Comments on mathematical proofs

> If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts, he shall end in certainties.
>
> Francis Bacon **(1561 – 1626)**, ***The Advancement of Learning*** **(1605), I.v. 8**

Since deductive reasoning and logical proofs play an extremely important role in the mathematical sciences, we shall summarize some of the main points here, describing some of the standard methods and strategies, illustrating these with some examples from high school mathematics and calculus, and pointing out some common mistakes and how to avoid them. Since we are simply trying to illustrate the techniques, our setting for now is informal, and in particular for the time being we shall not worry about things like how one proves the Intermediate Value Theorem that plays such an important role in calculus. This is technically an example of a concept called ***local deduction***, in which one only shows how to get from point **A** to point **B**, postponing questions about reaching point **A** to another time or place.

Some proofs use ***direct arguments***, while others use ***indirect arguments.*** The direct arguments are often the simplest, and many simple problem solving methods from elementary mathematics (algebra, in particular) are really just very simple examples of direct proofs.

Example. If $2x + 1 = 5$, show that $x = 4$. **SOLUTION:** If $2x + 1 = 5$, then by subtracting **1** from each side we obtain $2x = 4$. Next, if we divide both sides of the equation $2x = 4$ by 2, we obtain $x = 2$.

In contrast, an **indirect** argument usually involves considering the negation of either the hypothesis or the conclusion. This generally involves ***proof by contradiction*** or ***reductio ad absurdum***, in which one assumes the conclusion is false and then proves part of the hypothesis is false.

> Reductio ad absurdum ... is one of a mathematician's finest weapons. It is a far finer gambit than any chess gambit; a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game.
>
> G. H. Hardy (1877 – 1947), ***A Mathematician's Apology***

The indirect approach is related to the law of the ***contrapositive***: *A statement* **P** $\Rightarrow$ **Q** *is true if and only if the* ***contrapositive*** *statement* **not Q** $\Rightarrow$ **not P** *is true.*

> When you have eliminated the impossible, whatever remains, however improbable [or questionable it may seem], must be the truth.
>
> A. C. Doyle (1859 – 1930), ***Sherlock Holmes*** – ***Sign of the Four***

A general "rule of thumb" is to consider using an indirect argument if either no way of using a direct argument is apparent or if a direct approach seems to be getting very long and complicated. There is no guarantee that an indirect argument will be any better, but if you get stuck trying a direct approach there often is not much to lose by seeing what happens if you try an indirect approach; in some cases, attempts to give an indirect argument may even lead to a valid or better direct proof.

Example. Show that if **L** and **M** are two distinct lines, then they have at most one point in common.

SOLUTION: Suppose the conclusion is false, so that **x** and **y** are two distinct points on both **L** and **M**. Then both **L** and **M** are lines containing these two points. Since there is only one line **N** containing the two distinct points **x** and **y**, we know that **L** must be equal to **N** and similarly **M** must be equal to **N**, which means that **L** and **M** must be equal. This contradicts our original assumption; the problem arose because we added an assumption that **x** and **y** belonged to both lines. Therefore **L** and **M** cannot have two (or more) points in common.

An important step in such indirect arguments is to make sure that the negation of the conclusion is accurately stated. Mistakes in stating the negation usually lead to mistakes in arguments intended to prove the original result.

Forward and backwards reasoning. Very often it is helpful to work backwards as well as forwards (see the quotation below). For example, if you want to show that **P** implies **Q**, in some cases it might be easier to find some statement **R** that implies **Q**, and then to see if it is possible to prove that **P** implies **R**. Of course, there may be several intermediate steps of this type.

> In solving a problem of this sort, the grand thing is to be able to reason backwards. That is a very useful accomplishment, and a very easy one, but people do not practice it much ... and ... [it] comes to be neglected.
>
> ***Sherlock Holmes – A Study in Scarlet***

Example. Show that the polynomial $f(x) = x^5 - x - 1$ has a real root.

SOLUTION: We know that polynomials are continuous and that continuous functions have the Intermediate Value Property. Therefore if we can show that the polynomial is positive for some value of x and negative for another, then we can conclude that this polynomial has a real root. One way of doing this is simply to calculate the value of the polynomial for several different values of the independent variable. If we do so, then we see that $f(1) = -1$ and $f(2) = 29$. Therefore we know that $f(x)$ has a root, and in fact by the Intermediate Value Theorem from first year calculus we know there is a root which lies somewhere between 1 and 2.

Proofs by cases. Frequently it is convenient to break things up into all the different cases and to check them individually, and in some cases this is simply unavoidable.

Example. Let $\mathrm{sgn}(x)$ be the function whose value is 1 if x is positive, -1 if x is negative, and 0 if $x = 0$. Prove that $\mathrm{sgn}(xy) = \mathrm{sgn}(x)\,\mathrm{sgn}(y)$.

There are three possibilities for x (positive, negative, zero) and likewise for y, leading to the following list of nine possibilities for x and y:

$$[+,+],\ [+,0],\ [+,-],\ [0,+],\ [0,0],\ [0,-],\ [-,+],\ [-,0],\ [-,-]$$

One can then handle each case (or various classes of cases) separately; for example, the five cases where at least one number is zero follow because in all these cases we have $xy = \mathrm{sgn}(x)\mathrm{sgn}(y) = 0$. In the remaining cases, we can first establish and then use the identity $w = \mathrm{sgn}(w)|w|$ to complete the argument.

In all proofs by cases, ***it is important to be absolutely certain that ALL possibilities have been listed.*** The omission of some cases is an automatic mistake in any proof.

Interchanging roles of variables. This is a basic example of proofs by cases in which it is possible to "leverage" one case of the proof and obtain the other(s) with little or no additional work.

Example. Show that if x and y are real numbers which have opposite signs, then we have $|x - y| = |x| + |y|$.

SOLUTION: Suppose first that x is positive and y is negative. Then the left hand side is just $x + |y| = |x| + |y|$. Now suppose y is positive and x is negative. Then if we apply the preceding argument to y and x rather than to x and y we then obtain the equation $|y - x| = |y| + |x|$. Since the left hand side is equal to $|x - y|$ and the right hand side is equal to $|x| + |y|$, we get the same conclusion as before. In a situation of this type we often say that the second case follows from the first by reversing the roles of x and y.

Vacuous proofs. In some instances a statement is true because there are ***no examples where the hypothesis is valid.***

Example. Show that if x is a real number such that $x + 1 = x$, then $x^2 + 1 = x^2$.

SOLUTION: There is no real number satisfying the hypothesis, so whatever conclusion one states, there will be no number which satisfies the first but does not satisfy the second. Formally, the statement **P** $\Rightarrow$ **Q** (in words, "**P** implies **Q**") merely signifies that ***there are no situations in which*** **P** ***is true but*** **Q** ***is false***; if there are no situations where **P** is true, then there also cannot be any where **P** is true but **Q** is false.

How can this possibly be useful in mathematics? Sometimes the use of vacuously true statements allows one to state conclusions in a simpler or more uniform manner. For example, in elementary geometry one can show that the sum of the measures of the vertex angles for a regular $n-$gon is equal to $180\,(n-2)$. In some sense this is only valid if n is at least 3 because every regular polygon has at least three sides, but for some purposes it is convenient simply to state the formula for all positive integers n. The formula gives a negative angle measurement for $n = 1$, but this does not matter because this case of the formula does not apply if $n = 1$ since there are no $1-$gons. The point is that the statement of the formula is logically correct even if we omit the condition that n is at least 3. This is a simple example, but the concept of "vacuously true" also turns out to be useful in other situations where the hypothesis or conclusion is more complicated.

Adapting existing proofs. In all activities, it can be useful to use an idea that has worked to solve one problem in an attempt to solve another that may be somehow related. The same principle works for mathematical proofs. You might want to try modifying the first sample proof to show that if $3x + 1 = 10$, then $x = 3$ (modify the first proof above).

Disproving conjectures. Frequently one is faced with an unproven statement and the goal is to determine whether it is true or false. If you suspect the statement is false,

often the fastest way to confirm this is to construct a ***counterexample*** which satisfies the hypotheses but not the entire conclusion.

Illustration. If we are given real numbers a and b such that $a^3 - a = b^3 - b$, can we conclude that $a = b$?

SOLUTION: We should remark first that this is true if the absolute values of a and b are greater than 2, and someone who knows this might wonder if it is evidence that the result is always true. However, it is not; to show this, we need to find explicit distinct values of a and b for which the given equation holds. This can be done systematically, but the fastest way is to look at some examples and notices that the numbers 0 and 1 provide a counterexample because $1^3 - 1 = 0 = 0^3 - 0$.

On the other hand, ***it is important to recognize that one cannot prove a general statement by simply checking one, or even infinitely many examples that do not exhaust all the possibilities,*** and the preceding statement demonstrates this very convincingly (it is true whenever a and b are greater than 2).

Contrapositives, biconditionals and logical equivalences. In order to complete a proof of the biconditional (or logical equivalence) statement **P** $\Leftrightarrow$ **Q**, it suffices to prove the two separate statements **P** $\Rightarrow$ **Q** and (its "inverse" statement) **not P** $\Rightarrow$ **not Q**. **[**The reason for this rule is that the inverse statement **not P** $\Rightarrow$ **not Q** is the contrapositive of the converse statement **Q** $\Rightarrow$ **P.]**

Similarly, in order to complete a proof of **P** $\Leftrightarrow$ **Q**, it suffices to prove the contrapositive statement **not Q** $\Rightarrow$ **not P** and the inverse statement **not P** $\Rightarrow$ **not Q**.

Proofs of existence and uniqueness. It is *absolutely essential* to remember that *all such proofs have two parts*, one of which is an *existence proof* and the other of which is a *uniqueness proof*.

A symbolic approach to proofs. If it is difficult to decide how to start a proof, one suggestion is to put things into symbolic terms along the lines of the present section. This may provide enough insight into the question that a successful proof strategy can be found.

The use of definitions as a proof strategy. Another suggestion for finding a proof strategy is to recall all relevant definitions; it is very easy to overlook these or recall them inaccurately.

The do – something approach to finding proofs. This is simply trial and error, but it definitely should not be underestimated (recall Thomas Edison's comment about genius being 99 per cent perspiration and one per cent inspiration**!**). Even if no particular way of getting from the start to the finish is apparent, there is often little to lose by simply getting involved, doing something, trying different approaches, drawing pictures and proving everything that one can from the information given. Most of the proofs in print give no idea of the dead ends, incomplete arguments and otherwise unsuccessful efforts at proving something that took place before a valid proof was found. Trial and error is just as much a part of proofs in mathematics as it is of any other intellectual activity.

As suggested by the following quote, sometimes a point that seems very minor at first can provide the key to solving a problem.

You know my method. It is founded upon the observation of trifles.

Sherlock Holmes – The Boscombe Valley Mystery

Since the final method of proof is somewhat more complicated than the others, we shall discuss in in more detail.

Mathematical induction (also called ***Finite induction*** or ***Recursion).*** This is often a very powerful technique, and it is really more of a method to provide a formal verification of something that is suspected to be true rather than a tool for making intuitive discoveries, but it is absolutely essential. The use of mathematical induction dates back at least to some writings of F. Maurolico (1494 – 1575). There are many situations in discrete mathematics where this method is absolutely essential.

Most of the remaining material on mathematical induction is adapted from the following online references:

http://www.cut-the-knot.org/induction.shtml

http://en.wikipedia.org/wiki/Mathematical_induction

The similarity between the phrases "mathematical induction" and "inductive reasoning" may suggest that the first concept is a form of the second, but ***this is not the case.*** Inductive reasoning is different from deductive reasoning, but ***mathematical induction is actually a form of deductive reasoning.***

Proofs by mathematical induction involve a sequence of statements, one for each nonnegative integer n (sometimes it is impractical to start with $n = 0$, and one can begin instead with an arbitrary integer n_0), and it is convenient to let $P(n)$ denote the n^{th} statement. In the original example from the 16th century, $P(n)$ was the familiar formula for the sum of the first n odd positive integers:

$$1 + 3 + 5 + ... + (2n - 1) = n^2$$

In this case the first statement is $P(1)$, the statement $P(2)$ is $1 + 3 = 2^2$, the statement $P(3)$ is $1 + 3 + 5 = 3^2$, and so on.

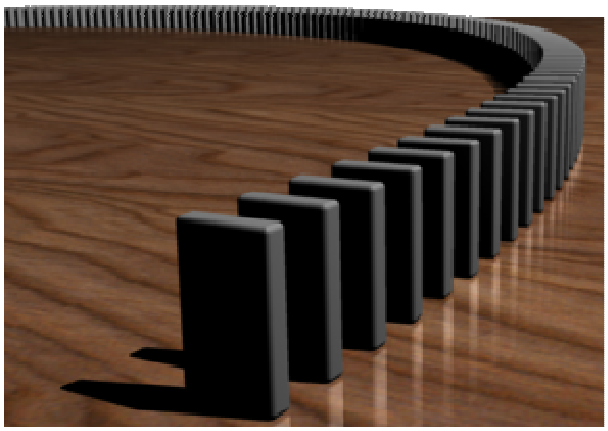
The method of proof by mathematical induction has two basic steps:

1. Proving that the first statement $P(n_0)$ is true.
2. Proving that for each value of k such that $k \geq n_0$, if $P(k)$ is true, then so is the next statement $P(k + 1)$.

In effect, *mathematical induction allows one to prove an* ***infinite list of statements***, say $P(1), P(2), P(3), ...$, *with an argument that has only* ***finitely many steps***. It may be helpful to visualize this in terms of the domino effect; if you have a long row of dominoes standing on end, you can be sure of two things:

1. The first domino can be pushed over.
2. Whenever a domino falls, then its next neighbor will also fall.

Under these conditions, we know that ***every one of the dominos in the picture below will eventually fall*** if the first one is nudged down in the right direction.



Incidentally, there is there is an animated version with Apple iPods at the following online site:

http://www.hemmy.net/2006/04/30/domino-ipod-commercial/

And here is a **YOUTUBE** video illustrating the same thing:

http://www.youtube.com/watch?v=IV68b0JlG9k&feature=related

There are some instances where one uses a variant of the principle of mathematical induction stated above; namely, one replaces the assumption in the second step with a stronger hypothesis that $\mathbf{P}(m)$ is true for ***all*** $m < k+1$ and not just for $m = k$.

Example of a proof by induction. Here is a proof of the summation formula for the first n odd integers. The statement $\mathbf{P}(1)$ merely asserts that $1 = 1^2$, and hence it is obviously true. Let's assume we know that $\mathbf{P}(k)$ is also true for some arbitrary k, so that we have the equation $1 + 3 + 5 + \ldots + (2k-1) = k^2$. The next step in mathematical induction is to derive $\mathbf{P}(k+1)$ from $\mathbf{P}(k)$. To do this, we note that

$$\begin{aligned} 1 + 3 + \ldots + (2k-1) + (2k+1) &= [1 + 3 + \ldots + (2k-1)] + (2k+1) \\ &= k^2 + (2k+1) \\ &= (k+1)^2 \end{aligned}$$

which shows that $\mathbf{P}(k+1)$ is also true because $2k + 1 = 2(k+1) - 1$. Therefore $\mathbf{P}(n)$ is true for all n and we have proven the general formula by mathematical induction.

Formally, the difference between mathematical induction and inductive reasoning sis that the latter would check the first few statements, say $\mathbf{P}(1)$, $\mathbf{P}(2)$, $\mathbf{P}(3)$, $\mathbf{P}(4)$, and then conclude that $\mathbf{P}(n)$ holds for all n. The crucial inductive step, "$\mathbf{P}(k)$ implies $\mathbf{P}(k+1)$," is missing. Needless to say, inductive reasoning does not constitute a proof in the strict sense of deductive logic.

Frequently the verification of the first statement in a proof by induction is fairly easy or even trivial, but ***it is absolutely essential to include an explicit statement about the truth of the initial case,*** and also ***it is important to be sure that the inductive step works for every statement in the sequence.*** If these are not done, the final conclusion may be false and in some cases downright absurd.

Example. (*Somewhat more difficult than the others*). Consider the following defective "proof" that a nonempty finite set contains as many elements as one of its proper

subsets. This is vacuously true for the empty set, so assume it is true for a set with **k** elements. Let **S** be a set with $k + 1$ elements; we need to show that some proper subset **T** contains the same number of elements as **S**. Let **T** be obtained from **S** by removing one element, and let **U** be obtained from **T** by removing one element. By the induction assumption we know that #(**T**) = #(**U**), and since we also know that #(**S**) = #(**T**) + 1 and #(**T**) = #(**U**) + 1 we conclude that #(**S**) = #(**T**). This is a ridiculous conclusion, so the point here is to ask, "***How did this happen?***" In fact, ***the inductive step we have given is valid for all values of*** k **except** ***for the case*** $k = 0$. However, when $k = 0$ this breaks down because **T** must be the empty set, so it is not possible to construct the subset **U** by removing an element from **T**.

Pólya's suggestions for solving problems. The classic book, ***How to solve it,*** by G. Pólya (1887 – 1985), discusses useful strategies for working problems in mathematics. A summary of his suggestions and a more detailed reference for the book appear in the following online document:

http://math.ucr.edu/~res/math133/polya.pdf

Avoiding and finding mistakes in proofs

Unfortunately, there is no simple way of doing these outside of checking things repeatedly and carefully, but we have already mentioned a few common causes of difficulties and how to prevent them and there are several more common errors that can be mentioned. This list is by no means exhaustive.

1. ***Eliminating distractions.*** It is absolutely essential to focus on **exactly** what was said and not to let one's mind wander or be distracted by irrelevant statements which may be included in a problem. Some examples along these lines are given in the following online sources:

   http://math.ucr.edu/~res/math133/braintest.pdf

   http://math.ucr.edu/~res/math133/braintest2.pdf

   http://math.ucr.edu/~res/math133/braintest3.pdf

Yet another Sherlock Holmes quotation is highly relevant here:

> It is of the highest importance in the art of detection to be able to recognize out of a number of facts which are incidental and which vital. Otherwise your energy and attention must be dissipated instead of being concentrated.
>
> ***Sherlock Holmes – The Reigate Puzzle***

2. ***Forgetting the assumptions.*** It is often easy to make mistakes in using assumptions, either adding to them or subtracting from them. The former often lead to mistakes in proofs, and the latter can often make it impossible to construct a valid argument.

3. ***Begging the question.*** Frequently one finds arguments in which a proof uses and relies upon some other auxiliary which has not been proven. In such instances all one has shown is that *if this auxiliary statement is true, then the original statement is true.* However, we may have no way of knowing whether the auxiliary statement is true or false.

4. ***Computational errors.*** Sometimes mistakes in arithmetic or algebra are embedded in arguments and destroy their validity. Plus and minus signs are particularly important examples. It is very easy to forget or confuse them if one is not careful (and sometimes even if one **is** very careful**!**), but getting them wrong can compromise or even destroy an entire proof.

5. ***Incorrect citations of other results.*** Of course, this can be deadly to a proof. ***Division by zero*** is a standard elementary example in which one neglects to recognize that $ax = ay$ *implies* $x = y$ **only** *if* a *is* **nonzero**.

6. ***Proving only half of biconditional or existence – uniqueness proofs.*** Half a proof may be better than none at all, but it is still just half a proof.

7. ***Proving the converse instead.*** Often one finds arguments which show that if the conclusion is true, then the hypothesis is true. This is the reverse of what is supposed to be established. An equivalent mistake is to show that if the hypothesis is false, then the conclusion is false. Such arguments may be valid, but they do not give prove of the statements which were supposed to be proved.

8. ***Overlooking alternative possibilities.*** Intuition is an extremely important tool in most areas of knowledge, but it is important to remember that intuition can be misleading and in many cases the expected outcome is not the actual one. Thus it is important in a logical proof to give equal weight to all the possibilities, even if some seem bizarre or counter – intuitive.

Here is a final Sherlock Holmes quotation, which is related to the preceding item:

> **[**Incomplete**]** evidence **...** may seem to point very straight to one thing, but if you shift your own point of view a little, you may find it pointing in an equally uncompromising manner to something entirely different.
>
> ***Sherlock Holmes – The Boscombe Valley Mystery***

9. ***Using unproven or false converses.*** This is a special case of the third item, but it is also one which plays an explicit role in elementary algebra.

The last of these is related to material on ***extraneous roots*** that one finds in elementary algebra courses. Here is a quick review of the underlying ideas. Suppose that we want to solve an equation like

$$x - 3 = \sqrt{30 - 2x}$$

The standard way to attack this type of problem is to eliminate the radical by squaring both sides and solving for x:

$$(x-3)^2 = \left(\sqrt{30-2x}\right)^2$$

$$x^2 - 6x + 9 = 30 - 2x$$

$$x^2 - 4x - 21 = 0$$

$$(x-7)(x+3) = 0$$

$$x = 7; x = -3$$

(**Source:** http://regentsprep.org/Regents/mathb/7D3/radlesson.htm)

This tells us that the only possible solutions are given by the two values above, but ***it does not guarantee that either is a solution.*** The reason for this is that the first step, in which we square both sides, shows that the first equation implies the second, but it does **not** follow that the second implies the first; for example, even though the squares of 2 and -2 are equal, these two numbers are clearly **not** the same. In order to complete the solution of the problem, we need to go back and determine which, if any, of these two possible solutions will work. It turns out that $x = 7$ is a solution, but on the other hand $x = -3$ is not (and hence it is an extraneous root).

The online site http://www.jimloy.com/algebra/square.htm has further examples of this type.

Ends of proofs. In classical writings mathematicians used the initials **Q. E. D.** (for the Latin phrase, *that which was to be demonstrated*) or **Q. E. F.** (for the Latin phrase, *that which was to be constructed*) to indicate the end of a proof or construction. Some writers still use this notation, but more often the end of a proof or line of reasoning is now indicated by a large black square, which is sometimes known as a "tombstone" or "Halmos (big) dot." We shall also use the symbol "■" to mark the end of an argument.

Proving mathematical impossibilities

There are many results in mathematics which state that something is impossible to do, and the meanings of such statements are often misunderstood. We shall illustrate the idea of impossibility theorems with a simple example.

> *It is mathematically impossible to find two odd (positive whole) numbers a and b such that $a + b$ is also odd.*

A statement of this sort might generate criticisms of the form, "Aren't you only saying that no one has found such a pair of odd numbers up to now? Isn't it possible that someone will find such a pair in the future?" The answer is that if such a pair existed, then it would lead to a consequence which is false, and here is the **impossibility proof**:

Suppose that we have such a pair of integers a and b, and express these numbers as $2m + 1$ and $2n + 1$ respectively. Then $a + b$ is equal to

$$(2m + 1) + (2n + 1) = 2(m + n + 1)$$

and hence $a + b$ is even. But we assumed that $a + b$ was odd, so we have a contradiction because every positive whole number is even or odd but not both. Since our assumption on a and b led to a false conclusion, the assumption that someone will find a pair of odd numbers a and b with $a + b$ odd must be false, and consequently no such pair can exist.■

The issues here are entirely different from past assertions like *no one can build an airplane* or *no one can survive a dive to the bottom of the ocean*. In these cases, the statements may have seemed correct in view of the methods and ideas which existed at the time, but are no longer so due to more recent advances. However, in mathematics the admissible methods and restrictions are all fixed in advance and the possibilities of inventing new methods or finding ways around restrictions are excluded. Sometimes one can circumvent these by modifying conditions (for example, the sum of two odd numbers is twice a ***rational*** number plus one), but such modifications generate an essentially different problem from a mathematician's point of view.