

$$5 \times 10^{47} < 3^{100} < 6 \times 10^{47}.$$

Hence write down the first digit of 3^{100}.

(ii) Find the first digit of each of the following numbers: 2^{1000}; $2^{10\,000}$; and $2^{100\,000}$.

2000 Paper I

Comments

This nice little question shows why it is a good idea to ban calculators from some mathematics examinations — though I notice that my calculator can't work out 2^{10000}.

When I was at school, before electronic calculators were invented, we had to spend quite a lot of time in year 8 (I think) doing extremely tedious calculations by logarithms. We were provided with a book of tables of four-figure logarithms and lots of uninteresting numbers to multiply or divide. The book also had tables of trigonometric functions. When it came to antilogging, to get the answer, we had to use the tables backwards, since no tables of inverse logarithms (exponentials) were provided.

A logarithm (to base 10, as always in this question) consists of two parts: the *characteristic* which is the number before the decimal point and the *mantissa* which is the number after the decimal point. It is the mantissa that gives the significant figures of the number that has been logged; the characteristic tells you where to put the decimal point. The important property of logs to base 10 is that $A \times 10^n$ and $A \times 10^m$ have the same mantissa, so log tables need only show the mantissa.

The characteristic of a number greater than 1 is non-negative but the characteristic of a number less than 1 is negative. The rules for what to do in the case of a negative characteristic were rather complicated: you couldn't do ordinary arithmetic because the logarithm consisted of a negative characteristic and a positive mantissa. In ordinary arithmetic, the number -3.4 means $-3 - 0.4$ whereas the corresponding situation in logarithms, normally written $\bar{3}.4$, means $-3 + 0.4$. Instead of explaining this, the teacher gave a complicated set of rules, which just had to be learned — not the right way to do mathematics.

This question is all about calculating mantissas and there are no negative characteristics, I'm happy to say.

Solution to problem 19

(i) For the very first part, we have

$\log 2 + \log 5 = \log 10 = 1$, so $\log 5 = 1 - 0.301029996 = 0.699$ (3 d.p.)

and

$\log 6 = \log 2 + \log 3 = 0.778$ (3 d.p.).

Now we have to show that

$$5 \times 10^{47} < 3^{100} < 6 \times 10^{47}\,. \qquad (*)$$

Taking logs preserves the inequalities (because $\log x$ is an *increasing* function), so we need to show that

$$47 + \log 5 < 100 \log 3 < 47 + \log 6$$

i.e. that

$$47 + 0.699 < 47.7121 < 47 + 0.778$$

which is true. We see from $(*)$ that the first digit of 3^{100} is 5.

(ii) To find the first digit of these numbers, we use the method of part (i).

We have (to 3 d.p.)

$\log 2^{1000} = 1000 \log 2 = 301.030 = 301 + 0.030 < 301 + \log 2$

Thus $10^{301} < 2^{1000} < 10^{301} \times 2$ and the first digit of $2^{1\,000}$ is 1.

Similarly,

$\log 2^{10\,000} = 10\,000 \log 2 = 3010.29996 < 3010 + \log 2$, so the first digit of $2^{10\,000}$ is 1.

Finally,

$\log 2^{100\,000} = 30102 + 0.9996$ (4 d.p.) and

$\log 9 = 2 \log 3 = 0.95$ (2 d.p.), so the first digit of $2^{100\,000}$ is 9.

Post-mortem

Although the ideas in this question are really quite elementary, you needed to understand them deeply. You should feel pleased with yourself it you get this one out.

Note that in part (i) I started with $(*)$, the result I was trying to prove and then showed it is true. This is of course dangerous. But provided you keep writing 'We have to prove that ... ' or 'RTP' (Required To Prove) you should not get muddled between what you have proved and what you are trying to prove.

Problem 20: Cosmological models (✓✓)

In a cosmological model, the radius R of the universe is a function of the age t of the universe. The function R satisfies the three conditions:

$$R(0) = 0, \qquad R'(t) > 0 \text{ for } t > 0, \qquad R''(t) < 0 \text{ for } t > 0, \qquad (*)$$

where R'' denotes the second derivative of R. The function H is defined by

$$H(t) = \frac{R'(t)}{R(t)}\,.$$

(i) Sketch a graph of $R(t)$. By considering a tangent to the graph, show that $t < \dfrac{1}{H(t)}$.

(ii) Observations reveal that $H(t) = \dfrac{a}{t}$, where a is constant. Derive an expression for $R(t)$. What range of values of a is consistent with the three conditions $(*)$?

(iii) Suppose, instead, that observations reveal that $H(t) = bt^{-2}$, where b is constant. Show that this is not consistent with conditions $(*)$ for any value of b.

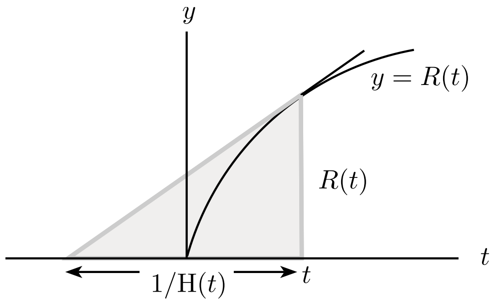
Note: $x^{\alpha}e^{-x} \to 0$ as $x \to \infty$ for any constant α.

2001 Paper I

Comments

The sketch just means any graph starting at the origin and increasing, but with decreasing gradient. The second part of (i) needs a bit of thought (where does the tangent intersect the x-axis?) so don't despair if you don't see it immediately. Parts (ii) and (iii) are perhaps easier than part (i).

Solution to problem 20



(i) The figure shows the curve $y = \mathrm{R}(t)$ and the tangent to the curve, which meets the t axis.
The height of the right-angled triangle in the figure is $\mathrm{R}(t)$ and the slope of the hypotenuse is $\mathrm{R}'(t)$. The length of the base is therefore $\mathrm{R}(t)/\mathrm{R}'(t)$, i.e. $1/\mathrm{H}(t)$. The figure shows that the tangent to the graph for $t > 0$ intersects the *negative* t–axis, so $1/\mathrm{H}(t) > t$.

(ii) One way to proceed is to integrate the differential equation:

$$\mathrm{H}(t) = \frac{a}{t} \Rightarrow \frac{\mathrm{R}'(t)}{\mathrm{R(t)}} = \frac{a}{t} \Rightarrow \int \frac{\mathrm{R}'(t)}{\mathrm{R}(t)}\mathrm{dt} = \int \frac{a}{t}\mathrm{dt} \Rightarrow \ln \mathrm{R(t)} = a \ln t + \text{constant} \Rightarrow \mathrm{R}(t) = At^a\,.$$

The first two conditions $(*)$ are satisfied if $a > 0$ and $A > 0$. For the third condition, we have $\mathrm{R}''(t) = a(a-1)At^{a-1}$ which is negative provided $a < 1$. The range of a is therefore $0 < a < 1$.

(iii) The obvious way to do this just follows part (ii). This time, we have

$$\mathrm{H}(t) = \frac{b}{t^2} \Rightarrow \frac{\mathrm{R}'(t)}{\mathrm{R}(t)} = \frac{b}{t^2} \Rightarrow \int \frac{\mathrm{R}'(t)}{\mathrm{R}(t)}\mathrm{dt} = \int \frac{b}{t^2}\mathrm{dt} \Rightarrow \ln \mathrm{R}(t) = -\frac{b}{t} + \text{constant} \Rightarrow \mathrm{R}(t) = Ae^{-b/t}\,.$$

Thus $\mathrm{R}'(t) = \mathrm{H}(t)\mathrm{R}(t) = Abt^{-2}e^{-b/t}$. Clearly $A > 0$ since $\mathrm{R}(t) > 0$ for $t > 0$ so $\mathrm{R}'(t) > 0$ provided $b > 0$. Furthermore $\mathrm{R}(0) = 0$ (think about this!), so only the condition $\mathrm{R}''(t) < 0$ remains to be checked. Differentiating $\mathrm{R}'(t)$ gives

$$\mathrm{R}''(t) = Ab(-2t^{-3})e^{-b/t} + Abt^{-2}e^{-b/t}(bt^{-2}) = Abt^{-4}e^{-b/t}(-2t+b)$$

This is positive when $t < \frac{1}{2}b$, which contradicts the condition $\mathrm{R}'' < 0$.

Instead of solving the differential equation, we could proceed as follows. We have

$$\frac{\mathrm{R}'}{\mathrm{R}} = bt^{-2} \Rightarrow \mathrm{R}' = b\mathrm{R}t^{-2} \Rightarrow \mathrm{R}'' = b\mathrm{R}'t^{-2} - 2b\mathrm{R}t^{-3} = b^2t^{-4}\mathrm{R} - 2b\mathrm{R}t^{-3} = b(b-2t)t^{-4}\mathrm{R}\,.$$

This is positive when $t < \frac{1}{2}b$, which contradicts the condition $\mathrm{R}'' < 0$.

Post-mortem

The interpretation of the conditions in $(*)$ of the question is as follows. The first condition $\mathrm{R}(0) = 0$ says that the universe started from zero radius — in fact, from the 'big bang'.

The second condition says that the universe is expanding. This was one of the key discoveries in cosmology in the last century. The quantity $\mathrm{H}(t)$ is called *Hubble's constant* (though it varies with time). It measures the rate of expansion of the universe. Its reciprocal is the Hubble time. The present-day value of the Hubble time has been a matter of great debate but it seems to be about 14.4 billion years, greater than the age of the universe (as calculated using cosmological models of this sort) by about 0.5 billion years.

The last condition says that the expansion of the universe is slowing down. This is expected on physical grounds because of the gravitational attraction of galaxies on one another. However, current observations indicate that the expansion of the universe is speeding up, and this is thought to be due to the presence of dark energy.

Problem 21: Melting snowballs (✓✓)

Frosty the snowman is made from two uniform spherical snowballs, initially of radii $2R$ and $3R$. The smaller (which is his head) stands on top of the larger. As each snowball melts, its volume decreases at a rate which is directly proportional to its surface area, the constant of proportionality being the same for each snowball. During melting, the snowballs remain spherical and uniform. When Frosty is half his initial height, show that the ratio of his volume to his initial volume is $37 : 224$.

What is this ratio when Frosty is one tenth of his initial height?

1991 Paper I

Comments

To start with, you have to set up a differential equation which gives the radii of the snowballs as a function of time. Don't worry if you have never solved a differential equation: you will be able to solve this one. Having solved it, you have to evaluate the constant of integration from the information in the question.

Setting up and solving the differential equation is not difficult in itself, but it is necessary to think what variables you want to use and then go through a number of steps in the dark, without any reassurance from the question.

Solution to problem 21

For either snowball, the rate of change of volume, call it v, at any time is related to the surface area a by

$$\frac{\mathrm{d}v}{\mathrm{d}t} = -ka\,, \qquad (*)$$

where k is a positive constant. For a sphere of radius r, this becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\tfrac{4}{3}\pi r^3\big) = -k(4\pi r^2)\,.$$

We can write $\dfrac{\mathrm{d}(r^3)}{\mathrm{d}t}$ as $3r^2\dfrac{\mathrm{d}r}{\mathrm{d}t}$, so equation $(*)$ is equivalent to (cancelling the factor of $4\pi r^2$)

$$\frac{\mathrm{d}r}{\mathrm{d}t} = -k\,.$$

Thus $r = -kt + C$, where C is a constant of integration.

Initially, Frosty's head has radius $2R$ and his body has radius $3R$, so the equations for the radii of the head and body at time t are respectively

$$r = -kt + 2R \qquad \text{and} \qquad r = -kt + 3R.$$

Frosty's height h is twice the sum of these radii, i.e. $h = 2(-2kt + 5R)$, which falls to half its original value of $10R$ when $kt = \frac{5}{4}R$. At this time, the radii of the head and body are $\frac{3}{4}R$ and $\frac{7}{4}R$, so the ratio of his volume to his initial volume is

$$\frac{\big(\frac{4}{3}\pi\big)\big(\frac{3}{4}R\big)^3 + \big(\frac{4}{3}\pi\big)\big(\frac{7}{4}R\big)^3}{\big(\frac{4}{3}\pi\big)(2R)^3 + \big(\frac{4}{3}\pi\big)(3R)^3} = \frac{\big(\frac{3}{4}\big)^3 + \big(\frac{7}{4}\big)^3}{2^3 + 3^3} = \frac{37}{224}\,.$$

When Frosty is just R high, all that remains of him, since $kt > 2R$, is the body, which is a sphere of radius $\frac{1}{2}R$. The ratio of his volume to his original volume is just $\frac{1}{240}$.

Post-mortem

As mentioned in the comments section, there is not a lot for you to do in this question, but it is a lot for you to do entirely on your own.

The key is to transcribe the words in the question into mathematics, the obvious starting point being equation $(*)$.

The original question had a different rider[18], asking about the maximum rate of change of volume with respect to area. I didn't much like this, because it seemed to be a new and not very interesting idea. My rider relates directly to the previous part, and requires a tiny bit of extra thought (a trap, some would say, but it is only a trap if you are on autopilot).

---

[18] The term *rider* seems a bit old-fashioned now. It referred to the final part at the end of an examination question in the days when examination questions often consisted of a relatively straightforward first part, perhaps the proof of a theorem, followed by a more tricky part extending or applying the first part.

Problem 22: Gregory's series (✓✓)

Give rough sketches of the function $\tan^k\theta$ for $0 \leqslant \theta \leqslant \frac{1}{4}\pi$ in the two cases $k = 1$ and $k \gg 1$.

(i) Show that for any positive integer n

$$\int_0^{\frac{1}{4}\pi} \tan^{2n+1}\theta \,\mathrm{d}\theta = (-1)^n \left(\tfrac{1}{2}\ln 2 + \sum_{m=1}^{n} \frac{(-1)^m}{2m} \right), \qquad (\dagger)$$

and deduce that

$$\ln 2 = -\sum_{m=1}^{\infty} \frac{(-1)^m}{m}\,. \qquad (\ddagger)$$

(ii) Show similarly that

$$\frac{\pi}{4} = -\sum_{m=1}^{\infty} \frac{(-1)^m}{2m-1}\,.$$

1991 Paper II

Comments

The symbol $\gg$ in the first paragraph means 'much greater than', so for the second sketch k is a large number.

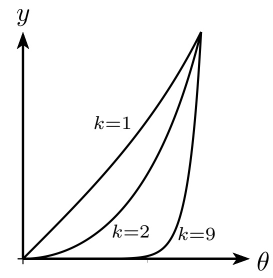
This is a good question. In part (i) you are told what to do (in not-very-easy stages) and part (ii) tests your understanding of what you have done and why you have done it by asking you to apply the method to a different but essentially similar problem.

In the first paragraph, you have to see how the function $\tan^k\theta$ changes when k increases. You need only a rough sketch to show that you have understood the important point. This should be done by thought, not by means of a calculator.

If you are stuck with the integral of the second paragraph, you might like to think in terms of a recurrence formula, i.e. a formula relating I_{2n+1} and I_{2n-1} (in the obvious notation).

The series derived in part (ii) for $\frac{1}{4}\pi$ is usually called Leibniz's formula, although the general series for $\tan^{-1}x$ was written down by Gregory in 1671, two years before Leibniz. It was one of the first explicit formulae for π, though Wallis had obtained a product formula in 1655 using a method similar to the method of this question, using $\sin^k x$ in the integral. Previously, the value of π could only be estimated geometrically, by (for example) approximating the circumference of a circle by the edges of an inscribed regular polygon. Using a square gives $\pi \approx 2\sqrt{2}$.

Solution to problem 22



For $0 \leqslant \tan\theta < \frac{1}{4}\pi$, we have $\tan\theta < 1$, so that the curve $y = \tan^k\theta$ is close to zero (i.e. much smaller than 1) when k is large. This is illustrated in the figure which shows three cases: for $k = 1$ the graph is mildly curved; for larger k the graph hugs the x-axis before taking off. The graphs all pass through the point $(\frac{1}{4}\pi, 1)$.

(i) To evaluate the integral, let

$$I_{2n+1} = \int_0^{\frac{1}{4}\pi} \tan^{2n+1}\theta \; d\theta\,. \qquad (**)$$

We shall express I_{2n+1} in terms of I_{2n-1} using the relation $\tan^2\theta = \sec^2\theta - 1$:

$$I_{2n+1} = \int_0^{\frac{1}{4}\pi} \tan^{2n-1}\theta\,(-1+\sec^2\theta)\,d\theta = -I_{2n-1} + \int_0^1 u^{2n-1}\,du = -I_{2n-1} + \frac{1}{2n}\,.$$

To evaluate the second integral, set $u = \tan\theta$ so that $du = \sec^2\theta\, d\theta$.

Repeating the process gives

$$I_{2n+1} = -I_{2n-1} + \frac{1}{2n} = I_{2n-3} - \frac{1}{2(n-1)} + \frac{1}{2n} = \cdots = (-1)^n I_1 + \frac{1}{2n} - \frac{1}{2(n-1)} + \cdots + (-1)^{n+1}\frac{1}{2}\,.$$

The above sum (starting with $1/(2n)$) is the same as that in (†) overleaf, so it only remains to evaluate I_1, corresponding to $n = 0$ in $(**)$:

$$I_1 = \int_0^{\frac{1}{4}\pi} \tan\theta \; d\theta = -\ln(\cos\theta)\Big|_0^{\frac{1}{4}\pi} = -\ln(1/\sqrt{2}) = \tfrac{1}{2}\ln 2\,,$$

as required.

We deduce the expression (‡) overleaf for $\ln 2$ using the first line of the question. When n is very large, I_{2n+1} is very small, being the area under a graph which is almost zero for almost all of the range of integration. In the limit $n \to \infty$, we set $I_{2n+1} = 0$ in (†) which leads immediately to (‡).

(ii) To obtain the formula for $\frac{1}{4}\pi$, we follow the above method using I_{2n} instead of I_{2n+1}. This time we have to calculate I_0:

$$I_0 = \int_0^{\frac{1}{4}\pi} 1 \; d\theta = \frac{\pi}{4}\,.$$

Post-mortem

You might think that the method of obtaining the formula for $\frac{1}{4}\pi$ in this question is rather indirect; one could instead just integrate the formula

$$\frac{d\tan^{-1}x}{dx} = \frac{1}{1+x^2} = 1 - x^2 + x^4 - \cdots \qquad (*)$$

term by term and get the result immediately by setting $x = 1$. The virtue of the method used in the question is that it gives an explicit form (an integral) of the remainder after n terms of the series. We were able to show, by means of a sketch, that the remainder tends to zero as n tends to infinity; in other words, we showed that the series converges. Although the sketch method of proof is a bit crude, it can easily be made more rigorous once the concept of integration is more carefully defined. On the other hand, integrating $(*)$ and setting $x = 1$ is a bit delicate, since the series only converges for $x^2 < 1$.

Problem 23: Intersection of ellipses (✓)

Show that the equation of any circle passing through the points of intersection of the ellipse

$$(x + 2)^2 + 2y^2 = 18$$

and the ellipse

$$9(x - 1)^2 + 16y^2 = 25$$

can be written in the form

$$x^2 - 2ax + y^2 = 5 - 4a\ .$$

2002 Paper I

Comments

You don't need to know anything about ellipses to do this question.

When this question was set it seemed too easy. But quite a high proportion of the candidates made no real progress. Of course, it is not easy to keep a cool head under examination conditions; but surely it is obvious that either x or y has to be eliminated from the two equations for the ellipses; and having decided that it is obvious which one to eliminate.

It is worth thinking about how many points of intersection of the ellipses we are expecting: a sketch might help if you know what shapes the ellipses are.

Solution to problem 23

First we have to find the intersections of the two ellipses by solving the simultaneous equations

$$\begin{aligned}(x+2)^2+2y^2 &= 18 \quad (1)\\ 9(x-1)^2+16y^2 &= 25\,. \quad (2)\end{aligned}$$

We can eliminate y by multiplying equation (1) by 8 and subtracting equation (2), so that at the intersections

$$8(x+2)^2-9(x-1)^2=144-25$$

i.e.

$$x^2-50x+96=0$$

i.e.

$$(x-2)(x-48)=0\,.$$

The two possible values for x at the intersections are therefore 2 and 48.

Next we find the values of y at the intersection. Taking $x=2$ and substituting into equation (1) gives $16+2y^2=18$, so $y=\pm1$. Taking $x=48$ gives $50^2+2y^2=18$ which has no (real) roots. Thus there are two points of intersection, at $(2,1)$ and $(2,-1)$.

Now we go after the circle. Suppose that a circle through the points of intersection has centre (p,q) and radius R. Then the equation of the circle is

$$(x-p)^2+(y-q)^2=R^2\,.$$

Setting $(x,y)=(2,1)$ and $(x,y)=(2,-1)$ gives two equations:.

$$(2-p)^2+(1-q)^2=R^2\,,\qquad (2-p)^2+(-1-q)^2=R^2\,.$$

Subtracting the two equations gives $q=0$ (it is obvious anyway, because of the symmetry of both ellipses under reflections in the x axis, that the centre of the circle must lie on the y axis). Thus the equation of any circle passing through the intersections is

$$(x-p)^2+y^2=(2-p)^2+1\,,$$

which simplifies to the given result with $p=a$.

Post-mortem

As commented earlier, there really wasn't much to this question apart from solving simultaneous equations and quadratic equations. I suppose that the daunting feature is that very little help is given in the way of intermediate steps of answers.

Another daunting feature is the unexplained parameter a in the equation of the circle. Two ellipses can intersect in four points, three points (if the ellipses touch rather than intersect at one of the points), two points, one point or not at all.

In general, we wouldn't expect to be able to draw a circle through four given points. We would expect exactly one circle through three given points (not lying on a line), but a whole family of circles through any two points. Our ellipses intersect in two points, which is why the circle depends on a parameter a.

Problem 24: Sketching $x^m(1-x)^n$ (✓✓✓)

> Let $\mathrm{f}(x) = x^m(x-1)^n$, where m and n are both integers greater than 1. Show that
>
> $$\mathrm{f}'(x) = \left(\frac{m}{x} - \frac{n}{1-x}\right)\mathrm{f}(x)\,.$$
>
> Show that the curve $y = \mathrm{f}(x)$ has a stationary point in the interval $0 < x < 1$. By considering $\mathrm{f}''(x)$, show that this stationary point is a maximum if n is even and a minimum if n is odd. Sketch the graphs of $\mathrm{f}(x)$ in the four cases that arise according to the values of m and n.

2002 Paper I

Comments

There is quite a lot in this question, but it is one of the best STEP questions on basic material that I came across.

First, you have to find $\mathrm{f}'(x)$. The very first part (giving a form of $\mathrm{f}'(x)$) did not occur in the original STEP question. This particular expression is extremely helpful when if comes to finding the value of $\mathrm{f}''(x)$ at the stationary point. In the actual exam, a handful of candidates successfully found a general expression for $\mathrm{f}''(x)$ in terms of x then evaluated it at the stationary point: first-class work.

You should find that the value of $\mathrm{f}''(x)$ at the stationary point can be expressed in the form $(\cdots)\mathrm{f}(x)$, where the factor in the brackets is quite simple and always negative, so that the nature of the stationary point depends only on the sign of $\mathrm{f}(x)$. Actually, this is intuitively obvious: $\mathrm{f}(0) = \mathrm{f}(1) = 0$ so the one stationary point with $0 < x < 1$ must be, for example, a maximum if $\mathrm{f}(x) > 0$ for $0 < x < 1$.

For the sketches, you really have only to think about the behaviour of $\mathrm{f}(x)$ when $|x|$ is large and the sign of $\mathrm{f}(x)$ between $x = 0$ and $x = 1$. That allows you to piece together the graph, knowing that there is only one stationary point between $x = 0$ and $x = 1$. You should then be able to identify the four cases referred to.

Note that, very close to $x = 0$, $\mathrm{f}(x) \approx x^m(-1)^n$ so it is easy to see how the graph there depends on n and m; you can use this to check that you have got the graphs right.

Post-mortem

Don't read this until you have worked through the question!

You may have noticed a little carelessness in the first line of the solution: what happens in the logarithmic differentiation if any of $\mathrm{f}(x)$ or x or $x-1$ are negative? The answer is that it doesn't matter. One way to deal with the problem of logs with negative arguments is to put modulus signs everywhere using the correct result

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln|\mathrm{f}(x)| = \frac{\mathrm{f}'(x)}{\mathrm{f}(x)}\,.$$

If you are not sure of this, try it on $\mathrm{f}(x) = x$ taking the two cases $x > 0$ and $x < 0$ separately.

A more sophisticated way of dealing with logs with negative arguments is to note that $\ln(-\mathrm{f}(x)) = \ln f(x) + \ln(-1)$. We don't have to worry about $\ln(-1)$ because it is a constant (of some sort) and so won't affect the differentiation. Actually, a value of $\ln(-1)$ can be obtained by taking logs of Euler's famous formula $\mathrm{e}^{i\pi} = -1$ giving $\ln(-1) = i\pi$.

Solution to problem 24

The first result can be established by differentiating $\mathrm{f}(x)$ directly, but the neat way to do it is to start with $\ln \mathrm{f}(x)$:

$$\ln \mathrm{f}(x) = m \ln x + n \ln(x-1) \Longrightarrow \frac{\mathrm{f}'(x)}{\mathrm{f}(x)} = \frac{m}{x} + \frac{n}{x-1} \Longrightarrow \mathrm{f}'(x) = \left(\frac{m}{x} - \frac{n}{1-x}\right)\mathrm{f}(x),$$

as required. Writing this as $\mathrm{f}'(x) = (m(x-1)+nx)x^{m-1}(x-1)^{n-1}$, we see that $\mathrm{f}(x)$ has stationary points at $x=0$ and $x=1$ (since $m-1>0$ and $n-1>0$) and when $m(x-1)+nx=0$. Solving this last equation for x gives $x = \dfrac{m}{m+n}$, which lies between 0 and 1 since m and n are positive.

Next we calculate $\mathrm{f}''(x)$. Starting with

$$\mathrm{f}'(x) = \left(\frac{m}{x} - \frac{n}{1-x}\right)\mathrm{f}(x),$$

we obtain

$$\mathrm{f}''(x) = \left(-\frac{m}{x^2} - \frac{n}{(1-x)^2}\right)\mathrm{f}(x) + \left(\frac{m}{x} - \frac{n}{1-x}\right)\mathrm{f}'(x).$$

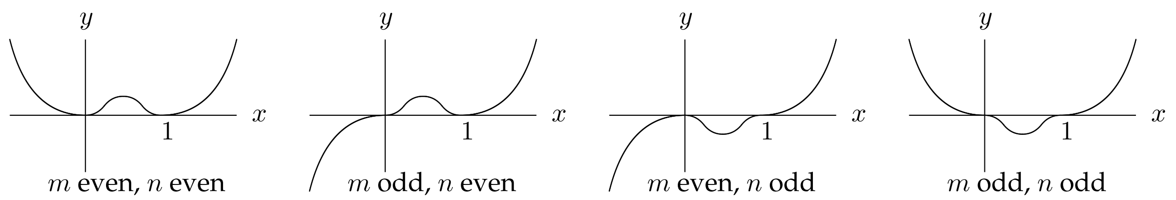
At a stationary point, the second of these two terms is zero because $\mathrm{f}'(x)=0$, leaving

$$\mathrm{f}''(x) = \left(-\frac{m}{x^2} - \frac{n}{(x-1)^2}\right)\mathrm{f}(x).$$

The bracketed expression is negative so at a stationary point $\mathrm{f}''(x) < 0$ if $\mathrm{f}(x) > 0$ and $\mathrm{f}''(x) > 0$ if $\mathrm{f}(x) < 0$.

The sign of $\mathrm{f}(x)$ for $0<x<1$ is the same as the sign of $(x-1)^n$, since $x^m>0$ when $x>0$. For the stationary point in the interval $0<x<1$, $(x-1)^n>0$ if n is even and $(x-1)^n<0$ if n is odd. Thus $\mathrm{f}''(x)<0$ if n is even and $\mathrm{f}''(x)>0$ if n is odd, which is the required result.

The four cases to sketch are determined by whether m and n are even or odd. The easiest way to understand what is going on is to consider the graphs of x^m and $(x-1)^n$ separately, then try to join them up at the stationary point between 0 and 1. If m is odd, then $x=0$ is a point of inflection, but if m is even it is a maximum or minimum according to the sign of $(x-1)^n$. You should also think about the behaviour for large $|x|$. All the various bits of information (including the nature of the turning point investigated above) should all piece neatly together.



Problem 25: Inequalities by area estimates (✓✓✓)

Give a sketch of the curve $y = \dfrac{1}{1+x^2}$, for $x \geqslant 0$.
Find the equation of the line that intersects the curve at $x = 0$ and is tangent to the curve at some point with $x > 0$. Prove that there are no further intersections between the line and the curve. Draw the line on your sketch.
By considering the area under the curve for $0 \leqslant x \leqslant 1$, show that $\pi > 3$.
By considering the volume formed by rotating the curve about the y axis, show also that $\ln 2 > \frac{2}{3}$.

Note: $\displaystyle\int_0^1 \frac{1}{1+x^2}\,\mathrm{d}x = \frac{\pi}{4}$.

2002 Paper I

Comments

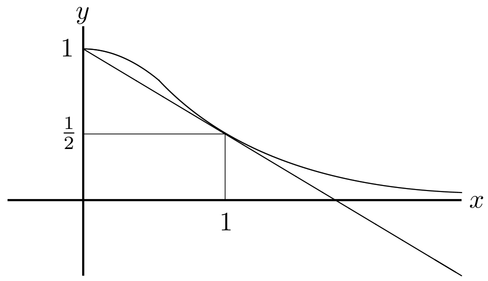
There is quite a lot to this question, so ✓✓✓ even though none of it is particularly difficult (very surprising to find it on Paper I). The most common mistake in finding the equation of the tangent is to muddle the y and x that occur in the equation of the line ($y = mx + c$) with the coordinates of the point at which the tangent meets the curve, getting 'constants' m and c that depend on x. I'm sure you wouldn't make this rather elementary mistake normally, but it is surprising what people do under examination conditions.

The reason for sketching the curve lies in the last parts: the shape of the curve relative to the straight line provides the inequality.

The note at the end of the question had to be given because integrals of that form (giving inverse trigonometric functions) are not in the A-level Mathematics specifications. Volumes of revolution are no longer in the A-level Mathematics specifications and so they are no longer in the STEP 1 specification. They are now included in the STEP 3 specification.

Solution to problem 25

Your sketch should show a graph which has gradient 0 at $(0,1)$ and which asymptotes to the x-axis for large x. The extra line in my sketch is the tangent to the graph from the point $(0,1)$ which is required later.



First we find the equation of the tangent to the curve $y=(1+x^2)^{-1}$ at the point (p,q). The gradient of the curve at the point (p,q) is $-2p(1+p^2)^{-2}$, i.e. $-2pq^2$, so the equation of the tangent is

$$y=-2pq^2x+c$$

where c is given by $q=-2pq^2p+c$. This line is supposed to pass through the point $(0,1)$, so $1=c$. Thus $1=q+2p^2q^2$, or (replacing q with $1/(1+p^2)$)

$$1=\frac{1}{1+p^2}+\frac{2p^2}{(1+p^2)^2}$$

which simplifies to

$$(1+p^2)^2=1+3p^2 \quad \text{i.e.} \quad p^4=p^2\ .$$

The only positive solution is $p=1$, so the equation of the line is $y=-\frac{1}{2}x+1$.

We can check that this line does not meet the curve again by solving the equation

$$-\frac{x}{2}+1=\frac{1}{1+x^2}\ .$$

Multiplying by $(1+x)^2$ gives

$$(1-\tfrac{1}{2}x)(1+x^2)=1 \quad \text{i.e.} \quad x^3-2x^2+x=0\ .$$

Factorising shows that $x=1$ and $x=0$ satisfy this equation (these are the known roots, $x=1$ being a double root corresponding to the tangency) and that there are no more roots.

The area under the curve for $0\leqslant x\leqslant 1$ is $\frac{1}{4}\pi$ as given. The sketch shows that this area is greater than the area under the tangent line for $0\leqslant x\leqslant 1$, which is $\frac{1}{2}+\frac{1}{4}$ (the area of the rectangle plus the area of the triangle above it). Comparing this with $\frac{1}{4}\pi$ gives the required result.

The volume formed by rotating the curve about the y axis is

$$\int_0^1 2\pi yx\,\mathrm{d}x=\pi\int_0^1\frac{2x}{1+x^2}\,\mathrm{d}x=\pi\ln 2\ .$$

This is greater than the volume formed by rotating the line about the y axis, which is

$$\int_0^1 2\pi yx\mathrm{d}x=2\pi\int_0^1(1-\tfrac{1}{2}x)x\mathrm{d}x=\tfrac{2}{3}\pi\ .$$

Comparison gives the required result.

Post-mortem

As I said before, no step of this question is particularly difficult, but *lots* of different ideas are required. The last part is very challenging, because you have to do and compare two separate volume integrals without any intermediate steer at all. You should be very pleased if you made good progress with it.

The important lesson to learn from this is that perseverance will eventually pay off.

Problem 26: Simultaneous integral equations (✓✓)

Let
$$I = \int_0^a \frac{\cos x}{\sin x + \cos x}\,\mathrm{d}x \quad \text{and} \quad J = \int_0^a \frac{\sin x}{\sin x + \cos x}\,\mathrm{d}x\,,$$
where $0 \leqslant a < \frac{3}{4}\pi$. By considering $I + J$ and $I - J$, show that $2I = a + \ln(\sin a + \cos a)$.

Find also:

(i) $\displaystyle\int_0^{\frac{1}{2}\pi} \frac{\cos x}{p\sin x + q\cos x}\,\mathrm{d}x$, where p and q are positive numbers;

(ii) $\displaystyle\int_0^{\frac{1}{2}\pi} \frac{\cos x + 4}{3\sin x + 4\cos x + 25}\,\mathrm{d}x$.

2002 Paper I

Comments

This is a model for a perfect STEP question: you are told how to do the first part, and you have to adapt the idea on your own for the later parts. Note the structure of the question: there is a 'stem' (the first paragraph) containing material that will be useful for both the later parts.

In the examination, candidates who were successful in parts (i) and (ii) nearly always started off with the statement 'Now let $I = \cdots$ and $J = \cdots$'. If you are stuck with part (ii), the choice of significant numbers (3, 4 and 25) should provide a clue.

When you arrive at an answer for part (i), you will of course check that it agrees with the given result in the opening paragraph when $a = \frac{1}{2}\pi$ and $p = q = 1$.

You will no doubt have noticed the restriction $0 \leqslant a < \frac{3}{4}\pi$ given in the first paragraph (and also the restrictions on p and q in part (i)). You should try to work out its purpose because it might provide an insight into the method of tackling the question, though in this case it doesn't.

Solution to problem 26

For the first part, we regard the two integrals essentially as a pair of simultaneous equations, adding and subtracting to simplify them. We have

$$\begin{aligned} I+J &= \int_0^a \frac{\cos x+\sin x}{\sin x+\cos x}\,\mathrm{d}x = \int_0^a \mathrm{d}x = a \\ I-J &= \int_0^a \frac{\cos x-\sin x}{\sin x+\cos x}\,\mathrm{d}x = \ln(\cos a+\sin a) \end{aligned}$$

(in the second integral, the numerator is the derivative of the denominator). Adding these equations gives the required expression for $2I$.

(i) Similarly, let $I = \displaystyle\int_0^{\frac12\pi} \frac{\cos x}{p\sin x+q\cos x}\,\mathrm{d}x$ and $J = \displaystyle\int_0^{\frac12\pi} \frac{\sin x}{p\sin x+q\cos x}\,\mathrm{d}x$. Then

$$\begin{aligned} qI+pJ &= \int_0^{\frac12\pi} \frac{q\cos x+p\sin x}{p\sin x+q\cos x}\,\mathrm{d}x = \tfrac12\pi \\ pI-qJ &= \int_0^{\frac{\pi}{2}} \frac{p\cos x-q\sin x}{p\sin x+q\cos x}\,\mathrm{d}x = \ln(p\sin\tfrac12\pi+q\cos\tfrac12\pi)-\ln(p\sin 0+q\cos 0) = \ln\frac{p}{q}\,. \end{aligned}$$

Now we solve these two equations simultaneously for I:

$$(p^2+q^2)I = \frac{q\pi}{2}+p\ln\frac{p}{q}\,.$$

(ii) This time, let $I = \displaystyle\int_0^{\frac12\pi} \frac{\cos x+4}{3\sin x+4\cos x+25}\,\mathrm{d}x$ and $J = \displaystyle\int_0^{\frac12\pi} \frac{\sin x+3}{3\sin x+4\cos x+25}\,\mathrm{d}x$. Then

$$\begin{aligned} 4I+3J &= \int_0^{\frac12\pi} \frac{4\cos x+3\sin x+25}{3\sin x+4\cos x+25}\,\mathrm{d}x = \frac{\pi}{2} \\ 3I-4J &= \int_0^{\frac12\pi} \frac{3\cos x-4\sin x}{3\sin x+4\cos x+25}\,\mathrm{d}x = \ln\tfrac{28}{29}\,. \end{aligned}$$

Solving simultaneously gives

$$25I = 2\pi+3\ln\tfrac{28}{29}\,.$$

Post-mortem

I very much like this question. The first part appeared on STEP Paper I in 1995 (the second part of the 1995 question was an integral completely unrelated to the first part — that wouldn't happen now) and I was completely taken by surprise.

In order to use it again for STEP in 2002, I added parts (i) and (ii). It took me some time to think of a suitable extension. I was disappointed to find that the basic idea is more or less a one-off: there are very few denominators, besides the ones given, that lead to integrands amenable to this trick. But I was pleased with what I came up with. The question leads you through the opening paragraph and the extra parts depend very much on your having understood why the opening paragraph works. The reason for restriction $0 \leqslant a < 3\pi/4$ is that the denominator should not be 0 for any value of x in the range of integration — otherwise, the integral is undefined. Writing $\sin x+\cos x = \sqrt{2}\sin(x+\pi/4)$ shows that the denominator is first zero when $x = 3\pi/4$.

You might like to think how you would evaluate these integrals without using the trick method of this question. Perhaps the easiest way is to use the substitution $t = \tan(\frac12 x)$ which converts the denominator to a quadratic in x.

Problem 27: Relation between coefficients of quartic for real roots(✓✓)

In this question you may assume that, if $k_1, \ldots, k_n$ are distinct positive real numbers, then

$$\frac{1}{n}\sum_{r=1}^{n} k_r > \left(\prod_{r=1}^{n} k_r\right)^{\frac{1}{n}},$$

i.e. their arithmetic mean is greater than their geometric mean.

Suppose that a, b, c and d are positive real numbers such that the polynomial

$$\mathrm{f}(x) = x^4 - 4ax^3 + 6b^2x^2 - 4c^3x + d^4$$

has four distinct positive roots.

(i) By considering the relationship between the coefficients of f and its roots, show that $c > d$.

(ii) By differentiating f, show that $b > c$.

(iii) Show that $a > b$.

1997 Paper III

Comments

This result is both surprising and pleasing.

The question looks difficult, but you don't have to go very far before you come across something to substitute into the given arithmetic mean/geometric mean (AM/GM) inequality.

To obtain the relationship between the coefficients and the roots for part (i), you need to write the quartic equation in the form $(x-p)(x-q)(x-r)(x-s) = 0$.

Watch out for the condition in the given form of the arithmetic/geometric inequality that the numbers are distinct: you will have to show that any numbers (or algebraic expressions) you use in the inequality are distinct.

I puzzled over this question for ages, not understanding the idea behind it[19].

It will be clear from the proof that it can be generalised to any equation of the form

$$x^N + \sum_{k=0}^{N-1}\binom{N}{k}(-a_k)^{N-k}x^k = 0,$$

where the numbers a_k are distinct and positive, which has positive distinct roots.

Eventually, I realised that the result of the question has really nothing to do with quartic equations, though quartic equations provide a neat method of proof. The inequalities, which apply to any positive numbers, form a sequence interpolating between the AM and the GM. These inequalities were first derived by Newton and Maclaurin in the 17th century but don't seem to be very well known now.

[19] I mentioned this in a talk I gave to sixth formers, and later saw irritatingly on www.thestudentroom.co.uk a comment to the effect that 'even Dr Siklos can't do this STEP question'. Of course I could *do* it; I wanted to *understand* it, as I hope you do too.

Solution to problem 27

(i) First write

$$f(x) \equiv (x-p)(x-q)(x-r)(x-s)$$

where p, q, r and s are the four roots of the equation (known to be real, positive and distinct). Multiplying out the brackets and comparing with $x^4 - 4ax^3 + 6b^2x^2 - 4c^3x + d^4$ shows that $pqrs = d^4$ and $pqr + qrs + rsp + spq = 4c^3$.

The required result, $c > d$, follows immediately by applying the AM/GM inequality to the positive real numbers pqr, qrs, rsp and spq:

$$c^3 = \frac{pqr + qrs + rsp + spq}{4} > [(pqr)(qrs)(rsp)(spq)]^{1/4} = [p^3q^3r^3s^3]^{1/4} = [d^{12}]^{1/4} .$$

Taking the cube root (c and d are positive) preserves the inequality.

The AM/GM inequality at the beginning of the question is stated only for the case when the numbers are distinct (though in fact it holds provided at least two of the numbers are distinct). To use the inequality as above, we must therefore show that no two of pqr, qrs, rsp and spq are equal. This follows immediately from the fact that the roots are distinct and non-zero. For if, for example, $pqr = qrs$ then $qr(p-s) = 0$ which means that $q = 0$, $r = 0$ or $p = s$, all of which are ruled out.

(ii) The polynomial $f'(x)$ is cubic so it has three zeros (roots). These are at the turning points of $f(x)$, which lie between the zeros of $f(x)$ and are therefore distinct and positive.

Now

$$f'(x) = 4x^3 - 12ax^2 + 12b^2x - 4c^3,$$

so at the turning points

$$x^3 - 3ax^2 + 3b^2x - c^3 = 0 .$$

Suppose that the roots of this cubic equation are u, v and w, all real, distinct and positive. Then comparing $x^3 - 3ax^2 + 3b^2x - c^3 = 0$ and $(x-u)(x-v)(x-w)$ shows that $c^3 = uvw$ and $3b^2 = vw + wu + uv$.

The required result, $b > c$, follows immediately by applying the AM/GM inequality to the positive distinct numbers vw, wu and uv.

(iii) Apply similar arguments to $f''(x)/12$.

Post-mortem

The general inequalities mentioned in the comments above are interesting. For example, for four numbers a_1, a_2, a_3 and a_4, we define

$$S_1 = \frac{a_1 + a_2 + a_3 + a_4}{4}, \qquad S_2 = \frac{a_2a_3 + a_3a_1 + a_1a_2 + a_4a_1 + a_4a_2 + a_4a_3}{6},$$

$$S_3 = \frac{a_2a_3a_4 + a_1a_3a_4 + a_1a_2a_4 + a_1a_2a_3}{4} \quad \text{and} \quad S_4 = a_1a_2a_3a_4 .$$

Except for the denominators and alternating signs, S_i is the coefficient of x^{4-i} in the equation $(x-a_1)(x-a_2)(x-a_3)(x-a_4) = 0$, but that doesn't matter. Maclaurin's inequalities are

$$S_1 \geqslant S_2^{\frac{1}{2}} \geqslant S_3^{\frac{1}{3}} \geqslant S_4^{\frac{1}{4}} .$$

What is really surprising is that we proved the three stronger inequalities from the apparently weaker AM>GM inequality ($S_1 \geqslant S_4^{\frac{1}{4}}$): a rare example of pulling yourself up by your bootstraps.

Problem 28: Fermat numbers (✓✓)

The nth Fermat number, F_n, is defined by

$$F_n = 2^{2^n} + 1 , \qquad n = 0,\ 1,\ 2,\ \dots ,$$

where 2^{2^n} means 2 raised to the power 2^n . Calculate F_0, F_1, F_2 and F_3 . Show that, for $k = 1$, $k = 2$ and $k = 3$,

$$F_0 F_1 \dots F_{k-1} = F_k - 2 \,. \qquad (*)$$

Prove, by induction, or otherwise, that $(*)$ holds for all $k \geqslant 1$. Deduce that no two Fermat numbers have a common factor greater than 1.
Hence show that there are infinitely many prime numbers.

2002 Paper II

Comments

Fermat (1601–1665) conjectured that every number of the form $2^{2^n} + 1$ is a prime number. F_0 to F_4 are indeed prime, but Euler showed in 1732 that F_5 (4294967297) is divisible by 641. As can be seen, Fermat numbers get very big and not many more have been investigated; but those that have been investigated have been found not to be prime numbers. It is now conjectured that in fact only a finite number of Fermat numbers are prime numbers.

The Fermat numbers have a geometrical significance as well: Gauss proved that a regular polygon of n sides can be inscribed in a circle using a Euclidean construction (i.e. only a straight edge and a compass) if and only if n is a power of 2 times a product of distinct Fermat primes.

The very last part of this question, showing that the number of primes is infinite, is completely unexpected and delightful. It comes from *Proofs from THE BOOK*[20], a set of proofs thought by Paul Erdös to be heaven-sent. Most of them are much less elementary this one. Erdös was an extraordinarily prolific mathematician. He had almost no personal belongings and no home[21]. He spent his life visiting other mathematicians and proving theorems with them. He collaborated with so many people that every mathematician is (jokingly) assigned an *Erdös number*; for example, if you wrote a paper with someone who wrote a paper with Erdös, you are given the Erdös number 2. Most mathematicians seem to have an Erdös number less than 8.

---

[20] M. Aigler and G.M. Ziegler (Springer, 1999). The idea behind the proof was given by Christian Goldbach (1690–1764), who is best known for his conjecture that there are an infinite number of *twin primes*, that is, prime numbers that are two apart such as 17 and 19.

[21] See *The Man Who Loved Only Numbers* by Paul Hoffman, published by Little Brown and Company in 1999 — a wonderfully readable and interesting book.

Solution to problem 28

To begin with, we have by direct calculation that $F_0 = 3$, $F_1 = 5$, $F_2 = 17$ and $F_3 = 257$, so it is easy to verify $(*)$ for $k = 1, 2$ and 3.

For the induction, we start by assuming that the result holds for $k = m$ so that

$$F_0F_1 \ldots F_{m-1} = F_m - 2\ .$$

We need to show that this implies that the result holds for $k = m + 1$, i.e. that

$$F_0F_1 \ldots F_m = F_{m+1} - 2\ . \qquad (*)$$

Starting with the left-hand side of this equation, we have

$$F_0F_1 \ldots F_{m-1}F_m = (F_m - 2)F_m = (2^{2^m} - 1)(2^{2^m} + 1) = \left(2^{2^m}\right)^2 - 1 = 2^{2^{m+1}} - 1 = F_{m+1} - 2,$$

as required. Since we know that the result holds for $k = 1$, the induction is complete.

To show that no two Fermat number have a common factor (other than 1), we proceed by contradiction. Suppose p divides F_l and F_m, where $l < m$. Then p divides $F_0F_1 \ldots F_l \ldots F_{m-1}$ and therefore p divides $F_m - 2$, by $(*)$, as well as F_m. But this is impossible: all Fermat numbers are odd so no number other than 1 divides both $F_m - 2$ and F_m. Hence no two Fermat numbers have a common factor greater than 1.

For the last part, note that every number can be written uniquely as a product of prime factors and that because the Fermat numbers are co-prime, each prime can appear in at most one Fermat number. Thus, since there are infinitely many Fermat numbers, there must be infinitely many primes.

Post-mortem

This question is largely about proof. The above solution has a proof by induction and a proof by contradiction. The last part could have been written as a proof by contradiction too.

The difficulty is in presenting the proof. It is not enough to understand your own proof: you have to be able to set it out so that it is clear to the reader. This is a vital part of mathematics and its most useful transferable skill.

At the meeting in which this question was first considered, one of the examiners said that he didn't see the point of the question, since there is already a well-known proof (due to Euclid, possibly) that there are infinitely many primes. As mentioned overleaf, this proof is due to Goldbach, and I could only respond that if Goldbach thought it worthwhile, that was good enough for me.

Problem 29: Telescoping series (✓✓)

(i) Show that the sum S_N of the first N terms of the series

$$\frac{1}{1.2.3} + \frac{3}{2.3.4} + \frac{5}{3.4.5} + \cdots + \frac{2n-1}{n(n+1)(n+2)} + \cdots$$

is

$$\frac{1}{2}\left(\frac{3}{2} + \frac{1}{N+1} - \frac{5}{N+2}\right).$$

What is the limit of S_N as $N \to \infty$?

(ii) The numbers a_n are such that

$$\frac{a_n}{a_{n-1}} = \frac{(n-1)(2n-1)}{(n+2)(2n-3)}.$$

Find an expression for $\dfrac{a_n}{a_1}$ and hence, or otherwise, evaluate $\sum\limits_{n=1}^{\infty} a_n$ when $a_1 = \frac{2}{9}$.

1998 Paper II

Comments

If you haven't the faintest idea how to do the sum, then look at the first line of the solution; but don't look without first having had a long think about it, picking up ideas from the form of the given answer.

Limits form an important part of first year university mathematics. The definition of a limit is one of the basic ideas in *analysis*, which is the rigorous study of calculus. At the end of part (i), no such definition is needed: you just see what happens when N gets very large (some terms get very small and eventually go away).

Part (ii) looks as if it might be some new idea. Since this is STEP, you will probably realise that the new series must be closely related to the series in part (i). The peculiar choice for a_1 ($= \frac{2}{9}$) should make you suspect that the sum will come out to some nice round number (not in fact round in this case, but straight and thin).

Having decided how to do the first part, please don't use the 'cover up' rule unless you understand why it works: mathematics at this level is not a matter of applying learned recipes. See the post-mortem for more thoughts on this matter.

Solution to problem 29

The given answer to the sum suggests partial fractions. It is difficult to think of any other way of starting, so let's convert the general term of the series to partial fractions in the hope that something good might happen. Set

$$\frac{2n-1}{n(n+1)(n+2)} \equiv \frac{A}{n} + \frac{B}{n+1} + \frac{C}{n+2}\,, \tag{$*$}$$

then use your favourite method to find $A = -\frac{1}{2}$, $B = 3$ and $C = -\frac{5}{2}$. Note that $A + B + C = 0$. The series can now be written

$$\left(\frac{A}{1} + \frac{B}{2} + \frac{C}{3}\right) + \left(\frac{A}{2} + \frac{B}{3} + \frac{C}{4}\right) + \left(\frac{A}{3} + \frac{B}{4} + \frac{C}{5}\right) +$$

$$+ \cdots + \left(\frac{A}{N-2} + \frac{B}{N-1} + \frac{C}{N}\right) + \left(\frac{A}{N-1} + \frac{B}{N} + \frac{C}{N+1}\right) + \left(\frac{A}{N} + \frac{B}{N+1} + \frac{C}{N+2}\right). \tag{$**$}$$

Now we collect up terms with the same denominators and find that all the terms in the series cancel, except those with denominators 1, 2, $N + 1$ and $N + 2$. These exceptions sum to the required answer.

The limit as $N \to \infty$ is $\frac{3}{4}$ since the other two terms obviously tend to zero.

For part (ii), we note that $\dfrac{a_n}{a_{n-1}} = \dfrac{b_n}{b_{n-1}}$, where b_n is the general term of the series in part (i). Thus

$$\frac{a_n}{a_1} = \frac{b_n}{b_1} \quad \text{and} \quad \sum_1^\infty a_n = \frac{a_1}{b_1}\sum_1^\infty b_n = \frac{2/9}{1/6} \times \frac{3}{4} = 1.$$

Alternatively, we can write out the nth term explicitly:

$$\begin{aligned} a_n &= \frac{(n-1)(2n-1)}{(n+2)(2n-3)}a_{n-1} = \frac{(n-1)(2n-1)}{(n+2)(2n-3)}\frac{(n-2)(2n-3)}{(n+1)(2n-5)}a_{n-2} \\ &= \frac{(n-1)(2n-1)}{(n+2)(2n-3)}\frac{(n-2)(2n-3)}{(n+1)(2n-5)}\frac{(n-3)(2n-5)}{(n)(2n-7)} \cdots \frac{5\times 11}{8\times 9}\frac{4\times 9}{7\times 7}\frac{3\times 7}{6\times 5}\frac{2\times 5}{5\times 3}\frac{1\times 3}{4\times 1}a_1 \\ &= \frac{2n-1}{n(n+1)(n+2)}\frac{3\times 2\times 1}{1}a_1 = \frac{12}{9}\frac{2n-1}{n(n+1)(n+2)}, \end{aligned}$$

all other terms cancelling. Now using the result of the first part gives $\sum\limits_1^\infty a_n = 1$.

Post-mortem

A small point of technique: equation ($**$) was made much clearer (and it saved writing) to stick with A, B and C instead of using $-\frac{1}{2}$, 3 and $-\frac{5}{2}$. The method would not depend on the arithmetic values of these constants.

There are various methods for finding A, B and C in ($*$).

One is to set $n = 1$, $n = 2$ and $n = 3$ consecutively and obtain three simultaneous equations.

Another is to multiply up and simplify, giving $2n - 1 = (A + B + C)n^2 + (3A + 2B + C)n + 2A$. You then equate coefficients of powers of n.

A better way is to multiply up without simplifying, giving $2n - 1 = A(n + 1)(n + 2) + Bn(n + 2) + Cn(n + 1)$. You then choose values for n that give quick results: for example, setting $n = -1$ gives $B = 3$ immediately. This is of course the method behind the iniquitous 'cover-up rule'. Note that the 'equivalence' sign, $\equiv$, indicates an identity (something that holds for all values of n) rather than an equation to solve for n, so n doesn't have to be a positive integer (you could set $n = -\frac{1}{2}$ if you fancied it).

Problem 30: Integer solutions of cubics (✓✓✓)

(i) Show that, if m is an integer such that

$$(m-3)^3 + m^3 = (m+3)^3, \qquad (*)$$

then m^2 is a factor of 54. Deduce that there is no integer m which satisfies the equation $(*)$.

(ii) Show that, if n is an integer such that

$$(n-6)^3 + n^3 = (n+6)^3, \qquad (**)$$

then n is even. By writing $n = 2m$ deduce that there is no integer n which satisfies the equation $(**)$.

(iii) Show that, if n is an integer such that

$$(n-a)^3 + n^3 = (n+a)^3\ , \qquad (***)$$

where a is a non-zero integer, then n is even and a is even. Deduce that there is no integer n which satisfies the equation $(***)$.

1998 Paper II

Comments

I slightly simplified the first two parts of the question, which comprised the whole of the original STEP question, and added the third part. This last part is conceptually tricky and very interesting: hence the ✓✓✓ rating.

It is a fairly standard technique in this sort of number theoretic problem to investigate whether the equation *balances*: for example, if the left-hand side is even, then the right-hand side must also be even. If this fails, then the equation can have no solutions.

Solution to problem 30

(i) Simplifying gives $m^2(m-18) = 54$.

Both m^2 and $(m-18)$ must divide 54, which is impossible since the only squares that divide 54 are 1 and 9, and neither $m = 1$ nor $m = 3$ satisfies $m^2(m-18) = 54$.

You could also argue that $m - 18$ must be positive so $m \geqslant 19$ and $m^2 \geqslant 361$ which is a contradiction.

(ii) The easiest method is a proof by contradiction. Suppose therefore that n is odd. The two terms on the left-hand side are both odd, which means that the left-hand side is even. But the right-hand side is odd so the equation cannot balance if n is odd.

Setting $n = 2m$ gives

$$(2m-6)^3 + (2m)^3 = (2m+6)^3.$$

Taking a factor of 2^3 out of each term leaves $(m-3)^3 + m^3 = (m+3)^3$ which is the same as the equation that was shown in part (i) to have no solutions.

(iii) First we show that n is even, dealing with the cases a odd and a even separately. The first case is n odd and a odd. In that case, $(n-a)^3$ is even, n^3 is odd and $(n+a)^3$ is even, so the equation does not balance. In the second case, n is odd and a is even, the three terms are all odd and again the equation does not balance. Therefore n cannot be odd.

Next, we investigate the case n is even and a odd. This time the three terms are odd, even and odd respectively so there is no contradiction. But multiplying out the brackets and simplifying gives

$$n^2(n-6a) = 2a^3 \, .$$

Since n is even the left-hand side is divisible by 4, because of the factor n^2 . That means that a^3 is divisible by 2 which cannot be the case since a is odd.

Now that we know n and a are both even, we can follow the method used in part (ii) and set $n = 2m$ and $a = 2b$. This gives

$$(2m-2b)^3 + (2m)^3 = (2m+2b)^3$$

from which a factor of 2^3 can be cancelled from each term. Thus m and b satisfy the same equation as n and a. They are therefore both even and we can repeat the process.

Repeating the process again and again will eventually result in an integer that is odd which will therefore not satisfy the equation that it is supposed to satisfy: a contradiction. There is therefore no integer n that satisfies $(***)$.

Post-mortem

The last proof is an example of what is known as the *method of infinite descent*. It was used by Fermat (1601 – 1665) to prove special cases of his Last Theorem[22], which of course is exactly what you are doing in this question. The method was probably invented by him and his faith in it sometimes led him astray. It is even possible that he thought he could use it to prove his Last Theorem in full. In fact, the proof of this theorem was only given in 1994 by Andrew Wiles; and it is 150 pages of pretty incomprehensible modern mathematics.

---

[22] The theorem says that the equation $x^n + y^n = z^n$, where x , y , z and n are positive integers, can only hold if $n = 1$ or 2.

Problem 31: The harmonic series (✓✓✓)

The function f satisfies $0 \leqslant \mathrm{f}(t) \leqslant K$ when $0 \leqslant t \leqslant 1$. Explain by means of a sketch, or otherwise, why

$$0 \leqslant \int_0^1 \mathrm{f}(t)\,\mathrm{d}t \leqslant K\,.$$

By considering $\displaystyle\int_0^1 \frac{t}{n(n-t)}\,\mathrm{d}t$ or otherwise show that, if $n > 1$, then

$$0 \leqslant \ln\left(\frac{n}{n-1}\right) - \frac{1}{n} \leqslant \frac{1}{n-1} - \frac{1}{n}$$

and deduce that

$$0 \leqslant \ln N - \sum_{n=2}^{N} \frac{1}{n} \leqslant 1\,.$$

Deduce that $\displaystyle\sum_{n=1}^{N} \frac{1}{n} \to \infty$ as $N \to \infty$.

Noting that $2^{10} = 1024$, show also that if $N < 10^{30}$ then $\displaystyle\sum_{n=1}^{N} \frac{1}{n} < 101$.

1999 Paper I

Comments

Quite a lot of different ideas are required for this question; hence ✓✓✓.

The first hurdle is to decide what the constant K in the first part has to be in order to make the second part work. You might try to maximise the integrand using calculus, but that would be the wrong thing to do (you should find if you do this that the integrand has no turning point in the given range: it increases throughout the range).

The next hurdle is the third displayed equation, which follows from the preceding result. Then there is still more work to do.

It is not at all obvious that the series $\displaystyle\sum_1^N \frac{1}{n}$ (which is called the *harmonic series*) tends to infinity as N increases. There are easier ways of proving this than the method used in this question, but this way tells us two interesting things. First it tells us that the sum increases very slowly indeed: the first 10^{30} terms only get to 100. Second it tells us that

$$0 \leqslant \sum_1^N \frac{1}{n} - \ln N \leqslant 1$$

for all N. (This is the third displayed equation in the question slightly rewritten.) Not only does the sum diverge, it does so logarithmically. In fact, in the limit $N \to \infty$,

$$\sum_1^N \frac{1}{n} - \ln N \to \gamma$$

where γ is *Euler's constant*. Its value is about $\frac{1}{2}$.

Solution to problem 31

Any sketch showing a squiggly curve all of which lies beneath the line $x = K$ and above the x-axis will do: area under curve (the integral) is less than the area of the rectangle (K).

Let's start the next part in the obvious way, by evaluating the integral.

Noting that $\dfrac{t}{n-t} = \dfrac{n}{n-t} - 1$, we have:

$$\int_0^1 \frac{t}{n(n-t)}\mathrm{d}t = \int_0^1 \left(\frac{1}{n-t} - \frac{1}{n}\right)\mathrm{d}t = \ln\left(\frac{n}{n-1}\right) - \frac{1}{n}\,.$$

The largest value of the integrand in the interval $[0, 1]$ occurs at $t = 1$, since the numerator increases as t increases and the denominator decreases as t increases (remember that $n > 1$). Using the result of the first paragraph gives

$$0 \leqslant \ln\left(\frac{n}{n-1}\right) - \frac{1}{n} \leqslant \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}\,.$$

Summing both sides from 2 to N and cancelling lots of terms in pairs gives

$$0 \leqslant \ln N - \sum_2^N \frac{1}{n} \leqslant 1 - \frac{1}{N}. \qquad (*)$$

Note that $1 - \dfrac{1}{N} < 1$. Since $\ln N \to \infty$ as $N \to \infty$, so also must $\displaystyle\sum_2^N \frac{1}{n}$ (it differs from the logarithm by less than 1).

Finally, rearranging the first inequality in ($*$) gives $\displaystyle\sum_2^N \frac{1}{n} \leqslant \ln N$, i.e. $\displaystyle\sum_1^N \frac{1}{n} \leqslant \ln N + 1$. Setting $N = 10^{30}$, and using the inequalities $1000 < 1024$ and $\mathrm{e} > 2$ (or $\ln 2 < 1$) gives:

$$\sum_1^{10^{30}} \frac{1}{n} \leqslant \ln(10^{30}) + 1 < \ln(1024^{10}) + 1 = 100 \ln 2 + 1 < 100 \ln \mathrm{e} + 1 = 100 + 1\,.$$

Post-mortem

This question seems very daunting at first, because you are asked to prove a sequence of completely unfamiliar results. However, you should learn from this question that if you keep cool and follow the hints, explicit and implicit, you can achieve some surprisingly sophisticated results. You might think that this situation is very artificial: in real life, you do not receive hints to guide you. But often in mathematical research, hints *are* buried deep in the problem if only you can recognise them.

Problem 32: Integration by substitution (✓✓)

Find $\dfrac{\mathrm{d}y}{\mathrm{d}x}$ if

$$y = \frac{ax+b}{cx+d}\,. \qquad (*)$$

By using changes of variable of the form $(*)$, or otherwise, show that

$$\int_0^1 \frac{1}{(x+3)^2}\,\ln\left(\frac{x+1}{x+3}\right)\mathrm{d}x = \frac{1}{6}\ln 3 - \frac{1}{4}\ln 2 - \frac{1}{12}\,,$$

and evaluate the integrals

$$\int_0^1 \frac{1}{(x+3)^2}\,\ln\left(\frac{x^2+3x+2}{(x+3)^2}\right)\mathrm{d}x \quad \text{and} \quad \int_0^1 \frac{1}{(x+3)^2}\,\ln\left(\frac{x+1}{x+2}\right)\mathrm{d}x\,.$$

1999 Paper II

Comments

You will find that the change of variable in each case is clearly signalled: it is really only the denominator of $(*)$ that matters.

For the first integral, you do the obvious thing, but for the second and third integrals you have to be quite ingenious to get the argument of the logs in a suitable form. Once you have got the idea for the second integral, you should be able to see the connection with the third integral, but it would be hard to do the third integral without having done the second. That's what I like about this question: one thing leads to another.

I have written below the version of this question that was proposed by the setter, because I thought you would be interested to see how a question evolves.

By changing to the variable y defined by

$$y = \frac{2x-3}{x+1}\,,$$

evaluate the integral

$$\int_2^4 \frac{2x-3}{(x+1)^3}\,\ln\left(\frac{2x-3}{x+1}\right)\mathrm{d}x\,.$$

Evaluate the integral

$$\int_9^{25} \left(2z^{-\frac{3}{2}} - 5z^{-2}\right)\ln\left(2 - 5z^{-\frac{1}{2}}\right)\mathrm{d}z\,.$$

Note in particular the way that the ideas in the final draft are closely knit and better structured: the first change of variable is strongly signalled but not explicit and the following parts, though based on the same idea, require increasing ingenuity. The first draft required quite a jump to evaluate the second integral.

Note also that the final integral of the first draft has an unpleasant contrived appearance, whereas the the integrals of the final draft are rather pleasing: beauty matters to mathematicians.