

Copyrighted Material

OXFORD
MATHEMATICS

solving mathematical problems

a personal perspective

TERENCE TAO

Copyrighted Material

Preface to the first edition

Proclus, an ancient Greek philosopher, said:

This therefore, is mathematics: she reminds you of the invisible forms of the soul; she gives life to her own discoveries; she awakens the mind and purifies the intellect; she brings to light our intrinsic ideas; she abolishes oblivion and ignorance which are ours by birth ...

But I just like mathematics because it is fun.

Mathematical problems, or puzzles, are important to real mathematics (like solving real-life problems), just as fables, stories, and anecdotes are important to the young in understanding real life. Mathematical problems are 'sanitized' mathematics, where an elegant solution has already been found (by someone else, of course), the question is stripped of all superfluousness and posed in an interesting and (hopefully) thought-provoking way. If mathematics is likened to prospecting for gold, solving a good mathematical problem is akin to a 'hide-and-seek' course in gold-prospecting: you are given a nugget to find, and you know what it looks like, that it is out there somewhere, that it is not too hard to reach, that it is unearthing within your capabilities, and you have conveniently been given the right equipment (i.e. data) to get it. It may be hidden in a cunning place, but it will require ingenuity rather than digging to reach it.

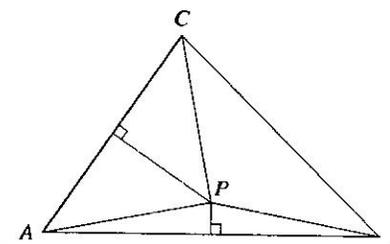
In this book I shall solve selected problems from various levels and branches of mathematics. Starred problems (*) indicate an additional level of difficulty, either because some higher mathematics or some clever thinking are required; double-starred questions (**) are similar, but to a greater degree. Some problems have additional exercises at the end that can be solved in a similar manner or involve a similar piece of mathematics. While solving these problems, I will try to demonstrate some tricks of the trade when problem-solving. Two of the main weapons—experience and knowledge—are not easy to put into a book: they have to be acquired over time. But there are many simpler tricks that take less time to learn. There are ways of looking at a problem that make it easier to find a feasible attack plan. There are systematic ways of reducing a problem into successively simpler sub-problems. But, on the other hand, solving the problem is not everything. To return to the gold nugget analogy, strip-mining the neighbourhood with bulldozers is clumsier than doing a careful survey, a bit of

geology, and a small amount of digging. A solution should be relatively short, understandable, and hopefully have a touch of elegance. It should also be fun to discover. Transforming a nice, short little geometry question into a ravaging monster of an equation by textbook coordinate geometry does not have the same taste of victory as a two-line vector solution.

As an example of elegance, here is a standard result in Euclidean geometry:

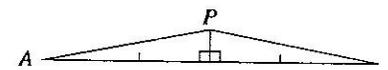
Show that the perpendicular bisectors of a triangle are concurrent.

This neat little one-liner *could* be attacked by coordinate geometry. Try to do so for a few minutes (hours?), then look at this solution:



PROOF. Call the triangle ABC . Now let P be the intersection of the perpendicular bisectors of AB and AC . Because P is on the AB bisector, $|AP| = |PB|$. Because P is on the AC bisector, $|AP| = |PC|$. Combining the two, $|BP| = |PC|$. But this means that P has to be on the BC bisector. Hence all three bisectors are concurrent. (Incidentally, P is the circumcentre of ABC .) \square

The following reduced diagram shows why $|AP| = |PB|$ if P is on the AB perpendicular bisector: congruent triangles will pull it off nicely.



This kind of solution—and the strange way that obvious facts mesh to form a not-so-obvious fact—is part of the beauty of mathematics. I hope that you too will appreciate this beauty.

Acknowledgements

Thanks to Peter O'Halloran, Vern Treilibs, and Lenny Ng for their contributions of problems and advice.

Special thanks to Basil Rennie for his corrections and ingenious shortcuts in solutions, and finally thanks to my family for their support, encouragement, spelling corrections, and put-downs when I was behind schedule.

Almost all of the problems in this book come from published collections of problem sets for mathematics competitions. These are sourced in the texts, with full details given in the reference section of the book. I also used a small handful of problems from friends or from various mathematical publications; these have no source listed.

Preface to the second edition

This book was written 15 years ago; literally half a lifetime ago, for me. In the intervening years, I have left home, moved to a different country, gone to graduate school, taught classes, written research papers, advised graduate students, married my wife, and had a son. Clearly, my perspective on life and on mathematics is different now than it was when I was 15. I have not been involved in problem-solving competitions for a very long time now, and if I were to write a book now on the subject it would be very different from the one you are reading here.

Mathematics is a multifaceted subject, and our experience and appreciation of it changes with time and experience. As a primary school student, I was drawn to mathematics by the abstract beauty of formal manipulation, and the remarkable ability to repeatedly use simple rules to achieve non-trivial answers. As a high-school student, competing in mathematics competitions, I enjoyed mathematics as a sport, taking cleverly designed mathematical puzzle problems (such as those in this book) and searching for the right 'trick' that would unlock each one. As an undergraduate, I was awed by my first glimpses of the rich, deep, and fascinating theories and structures which lie at the core of modern mathematics today. As a graduate student, I learnt the pride of having one's own research project, and the unique satisfaction that comes from creating an original argument that resolved a previously open question. Upon starting my career as a professional research mathematician, I began to see the intuition and motivation that lay behind the theories and problems of modern mathematics, and was delighted when realizing how even very complex and deep results are often at heart be guided by very simple, even common-sensical, principles. The 'Aha!' experience of grasping one of these principles, and suddenly seeing how it illuminates and informs a large body of mathematics, is a truly remarkable one. And there are yet more aspects of mathematics to discover; it is only recently for me that I have grasped enough fields of mathematics to begin to get a sense of the endeavour of modern mathematics as a unified subject, and how it connects to the sciences and other disciplines.

As I wrote this book before my professional mathematics career, many of these insights and experiences were not available to me, and so in many places the exposition has a certain innocence, or even naivety. I have been reluctant to tamper too much with this, as my younger self was almost

certainly more attuned to the world of the high-school problem solver than I am now. However, I have made a number of organizational changes: formatting the text into L^AT_EX, arranging the material into what I believe is a more logical order, and editing those parts of the text which were inaccurate, badly worded, confusing, or unfocused. I have also added some more exercises. In some places, the text is a bit dated (Fermat's last theorem, for instance, has now been proved rigorously), and I now realize that several of the problems here could be handled more quickly and cleanly by more 'high-tech' mathematical tools; but the point of this text is not to present the slickest solution to a problem or to provide the most up-to-date survey of results, but rather to show how one approaches a mathematical problem for the first time, and how the painstaking, systematic experience of trying some ideas, eliminating others, and steadily manipulating the problem can lead, ultimately, to a satisfying solution.

I am greatly indebted to Tony Gardiner for encouraging and supporting the reprinting of this book, and to my parents for all their support over the years. I am also touched by all the friends and acquaintances I have met over the years who had read the first edition of the book. Last, but not least, I owe a special debt to my parents and the Flinders Medical Centre computer support unit for retrieving a 15-year old electronic copy of this book from our venerable Macintosh Plus computer!

Terence Tao
 Department of Mathematics,
 University of California, Los Angeles
 December 2005



Strategies in problem solving

The journey of a thousand miles begins with one step.
 Lao Tzu

Like and unlike the proverb above, the solution to a problem begins (and continues, and ends) with simple, logical steps. But as long as one steps in a firm, clear direction, with long strides and sharp vision, one would need far, far less than the millions of steps needed to journey a thousand miles. And mathematics, being abstract, has no physical constraints; one can always restart from scratch, try new avenues of attack, or backtrack at an instant's notice. One does not always have these luxuries in other forms of problem-solving (e.g. trying to go home if you are lost).

Of course, this does not necessarily make it easy; if it was easy, then this book would be substantially shorter. But it makes it possible.

There are several general strategies and perspectives to solve a problem correctly; (Polya 1957) is a classic reference for many of these. Some of these strategies are discussed below, together with a brief illustration of how each strategy can be used on the following problem:

PROBLEM 1.1. A triangle has its lengths in an arithmetic progression, with difference d . The area of the triangle is t . Find the lengths and angles of the triangle.

Understand the problem. What kind of problem is it? There are three main types of problems:

- 'Show that ...' or 'Evaluate ...' questions, in which a certain statement has to be proved true, or a certain expression has to be worked out;
- 'Find a ...' or 'Find all ...' questions, which requires one to find something (or everything) that satisfies certain requirements;
- 'Is there a ...' questions, which either require you to prove a statement or provide a counterexample (and thus is one of the previous two types of problem).

The type of problem is important because it determines the basic method of approach. 'Show that ...' or 'Evaluate ...' problems start with given data and the objective is to deduce some statement or find the value of

an expression; this type of problem is generally easier than the other two types because there is a clearly visible objective, one that can be deliberately approached. 'Find a ...' questions are more hit-and-miss; generally one has to guess one answer that nearly works, and then tweak it a bit to make it more correct; or alternatively one can alter the requirements that the object-to-find must satisfy, so that they are easier to satisfy. 'Is there a ...' problems are typically the hardest, because one must first make a decision on whether an object exists or not, and provide a proof on one hand, or a counter-example on the other.

Of course, not all questions fall into these neat categories; but the general format of any question will still indicate the basic strategy to pursue when solving a problem. For example, if one tries to solve the problem 'find a hotel in this city to sleep in for the night', one should alter the requirements to, say 'find a vacant hotel within 5 kilometres with a room that costs less than 100\$ a night' and then use pure elimination. This is a better strategy than proving that such a hotel does or does not exist, and is probably a better strategy than picking any handy hotel and trying to prove that one can sleep in it.

In Problem 1.1 question, we have an 'Evaluate ...' type of problem. We need to find several unknowns, given other variables. This suggests an algebraic solution rather than a geometric one, with a lot of equations connecting d , t , and the sides and angles of the triangle, and eventually solving for our unknowns.

Understand the data. What is given in the problem? Usually, a question talks about a number of objects which satisfy some special requirements. To understand the data, one needs to see how the objects and requirements react to each other. This is important in focusing attention on the proper techniques and notation to handle the problem. For example, in our sample question, our data are a triangle, the area of the triangle, and the fact that the sides are in an arithmetic progression with separation d . Because we have a triangle, and are considering the sides and area of it, we would need theorems relating sides, angles, and areas to tackle the question: the sine rule, cosine rule, and the area formulas, for example. Also, we are dealing with an arithmetic progression, so we would need some notation to account for that; for example, the side lengths could be a , $a + d$, and $a + 2d$.

Understand the objective. What do we want? One may need to find an object, prove a statement, determine the existence of an object with special properties, or whatever. Like the flip side of this strategy, 'understand the data', knowing the objective helps focus attention on the best weapons to use. Knowing the objective also helps in creating tactical goals which we know will bring us closer to solving the question. Our example question has the objective of 'find all the sides and angles of the triangle'. This means, as mentioned before, that we will need theorems and results concerning sides and angles. It also gives us the tactical goal of 'find equations involving the sides and angles of the triangle'.

Select good notation. Now that we have our data and objective, we must represent it in an efficient way, so that the data and objective are both represented as simply as possible. This usually involves the thoughts of the past two strategies. In our sample question, we are already thinking of equations involving d , t , and the sides and angles of the triangle. We need to express the sides and angles in terms of variables: one could choose the sides to be a , b , and c , while the angles could be denoted α , β , γ . But we can use the data to simplify the notation: we know that the sides are in arithmetic progression, so instead of a , b , and c , we can have a , $a + d$, and $a + 2d$ instead. But the notation can be even better if we make it more symmetrical, by making the side lengths $b - d$, b , and $b + d$. The only slight drawback to this notation is that b is forced to be larger than d . But on further thought, we see that this is actually not a restriction; in fact the knowledge that $b > d$ is an extra piece of data for us. We can also trim the notation more, by labelling the angles α , β , and $180^\circ - \alpha - \beta$, but this is ugly and unsymmetrical—it is probably better to keep the old notation, but bearing in mind that $\alpha + \beta + \gamma = 180^\circ$.

Write down what you know in the notation selected; draw a diagram. Putting everything down on paper helps in three ways:

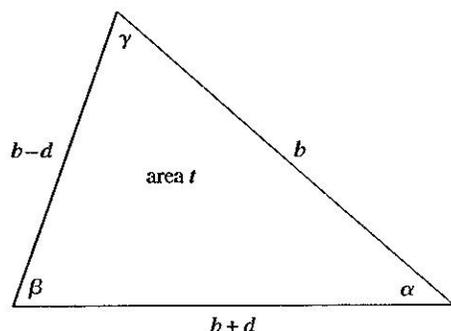
- you have an easy reference later on;
- the paper is a good thing to stare at when you are stuck;
- the physical act of writing down of what you know can trigger new inspirations and connections.

Be careful, though, of writing superfluous material, and do not overload your paper with minutiae; one compromise is to highlight those facts which you think will be most useful, and put more questionable, redundant, or crazy ideas in another part of your scratch paper. Here are some equations and inequalities one can extract from our example question:

- (physical constraints) $\alpha, \beta, \gamma, t > 0$, and $b \geq d$; we can also assume $d \geq 0$ without loss of generality;
- (sum of angles in a triangle) $\alpha + \beta + \gamma = 180^\circ$;
- (sine rule) $(b - d) / \sin \alpha = b / \sin \beta = (b + d) / \sin \gamma$;
- (cosine rule) $b^2 = (b - d)^2 + (b + d)^2 - 2(b - d)(b + d) \cos \beta$, etc.;
- (area formula) $t = (1/2)(b - d)b \sin \gamma = (1/2)(b - d)(b + d) \sin \beta = (1/2)b(b + d) \sin \alpha$;
- (Heron's formula) $t^2 = s(s - b + d)(s - b)(s - b - d)$, where $s = ((b - d) + b + (b + d))/2$ is the semiperimeter;
- (triangle inequality) $b + d \leq b + (b - d)$.

Many of these facts may prove to be useless or distracting. But we can use some judgement to separate the valuable facts from the unhelpful ones. The equalities are likely to be more useful than the inequalities, since our objective and data come in the form of equalities. And Heron's formula looks especially promising, because the semiperimeter simplifies to $s = 3b/2$. So we can highlight 'Heron's formula' as being likely to be useful.

We can of course also draw a picture. This is often quite helpful for geometry questions, though in this case the picture does not seem to add much:



Modify the problem slightly. There are many ways to vary a problem into one which may be easier to deal with:

- Consider a special case of the problem, such as extreme or degenerate cases.
- Solve a simplified version of the problem.
- Formulate a conjecture which would imply the problem, and try to prove that first.
- Derive some consequence of the problem, and try to prove that first.
- Reformulate the problem (e.g. take the contrapositive, prove by contradiction, or try some substitution).
- Examine solutions of similar problems.
- Generalize the problem.

This is useful when you cannot even get started on a problem, because solving for a simpler related problem sometimes reveals the way to go on the main problem. Similarly, considering extreme cases and solving the problem with additional assumptions can also shed light on the general solution. But be warned that special cases are, by their nature, special, and some elegant technique could conceivably apply to them and yet have absolutely no utility in solving the general case. This tends to happen when the special case is *too* special. Start with modest assumptions first, because then you are sticking as closely as possible to the spirit of the problem.

In Problem 1.1, we can try a special case such as $d = 0$. In this case we need to find the side length of an equilateral triangle of area t . In this case, it is a standard matter to compute the answer, which is $b = 2t^{1/2}/3^{1/4}$. This indicates that the general answer should also involve square roots or fourth roots, but does not otherwise suggest how to go about the problem. Consideration of similar problems draws little as well, except one gets further evidence that a gung-ho algebraic attack is what is needed.

Modify the problem significantly. In this more aggressive type of strategy, we perform major modifications to a problem such as removing data, swapping the data with the objective, or negating the objective (e.g. trying to disprove a statement rather than prove it). Basically, we try to push the problem until it breaks, and then try to identify where the breakdown occurred; this identifies what the key components of the data are, as well as where the main difficulty will lie. These exercises can also help in getting an instinctive feel of what strategies are likely to work, and which ones are likely to fail.

In regard to our particular question, one could replace the triangle with a quadrilateral, circle, etc. Not much help there: the problem just gets more complicated. But on the other hand, one can see that one does not really need a triangle in the question, but just the dimensions of the triangle. We do not really need to know the position of the triangle. So here is further confirmation that we should concentrate on the sides and angles (i.e. $a, b, c, \alpha, \beta, \gamma$) and not on coordinate geometry or similar approaches.

We could omit some objectives; for example, instead of working out all the sides and angles we could work out just the sides, for example. But then one can notice that by the cosine and sine rules, the angles of the triangle will be determined anyway. So it is only necessary to solve for the sides. But we know that the sides have lengths $b - d$, b , and $b + d$, so we only need to find what b is to finish the problem.

We can also omit some data, like the arithmetic difference d , but then we seem to have several possible solutions, and not enough data to solve the problem. Similarly, omitting the area t will not leave enough data to clinch a solution. (Sometimes one can *partially* omit data, for instance, by only specifying that the area is larger or smaller than some threshold t_0 ; but this is getting complicated. Stick with the simple options first.) Reversal of the problem (swapping data with objective) leads to some interesting ideas though. Suppose you had a triangle with arithmetic difference d , and you wanted to shrink it (or whatever) until the area becomes t . One could imagine our triangle shrinking and deforming, while preserving the arithmetic difference of the sides. Similarly, one could consider all triangles with a fixed area, and mold the triangle into one with the sides in the correct arithmetic progression. These ideas could work in the long run: but I will solve this question by another approach. Do not forget, though,

that a question can be solved in more than one way, and no particular way can really be judged the absolute best.

Prove results about our question. Data is there to be used, so one should pick up the data and play with it. Can it produce more meaningful data? Also, proving small results could be beneficial later on, when trying to prove the main result or to find the answer. However small the result, do not forget it—it could have bearing later on. Besides, it gives you something to do if you are stuck.

In a 'Evaluate ...' problem like the triangle question, this tactic is not as useful. But one can try. For example, our tactical goal is to solve for b . This depends on the two parameters d and t . In other words, b is really a function: $b = b(d, t)$. (If this notation looks out of place in a geometry question, then that is only because geometry tends to ignore the functional dependence of objects. For example, Heron's formula gives an explicit form for the area A in terms of the sides a , b , and c : in other words, it expresses the function $A(a, b, c)$.) Now we can prove some mini-results about this function $b(d, t)$, such as $b(d, t) = b(-d, t)$ (because an arithmetic progression has an equivalent arithmetic progression with inverted arithmetic difference), or $b(kd, k^2t) = kb(d, t)$ (this is done by dilating the triangle that satisfies $b = b(d, t)$ by k). We could even try differentiate b with respect to d or t . For this particular problem, these tactics allow us to perform some normalizations, for instance setting $t = 1$ or $d = 1$, and also provide a way to check the final answer. However, in this problem these tricks turn out to only give minor advantages and we will not use them here.

Simplify, exploit data, and reach tactical goals. Now we have set up notation and have a few equations, we should seriously look at attaining our tactical goals that we have established. In simple problems, there are usually standard ways of doing this: (For example, algebraic simplification is usually discussed thoroughly in high-school level textbooks.) Generally, this part is the longest and most difficult part of the problem: however, once can avoid getting lost if one remembers the relevant theorems, the data and how they can be used, and most importantly the objective. It is also a good idea to not apply any given technique or method blindly, but to think ahead and see where one could hope such a technique to take one; this can allow one to save enormous amounts of time by eliminating unprofitable directions of inquiry before sinking lots of effort into them, and conversely to give the most promising directions priority.

In Problem 1.1, we are already concentrating on Heron's formula. We can use this to attain our tactical goal of solving for b . After all, we have already noted that the sine and cosine rules can determine α, β, γ once b is known. As further evidence that this is going to be a step forward, note that Heron's formula involves d and t —in essence, it uses all our data (we have already incorporated the fact about the sides being in arithmetic progression

into our notation). Anyway, Heron's formula in terms of d, t, b becomes

$$t^2 = \frac{3b}{2} \left(\frac{3b}{2} - b + d \right) \left(\frac{3b}{2} - b \right) \left(\frac{3b}{2} - b - d \right)$$

which we can simplify to

$$t^2 = \frac{3b^2(b-2d)(b+2d)}{16} = \frac{3b^2(b^2-4d^2)}{16}.$$

Now we have to solve for b . The right-hand side is a polynomial in b (treating d and t as constants), and in fact it is a quadratic in b^2 . Now quadratics can be solved easily: if we put clear denominators and put everything on the left-hand side we get

$$3b^4 - 12d^2b^2 - 16t^2 = 0$$

so, using the quadratic formula,

$$b^2 = \frac{12d^2 \pm \sqrt{144d^4 + 196t^2}}{6} = 2d^2 \pm \sqrt{4d^2 + \frac{16}{3}t^2}.$$

Because b has to be positive, we get

$$b = \sqrt{2d^2 + \sqrt{4d^2 + \frac{16}{3}t^2}},$$

as a check, we can verify that when $d = 0$ this agrees with our previous computation of $b = 2t^{1/2}/3^{1/4}$. Once we compute the sides $b-d, b, b+d$, the evaluation of the angles α, β, γ then follows from the cosine laws, and we are done!

Authored by a leading name in mathematics, this engaging and clearly presented text leads the reader through the various tactics involved in solving mathematical problems at the Mathematical Olympiad level. Covering number theory, algebra, analysis, Euclidean geometry, and analytic geometry, *Solving Mathematical Problems* includes numerous exercises and model solutions throughout. Assuming only basic high-school mathematics, the text is ideal for general readers and students of 14 years and above with an interest in pure mathematics.

Terence Tao was born in Adelaide, Australia, in 1975. In 1987, 1988, and 1989 he competed in the International Mathematical Olympiad for the Australian team, winning a bronze, silver, and gold medal respectively, and being the youngest competitor ever to win a gold medal at this event. Since 2000, Terence has been a full professor of mathematics at the University of California, Los Angeles. He now lives in Los Angeles with his wife and son.

CONTENTS

- Preface
- 1 Strategies in problem solving
- 2 Examples in number theory
- 3 Examples in algebra and analysis
- 4 Euclidean geometry
- 5 Analytic geometry
- 6 Sundry examples
- References

"I must say I find it delightful . . . I have absolutely no doubt that this book will be popular."

Professor W T Gowers, University of Cambridge

ALSO AVAILABLE FROM OXFORD UNIVERSITY PRESS

The Mathematical Olympiad Handbook:

An introduction to problem solving based on the first 32 British Mathematical Olympiads 1965–1996

Tony Gardiner

Challenges in Geometry

for Mathematical Olympians past and present

Christopher J. Bradley

© Gower, Reprinted by permission of Oxford University Press

OXFORD
UNIVERSITY PRESS

www.oup.com

Copyrighted Material

ISBN 0-19-920560-4



9 780199 205608